

实证数据同化案例方法在新冠肺炎疫情分析中的应用

袁宏永¹, 梁漫春^{1*}, 黄全义¹, 苏国锋¹, 陈涛¹, 陈建国¹, 孙占辉¹, 杨思航¹,
邓李政¹, 黎崱¹, 秦泽生², 于淼淼¹, 程明¹, 李开远¹, 刘罡¹, 肖鑫鑫², 李文章²

1. 清华大学公共安全研究院, 北京 100094

2. 北京辰安科技股份有限公司中央研究院, 北京 100094

摘要 采用“实证数据同化SARS案例疫情趋势模型”方法,以各地市卫生健康委员会发布的新冠状病毒肺炎疫情数据为基础同化模型参数,总结研究团队2020年2月3—28日对全国(除湖北)、湖北省(除武汉)、武汉市的多次疫情趋势分析和决策建议,并提供3条预测曲线作为疫情防控指导线,预测疫情发展。该模型的疫情峰值线可作为评估当前防疫措施是否得当的标准线,用于各地市疫情趋势预警,指导防疫措施的制定,为医疗资源、生活应急物资的调度提供决策支持,为稳定民众情绪发挥作用。

关键词 新冠肺炎;疫情峰值线;疫情趋势预警

新型冠状病毒肺炎(简称新冠肺炎)疫情2019年末开始在中国31个省区市及港澳台地区、全球数十个国家迅速蔓延^[1]。当前对于该疫情的防控措施主要是限制大范围的人口流动和聚集,包括延长假期、推迟返工返学、在家上班等措施^[2];同时增大对于疫情重灾区湖北省(尤其是武汉市)的支援力度,包括医疗资源、生活应急物资等。探究疫情发展趋势,预测疫情峰值,对于评估疫情防控措施效果、制定疫情防控方案十分关键,有助于明确当前

疫情的严重程度,促进全社会科学认知新传染病,提高疫情防控期间的群体应对能力。

已有研究多通过动力学模型预测疫情拐点的方式评估疫情发展^[2-6]。这些模型对于当前疫情拐点预测不同,根据其定义和使用的数据不同,对拐点的解读也不同。以疫情数据最权威的国家卫生健康委员会官方公布的疫情数为例,主要包括5个衡量疫情的指标:确诊病例、治愈病例、死亡病例、疑似病例和密切接触者^[7],其他备受关注的指标还

收稿日期:2020-02-29;修回日期:2020-03-18

作者简介:袁宏永,研究员,研究方向为公共安全与应急管理,电子信箱:hy-yuan@tsinghua.edu.cn;梁漫春(通信作者),副研究员,研究方向为公共安全与应急管理,电子信箱:lmc@tsinghua.edu.cn

引用格式:袁宏永,梁漫春,黄全义,等.实证数据同化案例方法在新冠肺炎疫情分析中的应用[J].科技导报,2020,38(6):83-89;doi:10.3981/j.issn.1000-7857.2020.06.012

包括感染者和住院患者。每种指标又可以分为每日新增数和累计数两类。这样就演变出很多数据,例如,每日新增确诊病例、累计确诊病例、每日新增治愈病例和累计治愈病例等。不同指标可从不同维度对疫情发展趋势进行评估。每种指标都有对峰值和拐点的定义,且每种指标的每日新增数的峰值都会对应累计数的拐点,例如,确诊病例的拐点、疑似病例的拐点、住院患者数的拐点等。此外,每种指标可能存在多个峰值。不同指标的不同拐点对指导疫情防控工作有不同意义^[8],以下以3个峰值为例。

1) 新增确诊病例的峰值(累计确诊病例的拐点),该指标可以反映疫情防控工作的整体趋势。此拐点出现后,每日新增确诊病例数开始下降,预示疫情趋势向好;但此时每天确诊病例达到最大,确诊压力最大。

2) 住院患者的峰值,该指标可以反映救治工作的整体趋势。该峰值出现后,医疗救治队伍救治工作压力达到最大;但此后住院患者将逐天减少,疫情整体形势开始好转。住院患者的峰值出现时,出院患者大于或等于确诊入院患者。医疗救治队伍倾向于把这个峰值当成疫情的拐点。

3) 每日新增感染者的峰值(累计感染者的拐点),该指标可以掌握病毒传播的趋势。感染者没有直接观测数据,在传染病模型中多采用估算法。已有研究认为,此次疫情每日新增感染者的峰值日在武汉“封城”当天或者之后几天附近。

本文聚焦于当前疫情发展对社会整体的影响,选取新增确诊病例的峰值(累计确诊病例的拐点)作为疫情拐点的判断指标。清华大学公共安全研究院综合应急组和北京辰安科技股份有限公司中央研究院开发的“实证数据同化SARS案例疫情趋势模型”,以各地市卫生健康委员会发布的新冠疫情数据为基础同化模型参数,对全国(除湖北)、湖北省(除武汉)、武汉市及湖北其他各地市进行疫情趋势分析、总结和预测,探究该同化模型的疫情预测曲线在实际疫情防控中的应用,为制定全社会的停工、停学标准提供参考。

1 “实证数据同化SARS案例疫情趋势模型”方法介绍

1.1 模型假设与应用方法

“实证数据同化SARS案例疫情趋势模型”方法假设:1) 疫情防控措施和病毒传染力之间存在微妙的正反馈关系,使得最终有效再生指数符合全社会的代价利益平衡点,因此假设当前新型冠状病毒的有效再生指数变化规律与2003年SARS的有效再生指数变化规律相似性较高;2) 维持疫情控制措施强度,当每日新增确诊病例数下降为每日新增确诊病例峰值数的1/10后,防控措施才逐步减弱,因此假设新增案例数曲线为单峰形状。该方法通过假设新增确诊病例数的峰值日,采用新冠疫情实际数据同化SARS案例,可得到多条预测结果曲线。比较预测曲线与实际数据的拟合情况,则可近似认为,与实际数据最为吻合的预测曲线的假设峰值日为此疫情的峰值日。

1.2 模型应用流程

该方法的应用流程共包括3个环节:1) 选取与实际每日新增病例数和累计确诊病例数拟合度最好的3条预测曲线;2) 持续跟踪各地市卫生健康委员会公布的数据,分析实际数据的震荡规律及其与预测曲线的偏离度;3) 对实际每日新增病例数和累计确诊病例数超出预测曲线的省市进行预警,明确当前疫情趋势上升的主要原因,及时提出疫情防控的改善措施。本研究将“实证数据同化SARS案例疫情趋势模型”方法应用流程引入疫情防控策略决策中,并结合各地市疫情发展的特征对具体疫情防控措施提出针对性调整方案。

2 疫情趋势预测结果与分析

2.1 研究团队发表的多次疫情趋势分析结果

2月3日,研究团队在《科技导报》微信公众平台发表题为“预测:疫情一周内或现拐点!”的文章,提出基于实证数据同化SARS案例模型开展疫情趋势预测的方法,发布了预测曲线,指出预测疫情

的拐点在2020年2月6—7日出现的可能性较大,并提出“全面启动应急响应、信息的快速准确分析汇总和利用春节假期阻断病毒传播途径是疫情防控拐点到来的重要支撑;全国要提前决策、提前布局,并加大对湖北省(尤其是武汉市)的支援力度;各地实事求是地采取延长假期、推迟返工开学、在家上班等措施”的建议。

2020年2月5日,研究团队在《科技导报》微信公众平台发表题为“持续疫情分析:孝感数据现预警信号”的文章,发布了预测曲线,并指出湖北省疫情与其他地区疫情出现分化,湖北省和武汉市疫情拐点向8日后移,孝感市数据显现强烈预警信号,需要仔细研究如何进一步加强孝感市及同类型城市如黄冈市的防控和诊疗措施。

2月9日,研究团队在《科技导报》微信公众平台发表题为“防控成效明显,复工需关注重点城市”的文章,发布了预测曲线,并指出:全国疫情出现明显的峰值和拐点迹象,基本符合2月3日、5日的预测分析结果;峰值和拐点迹象的出现表明抗疫工作到了最艰难也是最有希望的阶段,真抓、细管、严控才是拐点到来和疫情好转的保障;北京、上海等大城市疫情数据保持相对稳定,为后续在科学防控有序恢复生产提供了较好条件。

2月14日,研究团队在《科技导报》微信公众平台发表题为“全国(除湖北省)防控形势持续向好,

河北省和长沙市预测曲线现强预警信号”的文章,发布了预测曲线,并对“各地有序复工对疫情的影响”做了详细分析。

2月16日,研究团队在《科技导报》微信公众平台发表题为“湖北疫情整体向好,孝感、鄂州、潜江3市预测曲线提示强预警信号”的文章,发布了预测曲线,预计武汉市最终确诊人数会落在41000~51000区间,中位数是46200(含临床诊断病例),对一些城市发出预警信号,给出相应建议。

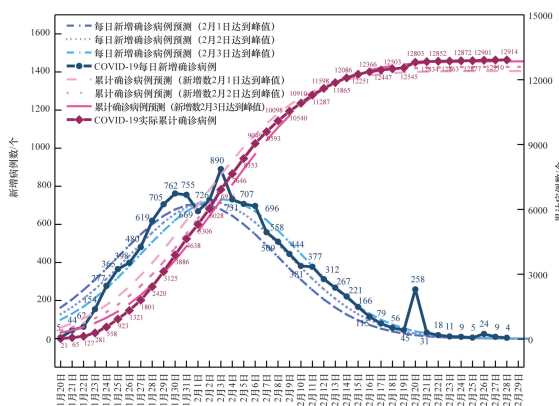
2月27日,研究团队在辰安科技微信公众平台发表题为“新冠肺炎疫情发展趋势分析(2020年2月27日)”的文章,发布了预测曲线,预计武汉市3月15日后,基本没有新增案例。

2.2 疫情趋势分析结果与实际疫情数据对比验证

针对2020年2月14日与16日在《科技导报》微信公众平台上发表的全国(除湖北)、湖北(除武汉)、武汉3个重要预测图与实际疫情数据进行对比验证。

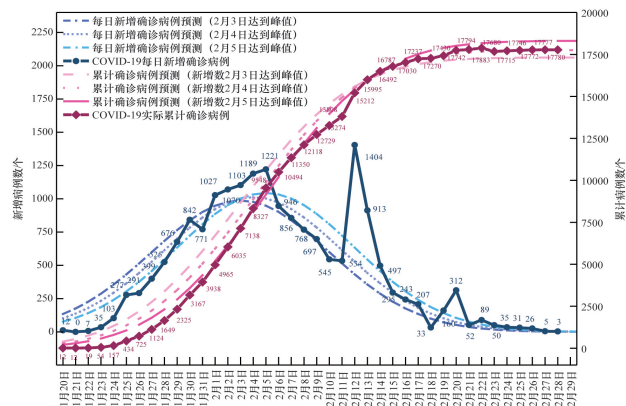
图1为全国(除湖北)地区疫情趋势预测曲线及其验证图,预测曲线采用2月13日前数据进行模型参数同化。可以看出,2月14—28日,疫情数据基本吻合模型估算的第3条预测曲线,即每日新增病例数在2月3日达到峰值的预测曲线。

图2为湖北(除武汉)地区疫情趋势预测分析结果及其验证图,预测曲线采用2月13日前数据进



注:预测使用2月13日24:00前数据,验证数据更新至2月28日24:00

图1 全国(除湖北)地区疫情趋势预测及其验证图

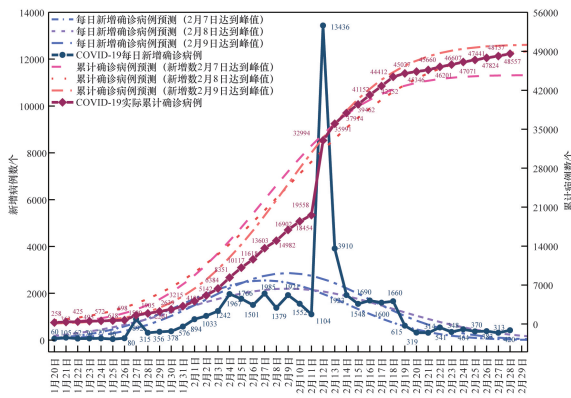


注:预测使用2月13日24:00前数据,验证数据更新至2月28日24:00

图2 湖北(除武汉)地区疫情趋势预测及其验证图

行模型参数同化。可以看出,2月14—28日疫情数据吻合模型估算的第2条预测曲线,即每日新增病例数在2月4日达到峰值的预测曲线。

图3为武汉疫情趋势预测及其验证图。采用2月13日前数据进行模型参数同化,结果显示2月14—28日疫情数据吻合模型估算的第2条预测曲线,即每日新增病例数在2月8日达到峰值的预测曲线。2月16日,武汉疫情前景十分模糊,模型较为准确地预测武汉市最终确诊人数会落在41000~51000区间,对抗疫决策有很强的参考意义。

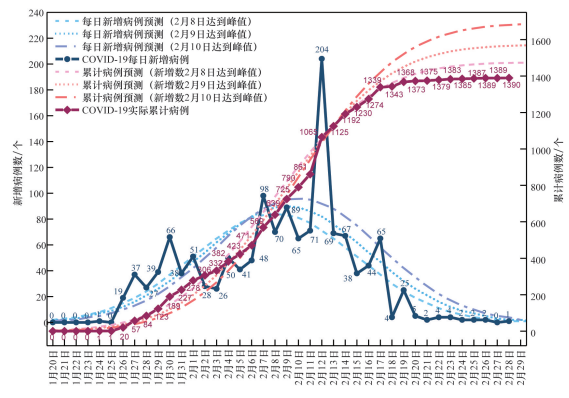


注:预测使用2月13日24:00前数据,验证数据更新至2月28日24:00

图3 武汉市疫情趋势预测及其验证图

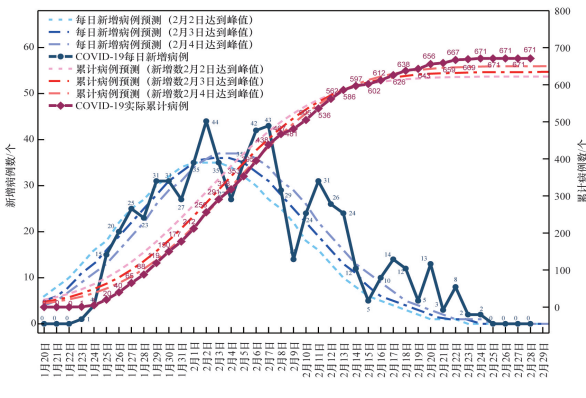
2.3 湖北各地市疫情预警实例

将本模型用于湖北各地市疫情预测,选择发出预警信号的地市进行分析,如图4—图7。对每日新增确诊病例和累计确诊病例均开始突破预测曲线的情况进行预警分析,结果显示,强烈预警信号在鄂州出现于1月30日、2月8日,在十堰出现于2月10日、2月11日,在潜江出现于2月12—16日。累计追踪数据显示,在以上各地市预警后数天均出现每日新增确诊病例的小高峰,表明疫情出现震荡,模型预警信号可靠。



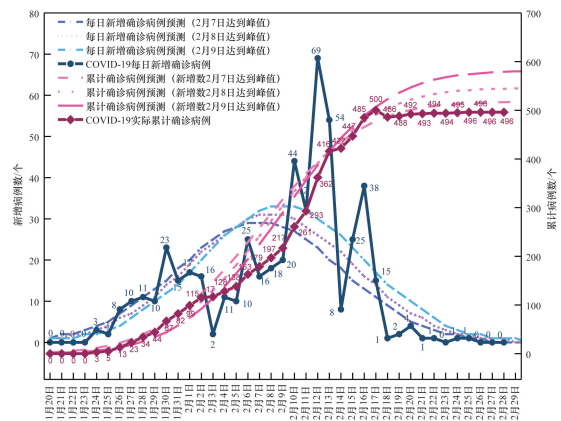
注:预测使用2月13日24:00前数据,验证数据更新至2月28日24:00

图4 鄂州市疫情趋势预测及其验证图



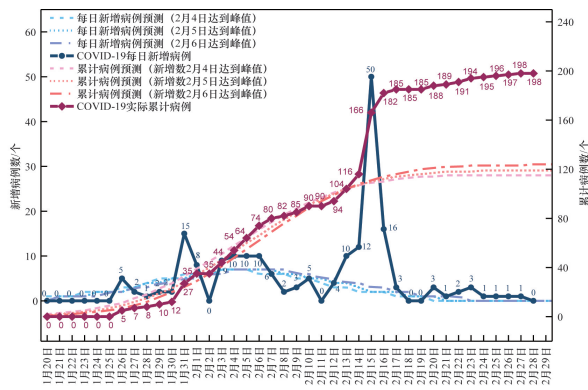
注:预测使用2月13日24:00前数据,验证数据更新至2月28日24:00

图5 十堰市疫情趋势预测及其验证图



注:预测使用2月13日24:00前数据,验证数据更新至2月28日24:00

图6 天门市疫情趋势预测及其验证图



注:预测使用2月13日24:00前数据,验证数据更新至2月28日24:00

图7 潜江市疫情趋势预测及其验证图

3 疫情趋势预测讨论

新冠肺炎是一种具有高传染性的急性呼吸道传染病,其疫情具有罹患率高、传染性强、扩散范围大、持续时间长等特点,属于突发公共卫生事件。运用预测模型指导疫情防控措施的制定具有十分重要的意义。本文采用“实证数据同化SARS案例疫情趋势模型”预测全国疫情趋势,选取新增确诊病例峰值(累计确诊病例拐点)作为拐点指标,以指导疫情防治工作。本研究中,预测模型提供3条预测曲线预测疫情发展趋势,通过比对每日新增病例数和累计病例数实际值与预测值的偏离程度,判断疫情是否出现爆发或者反复的现象,进而指导当前疫情防控措施的制定。当每日新增病例数和累计病例数吻合模型的预测曲线时,选择拟合度最高的预测曲线判断疫情拐点日期。在此日期之后,每日新增病例数开始下降,累计确诊病例数增速放缓。本文模型预测,全国(除湖北)地区拐点出现在2月1—3日,湖北(除武汉)地区拐点出现在2月3—5日,武汉市拐点出现在2月7—9日,基本符合实际疫情。

本文模型中,每日新增病例数和累计确诊病例数的预测曲线是期望值曲线,反映疫情发展趋势的平均水平。一般来说,实际每日新增病例数会小范围波动发展,其值应在模型预测曲线上波动。但

是,当每日新增病例数和累计确诊病例数均突破预测曲线时,表明当前疫情发展突破其期望值,可能出现疫情拐点后移、反复甚至爆发的情况,此时模型发出强烈预警信号,提醒有关部门加强防控措施。因此,模型中的3条预测曲线也可作为疫情的3道防线,提醒当前疫情发展的拐点是否突破预测日期向后移动,即每日新增病例数是否会出现新的峰值。

本文模型预测曲线与实际公布数据存在一定的偏离,主要原因在于:该肺炎潜伏期较长且疫情前期确诊能力不足,因此病例确诊和公布存在延迟^[9]。2月12日,国家卫生健康委员会增加临床确诊标准,实际每日新增病例数出现跳跃式增长,病例确诊和公布延迟的情况得到极大缓解,因此该数值的增加有助于增强对于当前疫情实际发展趋势的判断。2月12—13日,湖北省(尤其是武汉市)实际每日新增病例数出现明显峰值。本文模型中,这段峰值曲线与预测曲线之间的面积等于2月12日前实际新增病例数曲线与预测曲线之间的面积。这是由于2月12日前,湖北(尤其是武汉市)受医疗资源限制较大,每日确诊能力不能满足疫情需求,导致大量患者确诊延迟,因此每日公布的新增确诊数据偏低。这会造成疫情防控对当前形势的误判,低估当前疫情发展的严峻程度,影响防控效果,造成疫情拐点延迟、每日新增病例数和累计病例数突破预测值等现象。若能在2月12日前提升确诊能力,该峰值对应的新增病例则会分散到前几日,实际疫情病例数则与当前预测曲线拟合度更高,拐点迹象更明显。2月14日,新增确诊病例数和累计确诊病例数再次回归预测曲线,表明当前期积压患者基本得到确诊且病例及时公布后,本文模型对于疫情发展可进行准确预测,误差小于5%。

该模型在湖北各地市的疫情预测中,在鄂州、十堰、天门、孝感、潜江等地曾出现每日新增病例数和累计病例数均突破3条预测线的情况,此时模型发出预警信号,表示当时这些地市疫情防控出现反复迹象,疫情存在拐点后移的可能,这将导致累计确诊病例超出预测趋势的估计值。通过分析每日

新增病例(累计确诊病例)的预测值和实际值的差异,可建立一种灵敏的疫情预警机制,指导疫情防控措施及时调整。

综上,该模型可用于新冠肺炎在全国各省市疫情发展趋势预测和分析,其预测曲线可当作疫情防控的指导线进而建立实时预警机制。此外,2003年北京市SARS疫情曲线在当年5月下旬曾出现的反弹峰使得疫情延长10天左右,对本次疫情防控有很强的警示作用。因此,在疫情末期需要持续重点防控,保持密切接触者处于隔离状态,严格防止输入型案例和潜伏期较长案例造成的偶发聚集性疫情反复。疫情后期,要格外警惕偶发聚集性案例的发生。

4 结论

自新冠肺炎疫情暴发以来,湖北省尤其是武汉市受影响最严重。本文通过基于“实证数据同化SARS案例疫情趋势模型”分析,预测疫情拐点,准确判断全国、湖北、武汉疫情发展出现分化,并提出基于模型分析的每日疫情预警机制,为疫情防控提供有力支持。

对未来防疫工作提出以下建议。

1) 重视对累计确诊病例数拐点的预测分析和科学解读。采用新增确诊病例峰值(累计确诊病例拐点)预测疫情形势较为准确可靠,其预测曲线可以作为疫情防线建立预警机制,评估各地疫情防控措施是否有效。需要各界科学解读确诊病例数的拐点、治疗病例数的拐点以及其他不同拐点的意义。

2) 重视模型治理,提升数据价值。数据缺失和偏差会导致对疫情发展形势的误判,阻碍对疫情实时态势的预估,因此要重视对基于数据的模型治理,提升数据使用价值,为探索社会大范围内传染病的发展机制和应对措施提供依据。

3) 重视突发重大公共卫生事件的机理分析,完善应急体系建立。结合疫情发展下的社会实际情况,以突发事件的内在机理为基础,关注疫情的演化逻辑,才能发挥各方力量,从实际出发制定合理的疫情防控措施和政策,避免社会恐慌情绪,正确应对疫情发展。

参考文献(References)

- [1] Munster V, Koopmans M, van Doremalen N, et al. A novel coronavirus emerging in China—key questions for impact assessment[J]. *New England Journal of Medicine*, 2020, 382(8): 692–694.
- [2] Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia[J]. *The New England Journal of Medicine*, 2020, doi: 10.1056/NEJMoa2001316.
- [3] Liu T, Hu J, Kang M, et al. Transmission dynamics of 2019 novel coronavirus (2019-nCoV)[J]. *BioRxiv*, 2020, doi: 10.1101/2020.01.25.919787.
- [4] Shen M, Hu J, Kang M, et al. Modelling the epidemic trend of the 2019 novel coronavirus outbreak in China[J]. *BioRxiv*, 2020, doi: 10.1101/2020.01.23.916726.
- [5] Hermanowicz W. Forecasting the Wuhan coronavirus (2019-nCoV) epidemics using a simple (simplistic) model [J]. *MedRxiv*, 2020, doi: 10.1101/2020.02.04.20020461.
- [6] Ming W, Huang J, Zhang C. Breaking down of healthcare system: Mathematical modelling for controlling the novel coronavirus (2019-nCoV) outbreak in Wuhan, China[J]. *BioRxiv*, 2020, doi: 10.1101/2020.01.27.922443.
- [7] Tan W, Zhao X, Ma X, et al. A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019–2020[J]. *China CDC Weekly*, 2020, 2(4): 61–62.
- [8] Bauch C, Lloyd-Smith J, Coffee M, et al. Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future[J]. *Epidemiology*, 2005, 16(6): 791–801.
- [9] Funk S, Eggo R. Early dynamics of transmission and control of 2019-nCoV: A mathematical modelling study[J]. *MedRxiv*, 2020, doi: 10.1101/2020.01.31.20019901.

Application of empirical data assimilation method in trend analysis of COVID-19

YUAN Hongyong¹, LIANG Manchun^{1*}, HUANG Quanyi¹, SU Guofeng¹, CHEN Tao¹, CHEN Jianguo¹, SUN Zhanhui¹, YANG Sihang¹, DENG Lizheng¹, LI Ke¹, QIN Zesheng², YU Miaomiao¹, CHENG Ming¹, LI Kaiyuan¹, LIU Gang¹, XIAO Xinxin², LI Wenzhang²

1. Department of Engineering Physics, Tsinghua University, Beijing 100094, China

2. Environmental Safety Business Division, Beijing Safety Technology, Co., Ltd., Beijing 100094, China

Abstract The method of "empirical data assimilation for the SARS epidemic trend model" is used to assimilate the model parameters based on the new crown epidemic data released by the Health Committee of each city, and our research team's recent work is presented, including the multiple epidemic trend analysis and the decision-making suggestions for the whole country (except Hubei), Hubei Province (except Wuhan) and Wuhan city from February 3 to February 28, 2020. Three prediction curves are shown as the guideline of the epidemic prevention and control to predict the development of the epidemic. The epidemic peak line of the model can be used as a standard line to evaluate whether the current epidemic prevention measures are appropriate and be used for the early warning of the epidemic trend in various cities, to guide the proper taking of the epidemic prevention measures, to provide the decision support for the scheduling of medical resources and emergency supplies for life, and to play a role in stabilizing the public mood.

Keywords COVID-19; epidemic prediction curves; early warning of epidemic trend ●



(责任编辑 王丽娜)