• 复杂性科学 •

百度迁徙规模指数构造方法反演

王 聪[1]，严 洁[2*]

(1. 四川警察学院计算机科学与技术系 四川 泸州 646000；2. 四川警察学院道路交通管理系 四川 泸州 646000)

【摘要】百度提供了迁徙规模指数以反映迁入或迁出某一特定地域的人口规模，成为经济地理科学与流行病学的重要研究依据。然而该指数仅为一个无量纲数，其构造方法目前尚未公开。该文将此指数假定为实际迁徙人口的可微函数映射，利用迁徙数据中的一个内蕴等式，反演出实际迁徙人口与该指数满足简单正比关系 $y=kx$。通过迁徙人口的离散特征和费马–欧拉定理推导得到迁徙人口的高概率互质特性，结合真实数据进行参数估计，得到线性关系的比例系数 k 为 3.24×10^{-5}。在全部数据记录上考察了内蕴等式的可信程度：当考虑舍入误差时，93.81% 的市际迁徙记录、82.65% 的市–省迁徙记录和 84.87% 的省际迁徙记录完全支持内蕴等式；其余违例记录的误差峰值为 357 人，对应相对误差约 0.5%，轻微的违例程度显示这种线性映射模型是自洽的。

关 键 词 数据分析； 人口迁徙； 人口地理； 交通流

中图分类号 TP393; C922; O211.9 **文献标志码** A doi:10.12178/1001-0548.2020441

An Inversion of the Constitution of the Baidu Migration Scale Index

WANG Cong[1] and YAN Jie[2*]

(1. Department of Computer Science & Technology, Sichuan Police College Luzhou Sichuan 646000;
2. Department of Road Traffic Management, Sichuan Police College Luzhou Sichuan 646000)

Abstract Baidu migration scale index represents the human migration scale of a specific area in China, and it has been used widely in geo-economics, demography, and epidemiology. Nowadays, Baidu migration index is adopted as a key data source for studying epidemic models of COVID-19. But the index is just a dimensionless number, its constitution method is still ambiguous. In this paper, the index is assumed as an elementary function mapping result of the real human migrate populations. According to a hidden equation existing in the data set, the mapping function is deduced to be a linear function $y=kx$. Another key phenomenon in the data set is the minimum interval of the migration index. All the migration index values and their differentials are exactly divisible by this interval. Through Fermat-Euler Theorem, we prove the coprimeness of the human migrate populations, and then the relationship between the minimum interval and minimum counting unit of the migrate populations is built, which means $k=3.24\times10^{-5}$. In the experiments, the migration records between 01/01/2020−04/30/2020 are examined to verify the correctness of the hidden equation: while the rounding error is considered, there about 93.81% of the city-to-city migration records, 82.65% city-to-province migration records and 84.87% province-to-province migration records can support the equation exactly; the maximum absolute error of the violation records is 357 peoples, which corresponds to about 0.5% relative error. The verifications support the self-consistency of the proposed linear mapping function.

Key words data analysis; human migration; population geography; traffic flow

作为分析人口迁徙规律的重要工具，百度迁徙网站[1]提供了城市和省区间迁徙的人口比例和总体迁移规模估计，为 COVID-19 疫情防控提供了重要参考。然而，百度迁徙规模指数作为一个无量纲数，其构造方法并未公开，仅能从有限的信息推知该指数与实际迁徙人口可能存在正相关特征。目前国际疫情防控形势仍不乐观，输入性疫情在国内仍时有局部性传播。考虑到人类迁徙行为是 COVID-

收稿日期：2020－12－21；修回日期：2021－04－06
基金项目：国家自然科学基金 (61602331)
作者简介：王聪 (1981－)，男，博士，副教授，主要从事交通流模拟、复杂系统与复杂性科学方面的研究.
通信作者：严洁，E-mail：yan_jie@foxmail.com

19 迅速传播的主要驱动力[2]，从防范疫情全国性二次蔓延的立场出发[3]，分析百度迁徙的数据构造方法及与真实人类迁徙行为的对应关系，从中反推出迁徙行为的确切人数，可以为研究总结疫情防控规律提供有益参考。

来自移动通信网络的数据是公共卫生管理的重要研判依据[4]。文献 [5] 利用复杂网络理论拟合人类迁徙与流行病学传播的关系，发现相对于节点间的经纬度距离，疫情传播与节点的等效距离相关性更强，而节点间的人类迁徙流量是等效距离的核心构成部分。因此，在 COVID-19 疫情爆发初期，考虑人类迁徙特征的流行病传播研究就得到了广泛关注。文献 [6] 利用城市间航空流量数据和腾讯人类迁徙数据，以种群传播模型进行建模。由于航空并非中国大陆出行的首选工具，该研究对疫情初期传播过程的解释能力存在缺陷。曾在区域经济学[7]、城市经济学[8] 和人口地理学[9] 等领域得到应用的百度迁徙网站也已重新开放，公开了百度依托移动互联网采集的全国 300 余个地级市和 30 余个省 (直辖市，自治区) 的人类迁徙状况。百度迁徙网站的数据陆续更新至 2020 年 5 月初，并保留 1 月 10 日−3 月 15 日的数据以供参考。文献 [10] 利用百度迁徙的数据初步调查了湖北省外部分城市迁入人口与疫情数据的关系，然而该分析仅局限于百度标注的流量较大的 50 个城市，相对于全国 300 余个地级市而言，覆盖面有所欠缺。文献 [11] 从百度迁徙数据中发现，各地累积确诊量和自武汉流入的人口总数高度相关，且次生传播链基本被斩断，因此提出了一种考虑输入病例和地区人口效应的定量化评估新型冠状病毒地区防控效果的近似方法。文献 [12] 利用百度迁徙的数据，对国内前 50 天疫情管控措施的有效性进行了细致分析，评估了旅行限制和社会疏导措施在防止传染病传播方面的效果。文献 [13] 以百度迁徙数据为依据，分析了限制城际人口流动，筛查/诊断/隔离/疑似密切接触者，以及社交隔离与个人安全防护等非医学干预手段的效果。该研究指出，此类措施在付出高昂经济代价的同时，可能使得患病人数减少了 67 倍。文献 [14] 使用了百度迁徙公布的包括武汉市历史与实时人口流动数据，以说明病例输入在疫情城际传播中的作用，并评估了防控措施的效率。文献 [15] 则使用从百度迁徙数据中提取出武汉到河南的记录，将河南省的输入性病例视为对武汉市的无偏抽样，以此估算出 COVID-19 在武汉的传播情况。文献 [16] 利用百度迁徙的数据，结合我国疾控中心的每日确诊病例数据训练 SEIR 模型，参考 SARS 的部分流行特征，利用 LSTM 神经网络预测了 COVID-19 疫情在国内的峰值和演化趋势。文献 [17] 利用 2020 年 1 月 10 日−23 日的百度迁徙数据分析了中国大陆的疫情空间格局特征，指出在省域层面疫情严重程度主要受邻近特征与人口迁徙强度的影响。文献 [18] 利用百度迁徙数据分析了疫情对中国城市人口迁徙的影响和城市的恢复能力。以上工作存在的一个共同问题是将百度迁徙规模指数假定为每日铁路、公路和航空人口流量的近似拟合，而这一假设目前并没有明确的依据。因此，本文前期工作[19] 利用公开新闻报道中的春运数据，证实了迁徙规模指数与实际迁徙人数呈粗略线性关系，并给出了一个线性系数的大致估计，以此为依据分析了 COVID-19 在早期的时空传播特征。

随着疫情在全世界的蔓延，部分研究人员也利用人类迁徙数据研究疫情在国外的传播与控制。文献 [20] 使用了包含 547 166 次航班，总计 101 455 913 名乘客的人类迁徙数据集，分析了遍及六大洲 22 个国家的人口迁徙与疫情流行状况的潜在关联性，并建议在限制高感染地区人口流动的同时，亦应对全球范围内的人口迁徙进行必要管控。涉及具体国家和地区的人口迁徙与疫情防控研究也普遍展开。文献 [21] 使用了由 Teralytics 提供的 2020 年 1 月 1 日−4 月 20 日匿名手机漫游数据捕获美国每个县的实时移动趋势，利用这些数据来生成社交隔离评价指标，并结合流行病学数据来探索 COVID-19 的疫情增长规律；文献 [22] 利用超过 2 700 万个移动设备的漫游记录，结合社交网站公开的数据，估计了美国不同区域社交隔离政策造成的地理和社会网络溢出效应；文献 [23] 将移动迁徙数据与人口普查统计数据相结合，建立了 COVID-19 在波士顿市区的精细传播模型。文献 [24]利用一个包含意大利 107 个大区的人类迁徙网络数据集估计了改进 SEIR 传播模型的参数后指出，对人类迁徙与社交隔离的有效限制已将该国疫情严重程度降低了 45%。文献 [25] 利用社交网站提供的近似实时的意大利人口迁徙数据进行了大规模分析，以研究交通管制策略对个人和地方政府经济状况的影响；文献 [26] 则关注了另一个疫情严重的国家巴西：通过航空数据的分析显示，约 76% 的巴西毒株可能在 2020 年 2 月 22 日−3 月 11 日期间自欧洲传入，并主要在本地和本州内传播。此后尽管航空旅

行人数急剧下降，但大型城市的输出效应不容忽视，当前该国的干预措施仍不足以控制疫情传播。文献 [27] 利用巴西数百万匿名移动漫游数据分析了 COVID-19 在巴西各州内最可能的传播方式，为公共管理计划制定与资源分配提供了参考。人类迁徙数据同样被应用于英国[28] 和印度[29] 等国家的疫情防控研究。

概览近期文献和成果，百度迁徙提供的数据已成为 COVID-19 疫情传播研究的核心数据来源之一。然而可能出于商业原因，百度迁徙提供的反映迁徙人口绝对规模的指数仅为无量纲数，公开的信息仅能表明该指数的构成与人口迁徙量正相关，仅能回答如“区域 A 的在某日的迁徙规模指数相对于区域 B 高约 1.25”，该指数代表的物理意义不够明确，对于迁徙人口的绝对数量刻画存在缺陷。考虑到流行病学模型对参量的敏感性，这一概要性质的表述限制了相关研究的可靠性。因此，有两个问题是不得不回答的：1) 百度迁徙的数据与真实人类迁徙流量满足什么映射关系?2) 如何从百度迁徙数据反推出真实的人口迁徙流量?

为了解答这两个问题，本文首先概要阐述了百度迁徙的数据来源与获取，然后以一个具体行政区划为例，挖掘了百度迁徙数据中内蕴的一个恒等关系。在此基础上，从理论上反演了实际迁徙人口和百度迁徙指数的函数表达式。基于费马 – 欧拉定理 (Fermat-Euler theorem) 证明得到了真实迁徙人数的高概率互质特征，以此为基础对映射函数的参数进行了有效估计，最终得到了一个自洽的线性函数映射模型。真实数据集上对内蕴恒等式的验证结果支持了该模型的有效性。

1 百度迁徙数据概览

百度慧眼是百度推出的一个商业地理智能数据平台。作为商业数据中面向公众开放的部分，百度迁徙网站展示了中国大陆省市两级全部行政区划的迁入/迁出迁徙规模指数以及与上一年度同一时间节点的对比，并针对每个行政区划，分别按照地市级和省级级别提供了最热门的 100 个迁入来源区划和迁出目的区划，以及迁自/迁入对应区划的人口百分比。其迁徙边界定义为某一区划的行政管理地域，包括该行政区划所管辖的所有下级区划。

百度迁徙数据总体可以分为两部分：迁徙规模指数和热门迁徙区划的迁徙人口百分比。百度将这两个参量解释为：1) 迁徙规模指数：反映迁入或迁出人口规模，城市间可横向对比；2) 热门迁入/迁出地比例：迁入/迁出到某城市的人口与全国迁入/迁出总人口的比值。

典型的百度迁徙数据的核心内容可以整理如表 1 和表 2 所示。

表 1　人口迁徙百分比

日期	迁徙类型	地域	迁入/迁出地	百分比/%
20200101	move_in	北京市	上海市	1.62
…	…	…	…	…
20200101	move_in	保定市	湖北省	0.42

表 2　特定日期迁徙规模指数列表

日期	迁徙类型	地域	值
20200101	move_in	天津市	2.480 868
…	…	…	…
20200101	move_out	茂名市	0.739 951 2

其中，表 1 的核心数据是特定区划迁徙人口的百分比。如表 1 的第一条目可解读为：2020 年 1 月 1 日自上海市迁入北京市的人口占北京市总体迁入人口的 1.62%；表 2 的值项是指定区划和指定方向的迁徙指数。如表 2 的第一条目表明，天津市在 2020 年 1 月 1 日的迁入规模指数为 2.480 868。

2 百度迁徙数据中的内蕴等式

在时刻t，定义行政区划i的迁入规模指数为$M^t_{i\leftarrow}$，区划j迁入i的人数占i总体迁入人数的百分比为$P^t_{i\leftarrow j}$；定义迁出规模指数为$M^t_{i\rightarrow}$；定义i迁向区划j的人数占总体迁出人数百分比为$P^t_{i\rightarrow j}$。定义区划i在时刻i的总体迁入人数为$H^t_{i\leftarrow}$，总体迁出人数为$H^t_{i\rightarrow}$，这两个参量为非负整数。

迁徙数据的重要核心部分是迁徙人数和流向。从表 1 和表 2 可知，迁徙流向可以通过百分比直接获得，而迁徙人数$H^t_{i\leftarrow}$和$H^t_{i\rightarrow}$是未知量，仅能通过迁徙规模指数推测。简化问题起见，假定不同日期，不同方向和不同区划的迁徙规模指数与迁徙人数间的函数映射方法相同，且该函数映射可用可微函数表达。显然有：

$$M^t_{i\leftarrow} \propto H^t_{i\leftarrow} \tag{1.a}$$

$$M^t_{i\rightarrow} \propto H^t_{i\rightarrow} \tag{1.b}$$

即迁徙规模指数与实际迁徙人数正相关。将迁徙规模指数的构造方法定义为真实迁徙人数的函数：

$$M^t_{i\leftarrow} = f(H^t_{i\leftarrow}) \tag{2.a}$$

$$M^t_{i\rightarrow} = f(H^t_{i\rightarrow}) \tag{2.b}$$

于是构造方法反演问题可以定义为以上函数的反函数求解，即给定任一方向的迁徙规模指数M_{i*}^t，求$f^{-1}(\cdot)$，使得对应方向的迁徙人口H_{i*}^t和迁徙指数满足：

$$H_{i*}^t = f^{-1}(M_{i*}^t) \tag{3}$$

对于任意行政区划对$\langle\alpha,\beta\rangle$，显然有：

$$P_{\alpha\to\beta}^t H_{\alpha\to}^t = P_{\beta\leftarrow\alpha}^t H_{\beta\leftarrow}^t \tag{4}$$

式中，以区划α的视角统计迁至区划β的人口数量，应等同于以区划β视角统计的自区划α迁入的人口数量。然后从真实数据中观察是否存在其他等式。对美元流通数据[30]、手机信令数据[31]、GPS 漫游数据[32] 和小样本的问卷调查[33] 研究证实，群体视角下人类出行距离呈现出显著的幂律分布，或带指数截断的幂律分布特征，出行人数随出行距离增长将显著衰减。因此同省内的区划更有可能出现于彼此的 Top100 迁徙目的地中。宁夏回族自治区仅辖有 5 个地级市，是全国下辖地级市最少的省区之一，为缩短行文，在此将其作为示例进行考察。抽取 2020 年 1 月 1 日宁夏及所辖地级市的人口迁徙情况如表 3～表 5 所示。

表 3　宁夏所辖区划 2020 年 1 月 1 日迁徙规模指数统计

迁徙方向	行政区划	迁徙规模指数
move_in	银川	0.877 521 6
move_out	银川	0.911 898
move_in	石嘴山	0.250 030 8
move_out	石嘴山	0.248 054 4
move_in	吴忠	0.487 684 8
move_out	吴忠	0.473 688
move_in	固原	0.206 712
move_out	固原	0.200 005 2
move_in	中卫	0.286 578
move_out	中卫	0.270 637 2

表 4　宁夏所辖区划 2020 年 1 月 1 日迁入百分比统计

%

行政区划	迁入				
	银川	石嘴山	吴忠	固原	中卫
银川	0	18.13	31.06	6.78	10.00
石嘴山	63.19	0	4.25	2.52	2.28
吴忠	59.92	2.05	0	4.26	12.98
固原	33.77	3.29	10.17	0	19.04
中卫	34.33	2.01	21.92	14.65	0

其中表 3 可解读如：2020 年 1 月 1 日，银川市迁入规模指数为 0.877 521 6，迁出规模指数为 0.911 898；表 4 可解读如：银川市迁入人口中有 18.13% 来自石嘴山市，有 31.06% 来自吴忠市；表 5 可解读如：银川市迁出人口中有 17.32% 前往石嘴山市，有 32.04% 前往吴忠市。

表 5　宁夏所辖区划 2020 年 1 月 1 日迁出百分比统计

%

行政区划	迁出				
	银川	石嘴山	吴忠	固原	中卫
银川	0	17.32	32.04	7.65	10.79
石嘴山	64.17	0	4.04	2.74	2.32
吴忠	57.55	2.24	0	4.43	13.26
固原	29.75	3.15	10.40	0	20.99
中卫	32.44	2.10	23.39	14.54	0

观察发现，表 3～表 5 中的内蕴等式为：

$$P_{\alpha\to\beta}^t M_{\alpha\to}^t = P_{\beta\leftarrow\alpha}^t M_{\beta\leftarrow}^t \tag{5}$$

为校验该内蕴等式是否成立，首先定义相对误差 RE(relative error)：

$$\mathrm{RE}_{\alpha\to\beta} = \frac{\mathrm{abs}\left(P_{\alpha\to\beta}^t M_{\alpha\to}^t - P_{\beta\leftarrow\alpha}^t M_{\beta\leftarrow}^t\right)}{P_{\alpha\to\beta}^t M_{\alpha\to}^t} \tag{6}$$

相对误差 RE 的作用是评价迁徙数据相对于式 (5) 的偏离程度。将表 3～表 5 的数据代入式 (6)，以迁入数据为基准，得到以百分比表示的相对误差统计如表 6 所示。

表 6　宁夏所辖区划 2020 年 1 月 1 日迁徙指数相对误差统计

%

行政区划	银川	石嘴山	吴忠	固原	中卫
银川	–	0.03	0.02	0.07	0.01
石嘴山	0.05	–	0.24	0.06	0.09
吴忠	0.02	0.15	–	0.18	0.01
固原	0.01	0.01	0.12	–	0.01
中卫	0.05	0.30	0.00	0.02	–

表中可见，最大的相对误差值仅为 0.3%，平均相对误差也仅为 0.07%。因此，从小样本数据来看，可以认为内蕴等式得到了有效验证。

3　迁徙规模指数构造反演与参数估计

3.1　迁徙规模指数构造过程推导

注意到式 (1) 对迁徙规模指数特征的刻画仍是极为粗略的，满足该式的函数形式也不是唯一的。因此有必要推导出迁徙规模指数的确定表达式，即式 (2) 的确切形式。

将式 (2) 代入式 (5)，可得：

$$P_{\alpha\to\beta}^t f(H_{\alpha\to}^t) = P_{\beta\leftarrow\alpha}^t f(M_{\beta\leftarrow}^t) \tag{7}$$

当$P_{\alpha\to\beta}^t \neq 0$时，式 (4) 可化为：

$$H^t_{\alpha\to}=\frac{P^t_{\beta\leftarrow\alpha}}{P^t_{\alpha\to\beta}}H^t_{\beta\leftarrow} \tag{8}$$

将式 (8) 代入式 (7) 可得：

$$P^t_{\alpha\to\beta}f\left(\frac{P^t_{\beta\leftarrow\alpha}}{P^t_{\alpha\to\beta}}H^t_{\beta\leftarrow}\right)H^t_{\alpha\to}=P^t_{\beta\leftarrow\alpha}f(H^t_{\beta\leftarrow}) \tag{9}$$

根据上文给出的可微假设，式 (9) 显然是一个连续可导函数。因此对式 (9) 两边分别求导并化简可得：

$$f'\left(\frac{P^t_{\beta\leftarrow\alpha}}{P^t_{\alpha\to\beta}}H^t_{\beta\leftarrow}\right)H^t_{\alpha\to}=f'(H^t_{\beta\leftarrow}) \tag{10.a}$$

由于迁徙人口百分比，即$P^t_{\beta\leftarrow\alpha}$和$P^t_{\alpha\to\beta}$的随机特征，通常有$(P^t_{\beta\leftarrow\alpha}/P^t_{\alpha\to\beta})H^t_{\beta\leftarrow}\neq H^t_{\beta\leftarrow}$。因此式 (10.a) 可等价为如下问题：对于任意给定的未知量x_1, x_2，有：

$$f'(x_1)\equiv f'(x_2)=k \tag{10.b}$$

因此必然有：

$$M^t_{i\leftarrow}=kH^t_{i\leftarrow}+b \tag{11.a}$$

对应地，有：

$$M^t_{i\to}=kH^t_{i\to}+b \tag{11.b}$$

将式 (11) 代入式 (5) 可得：

$$P^t_{\alpha\to\beta}(kH^t_{\alpha\to}+b)=P^t_{\beta\leftarrow\alpha}(kH^t_{\beta\leftarrow}+b) \tag{12}$$

利用式 (4) 约去式 (12) 的恒等项，有：

$$P^t_{\alpha\to\beta}b\equiv P^t_{\beta\leftarrow\alpha}b \tag{13}$$

同样考虑迁徙的随机特征，$P^t_{\alpha\to\beta}\equiv P^t_{\beta\leftarrow\alpha}$的条件显然不满足，因此必然有$b=0$。于是对于任一时间与迁徙方向上的迁徙规模指数$M^t_{i*}$和对应的实际迁徙人数$H^t_{i*}$，必然有：

$$M^t_{i*}=kH^t_{i*} \tag{14}$$

即，迁徙规模指数可表达为实际迁徙人数的线性函数。

3.2 参数估计

在爬取的数据中，迁徙指数至多保留至小数点后 7 位，因此首先排除迁徙指数上的舍入误差问题。考虑人口迁徙的随机性，若指数存在舍入误差，则尾数的最后一位的取值应近似服从均匀分布。抽取 2020 年 1 月−4 月迁徙规模指数共 95 590 条，最后一位实际取值分布如表 7 所示：

其中，原生数据中小数点后有效数字不满 7 位的取值，以 0 补足。表中可见末位尾数全部为偶数，难以满足均匀分布推论，不应认为是偶然因素所致。因此有理由认为爬取的指数是一个精确的数值，可以排除舍入误差问题。

表 7 迁徙规模指数尾数统计

末位尾数	频数	末位尾数	频数
0	19 228	5	0
1	0	6	18 981
2	19 151	7	0
3	0	8	19 196
4	19 034	9	0

注意到实际迁徙人数必然为非负整数，即$H^t_{i\to}$和$H^t_{i\leftarrow}$的值域必然是离散的。由此可得如下递进的推论。

推论 1 $H^t_{i\to}$和$H^t_{i\leftarrow}$的离散取值映射在迁徙指数$M^t_{i\to}$和$M^t_{i\leftarrow}$上，使得$M^t_{i\to}$和$M^t_{i\leftarrow}$的值域同样应是离散的；

推论 2 若推论 1 成立，则$M^t_{i\to}$和$M^t_{i\leftarrow}$的所有可能取值之间必然存在一个最小间距τ，其物理意义可推断为最小迁移统计单位。不引入过多复杂性的前提下，可推断为一个自然人在迁徙规模指数上的映射；

推论 3 若推论 2 成立，则最小间距τ应能被任一$M^t_{i\to}$和$M^t_{i\leftarrow}$的可能取值整除，即τ必然为$M^t_{i\to}$和$M^t_{i\leftarrow}$的公约数。

对 181 701 条迁徙规模指数记录 (包含 2020 年数据，及对应的 2019 年历史数据) 进行统计，其中仅包含 44 703 个不同的取值。因此有理由认为，该指数的取值是离散的，即推论 1 是成立的。于是将 44 703 个出现过的指数值进行排序并取级差，结果如图 1 所示。

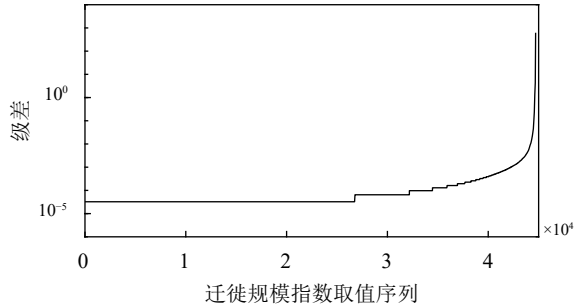


图 1 迁徙规模指数取值级差

图中可以看到鲜明的离散特征，即不同取值之间的差值集中在有限个离散的值上，这为推论 2 的成立提供了可靠的依据。更为关键的现象是，无论是级差还是迁徙规模指数取值，都是最小间隔 3.24×10^{-5} 的正整数倍，有理由认为是一个或多个

自然人在迁徙规模指数上映射的结果。

接下来讨论实际迁徙人数的互质特征。根据费马–欧拉定理，s 条记录值互质的概率$P(s)$可利用黎曼ζ函数 (Riemann ζ function, 当 s 为正整数时，$\zeta(s)$退化为欧拉乘积公式) 表示为[34]：

$$P(s)=\prod_{\forall p\in\text{prime}}(1-p^{-s})=\zeta(s)^{-1} \tag{15}$$

式中，p 的值域被定义为质数集合。根据ζ函数性质可知，当$s\geqslant 1$时，$P(s)$单调递增。特殊地，当 s 为正偶数时，$\zeta(s)$的取值可解析地表达为：

$$\zeta(s)=(-1)^{\frac{s}{2}-1}\frac{(2\pi)^s}{2s!}B_s \tag{16}$$

式中，B_s为第 s 项伯努利数 (Bernoulli number)。

当 s=10 时，$P(s)$的值收敛至约 99.9%；当 s=14 时，$P(s)$收敛至高于 99.99%。即随机抽取不少于 14 条不同的迁徙人口值，其互质的概率超过 99.99%，且随着抽取记录数量的增加，这一概率仍会进一步提升。而统计得到指数的取值高达 4 万余条，因此有理由认为，迁徙指数记录所代表的实际迁徙人数极高概率是互质的，其最大公约数为 1。因此，可以认为当一个自然人映射到迁徙规模指数上时，有：

$$M_{i*}^{t}=k\times 1=3.24\times 10^{-5} \tag{17}$$

于是，将斜率 k 代入式 (13)，可得任一方向上百度迁徙规模指数的构造方法为：

$$M_{i*}^{t}=H_{i*}^{t}\times 3.24\times 10^{-5} \tag{18}$$

4　数据获取方法

4.1　数据访问接口

通过对百度迁徙网站 Web 页面的分析可知，迁徙规模指数数据来自接口：http://huiyan.baidu.com/migration/historycurve.json，该接口以 HTTP GET 方法访问，并携带必要参数如表 8 所示。

表 8　迁徙规模指数数据访问必要参数

参数	取值范围	意义
dt	city, province, country	行政区划级别
id	GB/T2260-2007[32]	行政区划代码
type	move_in, move_out	迁徙方向

其中的 id 参数定义为以国家标准 GB/T2260-2007 定义的中华人民共和国行政区划代码[35]，涵盖了所有省级区划及其 (除直辖市) 直管的下级区划。正常情况下返回 JSON 格式文本形如：


```
{
  "errno": 0,
  "errmsg": "SUCCESS",
  "data": {
    "list": {
      "20190112": 7.6650624,
      "20190113": 7.8045444,
      …
      "20200314": 2.0624544,
      "20200315": 2.2150584
    }
  }
}
```


其中的有效数据为 list 字段，记录了 2020 年春运期间特定区划在特定日期的迁徙规模指数，以及以农历日期对齐的 2019 年同期数据作为对比。

地级市迁徙人口比例数据来自接口：

http://huiyan.baidu.com/migration/cityrank.json

省级迁徙人口比例数据来自接口：

http://huiyan.baidu.com/migration/provincerank.json

以上接口以 HTTP GET 方法访问，并携带必要参数如表 9 所示。

表 9　迁徙百分比数据访问必要参数

参数	取值范围	意义
dt	city, province, country	行政区划级别
id	GB/T2260-2007	行政区划代码
type	move_in, move_out	迁徙方向
date	20200110~20200315	数据统计日期

正常情况下返回 JSON 格式文本形如：


```
{
  "errno": 0,
  "errmsg": "SUCCESS",
  "data": {
    "list": [
    {
    "city_name": "\u5eca\u574a\u5e02",
    "province_name": "\u6cb3\u5317\u7701",
    "value": 21.72
    },
    …
    {
    "city_name": "\u5357\u901a\u5e02",
```

"province_name": "\u6c5f\u82cf\u7701",
"value": 0.12
}
]
}
}

其中有效数据为 list 字段。"city_name"等字段以 Unicode 转义字符形式编码，使用时应进行解码。

4.2 数据污染与有效性校验

百度迁徙网站一种可能的反爬虫策略为随机投放污染数据。举例而言，本文初次爬取的三亚市在 2020 年 2 月 2 日迁出至地级市的数据即可能存在污染。与真实数据对比如表 10 所示。

限于篇幅，表 10 仅枚举前 3 位数据。因此为了确保爬取数据的准确性，采用了一种主−从爬虫框架，首先确保主从节点使用不同的 IP 地址，由主节点按日期爬取数据并进行校验。对于校验失败的数据，交由从节点重新爬取，以避免主从节点同时被远程主机屏蔽。

表 10 污染数据与真实数据对比示例

位次	污染数据		真实数据	
	区划	百分比/%	区划	百分比/%
1	重庆市	1.56	乐东黎族自治县	10.77
2	成都市	1.36	陵水黎族自治县	7.55
3	广州市	1.25	海口市	6.74
…	…	…	…	…

*注：不同爬取节点的污染数据可能有差异。

数据有效性的校验规则是隔离污染数据的关键。一方面，对于某一特定区划 α，仅有前 100 位的迁徙人口流量数据被公开，因此存在 $P^t_{\alpha\to\beta}$ 和 $P^t_{\beta\leftarrow\alpha}$ 单向或双向缺失的可能；另一方面，对于区划 α 而言，污染数据投放至特定日期和特定方向的全部数据，目前未发现针对特定区划对$\langle\alpha,\beta\rangle$的污染策略。由此，设计数据有效性校验算法如下：

算法 1 数据有效性校验算法

For each day t and area tuple $<\alpha,\beta>$:

If both $P^t_{\alpha\to\beta}$ and $P^t_{\beta\leftarrow\alpha}$ exist:

If abs$\left(P^t_{\alpha\to\beta}M^t_{\alpha\to}-P^t_{\beta\leftarrow\alpha}M^t_{\beta\leftarrow}\right)\geqslant\varepsilon$:

Report error $<\alpha,\beta>$

本文实验取$\varepsilon=P^t_{\alpha\to\beta}M^t_{\alpha\to}\times0.05$。对于多次爬取仍无法通过校验的记录，改由人工校验和爬取。

5 内蕴等式有效性验证

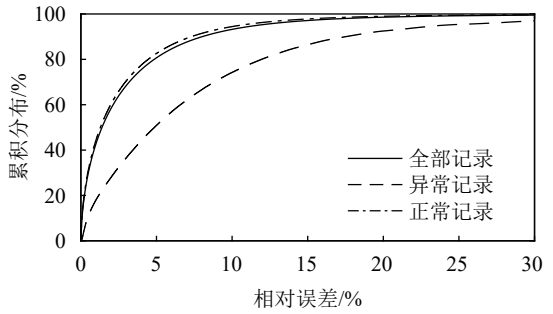
百度慧眼通过移动互联网进行数据采集。受网络质量和用户行为等因素影响，数据测量过程本身产生的误差并不能完全排除。而本文提出的初等函数映射成立的基础是内蕴等式 (4) 的成立，因此在考察式 (5) 能否得到满足时，除因人口迁徙百分比仅保留至小数点后 2 位有效数字所引起的舍入误差外，亦不能忽视测量误差的存在，误差的严重程度应进行准确判断。在此，取 2020 年 1 月 1 日 − 4 月 30 日共 4 个自然月的数据，将迁徙百分比$P^t_{\alpha\to\beta}$的取值松弛到区间$[P^t_{\alpha\to\beta}-0.005\%,P^t_{\alpha\to\beta}+0.005\%)$以解释舍入误差。当$P^t_{\alpha\to\beta}M^t_{\alpha\to}$与$P^t_{\beta\leftarrow\alpha}M^t_{\beta\leftarrow}$取值区间的交集为$\phi$，即存在无法以四舍五入解释的误差时，将此类记录归为异常记录。

首先考察市际迁徙流量是否满足本文提出的线性关系。在数据中，北京、上海等 4 个直辖市，以及湖北省潜江市、天门市和新疆维吾尔自治区石河子市、图木舒克市等直辖县级行政区划均被纳入城市区划进行采集和统计。数据中，约 93.81% 的记录误差位于舍入误差区间内，异常记录仅占约 6.19%。意味着在城市间交通流量这个层面，线性映射模型的基本假定可以得到满足，数据测量误差对于函数映射模型有效性的影响是有限的。正常记录、异常记录和全部记录的相对误差累积分布如图 2a 所示。图中可见，大约 81.2% 的记录相对误差在 5% 以内；而由于异常记录占比较低，过滤异常记录后，这一指标微升到 82.8%。对于异常记录而言，这一百分比则有 51.1%。然而仅仅考察相对误差是不够全面的，误差的绝对差值，抑或就本文述及的模型而言，误差的绝对人口数，也是评价模型有效性的重要指标。定义绝对误差 AE(absolute error)：

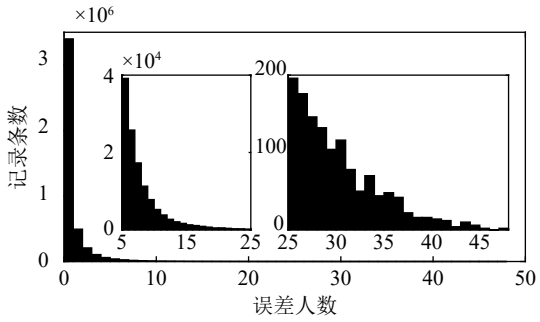
$$\mathrm{AE}_{\alpha\to\beta}=\mathrm{abs}(P^t_{\alpha\to\beta}M^t_{\alpha\to}-P^t_{\beta\leftarrow\alpha}M^t_{\beta\leftarrow}) \tag{19}$$

迁入流量的绝对误差与式 (19) 类似，不再赘述。绝对误差的含义显然是经由线性映射模型换算后城市α和β统计视角下迁徙人口的差值。图 2b 是正常节点绝对误差统计直方图。图中可见，对于正常记录而言，当不考虑舍入误差时，有约 87.44% 的记录绝对误差不多于 3 人；约 93.44% 的记录绝对误差不多于 5 人。绝对误差的极值出现在 1 月 20 日：当日汕头视角下自深圳迁入人口及对应的反向记录的误差达到了 79 人的极值，但对应的相对误差仅为 0.48%。因此有理由认为，相较于测量误差，舍入误差具备压倒性的影响。当考虑舍入误差时，迁徙人数的取值将松弛为某个特定区间，因此记录的绝对误差显著减小。图 2c 统计了异常记

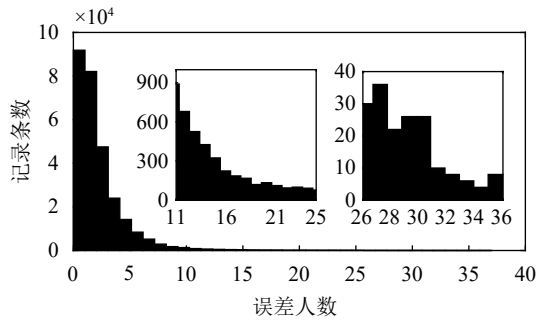
录绝对误差人数。图中可以看到，即使是异常记录，其最大绝对误差人数相对于舍入误差区间也仅偏出 36 人。在异常记录中，有 82.98% 的记录误差人数在 3 人以内，有 98.65% 的记录绝对误差人数在 10 人以内。可见，少量的违例现象对线性映射模型不产生本质影响，将其假定为数据测量误差是自洽的。



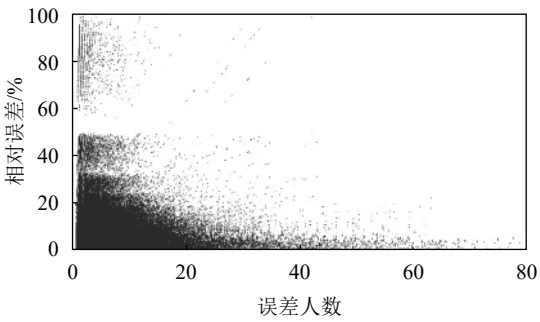
a. 迁徙记录相对误差累积分布



b. 正常记录误差直方图



c. 异常记录误差直方图

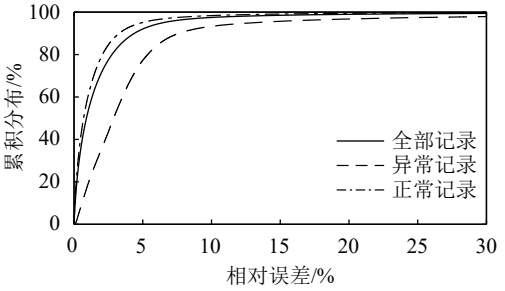


d. 相对误差与绝对误差关系

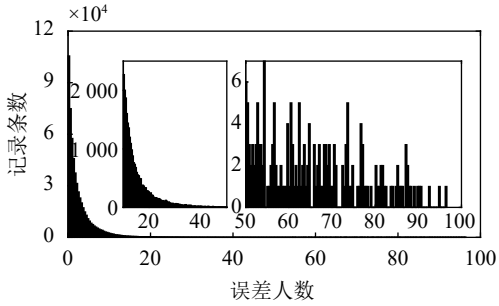
图 2　市际迁徙流量校验

注意到一个现象，即较多的绝对误差人数未必对应于较高的相对误差。因此，通过图 2d 分析异常记录的相对误差和绝对误差的对应关系。该图可分为 4 个逻辑象限：高相对误差高绝对误差；高相对误差低绝对误差；低相对误差高绝对误差和高相对误差高绝对误差。在图中，高相对误差高绝对误差区域几乎为空白。此外，除在低相对误差低绝对误差象限集中了大部分记录外，另外两个象限也存在一定比例的记录分布。分析可知，当两地人口迁徙流量悬殊时，以低流量区划视角统计的记录易出现高相对误差低绝对误差的情况；而两地人口流量均较大时，则易出现低相对误差高绝对误差的违例数据。

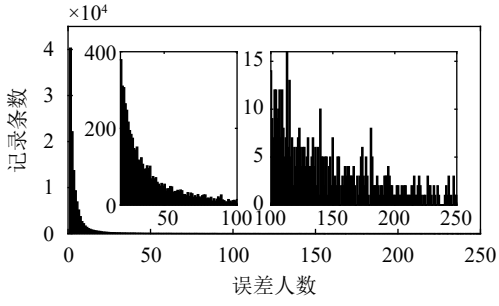
市–省间迁徙流量数据同样可以印证线性映射模型的有效性。利用与市际迁徙流量相同的统计方法进行分析。如图 3a，有 82.65% 的数据记录误差位于舍入误差区间内。该数据虽较城市间流量数据偏低，但全部记录的相对误差同时亦有显著降低：有约 92.06% 的记录相对误差不高于 5%；这一指标在正常记录中达到了 97.13%，在异常记录中同样达到了 77.3%，说明在市省流量层面的测量误差影响同样是有限的。图 3b 是正常记录的绝对误差统计。其中有 73.86% 的绝对误差人数在 3 人以内，有 95.77% 的绝对误差人数在 10 人以内。在正常记录中误差人数极值为 97 人，出现于 1 月 20 日北京市视角下自广东省迁入人数，此时相对误差为 1.32%，仍处于舍入误差松弛区间。如图 3c，当将考察视角迁移到异常记录时，发现擦除舍入误差后最大误差人数为 250 人，出现于 1 月 17 日濮阳市视角下自山东省迁入数据，此时对应的相对误差也仅为 2.64%。注意到即使仅考虑异常记录，也有约 98.6% 的绝对误差人数仍不多于 50 人——对于少则数百万，多则近亿人口的省级行政区划而言，可以认为这个量级的测量误差影响仍是有限的。相对误差与绝对误差的对应关系如图 3d 所示。可见在市–省层面表现出了与市际迁徙相似的分布特征，但其低相对误差低绝对误差象限的记录更加贴近相对误差坐标轴。一个合理的解释是，省级区划的迁徙记录来自下辖市级区划对应记录的简单加和，因此下属区划间测量误差的累积会抬高绝对误差；但由于测量误差存在部分相互抵消的现象，而市级区划的流量基数不变，因此随着迁徙流量的累加，相对误差反而会有所下降。



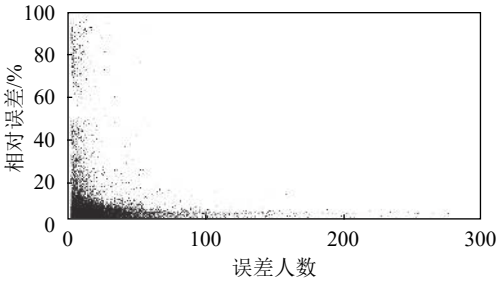
a. 迁徙记录相对误差累积分布



b. 正常记录误差直方图



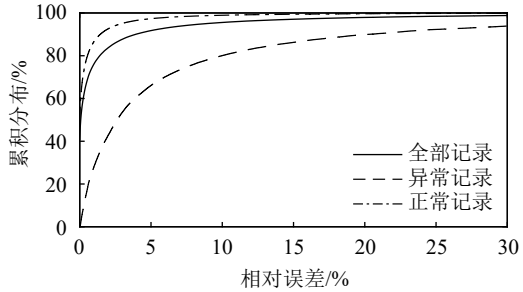
c. 异常记录误差直方图



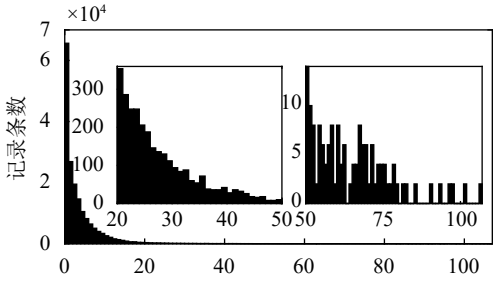
d. 相对误差与绝对误差关系

图 3 市–省迁徙流量校验

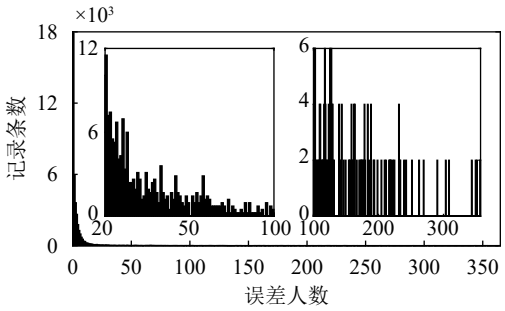
将同样的分析方法应用于省际迁徙数据进行验证。在图 4a 中，有 84.87% 的记录误差可被舍入误差区间覆盖。同时，由于记录两端的节点均为省级区划，人口迁徙基数较大，降低了迁徙记录的相对误差：有 50.73% 的记录相对误差小于 0.5%；89.43% 的记录相对误差小于 5%。图 4b 与 4c 分别统计了正常记录与擦除舍入误差后异常记录的绝对误差。可以看出，即使在省级区划这个层面，绝对误差仍可控制在相对很低的水平。对 4 个月的迁徙记录统计显示，正常记录中的极值出现于 1 月 12 日江西视角下自广东迁入记录，与其对应的反向记录差值为 107 人，对应的相对误差仅为 0.05%。异常记录中的极值出现在 1 月 23 日北京视角下迁往山东的记录及对应的反向记录，此时绝对误差达到 357 人。相对于两地当日该方向上 70 871～71 337 人的迁徙人数而言，其相对误差仅为约 0.5%。如图 4d 所示，相对误差与绝对误差的关系也体现出与市际流量和市省流量相似的特征。但随着流量基数的增加，低相对误差高绝对误差象限汇聚了相对更多的记录。总的来看，省际迁徙流量的数据同样可以给予线性映射模型有力的支持。



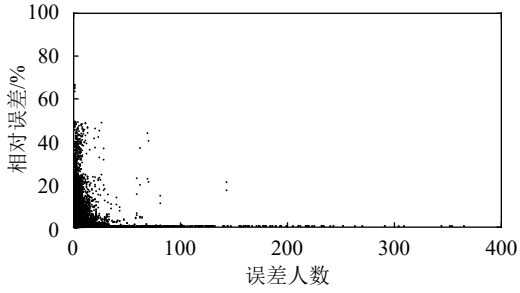
a. 迁徙记录相对误差累积分布



b. 正常记录误差直方图



c. 异常记录误差直方图



d. 相对误差与绝对误差关系

图 4 省际迁徙流量校验

6　结束语

人类迁徙活动是人口经济学、人类地理学乃至流行病学研究的重要依据。本文针对百度慧眼提供的国内长达 4 个月的人类迁徙数据进行分析归纳出数据中的一个内蕴等式。结合给出的基本假设：不同日期，方向与区划的实际迁徙人口与百度迁徙规模指数映射关系可用相同的初等函数表达，反演出迁徙规模指数的一个自洽的线性映射生成方法，即任一区划 i 在日期 t 的任一方向迁徙规模指数 M_{i*}^{t} 与当日在该方向上的实际迁徙人口 H_{i*}^{t} 满足简单线性映射关系 $M_{i*}^{t}=kH_{i*}^{t}$。通过迁徙人口的离散特征和费马−欧拉定理推导得到迁徙人口的高概率互质特性，结合真实数据进行参数估计，得到待定参数的有效估计 k=3.24×10^{-5}。为了验证线性映射模型的有效性，在全部数据记录上考察了内蕴等式的可信程度。验证结果对线性映射模型的支持显著：在市际迁徙流量层面，有 93.81% 的迁徙记录完全支持内蕴等式的成立，其误差可以完全由舍入误差解释；在市−省迁徙流量层面，有 82.65% 的记录完全支持内蕴等式；在省际迁徙流量层面，有 84.87% 的记录完全支持内蕴等式。少量违例记录相对于内蕴等式计算结果的偏差均较轻微，一般可认为是移动互联网测量手段限制所导致的误差。内蕴等式的可靠验证有效佐证了线性映射模型的自洽，使得从百度迁徙数据反推出区划间的实际迁徙人数成为可能。

参　考　文　献

[1] Baidu. com. 全国迁徙详情——百度地图迁徙大数据[EB/OL]. [2020-5-02]. https://qianxi.baidu.com/.

[2] JIANG Jun-feng, LUO Li-sha. Influence of population mobility on the Novel Coronavirus Disease (COVID-19) epidemic: Based on panel data from Hubei, China[J]. Global Health Research and Policy, 2020, 5(1): 30.

[3] XU Shun-qing, LI Yuan-yuan. Beware of the second wave of COVID-19[J]. Lancet, 2020, 395(10233): 1321-1322.

[4] OLIVER N, LEPRI B, STERLY H, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle[J]. Science Advances, 2020, 6(23): eabc0764.1-eabc0764.6.

[5] DIRK B, DIRK H. The hidden geometry of complex, network-driven contagion phenomena[J]. Science, 2013, 342(6164): 1337-1342.

[6] LI Qun, MED M, GUAN Xu-hua, et al. Early transmission dynamics in Wuhan, China of novel coronavirus-infected pneumonia[J]. The New England Journal of Medicine, 2020, 382(13): 1199-1207.

[7] 叶强, 张俪璇, 彭鹏, 等. 基于百度迁徙数据的长江中游城市群网络特征研究[J]. 经济地理, 2017, 37(8): 53-59.
YE Qiang, ZHANG Li-xuan, PENG Peng, et al. The network characteristics of urban agglomerations in the middle reaches of the Yangtze river based on Baidu migration data[J]. Economic Geography, 2017, 37(8): 53-59.

[8] 徐腾, 姚洋. 城际人口迁移与房价变动——基于人口普查与百度迁徙数据的实证研究[J]. 江西财经大学学报, 2018(1): 11-19.
XU Teng, YAO Yang. Urban population migration and housing price fluctuation: An empirical research based on the census data and baidu migration data[J]. Journal of Jiangxi University of Finance and Economics, 2018(1): 11-19.

[9] 蒋小荣, 汪胜兰, 杨永春. 中国城市人口流动网络研究——基于百度 LBS 大数据分析[J]. 人口与发展, 2017, 23(1): 13-23.
JIANG Xiao-rong, WANG Sheng-lan, YANG Yong-chun. Research on China's urban population mobility network based on Baidu LBS big data[J]. Population and Development, 2017, 23(1): 13-23.

[10] 许小可, 文成, 张光耀, 等. 新冠肺炎爆发前期武汉外流人口的地理去向分布及影响[J]. 电子科技大学学报, 2020, 49(3): 324-329.
XU Xiao-ke, WEN Cheng, ZHANG Guang-yao, et al. The geographical destination distribution and effect of outflow population of Wuhan when the outbreak of COVID-19[J]. Journal of the University of Electronic Science and Technology of China, 2020, 49(3): 324-329.

[11] 李冀鹏, 洪峰, 白薇, 等. 评估新型冠状病毒地区防控效果的一种近似方法[J]. 物理学报, 2020, 69(10): 99-106.
LI Ji-peng, HONG Feng, BAI Wei, et al. Approximate method to evaluate the regional control efficacy of COVID-19[J]. Acta Physica Sinica, 2020, 69(10): 99-106.

[12] TIAN Huai-yu, LIU Yong-hong, LI Yi-dan, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China[J]. Science, 2020, 368(6491): 638-642.

[13] LAI Sheng-jie, RUKTANONCHAI W, ZHOU Liang-cai, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China[J]. Nature, 2020, 585(7825): 410-413.

[14] KRAEMER M, YANG Chia-Hung, GUTIERREZ B, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China[J]. Science, 2020, 368(6490): 493-497.

[15] SONG Hai-tao, LI Feng, JIA Zhong-wei, et al. Using traveller-derived cases in Henan province to quantify the spread of COVID-19 in Wuhan, China[J]. Nonlinear Dynamics, 2020, 101(3): 1-11.

[16] YANG Zi-feng, ZENG Zhi-qi, WANG Ke, 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.

[17] 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.
LI Gang, WANG Jiao-bei, XU Ting-ting, et al. Spatio-

Temporal evolution process and integrated measures for prevention and control of COVID-19 epidemic in China[J]. Acta Geographica Sinica, 2020, 75(11): 2475-2489.

[18] 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.
TONG Yun, MA Yong, LIU Hai-meng. The short-term impact of COVID-19 epidemic on the migration of Chinese urban population and the evaluation of Chinese urban resilience[J]. Acta Geographica Sinica, 2020, 75(11): 2505-2520.

[19] 王聪, 严洁, 王旭, 等. 新型冠状病毒肺炎早期时空传播特征分析[J]. 物理学报, 2020, 69(8): 080701-1-080701-10.
WANG Cong, YAN Jie, WANG Xu, et al. Analysis on early spatiotemporal transmission characteristics of COVID-19[J]. Acta Physica Sinica, 2020, 69(8): 080701-1-080701-10.

[20] ZHANG Cheng, QIAN Li-xian, HU Jian-qiang. COVID-19 pandemic with human mobility across countries[J]. Journal of the Operations Research Society of China, 2021(9): 229-244.

[21] BADR H, DU Hong-ru, MARSHALL M, et al. Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study[J]. Lancet Infectious Diseases, 2020, 20(11): 1247-1254.

[22] HOLTZ D, ZHAO M, BENZELL S, et al. Interdependence and the cost of uncoordinated responses to COVID-19[J]. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117(33): 19837-19843.

[23] ALETA A, MARTIN-CORRAL D, PIONTTI A, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19[J]. Nature Human Behaviour, 2020, 4(9): 964-971.

[24] GATTO M, BERTUZZO E, MARI L, et al. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures[J]. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117(19): 10484-10491.

[25] BONACCORSI G, PIERRI F, CINELLI M, et al. Economic and social consequences of human mobility restrictions under COVID-19[J]. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117(27): 15530-15535.

[26] CANDIDO D, CLARO I, JESUS J, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil[J]. Science, 2020, 369(6508): 1255-1260.

[27] PEIXOTO P, MARCONDES D, PEIXOTO C, et al. Modeling future spread of infections via mobile geolocation data and population dynamics an application to COVID-19 in Brazil[J]. PLOS ONE, 2020, 15(7): e0235732.

[28] JEFFREY B, WALTERS C, AINSLIE K, et al. Anonymised and aggregated crowd level mobility data from mobile phones suggests that initial compliance with COVID-19 social distancing interventions was high and geographically consistent across the UK[J]. Wellcome Open Research, 2020, 5(5): 170.

[29] SAHA J, BARMAN B, CHOUHAN P. Lockdown for COVID-19 and its impact on community mobility in India: An analysis of the COVID-19 community mobility reports, 2020[J]. Children and Youth Services Review, 2020, 116(116): 105160.

[30] BROCKMANN D, HUFNAGEL L, GEISEL T. The scaling laws of human travel[J]. Nature, 2006, 439: 462-465.

[31] GONZÁLEZ M C, HIDALGO C A, BARABÁSI A L. Understanding individual human mobility patterns[J]. Nature, 2008, 453: 779-782.

[32] JIANG B, YIN J J, ZHAO S J. Characterizing the human mobility pattern in a large street network[J]. Physical Review E, 2009, 80: 021136.

[33] 闫小勇. 人类个体出行行为的统计实证[J]. 电子科技大学学报, 2011, 40(2): 168-173.
YAN Xiao-yong. Empirical statistics on individual human travel behavior[J]. Journal of the University of Electronic Science and Technology of China, 2011, 40(2): 168-173.

[34] 潘承洞, 潘承彪. 初等数论[M]. 第二版. 北京: 北京大学出版社, 2002.
PAN Cheng-dong, PAN Cheng-biao. Elementary number theory[M]. The 2nd edition. Beijing: Peking University Press, 2002.

[35] 全国信息分类与编码标准化技术委员会. 中华人民共和国行政区划代码: GB/T2260-2007[S]. 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 北京: 中国标准出版社, 2007.
Information Classifying and Coding. Codes for the Administrative Divisions of the People's Republic of China: GB/T2260-2007[S]. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Beijing: Standards Press of China, 2007.

编 辑 蒋 晓