

刘俊,王胜宏,金朦朦,等.基于微博大数据的2010~2018年中国桃花观赏日期时空格局研究[J].地理科学,2019,39(9):1446-1454.[Liu Jun, Wang Shenghong, Jin Mengmeng et al. Reconstruction of Peach Blossom-viewing Date of China Using Weibo Big Data. Scientia Geographica Sinica,2019,39(9):1446-1454.] doi: 10.13249/j.cnki.sgs.2019.09.010

基于微博大数据的2010~2018年中国 桃花观赏日期时空格局研究

刘俊,王胜宏,金朦朦,李宁馨

(四川大学旅游学院,四川 成都 610065)

摘要:从新浪微博中检索到2 952 828条有关桃花观赏活动的微博,按照数据筛选、时空匹配、检验校准的流程,筛选出843 034条表述“桃花开了”的可采信微博,从中提取并构建了2010~2018年全国293个城市的桃花观赏日期数据集。之后使用物候站点观测数据、气温数据和物候模型模拟数据验证了桃花观赏日期数据集,得出基于微博大数据提取的中国桃花观赏日期数据集符合物候站点观测的桃花花期变化规律,符合桃花开花前气温升高则花期提前的基本规律,并与物候模型模拟的桃花花期基本一致,可以作为物候站点观测数据的重要补充。基于提取出的桃花观赏日期数据集,进一步分析了中国桃花观赏日期的时空格局,结果表明,纬度每升高1°,桃花观赏日期推迟1.78 d,其中北亚热带桃花观赏日期对纬度变化最为敏感,纬度每升高1°,桃花观赏日期推迟2.56 d。过去9 a全国大部分城市的桃花观赏日期提前。

关键词:物候;大数据;桃花;赏花旅游

中图分类号:K903 **文献标识码:**A **文章编号:**1000-0690(2019)09-1446-09

物候现象是陆地生态系统对环境变化最敏感、最直观和最综合的反映,已经成为全球气候变化的“诊断指纹”(fingerprints)^[1]。作为陆地生态系统的重要组成部分,植被在全球变化碳循环研究中发挥着重要作用,其物候期的变化状况是众多生态系统模型的重要参数^[2-4]。与此同时,植被物候期中的开花和叶变色现象还是非常重要的旅游资源^[5],其起止时间和变化趋势直接决定了赏花旅游^[6,7]和观叶旅游活动的组织开展^[8,9]。因此,对观赏类植物开展多尺度的物候监测,研究其物候期变化规律具有重要意义。

在著名地理学家竺可桢先生的倡导下,“中国物候观测网”(Chinese Phenological Observation Network, CPON)于1963年成功创建。该科技基础条件平台提供的物候观测数据具有长时间序列和高可靠性的优点,多年来中国物候学研究一直以此作为基础数据。但由于目前全国仅有53个物

候观测站点,观测数据的覆盖范围还十分有限^[10],而建立新的观测站点需要持续投入大量的人力物力,因此,拓宽物候观测数据渠道,寻找传统站点观测数据的补充数据变得尤为紧迫。此时,社会大众发布在自媒体平台上的物候信息就成为了重要的数据来源。

有鉴于此,本文以中国传统观赏植物桃花为研究对象,以新浪微博为数据来源,通过对检索到的物候信息进行分析和提取,制作了中国桃花观赏日期数据集,并结合站点观测数据、气温数据和物候模型对数据集进行了验证,在此基础上进一步分析了中国桃花观赏日期的时空格局及其变化规律。此研究一方面验证利用大数据技术提取植物物候数据的可行性,为站点观测数据提供数据补充,另一方面为中国桃花赏花旅游活动的组织和开展提供数据支撑。

收稿日期:2018-11-29;**修订日期:**2019-01-18

基金项目:国家自然科学基金项目(41771163)、四川省重点研发项目(2018SZ0373)、四川大学创新火花项目库(2018hhs-44)资助。
[Foundation: National Natural Science Foundation of China (41771163), Key Research and Development Program of Sichuan Province (2018SZ0373), Innovation Inspiration Fund of Sichuan University (2018hhs-44).]

作者简介:刘俊(1979-),男,四川乐山人,教授,主要从事可持续旅游研究。E-mail: liujun_ignrr@126.com

1 数据来源与研究方法

1.1 数据来源

微博数据来源于新浪微博网页端(<https://weibo.com/>)。据《中国植物志》记载,桃(*Amygdalus persica*)的花期为3~4月,为保证检索结果能够代表全国范围内桃花花期状况,检索了2010~2018年每年1~6月的微博。以“桃花”为关键词,共计检索到2 952 828条微博记录,每条微博包含了发布时间、微博正文、发布城市和微博网址信息。

成都市桃花站点观测数据来源于中国物候观测网在成都设立的望江观测站和江安观测站,由四川大学旅游学院观测和维护,观测工作开始于2015年,观测物种47种。

气象数据来源于中国气象数据网(<http://data.cma.cn/>)提供的中国地面气候资料日值数据集(V3.0)。为验证冬季低温刺激对桃花花期的影响,选取了2010~2018年每年1~6月以及前一年12月的气温数据。

中国气候区划数据来源于中国科学院资源环境科学数据中心(<http://www.resdc.cn>),用以研究不同气候区内桃花观赏日期的时空格局差异。

1.2 研究方法

1.2.1 桃花观赏日期数据提取

大数据技术为物候观测打开了全新视角^[11,12]。国外已有使用通讯技术收集广大志愿者物候观测结果的经验,中国物候观测网也于2014年发布了物候观测志愿者招募信息,通过互联网等工具收集整理物候观测志愿者的业余物候观测结果后加以利用^[10]。在互联网技术的发展和推动下,自媒体时代已经到来,微博用户自发记录和分享诸如“小区楼下桃花开了”的物候信息,使其加入到物候观

测志愿者队伍中来。截至2018年底,新浪微博月活跃用户达4.62亿人次^[13],广大用户群体创造出覆盖全国、数量庞大、全网共享且持续更新的物候大数据。国内学者已经开始使用微博大数据进行旅游研究和探索,例如张子昂、唐佳、闫闪闪等借助新浪微博的位置服务(location based service, LBS),分别以南京钟山风景区、西安市和洛阳市为案例地对旅游活动的时空规律进行了分析^[14-16],但鲜有研究将大数据技术应用到赏花旅游物候数据的提取中来。有鉴于此,本文使用新浪微博LBS,从表述桃花物候信息的微博中提取出地理位置,将其与微博发布时间相结合,从而获取到包含时空信息的桃花物候数据。

不同于物候观测站的专业观测人员,微博发布者很难在微博正文中对桃花的种属及所处物候期阶段进行准确判读,但是由于有关桃花物候信息的微博数量众多且持续更新,覆盖了桃花多个种属的整个物候期,而且人们集中在桃花适宜观赏的时节前往观赏并发布微博,因此基于微博大数据提取出的桃花观赏日期能够从赏花旅游的视角反映常见桃花种属的物候变化状况。所以本文按照数据筛选、时空匹配、检验校准的流程,对检索到的2 952 828条有关桃花物候的微博进行处理,制作出基于微博大数据的中国桃花观赏日期数据集。

数据筛选通过设置黑名单和白名单的方法完成。自然语言识别是文本分析的难点,丰富的汉语词汇和语法组合形成了人们对同一事件多样的表述方式,但是事件表述的核心词汇和语法有一定的规律。通过设置微博表述“桃花开了”这一事件的黑名单和白名单(表1),保留了诸如“我家门口的桃花开了”的可采信微博,剔除了诸如“梦到

表1 微博筛选白名单和黑名单
Table 1 Weibo filters whitelist and blacklist

设置	文本关键词
白名单	桃花开;看桃花,赏桃花,摘桃花,送桃花;竹外桃花三两枝,桃花依旧笑春风,山寺桃花始盛开,人面桃花相映红,桃花流水鳜鱼肥,桃花一簇开无主;桃花娇,桃花红,桃花艳,桃花香,桃花美;桃花盛开,桃花复盛,桃花灼灼,桃花开得正好,桃花格外多,桃花红过了头,桃花影落,桃花朵朵开,桃花始盛开,桃花三两支,桃花真的太美了,桃花真灿烂,一片桃花,小小的花朵,满目桃花,桃花一簇,全开了,都开了,遍地开,开了满天,花开成海,竞相绽放;今年的桃花,桃花季,桃花节,桃花公园
黑名单	梦到桃花开了,以为桃花开了,以后看桃花,等着桃花开,待到桃花开,不见桃花开,桃花开了没,桃花开了吗;等着你回来看那桃花开,大寒桃花开,三生三世十里桃花,人缘爆棚桃花开;烂桃花,桃花运,桃花劫,旺桃花,犯桃花,眼若桃花,面若桃花,桃花手绘,桃花妆,桃花妖,星座,打榜,转发

桃花开了”的不可采信微博。按照满足白名单但不满足黑名单的原则,对微博正文进行逐条筛选,从而得到可采信的表述桃花物候期信息的微博共计843 034条。

在提取出可采信微博的发布时间和城市后,根据发布年份和城市对所有可采信微博进行时空匹配。匹配到具体的年份和城市后,将可采信微博的发布时间转换为距离当年1月1日的第几天天数(day of year)作为提取出的适宜开展赏花旅游活动的桃花观赏日期。按照2倍标准差剔除每个城市每年桃花观赏日期中的异常值,并逐年计算该城市所有可采信微博的桃花观赏日期平均值,从而得出所有城市历年的桃花观赏日期。

由于微博用户数量和桃花物种分布存在地域差异,导致了不同城市保留的可采信微博数量存在差距^[17]。为保证数据筛选结果的准确性,对可采信微博数量少于10条的城市年份,采用人工判读的方式对该城市当年所有检索到的原始微博进行筛选,从而确定该城市当年准确的桃花观赏日期。受到新浪微博上线时间和检索结果的影响,部分城市缺失2010年等年份的微博数据。对于时间序列长度超过5 a的城市,求取缺失值前后两年桃花观赏日期的平均值进行补充,时间序列长度小于5 a的城市则直接剔除。

1.2.2 观赏日期数据集验证

中国物候观测网拥有的植物物候观测数据,是验证桃花观赏日期数据准确性的重要依据。通过对比站点观测的花期数据和基于微博提取的桃花观赏日期数据,能够得出2套数据之间的差异,从而验证基于微博提取的桃花观赏日期的准确性。成都市现有植物物候观测站2个,分别为江安站(郊区)和望江站(市区),站点的地理位置分布使其能够反映整个成都市范围内植物的物候情况。由于站点观测数据起始时间不同,选取了2个站点同时开始观测的2016年作为起始年份,对比了2016~2018年站点观测数据与基于微博提取的桃花观赏日期数据的差异。

现代生物学与物候学研究表明,温度是决定植物物候期演变的主导因素,植物物候期的更替需要一定的临界温度强度和积温^[18],前1个月和前3个月的温度通常对植物物候的影响最大^[19]。对每个城市基于微博提取的桃花观赏日期和该城市气温数据做相关性分析,如果呈现显著的负相关

关系,则可以验证基于微博提取的桃花观赏日期满足桃花开花前气温升高则花期提前的基本规律。由于部分城市观赏日期和气温数据不对称,且部分城市地域范围内有多个气象站点,在剔除了桃花观赏日期缺失或气温数据缺失的城市后,求取城市范围内多个气象站点数据的平均值作为该城市的气温数据,将其与桃花观赏日期做相关性分析。

物候模型是模拟和预测植物物候期的重要手段,国内外学者利用多种物候模型对植物物候期进行了模拟和预测^[20-22]。仲舒颖选择桃(*Amygdalus persica*)、杏(*Armeniaca vulgaris*)、紫荆(*Cercis chinensis*)和紫丁香(*Syringa oblata*)4种典型观赏植物,通过这些植物在中国42个站点1963~2012年的物候观测资料及对应的气象资料,建立并验证了模拟不同站点植物始花期和末花期的时空物候模型^[23]。该模型中对桃的物候模拟基于36个站点共计184条观测记录,使用GDD(Growing Degree Days)模型模拟花期变化,模型对桃的始花期和末花期内部检验和交叉检验的 R^2 分别为0.90和0.85,拟合优度较高。将这36个观测站所在城市的气温数据代入该模型,计算出每个城市模型模拟出的桃花始花期和末花期时间,将其与基于微博提取的桃花观赏日期进行对比,观察基于微博提取的桃花观赏日期是否在模型模拟的桃花始花期至末花期时间区间内,以此验证观赏日期数据的准确性。

1.2.3 观赏日期时空格局

桃花观赏日期的空间格局主要体现于其纬度地带性,对每个城市历年赏花日期求平均值后,采用线性回归模型研究不同气候区内桃花观赏日期随纬度的梯度变化。时间格局则采用线性回归模型对每个城市历年桃花观赏日期进行线性拟合,以此分析近年来桃花观赏日期是否显著提前或滞后。

2 结果

2.1 观赏日期数据集

对检索到的以“桃花”为主题的微博进行采信筛选,提取出发布时间和发布城市后按年份和城市分类匹配,剔除异常值并对采信微博数量少于10条的城市进行人工判读校准,按平均值处理缺失值后制作出基于微博大数据提取的桃花观赏日

期数据集。该数据集共计剔除2 109 794条无关微博,筛选出843 034条表述“桃花开了”的可采信微博,提取出293个城市的桃花观赏日期数据,覆盖中国除香港、澳门以外的所有省、市、区(图1),平均每个城市采信微博2 877条,数据起止时间为2010~2018年。数据集中提取到的最晚桃花观赏日期出现于黑龙江省佳木斯市,日期为2013年5月27日,最早的桃花观赏日期出现于海南省海口市,2010~2018年桃花观赏日期均稳定在每年的2月2日前后。全国范围内桃花观赏日期的平均值为第79天,即每年的3月20日前后,众数为第87天,即每年的3月28日前后。

2.2 数据集验证

将基于微博提取出的桃花观赏日期数据集分别与成都市站点观测数据、桃花开花前期气温数据、物候模型模拟数据作对比分析,验证该数据集准确性。

2.2.1 通过站点观测数据验证

对比江安站和望江站观测记录的桃花花期数据可以看出(图2),同期桃花开花时间存在差异,2017年2个站点始花期相差13 d;同一站点不同时期桃花开花时间出现推迟或提前,江安站2017年始花期较2016年推迟14 d,2018年始花期较2017年提前9 d;郊区桃花物候期变化较市区波动更大,

江安站(郊区)桃花花期变化更加活跃。对比基于微博提取的桃花观赏日期和观测站点记录的桃花盛花期,可以看出基于微博提取出的桃花观赏日期均在观测站点记录的盛花期时间区间内,2016年二者几乎完全一致,表明基于微博提取的桃花观赏日期符合观测站点记录的桃花花期变化规律,且能够较好均衡观测站点地理位置差异从而代表整个城市桃花花期状况。同时由于基于微博提取的桃花观赏日期数据来源于大量的群众观测记录,桃花观赏日期的数据波动更加稳定。

2.2.2 通过气温数据验证

桃花开花前温度升高则花期提前是其基本的物候规律。从基于微博提取的桃花观赏日期数据集中筛选出桃花观赏日期数据与气温数据对称的235个城市,将这些城市历年的桃花观赏日期数据与气温数据加以整合。选择桃花观赏日期前一句积温(P_1)、前第二句积温(P_2)、前第三句积温(P_3)、前两句积温(P_{12})、前三句积温(P_{123})、前一个月平均气温(T_{ave})共计6个气温指标,分别与桃花观赏日期(F)做相关性分析,相关系数分别为 $-0.45, -0.53, -0.59, -0.52, -0.56, -0.55$ 。结果表明,桃花观赏日期(F)与6个气温指标为 P 在0.01的显著性水平上均呈现负相关关系,符合桃花开花前气温升高则花期提前的基本规律,验证了基于微博提取的桃

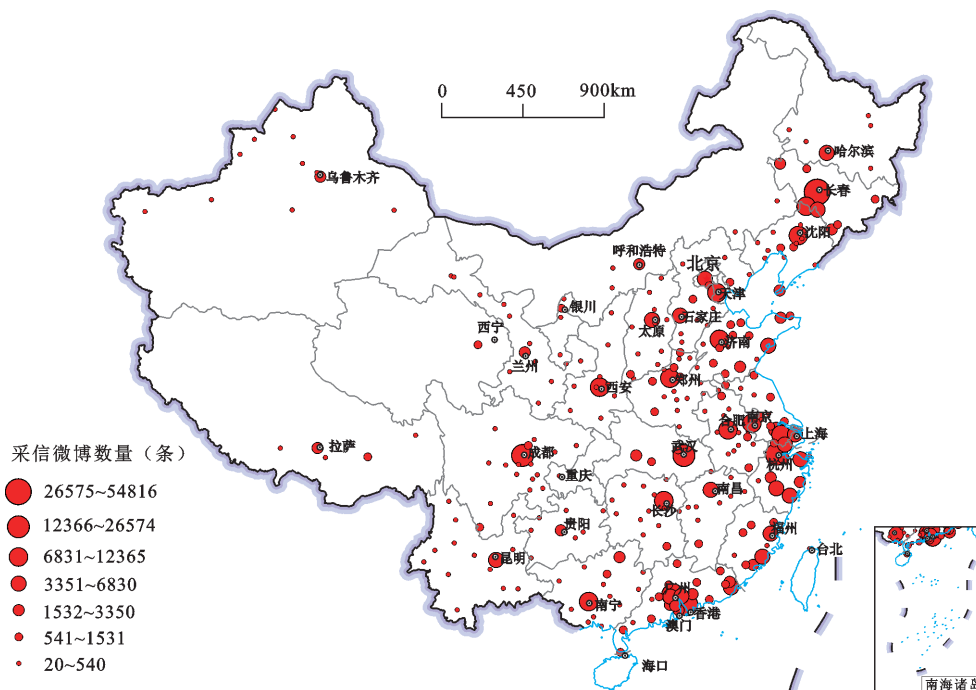


图1 桃花观赏日期数据集城市分布及采信微博数量

Fig.1 City distribution and density of peach blossom viewing date dataset

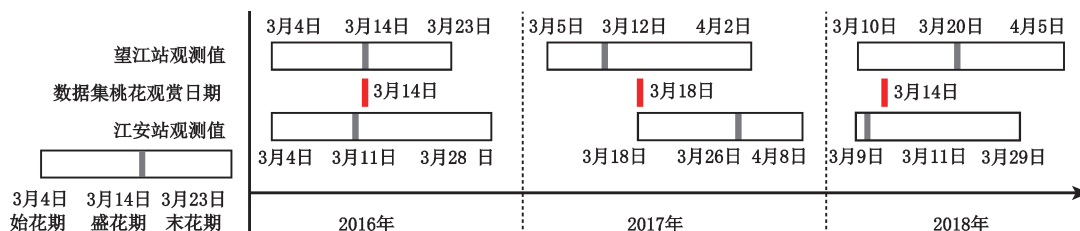


图2 成都市物候观测站点数据验证

Fig.2 Data verification of Chengdu phenological observation site

花观赏日期满足桃花物候期变化所需的气温条件基础。

2.2.3 通过物候模型验证

植物花期的变化是多个因素互相制约的结果,所选物候模型虽基于36个站点的桃花花期观测数据,然而在模拟花期物候时仍有3~8 d的误差。选取出物候模型选用的36个站点所在城市中有微博桃花观赏日期数据的29个城市,按照模型给出的参数代入气温数据计算出每个城市的历年模拟始花期和末花期时间,将其与基于微博提取出的桃花观赏日期作相关性分析,结果表明基于微博提取的桃花观赏日期和模拟始花期、模拟末花期均在0.01的显著性水平上正相关。在29个城市9 a共计261个对比样本中,基于微博提取的桃花观赏日期有36%的样本在模拟始花期和末花期时间区间内,有63.2%的样本在物候模型的误差范

围内,验证了基于微博提取的桃花观赏日期与物候模型模拟的花期基本一致。

2.3 桃花观赏日期时空格局

求出每个城市历年桃花观赏日期的平均值后,使用反距离权重法空间插值得到中国桃花观赏日期等值线图(图3)。从图中可以看出中国桃花观赏日期由南到北逐渐推迟,其中桃花观赏日期最早的是中国南部的海南、广东、广西和云南等省区,桃花观赏日期为2月中下旬,最晚的是新疆西北部以及黑龙江北部的三江平原地区,桃花观赏日期为4月下旬,全国桃花观赏日期的南北差异达2个月以上。赏花旅游活动的组织和开展时间在全国范围内呈现明显的区域差异,游客也可据此选择合适的时间和地点出游赏花。

桃花观赏日期随纬度的变化有以下规律。纬度每升高1°,桃花观赏日期推迟1.78 d(图4)。不

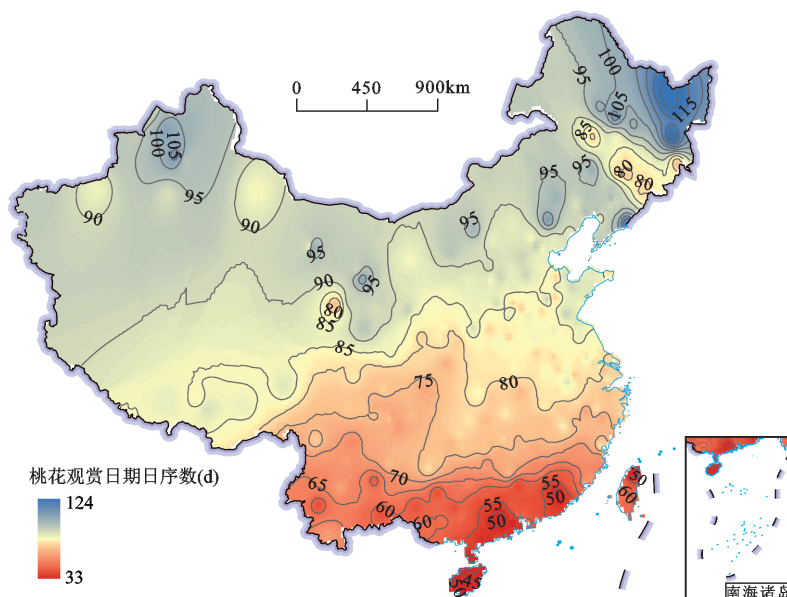


图3 桃花观赏日期空间格局

Fig.3 Spatial pattern of peach blossom viewing date

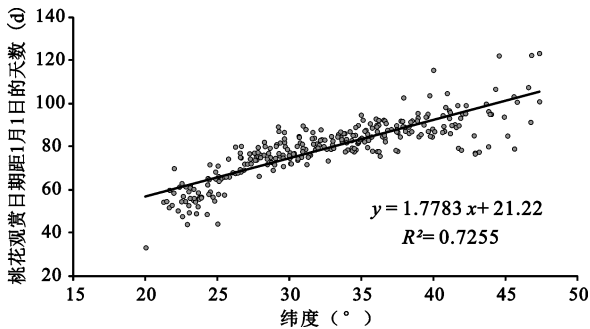


图4 桃花观赏日期与纬度的线性拟合结果

Fig.4 Linear fitting results of peach blossom viewing date and latitude

同气候区内桃花观赏日期对纬度变化的敏感程度有所差异,其中高原气候区和南亚热带桃花观赏

日期的变化和纬度的相关性并不显著,南温带、中温带、中亚热带和北亚热带均显著相关,其中北亚热带桃花观赏日期随纬度变化最快,纬度每升高1°,桃花观赏日期推迟2.56 d(表2)。

历年桃花观赏日期的线性回归分析表明,293个提取到桃花观赏日期的城市在2010~2018年期间,225个(76.8%)城市的桃花观赏日期提前,其中22个通过显著性检验,平均每年提前2.3 d,68个(23.2%)城市的桃花观赏日期推迟,其中深圳市通过显著性检验,每年推迟1 d。虽然同一气候区内不同城市间的桃花观赏日期变化存在差异,桃花观赏日期的提前和推迟均有出现,但是大部分气候区内桃花观赏日期均呈现出提前趋势(图5)。

表2 不同气候区桃花观赏日期线性拟合结果

Table 2 Linear fitting results of peach blossom viewing date in different climate zones

	斜率	截距	R值	P值	标准差	决定系数
高原气候区	0.537	67.907	0.258	0.444	0.670	0.066
中温带	1.180	44.535	0.309*	0.027	0.519	0.095
南温带	1.053	47.517	0.483**	0.000	0.216	0.234
北亚热带	2.563	0.371	0.831**	0.000	0.256	0.691
中亚热带	2.159	12.348	0.656**	0.000	0.308	0.431
南亚热带	1.415	23.622	0.214	0.203	1.091	0.046

注:**在 0.01 级别(双尾),相关性显著;*在 0.05 级别(双尾),相关性显著。

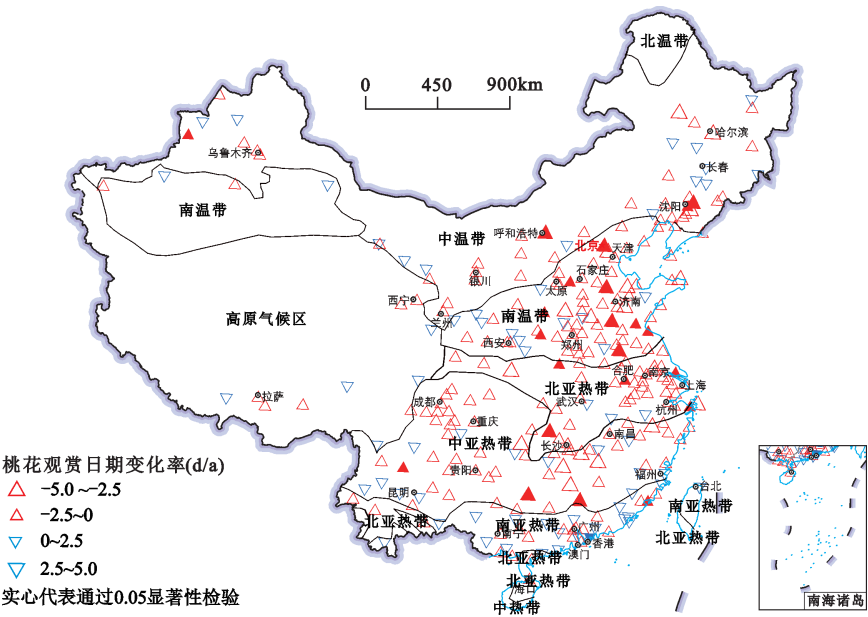


图5 桃花观赏日期时间格局

Fig.5 Time pattern of peach blossom viewing date

3 结论与讨论

3.1 结论

使用大数据技术从微博中提取桃花观赏日期的方法有效且可靠,使用该方法制作出了中国桃花观赏日期数据集,筛选出 843 034 条可采信微博,数据集包含的 293 个城市覆盖了中国大部分地区。同时,基于该方法提取制作的数据集通过了站点观测数据、气温数据以及物候模型的验证,是对现有物候观测数据的有效补充。

中国桃花观赏日期有显著的时空格局规律。在空间尺度上呈现明显的纬度地带性,不同气候区观赏日期的纬度敏感度存在差异,北亚热带桃花观赏日期对纬度变化最为敏感,纬度每提高 1° ,观赏日期推迟 2.56 d。在时间尺度上,2010~2018 年全国大部分城市(76.8%)桃花观赏日期提前。

3.2 讨论

基于微博大数据提取的桃花观赏日期数据反映的是城市尺度的契合赏花旅游需要的桃花物候信息,与物候观测站记录的植株尺度反映桃花生物学状态的始花期、盛花期以及末花期时间有所区别,同时与遥感技术监测到的地面宏观尺度反映植被生长状态的物候数据有所不同。虽然大数据技术具有覆盖面广、样本量大、时效性强的特质,但传统的站点观测和遥感监测仍旧非常重要,且二者在旅游研究中的应用也日益丰富^[24-26]。该数据集的价值在于能够从不同的尺度和视角为已有物候数据提供数据补充和参考,以期更加全面地掌握植被物候状态。同时可以用于进行桃花赏花日期的模拟预测,结合现有花期预测方法提升预测精度,以及为桃花赏花旅游活动的开展提供数据支撑。更为重要的是,该数据集通过了站点观测数据、气温数据以及物候模型模拟数据的验证,表明基于微博大数据提取桃花物候数据的方法有效且可靠,这就为重构中国植物观赏日期历史数据提供了方法上的参考。另一方面,由于该方法基于大量的来自群众的物候观测信息,因此开展物候观测全民科普教育活动,动员对物候观测感兴趣的人员或单位积极参与物候观测就显得尤为重要。同时需要进一步扩充使用该方法提取物候信息的数据来源,对更多种类的植物开展物候信息的提取,并不断优化数据筛选和清洗的程序和算法。

使用物候模型验证桃花观赏日期数据集的方法客观上存在不足。一方面所选模型由于仅考虑了温度这一个影响桃花花期的因子,没有综合考虑光照、水分条件等其他环境因子,模型本身存在 3~8 d 的误差^[23]。另一方面,该物候模型是基于 36 个站点的定点观测数据建立的,反映的是单个选定位置内桃花的花期状态,而基于微博大数据提取出的桃花观赏日期数据集则在城市尺度上对该城市内所有位置的桃花观赏日期记录求取平均值得出。在模型误差和数据尺度差异的共同影响下,仍有 63.2% 的样本通过验证,在没有找到更好的物候模型来进行数据验证之前,可以认为基于微博大数据提取的桃花观赏日期与物候模型模拟的日期基本一致。

观赏类植物的花期数据是赏花旅游活动组织和开展的重要支撑,能够为游客制定赏花旅游计划提供指导,为景区策划赏花活动提供参考,为周边社区参与旅游经营提供借鉴。同时旅游管理部门、景区管理者以及旅游从业人员应该注意到,有 23 个城市的桃花观赏日期在近 9 a 来已经明显提前或推迟,花期的变化将对赏花旅游活动产生直接影响。在全球气候变化的背景下,提高对桃花物候变化的认识,提前采取适应性措施进行应对,是减少物候变化对赏花旅游影响的重要举措。一方面,景区管理者需要提前关注当年的气候状况,结合历年桃花花期记录对当年桃花开花日期做出预估,并根据周末及节假日情况制定出赏花旅游活动的组织方案,同时做好应对花期提前或推迟的预备方案;另一方面,景区周边的旅游从业者需要时刻关注景区赏花旅游活动的开展进度,并根据桃花花期的变化以及赏花游客的消费习惯及时调整营销方案,在为游客提供优质服务的同时获取旅游收入。与此同时,旅游管理部门需要联合气象机构和科研院所,采用多种途径提升花期预报精度,并通过多种渠道和平台进行实时播报,为赏花旅游活动的组织和开展提供指导。

参考文献(References):

- [1] Root Terry L, Price Jeff T, Hall Kimberly R et al. Fingerprints of global warming on wild animals and plants[J]. *Nature*, 2003, 421(6918):57-60.
- [2] Wu Chaoyang, Wang Xiaoyue, Wang Huanjiong et al. Contrasting responses of autumn-leaf senescence to daytime and night-time warming[J]. *Nature Climate Change*, 2018, DOI:

- 10.1038/s41558-018-0346-z.
- [3] 神祥金,周道玮,李飞,等.中国草原区植被变化及其对气候变化的响应[J].地理科学, 2015, 35(5). [Shen Xiangjin, Zhou Daowei, Li Fei et al. Vegetation change and its response to climate change in grassland region of China. *Scientia Geographica Sinica*, 2015, 35(5): 622-629.]
- [4] 陆晴,吴绍洪,赵东升.1982~2013年青藏高原高寒草地覆盖变化及与气候之间的关系[J].地理科学,2017, 37(2):135-143. [Lu Qing, Wu Shaohong, Zhao Dongsheng. Variations in Alpine grassland cover and its correlation with climate variables on the Qinghai-Tibet Plateau in 1982-2013. *Scientia Geographica Sinica*, 2017, 37(2): 292-300.]
- [5] Tao Zexing, Ge Quansheng, Wang Huanjiong et al. Phenological basis of determining tourism seasons for ornamental plants in central and eastern China[J]. *Journal of Geographical Sciences*, 2015, 25(11):1343-1356.
- [6] 刘俊,李云云,刘浩龙,等.气候变化对成都桃花观赏旅游的影响与人类适应行为[J].地理研究,2016,35(3):504-512. [Liu Jun, Li Yunyun, Liu Haolong et al. Climate change and peach blossom viewing: Impact and adaptation. *Geographical Research*, 2016, 35(3):504-512.]
- [7] 高新月,戴君虎,张明庆.1965-2014年北京西郊地区植物观赏期对气候变化的响应[J].地理研究, 2018,37(12):2420-2432. [Gao Xinyue, Dai Junhu, Zhang Mingqing. Responses of variations of plant ornamental period to climate change in the west suburbs of Beijing from 1965-2014. *Geographical Research*, 2018, 37(12):2420-2432.]
- [8] Liu Jun, Cheng Han, Jiang Di et al. Impact of climate-related changes to the timing of autumn foliage colouration on tourism in Japan[J]. *Tourism Management*, 2019, 70:262-272.
- [9] Ge Quansheng, Dai Junhu, Liu Jun et al. The effect of climate change on the fall foliage vacation in China[J]. *Tourism Management*, 2013, 38(Complete):80-84.
- [10] 葛全胜,戴君虎,郑景云.物候学研究进展及中国现代物候学面临的挑战[J].中国科学院院刊, 2010, 25(3):310-316. [Ge Quansheng, Dai Junhu, Zheng Jingyun. The progress of Phenology studies and challenges to modern phenology research in China. *Bulletin of The Chinese Academy of Sciences*, 2010, 25(3): 310-316.]
- [11] Rein Ahas, Anto Aasa, Ülar Mark et al. Seasonal tourism spaces in Estonia: Case study with mobile positioning data[J]. *Tourism Management*, 2007, 28(3):898-910.
- [12] Sun Shaolong, Wang Shouyang, Wei Yunjie et al. Forecasting tourist arrivals with machine learning and internet search index [C]// IEEE International Conference on Big Data. Beijing: IEEE, 2018.
- [13] 新浪科技.微博月活跃用户增至4.62亿 年度营收破百亿[EB/OL].<https://tech.sina.com.cn/i/2019-03-05/doc-ihrfqzkc1446626.shtml>, 2019-03-05/2019-03-06. [Sina. Monthly active users of weibo increased to 462 million, and annual revenue exceeded 10 billion yuan. <https://tech.sina.com.cn/i/2019-03-05/doc-ihrfqzkc1446626.shtml>, 2019-03-05/2019-03-06.]
- [14] 张子昂,黄震方,靳诚,等.基于微博签到数据的景区旅游活动时空行为特征研究——以南京钟山风景名胜区为例[J].地理与地理信息科学, 2015, 31(4):121-126. [Zhang Zi'ang, Huang Zhenfang, Jin Cheng et al. Research on spatial-temporal characteristics of scenic tourist activity based on sina microblog: A case study of Nanjing Zhongshan Mountain National Park. *Geography and Geo-Information Science*, 2015, 31(4):121-126.]
- [15] 唐佳,李君轶.基于微博大数据的西安国内游客日内时间分布模式研究[J].人文地理, 2016, 31(3):151-160. [Tang Jia, Li Junyi. A study on the temporal distribution pattern of domestic tourists in Xi'an by hours based on the microblog big data. *Human Geography*, 2016, 31(3):151-160.]
- [16] 闫闪闪,梁留科,索志辉,等.基于大数据的洛阳市旅游流时空分布特征[J].经济地理, 2017, 37(8):216-224. [Yan Shanshan, Liang Liuke, Suo Zhihui et al. Big data based spatiotemporal distribution characters of tourism flow in Luoyang. *Economic Geography*, 2017, 37(8):216-224.]
- [17] 刘大均,胡静,程绍文,等.中国旅游微博空间分布格局及影响因素——以新浪旅游微博为例[J].地理科学, 2015, 35(6): 717-724. [Liu Dajun, Hu Jing, Cheng Shaowen et al. Spatial pattern and influencing factors of tourism micro-blogs in China: A case of tourism sina micro-blogs. *Scientia Geographica Sinica*, 2015, 35(6): 717-724.]
- [18] 葛全胜,郑景云,满志敏,等.过去2000a中国东部冬半年温度变化序列重建及初步分析[J].地学前缘, 2002, 9(1):169-181. [Ge Quansheng, Zheng Jingyun, Man zhimin et al. Reconstruction and analysis on the series of winter-half-year temperature changes over the past 2000 years in Eastern China. *Earth Science Frontiers*, 2002, 9(1):169-181.]
- [19] Zheng Jingyun, Ge Quansheng, Hao Zhixin et al. Spring phenophases in recent decades over eastern China and its possible link to climate changes[J]. *Climatic Change*, 2006, 77(3-4): 449-462.
- [20] 王焕炯,戴君虎,葛全胜.1952~2007年中国白蜡树春季物候时空变化分析[J].中国科学:地球科学, 2012, 42(5):701-710. [Wang H J, Dai J H, Ge Q S. The spatiotemporal characteristics of spring phenophase changes of *Fraxinus chinensis* in China from 1952 to 2007. *Science China: Earth Science*, 2012, 42(5): 701-705.]
- [21] Ge Quansheng, Wang Huanjiong, Dai Junhu. Simulating changes in the leaf unfolding time of 20 plant species in China over the twenty-first century[J]. *International Journal of Biometeorology*, 2014, 58(4):473-484.
- [22] Aono, Yasuyuki. Cherry blossom phenological data since the seventeenth century for Edo (Tokyo), Japan, and their application to estimation of March temperatures[J]. *International Journal of Biometeorology*, 2015, 59(4):427-434.
- [23] 仲舒颖,葛全胜,戴君虎,等.中国典型观赏植物花期模型建立及过去花期变化模拟[J].资源科学, 2017, 39(11):2116-2129. [Zhong Shuying, Ge Quansheng, Dai Junhu et al. Development

- of phenological models for simulating past flowering phenology of typical ornamental plants in China. *Resources Science*, 2017, 39(11):2116-2129.]
- [24] Ge Quansheng, Yang Xi, Qiao Zhi et al. Monitoring grassland tourist season of Inner Mongolia, China using remote sensing data[J]*Advances in Meteorology*,2014:1-5.
- [25] Yang Jun ,Sun Jing , Ge Quansheng, Li Xueming. Assessing the impacts of urbanization-associated green space on urban land surface temperature: A case study of Dalian, China[J]. *Urban Forestry & Urban Greening*,2017, 22:1-10.
- [26] 张爱英,王焕炯,戴君虎,等.物候模型在北京观赏植物开花期预测中的适用性[J].*应用气象学报*,2014,25(4):483-492. [Zhang Aiying, Wang Huanjiong, Dai Junhu et al.Applicability of phenological models in the prediction of flowering stage of ornamental plants in Beijing.*Journal of Applied Meteorological Science*,2014,25(4):483-492.]

Reconstruction of Peach Blossom-viewing Date of China Using Weibo Big Data

Liu Jun, Wang Shenghong, Jin Mengmeng, Li Ningxin

(*Tourism School, Sichuan University, Chengdu 610065, Sichuan, China*)

Abstract: The big data of Web site are important complementary sources for phenological site observation data. Using Sina Weibo big data, 2.95×10^6 peach blossom timing data were retrieved for 293 Chinese cities. Through data screening, space-time matching, and verification, 843 034 trusted microblogs were related to the peach blossom-viewing season; from those data, a peach blossom-viewing dataset was extracted and constructed for those cities in 2010-2018. The dataset was validated with respect to the following: whether it matched phenological site observation data; whether the peach blossom-viewing period was preceded by an increase in temperature; and whether the results matched the simulation produced using a phenological model. The results with the dataset were as follows: it was found to be in accordance with the change in the peach blossom period observed at phenological sites; the beginning of the peach blossom-viewing period was indeed preceded by an increase in temperature; and the findings were basically consistent with the peach blossom-viewing period simulated using the phenological model. Thus, the peach blossom-viewing dataset could be employed as an important supplement to observation data observed at phenological sites. Further analysis showed that for every 1° increase in latitude, the peach blossom-viewing date in China was delayed by 1.78 days. Among all the climate zones, the peach blossom-viewing date in the northern subtropics was most sensitive to latitude: for each increase in latitude, the peach blossom-viewing date was delayed by 2.56 days in that zone. Over the past 9 years, the peach blossom-viewing date with most Chinese cities has become earlier.

Key words: phenology; big data; peach blossom; flower viewing tourism