光温驱动下中国1979~2018年春季物候时空分异规律

吴晓静^{1,2,3,4}, 程昌秀^{1,2,3,4*}, 乔灿灿³, 宋长青^{1,3,4}

1. 北京师范大学地表过程与资源生态国家重点实验室, 北京 100875;
2. 北京师范大学环境演变与自然灾害教育部重点实验室, 北京 100875;
3. 北京师范大学地理科学学部, 北京 100875;
4. 北京师范大学地理数据与应用分析中心, 北京 100875

* 通讯作者, E-mail: chengcx@bnu.edu.cn

收稿日期: 2019-09-16; 收修改稿日期: 2019-12-17; 接受日期: 2020-01-10; 网络版发表日期: 2020-03-24

国家重点研发计划重点专项项目(编号: 2019YFA0606901)、国家自然科学基金项目(批准号: 41901317)和中国博士后科学基金项目(编号: 2018M641246)资助

摘要 生态和物候学研究领域积累的大数据迫切需要分析其中有用的规律和信息. 基于双向聚类方法, 文章使用光温驱动物候模型对长时间序列全国春季物候的时空分异规律开展研究. 首先, 基于中国区域高时空分辨率气象要素驱动数据集, 利用拓展春季指数物候模型获取1979~2018年全国开花始期数据集. 之后, 使用双向聚类和 k 均值算法分析开花始期数据集, 获得双向聚类结果. 本研究实现了全国范围内空间连续的开花始期物候区划. 研究结果显示三种遍布全国的开花始期空间格局及其在40年间(1979~2018年)的动态演化. 具体来说, 1979~1996年开花始期较晚, 江西、新疆北部和内蒙古中部开花始期变化最为频繁. 1997~2012年, 全国开花始期呈现持续提前趋势, 福建、湖南和黑龙江东部开花始期变化最为频繁. 2012年后, 全国开花始期呈现稳定较早状态. 此外, 研究结果同时显示出15种延续40年的开花始期时间格局及其在全国范围内的空间区划. 全国大部分区域处于持续40年开花始期不变状态, 而贵州北部、湖南和湖北南部1979~1997年持续不变而后出现波动. 文章的研究结果对中国物候观测网现有站点布设具有一定的指导意义.

关键词 开花始期, 双向聚类, 大数据, 时空分异, 光温驱动物候模型

1 引言

随着现代数据采集技术例如传感器等的迅速发展, 在生态学及相关环境科学领域生成了前所未有的海量数据(Kelling等, 2009; Zhu等, 2017; 程昌秀等, 2018; Li等, 2019). 传统的遥感卫星观测系统以及新兴

的近地面遥感观测平台能够实时收集时空分辨率各异的数据(La Salle等, 2016; 郭庆华等, 2018). 另外, 模型和实地测量技术也加快了从各种来源积累数据的速度(Chen等, 2011). 因此, 了解大数据特性并分析数据背后隐藏的信息成为大数据时代首要任务(Kelling等, 2009; Kisilevich等, 2010).

中文引用格式: 吴晓静, 程昌秀, 乔灿灿, 宋长青. 2020. 光温驱动下中国1979~2018年春季物候时空分异规律. 中国科学: 地球科学, 50(10): 1386–1400, doi: 10.1360/SSTe-2019-0212

英文引用格式: Wu X, Cheng C, Qiao C, Song C. 2020. Spatio-temporal differentiation of spring phenology in China driven by temperatures and photoperiod from 1979 to 2018. Science China Earth Sciences, 63(10): 1485–1498, <https://doi.org/10.1007/s11430-019-9577-5>

空间和时间分异现象是存在于生态物候学研究中的重要特性(Dutilleul, 2011; Wang等, 2016). 植物物候学是研究环境因子驱动下植物生命周期规律的科学, 监测并研究长时间序列植物物候动态演化对理解生态系统的变化特征有重要意义(竺可桢和宛敏渭, 1973; Schwartz等, 2006; 方修琦和陈发虎, 2015; Wang等, 2019). 作为反映气候变化最为敏感的生物指标, 春季植物物候在空间和时间上的分异性在大量研究中得到证实(White等, 2009; Gordo和Sanz, 2010; Jeong等, 2011; Wu和Liu, 2013). 例如, 春季物候的提早到来在北半球中高纬度相比其他区域较为明显, 并且在20世纪80和90年代相比其他时期较为显著(Piao等, 2006). 这些发现使得研究春季物候现象在空间上和时间上的分异规律成为必要.

春季物候资料来源主要有以下三种方法. 首先是地面物候观测. 实地物候观测站点通常会记录特定物种及其物候如展叶和开花等. 然而实地观测资料虽然时间跨度长但其空间覆盖度较差(White等, 2005). 另一种方法是使用遥感卫星图像推测春季开始日期. 然而, 截至目前尚未有被普遍认可的方法并且结果过于依赖传感器和分析方法(White等, 2009). 此外, 卫星数据自20世纪80年代才大量出现, 并不能满足物候对气候变化响应的长期研究需要. 第三种方法是物候模型. 在缺乏实地观测资料的地区, 物候模型是唯一能够预测和模拟春季物候的方法. 研究表明, 温度和光周期是影响植物物候的两个重要环境因素(Schwartz等, 2006; Gordo和Sanz, 2010; Piao等, 2015). 因此, 本文拟使用光温驱动物候模型(2.2节)模拟长时间尺度下的春季物候(Schwartz等, 2013).

作为研究物候时空分异规律的常用方法之一, 聚类将相似的数据要素划分为子集或聚簇, 从而在更高层次上分析事物现象(Andrienko等, 2009). 聚类方法在生态及相关环境学领域研究中尤其适用, 因为在大区域和长时间尺度下的聚类结果比短时间观察若干地点总结出的结果更有实用意义(Zirlewagen和von Wilpert, 2010). 目前已经有许多学者使用空间聚类方法(Has和Aasa, 2003; White等, 2005; Kumar等, 2011; Zhang等, 2012). 比如, Has和Aasa(2003)基于物候季节节奏使用时间聚类方法划分年份来研究春季物候的时间分异规律. Zhang等(2012)则结合主成分分析和k-means++方

法研究科罗拉多河上游流域物候现象的空间分异规律. 然而, 仅使用空间聚类方法获取的空间分异规律无法显示物候现象随时间变化的行为, 反之亦然(Deng等, 2011). 因而需要一种能够同时获取时空分异规律的聚类方法来显示物候现象随时空变化的行为. 双向聚类方法通过对物候现象同时进行空间聚类和时间聚类, 能够展现其随空间和时间的变化. 因此, 本文使用双向聚类方法研究物候时空分异规律(Wu等, 2015, 2016).

鉴于上述认识, 本研究基于双向聚类方法对春季物候现象中复杂的时空分异规律进行分析: 首先使用光温驱动物候模型和长时间序列气象数据集计算得到开花始期数据集, 然后使用双向聚类方法分析开花始期数据集并进一步对结果进行可视化分析用来展示春季物候的时空分异规律.

2 材料与方法

2.1 数据和计算平台

本文所用气象数据来自中国区域高时空分辨率地面气象要素驱动数据集(<http://westdc.westgis.ac.cn/data/7a35329c-c53f-4267-aa07-e0037d913a21>; Chen等, 2011). 该数据集空间覆盖范围为18.2°N~53.6°N和73.4°E~135°E, 持续时间为40年(1979~2018年), 其空间分辨率为0.1°×0.1°, 时间分辨率为3h. 本研究利用数据集中瞬时近地面(2m)气温计算得到日最高温和日最低温作为光温驱动物候模型的输入参数.

为了保证数据的完整性, 首先对日最高温和日最低温数据集进行检验, 剔除掉使用该数据集及物候模型计算得到开花始期空值大于20年的网格. 以此筛选出的气温数据集共74154个网格, 其空间范围如图1所示, 以任意选定一天(2018-02-01)日最高温空间分布为例.

用于检验物候模型模拟结果的物候观测数据来自中国物候观测网, 下载于国家地球系统科学数据共享服务平台(<http://www.geodata.cn>). 中国物候观测网由中国科学院组织管理, 共包括全国44个观测站点, 物候记录时间为1963~2008年(Ge等, 2014, 2015; Wang等, 2015a, 2015b)(图2). 因为紫丁香(*Lilac Syringa oblata*)在全国范围内被广泛观测(Schwartz和Chen, 2002), 本文选择该物种用于检验物候模型的适用性.

图1 气象数据集空间范围

以任意选定一天(2018-02-01)日最高温空间分布为例

图2 中国物候观测网的44个观测点分布图

另外紫丁香已经被成功地用于检验本文所使用物候模型的前一版本在中国的适用性(Schwartz等, 2006).

为了对全国高时空分辨率气温数据集以及长时间序列全国开花始期数据集进行高效处理, 本研究中实验均在高性能计算平台上并行计算中的3个刀片计算节点上运行. 其中每个计算节点配置两个Intel Xeon

Gold 6132 CPU(2.60GHz)、96GB的内存和240GB硬盘空间. 运行过程中观察到每个节点内有10个核心在进行并行计算.

2.2 拓展春季指数模型

拓展春季指数(Extended spring indices, SIx)模型是一系列基于回归分析的光温驱动物候模型, 被广泛地用于预测春季物候的开始(Schwartz等, 2013; Izquierdo-Verdiguier等, 2018). 该模型无需春季指数模型(SIx前版本)中对寒冷和温暖条件的要求, 因此适用于预测和模拟更广区域的春季物候(Schwartz等, 2013). SIx主要基于三个关键物种预测其展叶期(first leaf dates, FLD)和开花始期(first bloom dates, FBD), 分别为紫丁香(*Syringa chinensis*)、新疆忍冬(*Lonicera tatarica*)和蓝叶忍冬(*Lonicera Korolkowii*).

SIx模型的输入参数为研究地点的日最高温、日最低温和纬度, 其中纬度代表昼长来表示光周期变化(Basler和Körner, 2012). 因为开花始期指示性好并易于观察(陆佩玲, 2006), 本研究选择FBD作为春季物候指标. SIx模型开花始期(SIx FBD)是晚春到来的物候指标, 为上述三个物种第一朵花开放的平均日期, 与主要森林植被的大范围开始生长密切相关(Schwartz和Chen, 2002). SIx FBD的计算公式如下所示:

$$FBD_{ij} = SIx(TX_{ij}, TN_{ij}, lat_i), \quad (1)$$

$$i = 1, \dots, 74154; j = 1, \dots, 40$$

式中, TX为日最高温度, TN是日最低温度, lat为纬度. 索引*i*表示地点, 即研究区域的74154个网格, *j*表示时间, 即1979~2018年. 公式中TX与TN的单位均为华氏度. 关于SIx的详细信息请参考文献(Ault等, 2015).

2.3 双向聚类分析

空间聚类和时间聚类均为单向聚类分析. 如图3a中所示开花始期示例数据集, 空间聚类(图3b)将数据集中的地点和时间分别看作对象及其属性, 聚类结果是在所有时间上均具有相似值的地点聚簇, 例如地点聚簇1(图3b中加粗矩形). 而时间聚类(图3c)则将数据集中的时间和地点分别看作对象及其属性, 聚类结果是在所有地点上均具有相似值的时间聚簇, 例如时间聚簇1(图3c中加粗矩形). 一般可用于空间聚类的方法也适用于时间聚类, 例如*k*-means.

图3 对开花始期示例数据集进行空间/时间聚类分析

(a) 开花始期示例数据集; (b) 空间聚类分析得到在所有时间上均具有相似值的地点簇, 例如地点簇1(加粗矩形); (c) 时间聚类分析得到在所有地点上均具有相似值的时间簇, 例如时间簇1(加粗矩形); (d) 双向聚类方法分析得到在任意地点和时间上均具有相似值的时空簇, 例如地点簇1和时间簇1的交叉(加粗矩形)

和单向聚类不同, 双向聚类(图3d)能够同时对地点和时间进行划分(Dhillon等, 2003; Han等, 2012; Shen等, 2018). 通过将数据集中的地点和时间同时划分到地点簇和时间簇, 双向聚类得到数据元素相似的时空双向簇, 也就是每个地点簇和时间簇的交叉. 在双向簇内部任意地点和时间上的值均相似, 例如地点簇1和时间簇1的交叉(图3d中加粗矩形).

国内外已经有很多学者使用双向聚类方法在众多领域开展研究(Dhillon等, 2003; Banerjee等, 2007; Wu等, 2015, 2016). Dhillon等(2003)提出信息论双向聚类算法(Information Theoretic Co-clustering algorithm, ITCC)用于文档分析, 该算法使用信息散度作为度量标准优化数据集中的双向簇. Banerjee等(2007)提出布莱格曼双向聚类算法(Bregman co-clustering algorithm), 该算法使用信息散度和欧氏距离等多种度量标准优化双向簇, 其中信息散度的优越性在文档分析实验中得到证实. 最近, Wu等(2015, 2016)均使用布雷格曼块平均双向聚类算法(Bregman block average co-clustering algorithm with I-divergence, BBAC_I)分析时空数据集. 作为布雷格曼双向聚类算法中的一种, BBAC_I能够有效地识别双向簇从而展现出数据集中的时空格局及其动态演化. 因此, 本文使用该算法分析开花始期数据集中的时空格局并进一步展现春季物

候的时空分异规律.

BBAC_I一般用于分析元素是实值且为正数的数据矩阵, 该矩阵表示两个随机变量间的共生矩阵或联合概率(Wu等, 2015). 该算法把双向聚类问题看作信息论中的优化问题, 而最佳双向聚类结果则使原始数据矩阵和双向聚类后矩阵间的互信息量丢失最少. 将开花始期数据集组织为一个空间变量和一个时间变量间的共生矩阵—— O_{FBD} , 其中空间和时间变量分别在74154个网格和40年取值, 即 O_{FBD} 是一个大小为74154×40的矩阵. 以下分为3步对BBAC_I将开花始期数据矩阵优化为双向簇的过程进行阐述(图4).

(i) 随机初始化. 即将74154个网格和40年随机地划分到 g_c 个网格簇和 y_c 个年簇, 并由此产生双向聚类后的数据矩阵, $\bar{O}_{FBD}$, 其大小为 $g_c \times y_c$.

(ii) 计算互信息量丢失. 建立目标函数用于计算原始矩阵与双向聚类后矩阵间的互信息量的丢失:

$$f_{\text{loss}} = D_I(O_{FBD} \| \bar{O}_{FBD}), \quad (2)$$

式中, $D_I(\cdot \| \cdot)$ 表示两矩阵间的信息散度, 可用于计算两矩阵间的互信息量丢失.

(iii) 迭代以优化目标函数. 即进行迭代用于更新网格和年簇的划分, 并使互信息量的丢失达到极小值. 这一步又分为两个子步骤:

(1) 更新网络聚簇划分. 根据目标函数找出使互信

布雷格曼块平均双向聚类算法(BBAC_I)	
输入:	O_{FBD} , gc (网格聚簇数目), yc (年聚簇数目)
输出:	最优双向聚簇
开始	
1.	初始化: 将74154个网格和40年随机划分到 gc 个网格聚簇和 yc 个年聚簇;
2.	计算互信息量丢失函数 $f_{loss} = D_1(O_{FBD} \hat{O}_{FBD})$
3.	迭代:
3.1	更新网格聚簇划分 $i = \arg \min_{i \in \{1, \dots, gc\}} D_1(O_{FBD} \hat{O}_{FBD})$
3.2	更新年聚簇划分 $j = \arg \min_{j \in \{1, \dots, yc\}} D_1(O_{FBD} \hat{O}_{FBD})$
结束	

图4 布雷格曼块平均双向聚类算法

息量丢失最小的网络聚簇划分:

$$i = \arg \min_{i \in \{1, \dots, gc\}} D_1(O_{FBD} || \hat{O}_{FBD}). \quad (3)$$

(2) 更新年聚簇划分. 根据目标函数找出使互信息量丢失最小的年聚簇划分:

$$j = \arg \min_{j \in \{1, \dots, yc\}} D_1(O_{FBD} || \hat{O}_{FBD}). \quad (4)$$

研究表明, 每次迭代后互信息量丢失均会减少(Banerjee等, 2007). 当丢失函数收敛到局部最小值, 即丢失的变化低于设定阈值时, 迭代过程停止. 因为无法确保全局最优, 通常将上述过程重复若干次以便能最大程度上找到最优双向聚类结果. 关于BBAC_I的详细优化过程请参考文献(Wu等, 2015).

因为在上述过程中BBAC_I对整行和整列进行划分, 双向聚类后矩阵通常呈现相邻聚簇相似的棋盘格式(Wu等, 2016). 为了对数据集中的时空分异规律进行彻底分析, 本研究在进行双向聚类分析时首先定义较大的双向聚簇数目, 然后再使用 k 均值算法将双向聚簇进一步划分为仅内部元素相似的 k 个与坐标轴平行但非矩形形状的不规则双向聚簇. k 均值的输入参数为每个双向聚簇中开花始期平均值, 聚簇数目通过轮廓法确定(Rousseeuw, 1987). 另外, 双向聚类和 k 均值算法均为局部最优化算法, 因此需多次运行以获得稳定

和最优化结果. 最后, 我们使用地理探测器的 q 值来评价双向聚类结果(王劲峰和徐成东, 2017).

2.4 春季物候时空分异规律

为了展现开花始期数据集中的时空分异规律, 需要使用并列地图(small multiples)¹⁾、线性时间线(linear timeline)²⁾以及地图对不规则双向聚簇进行可视化分析. 首先, 将双向聚簇结果中落到相同年份的春季物候区, 即开花始期相似的区域, 进行合并组成多种空间格局. 并列地图中的每一幅地图用于对不同的空间格局进行可视化. 而线性时间线则用于可视化研究时段内多种空间格局的动态演化. 然后, 将双向聚簇结果中在开花始期变化相同的研究时段进行特征提取并按照时间顺序排列得到多种时间格局. 一组线性时间线中的每一个时间线表示一种时间格局. 而地图则用于可视化具有相同时间格局的网格, 即有相同开花始期变化特征的空间区划.

3 结果

3.1 开花始期

基于筛选后的气温数据集, 使用SIx模型模拟得到全国74154个网格40年(1979~2018年)的开花始期数据集. 其空间分布如图5所示(以2018年为例), 其中紫色越深的区域开花始期越早. 从图中可以观察到全国的开花始期从南到北呈现逐渐推迟的变化趋势. 南部省份如云南、广西和四川南部的开花始期发生在三月份初期; 中部地区如山东和河南的开花始期发生在四月份; 东三省开花始期较晚, 一般在五月份; 西藏地区开花始期最晚, 一般在六月份和以后. 值得注意的是, 虽然开花始期数据集(图5)和日最高温数据集(图1)的空间变化趋势相似, 但二者在本质上有所区别. 这种区别不仅在于前者展示的是某一年的春季物候而后者展示的是该年某一天的日最高温, 更重要的是SIx模型考虑了积温的非线性特征(Schwartz等, 2013).

使用中国物候观测网紫丁香开花始期观测记录检验SIx模型模拟的开花始期数据集(SIx FBD). 按照中国物候观测网的观测标准(宛敏渭和刘秀珍, 1979), 开

1) 并列地图是将一系列地图并排放在一起的可视化方法, 主要用于帮助了解多变量信息的变化特征

2) 线性时间线是时间线的一种, 即使用一条直线用于可视化按照线性时间(比如年等)顺序排列的事件

图5 SIx模型开花始期数据集空间分布(以2018为例)

花始期是指在选定的同种寄住植物上, 看到一半以上的植株有一朵或同时有3朵以上花的花瓣开始完全开放的日期. 本研究收集了270条紫丁香开花始期观测记录, 并基于观测地点和年份利用地理坐标和年份提取对应的SIx模型模拟的开花始期数据. 然后绘制散点图用于对比模型模拟结果和实测记录(图6). 结果表明, SIx开花始期日期和紫丁香观测记录开花始期日期具有强相关性($r=0.9543$, $p<0.01$).

3.2 双向聚类结果

基于上述结果, 将开花始期数据集组织为数据矩阵, 其中行代表74154个网格, 列代表40年(1979~2018年). 然后使用BBAC_I和 k 均值算法进行分析. 为确定网格聚簇和年聚簇数目, 分别使用从300到700间隔为50和从3到7间隔为1的数目进行实验. 结果表明, 当网格和年聚簇数目分别为495和3时, 原始矩阵和双向聚类后矩阵间的互信息量丢失最小. 实验结果还表明, 当迭代过程重复次数、迭代次数和阈值分别设置为200、3000和 10^{-6} 时, BBAC_I丢失函数能够收敛到最小值并取得稳定结果. 使用轮廓法确定 k 均值中不规则聚簇数目为5. 最终评价聚类结果的地理探测器 q 值为0.9141, 说明双向聚类结果有很好的分异性(董玉祥等, 2017).

双向聚类分析将开花始期数据集划分为五类. 为描述每类特征, 将其按照开花始期序日值分别命名为

图6 使用物候观测站开花始期观测记录检验拓展春季指数物候模型

“特早”“较早”“居中”“较晚”和“特晚”开花始期. 五类开花始期对应日期为: 二月末(第58序日)、四月初(第102序日)、五月初(第124序日)、五月末(第148序日)和第199序日. 上述五类开花始期如图7所示. x 坐标轴表示数据集中的40年(1979~2018年), 按照属于相同年聚簇的顺序从右到左进行升序排列. y 坐标轴表示数据集中的74154个网格, 按照属于相同网格聚簇的顺序从下到上进行升序排列. 因为双向聚类方法按照整行(网格)或整列(年)的平均值进行排列, 图中不同开花始期有所嵌套, 比如某行网格在40年间的平均值高于另一行, 但是在某些年份较低. 每类开花始期的空间分布如图8所示, 因为在不同的年聚簇中空间范围不同, 每类开花始期均有三个春季物候区.

从图7可以看出, “居中”开花始期占据数据集最大比例, 其次是“较早”开花始期; “特晚”开花始期占比最小, 次小为“特早”开花始期. 对比显示, “特早”开花始期中占据最大比例的年份为近期年份. 而“特早”开花始期中占据最小比例和“特晚”开花始期中占据最大比例年份均在2000年以前, 属于研究时段的前半段. 结合图7和图8可以看出, “特早”开花始期大多发生在中国南部区域, 例如海南、广东和四川南部. 纵观整个研究时段, 该类别近年来有向北覆盖更多区域的趋势; “较早”开花始期大多发生在中东部地区, 例如江苏和安徽省; “居中”开花始期多发生在中部和北部, 例如宁夏和东三省; “较晚”开花始期发生在中国北部省份; 而“特晚”开花始期主要发生在青藏高原地区.

图7 双向聚类分析将开花始期数据

数据集划分为五类, 分别命名为“特早”、“较早”、“居中”、“较晚”和“特晚”开花始期。x坐标轴表示数据集中的40年(1979~2018), 按照属于相同年聚簇的顺序进行排列。y坐标轴表示数据集中的74154个网格, 按照属于相同网格聚簇的顺序进行排列

3.3 开花始期的时空分异规律

对上述5类开花始期进行进一步可视化分析可显示出开花始期数据集中多种空间格局及其动态演化, 并可同时显示出多种时间格局及其空间分布, 由此展示春季物候的时空分异规律。

3.3.1 空间格局及其动态演化

将图7中整个研究区域上变化相同的网格(即落到相同年聚簇的开花始期物候区)进行合并后得到三种遍布全国的空间格局, 分别命名为空间格局1到3。三种空间格局如图9左侧的并列地图所示, 1979~2018年空间格局的动态演化如图9右侧的线性时间线所示。各空间格局中每类开花始期物候区占比情况及所占年份数目如表1所示。

在空间格局3中, “特早”开花始期占比最大, “较

晚”和“特晚”开花始期则占比最小; 相反在空间格局1中, “特早”开花始期占比最小, 而“较晚”和“特晚”开花始期占比最大。据此, 在空间格局1到3的变化中, 所属“特早”开花始期的物候区占比上涨7.53%, 新增地区主要集中在贵州、湖南、江西、福建和四川东部。同时, 所属“较晚”和“特晚”开花始期的物候区占比分别下降6.41%和1.56%, 缩减地区主要集中在新疆北部、内蒙古中部和黑龙江南部。因此该变化表明开花始期呈现持续提前趋势。从空间格局1到2的变化中, 所属“特早”开花始期的物候区占比上升2.76%, 主要分布在江西省, 所属“较晚”和“特晚”的物候区占比分别下降1.9%和1.09%, 主要在新疆北部和内蒙古中部, 该变化也表明了开花始期呈现提前趋势。相反地, 在空间格局3到2的变化中, 所属“特早”开花始期的物候区占比下降4.59%, 主要在湖南和福建, 而所属“较晚”开花始期物候区占比扩大4.51%, 主要在黑龙江东部, 该变化则表明开花始期呈现推迟趋势。

三种空间格局在研究时段的动态演化显示近年来开花始期呈现整体提前的趋势。纵观整个研究时段(1979~2018年), 空间格局3占比近半(16/40)且集中在1997年后的近期年份。而空间格局1占比不到1/4(9/40)且集中在1997年之前。如此一来, 开花始期在1979~1996年、1997~2012年以及2012年后三个时期呈现出不同的变化。其中, 1979~1996年开花始期在空间格局1和2间交替变换, 表明全国开花始期较晚, 在此时期江西、新疆北部和内蒙古中部的开花始期变化最为频繁。1997~2012年开花始期在空间格局2和3间交替变换, 表明开花始期到来呈现持续提前趋势, 在此时期福建、湖南和黑龙江东部的开花始期变化最为频繁。2012年后集中在空间格局3, 表明全国开花始期均较早且呈现稳定状态。此外, 最剧烈的开花始期变化发生在1996~1998年, 在此时期贵州、湖南、江西、福建和四川省东部, 以及新疆北部、内蒙古中部和黑龙江南部地区开花始期均呈现持续提前趋势。

表1 各空间格局中每类开花始期物候区占比情况及所占年份数目

空间格局	“特早”开花始期	“较早”开花始期	“居中”开花始期	“较晚”开花始期	“特晚”开花始期	所占年份数目
空间格局1	11.42%	23.65%	25.62%	25.41%	13.90%	9
空间格局2	14.18%	23.57%	25.93%	23.51%	12.81%	15
空间格局3	18.77%	23.22%	26.67%	19.00%	12.34%	16

图 8 每类开花始期的三个生态物候区

图 9 开花始期数据集中三种空间格局及其动态演化

3.3.2 时间格局及其空间分布

将图7中开花始期变化相同的研究区进行特征提

取并排列后得到15种时间格局, 如图10a所示. 根据开花始期变化状态可将15种时间格局分为5类: 平稳(第一排)、先平稳后波动(第二排)、先波动后平稳(第三

(a) 时间格局

(b) 空间区划

图 10 开花始期数据集中15种时间格局(a)及其空间区划(b)

排)、频繁波动(第四排第一列)和剧烈波动(第四排第二列)。每种时间格局的空间分布如图10b中地图所示,其中每类开花始期变化状态使用不同颜色表示。空间区划地图中每类开花始期变化状态及其包含的时间格局占比情况如表2所示。

在5类开花始期变化状态中,平稳状态占比最大(超过75%),其分布范围最广;先平稳后波动状态次之;剧烈波动状态占比最小(低于0.02%)。平稳开花始期变化状态基本不受气候变化影响,其中占比最大的时间格局为持续40年“较晚”开花始期(18.58%),主要分布

表 2 空间区划中每类开花始期变化状态以及每种时间格局占比情况

开花始期变化状态	时间格局	比例	比例总和
平稳	持续特晚	12.67%	75.06%
	持续较晚	18.58%	
	持续居中	17.14%	
	持续较早	13.49%	
	持续特早	13.18%	
先平稳后波动	平稳: 特晚 波动: 特晚&较晚	0.72%	14.17%
	平稳: 较晚 波动: 较晚&居中	4.94%	
	平稳: 居中 波动: 居中&较早	3.23%	
	平稳: 较早 波动: 较早&特早	5.28%	
	波动: 特晚&较晚 平稳: 较晚	0.88%	
	波动: 较晚&居中 平稳: 居中	2.96%	
先波动后平稳	波动: 居中&较早 平稳: 较早	2.17%	9.19%
	波动: 较早&特早 平稳: 特早	3.18%	
	频繁波动	0.23%	
	剧烈波动	0.0186%	

在内蒙古东部和黑龙江; 占比次之的时间格局为持续40年“居中”开花始期, 主要分布在辽宁和吉林中西部; 接下来是持续40年“较早”和“特早”开花始期, 前者主要分布在河北南部、山东西部、河南、安徽和江苏, 而后者分布在云南南部、广西、广东和福建南部; 占比最小的时间格局为持续40年“特晚”开花始期, 主要分布在青藏高原、四川西部和青海东部。

先平稳后波动和先波动后平稳状态受气候变化影响较小。在先平稳后波动状态中, 占比最大(5.28%)的时间格局为1979~1997年持续“较早”开花始期而后在“较早”和“特早”开花始期间交替变换, 其主要分布在贵州北部、湖南以及湖北南部; 占比其次的时间格局为1979~1997年“较晚”开花始期而后在“较晚”和“居中”开花始期间交替变换, 其主要分布在黑龙江南部和内蒙古中部部分地区。在先波动后平稳状态中, 占比最大的时间格局为1979~1997年在“较早”和“特早”开花始期间交替变换而后持续“特早”开花始期(3.18%), 其主要分布在四川东部部分地区、湖南东南部和江西

北部。

频繁波动和剧烈波动开花始期变化状态受气候变化影响较大, 但其占比极小, 分布范围很窄。其中前者40年间在“特晚”和“较晚”开花始期间频繁交替变换, 主要分布在西藏东部和西部、新疆中部和内蒙古东北部的少数网格区域内。后者1979~1997年在“较晚”和“居中”开花始期间、1998~2018年在“居中”和“较早”开花始期间频繁交替变换, 分布在西藏东北部和台湾东部地区的极少数网格区域内。

4 讨论

目前已经有很多学者利用遥感地面物候和物候观测资料研究全国春季物候分异规律。例如, Ge等(2015)使用1960~2011年全国145个物候站点观测记录开展荟萃分析(meta-analysis), 结果表明大部分记录显示春季提前到来的趋势。本文研究结果同样显示出该现象的发生(图9)。另外, Cong等(2013)使用Polyfit方法对

1982~2010年全国温带地区AVHRR NDVI数据集进行分析,发现内蒙古地区的春季物候普遍提前.这与本文研究结果基本一致,1979~1996年全国开花始期从空间格局1变化为空间格局2(图9),内蒙古中部开花始期呈现提前趋势. Wang等(2015b)利用1963~2012年全国22个物候站点观测记录对春季物候空间格局进行分析,结果发现中国西南部地区物候变化最大.本文研究结果发现同样现象,1979~1996年和1997~2012年全国空间格局交替变换均导致中国西南部地区例如贵州、湖南、江西和四川开花始期变化频繁(图9).此外,很多研究发现在20世纪90年代中期或者后期春季物候变化出现拐点.例如, Peng等(2011)利用1982~2012年全国NOAA/AVHRR NDVI数据集分析春季物候的时空分异,结果发现春季物候1982~1998年呈现显著提前趋势而1998~2010年提前趋势小于上一个始期.这与本文研究结果几乎一致.本文结果发现拐点出现在1997年,两者结果中的差异可能因为本文研究的是开花始期. Ge等(2016)使用1982~2012年全国AVHRR NDVI数据集分析春季物候格局,结果发现1982~1995年春季到来较晚而1996~2008年春季到来较早.本文研究结果表明1979~1996年开花始期整体较晚而在1997~2012年后开花始期整体到来较早.两者结果中拐点的差异可能与本文中较长研究时段(1979~2018年)有关.上述结果均证实,使用SIx模型和双向聚类方法能够有效地分析开花始期的空间格局及其动态演化和时间格局及其空间分布,并展现春季物候的时空分异规律.

与上述研究相比较,本文通过使用SIx模型扩大了研究区域并延长研究时段.通过这种方式,本文首次实现了覆盖全国范围的空间上连续的开花始期物候区划,而在此之前已有相关研究划分欧洲和美国物候区(Zhang等, 2012; Wu等, 2016).和空间上不均匀分布的气象或物候观测站点相比较,空间上连续的生态物候区能够提供更加详细的信息(Kumar等, 2011; Wu等, 2016).此外,通过使用基于双向聚类的方法分析全国春季物候,本研究成果显示出全国多种开花始期空间格局及其在研究时段内的动态演化.进一步分析不仅显示出一定时期内开花始期变化最频繁的区域,还发现物候变化最剧烈的区间及其发生地点;并且研究结果还同时显示出全国多种开花始期时间格局及其在研究区域上的空间分布.

另外,本文研究结果还可以用于评价和建议中国物候观测网现有站点的布设.目前在其他国家已经开展了评估生态监测网站点布设的研究,例如美国等(Hoffman等, 2013).通过将物候观测网44个观测站点叠加到开花始期时间格局的空间区划地图上,可以看到现有站点在各类开花始期变化状态的分布情况(图11).现有44个观测站点中将近60%(26个站点)属于平稳开花始期变化状态,比如黑龙江嫩江、辽宁沈阳、山东德州和广东佛山等;其他站点均属于先平稳后波动开花始期变化状态,比如黑龙江佳木斯、湖北武汉和湖南常德,或者先波动后平稳开花始期变化状态,比如四川仁寿和江西南昌;且西部地区仅有极少数站点.因此,结合各类开花始期变化状态的特征(图10a),本文建议处于先平稳后波动以及先波动后平稳开花始期变化状态的地区,如江西、湖南、湖北、重庆、四川东北部和黑龙江南部增设开花始期观测站点.同时在全国西部地区增设站点,比如新疆北部和青海,所增设观测站点还应考虑离城市的远近(Chu等, 2017).此外,上述处于波动的地区还应加强对农林业的管理(Ruml和Vulić, 2005).

关于本文中分析得到的多种春季物候空间格局和时间格局的驱动因素,我们认为主要是日最低/高温和光周期.本文使用拓展春季指数模型来模拟开花始期数据集.该模型的输入参数为日最低温、日最高温和代表昼长,用来表示光周期变化的纬度(Schwartz等, 2013).但是我们会在其他后续研究中考虑气候带、高

图11 开花始期数据集空间格局和中国物候观测网现有站点布设

程带、季风带、降水及其交互作用对春季物候时空格局的影响(方修琦和陈发虎, 2015).

5 结论

本文基于双向聚类方法使用光温驱动物候模型研究了长时间序列全国春季物候的时空分异规律. 首先, 基于中国区域高时空分辨率气象要素驱动数据集, 使用SIx模型获取1979~2018年全国开花始期数据集, 并使用中国物候观测网资料进行检验. 之后, 使用双向聚类法和 k 均值算法对开花始期数据集进行分析得到开花始期在空间和时间维度上均相似的不规则双向聚类簇. 最后, 对双向聚类结果进行可视化分析展示春季物候的时空分异规律.

本研究基于双向聚类分析首次实现全国空间上连续的开花始期物候区划. 分析结果显示三种遍布全国的开花始期空间格局及其在40年间的(1979~2018年)动态演化. 具体来说, 开花始期在1979~1996年、1997~2012年以及2012年后三个时期呈现出不同的变化. 1979~1996年全国开花始期普遍较晚, 在此期间江西、新疆北部和内蒙古中部开花始期变化频繁; 1997~2012年开花始期呈现持续提前趋势, 在此期间福建、湖南和黑龙江东部开花始期变化频繁; 2012年后开花始期呈现稳定较早状态. 其中开花始期变化最剧烈时期发生在1996~1998年, 此期间内中国西南和东北区域开花始期均呈现提前趋势. 分析结果还同时显示出15种延续40年的开花始期时间格局及其在全国范围内的空间区划. 具体来说, 全国大部分地区处于平稳状态, 即40年间持续同一种开花始期; 而贵州北部、湖南、湖北南部、黑龙江南部和内蒙古中部部分地区1979~1997年处于平稳而后出现波动.

致谢 北京师范大学地理科学学部地理数据与应用分析中心提供地学高性能计算平台(<https://gda.bnu.edu.cn/>), 中国科学院青藏高原研究所提供气象数据(<http://westdcwestgis.ac.cn>), 国家科技基础条件平台——国家地球系统科学数据共享平台(<http://www.geodata.cn>)提供物候数据, 在此一并致谢. 此外, 感谢Prof. Raul Zurita-Milla对SIx模型的帮助, 感谢朱文泉教授和黄咏梅教授对本工作提出的宝贵意见.

参考文献

- 程昌秀, 史培军, 宋长青, 高剑波. 2018. 地理大数据为地理复杂性研究提供新机遇. *地理学报*, 73: 5–14
- 董玉祥, 徐茜, 杨忍, 徐成东, 王钰莹. 2017. 基于地理探测器的中国陆地热带北界探讨. *地理学报*, 72: 135–147
- 方修琦, 陈发虎. 2015. 植物物候与气候变化. *中国科学: 地球科学*, 45: 707–708
- 郭庆华, 胡天宇, 姜媛茜, 金时超, 王瑞, 关宏灿, 杨秋丽, 李玉美, 吴芳芳, 翟秋萍, 刘瑾, 苏艳军. 2018. 遥感在生物多样性研究中的应用进展. *生物多样性*, 26: 9–26
- 陆佩玲. 2006. 中国木本植物物候对气候变化的响应研究. 博士学位论文. 北京: 北京林业大学
- 宛敏渭, 刘秀珍. 1979. 中国物候观测方法. 北京: 科学出版社
- 王劲峰, 徐成东. 2017. 地理探测器: 原理与展望. *地理学报*, 72: 116–134
- 竺可桢, 宛敏渭. 1973. 物候学. 北京: 科学出版社
- Ahas R, Aasa A. 2003. Developing comparative phenological calendars. In: Schwartz M D, ed. *Phenology: An Integrative Environmental Science*. Springer Netherlands. 39: 301–318
- Andrienko G, Andrienko N, Rinzivillo S, Nanni M, Pedreschi D, Giannotti F. 2009. Interactive visual clustering of large collections of trajectories. Atlantic City: IEEE Symposium on Visual Analytics Science and Technology (VAST)
- Ault T R, Zurita-Milla R, Schwartz M D. 2015. A Matlab® toolbox for calculating spring indices from daily meteorological data. *Comput Geosci*, 83: 46–53
- Banerjee A, Dhillon I, Ghosh J, Merugu S, Modha D S. 2007. A generalized maximum entropy approach to bregman co-clustering and matrix approximation. *J Mach Learn Res*, 8: 1919–1986
- Basler D, Körner C. 2012. Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agric For Meteorol*, 165: 73–81
- Chen Y Y, Yang K, He J, Qin J, Shi J C, Du J Y, He Q. 2011. Improving land surface temperature modeling for dry land of China. *J Geophys Res*, 116: D20104
- Chu H, Baldocchi D D, John R, Wolf S, Reichstein M. 2017. Fluxes all of the time? A primer on the temporal representativeness of FLUXNET. *J Geophys Res-Biogeosci*, 122: 289–307
- Cong N, Wang T, Nan H J, Ma Y C, Wang X H, Myneni R B, Piao S L. 2013. Changes in satellite-derived spring vegetation green-up date and its linkage to climate in China from 1982 to 2010: A multimethod analysis. *Glob Change Biol*, 19: 881–891
- Deng M, Liu Q L, Wang J Q, Shi Y. 2011. A general method of spatio-temporal clustering analysis. *Sci China Inf Sci*, 56: 1–14
- Dhillon I S, Mallela S, Modha D S. 2003. Information-theoretic co-clustering. Washington D C: The 9th International Conference on

- Knowledge Discovery and Data Mining (KDD). 89–98
- Dutilleul P. 2011. Spatio-Temporal Heterogeneity: Concepts and Analyses. Cambridge University Press
- Ge Q S, Dai J H, Cui H J, Wang H J. 2016. Spatiotemporal variability in start and end of growing season in China related to climate variability. *Remote Sens*, 8: 433
- Ge Q S, Wang H J, Zheng J Y, Rutishauser T, Dai J H. 2014. A 170 year spring phenology index of plants in eastern China. *J Geophys Res-Biogeosci*, 119: 301–311
- Ge Q S, Wang H J, Rutishauser T, Dai J H. 2015. Phenological response to climate change in China: A meta-analysis. *Glob Change Biol*, 21: 265–274
- Gordo O, Sanz J J. 2010. Impact of climate change on plant phenology in Mediterranean ecosystems. *Glob Change Biol*, 16: 1082–1106
- Han J W, Kamber M, Pei J. 2012. Data Mining Concepts and Techniques. Waltham: Morgan Kaufman MIT Press
- Hoffman F M, Kumar J, Mills R T, Hargrove W W. 2013. Representativeness-based sampling network design for the State of Alaska. *Landscape Ecol*, 28: 1567–1586
- Izquierdo-Verdiguier E, Zurita-Milla R, Ault T R, Schwartz M D. 2018. Development and analysis of spring plant phenology products: 36 years of 1-km grids over the conterminous US. *Agric For Meteorol*, 262: 34–41
- Jeong S J, Ho C H, Gim H J, Brown M E. 2011. Phenology shifts at start vs. end of growing season in temperate vegetation over the Northern Hemisphere for the period 1982–2008. *Glob Change Biol*, 17: 2385–2399
- Kelling S, Hochachka W M, Fink D, Riedewald M, Caruana R, Ballard G, Hooker G. 2009. Data-intensive science: A new paradigm for biodiversity studies. *BioScience*, 59: 613–620
- Kisilevich S, Mansmann F, Nanni M, Rinzivillo S. 2010. Spatio-temporal clustering. In: Maimon O, Rokach L, eds. Data Mining and Knowledge Discovery Handbook. Boston: Springer. 855–874
- Kumar J, Mills R T, Hoffman F M, Hargrove W W. 2011. Parallel k -means clustering for quantitative ecoregion delineation using large data sets. *Procedia Comput Sci*, 4: 1602–1611
- La Salle J, Williams K J, Moritz C. 2016. Biodiversity analysis in the digital era. *Phil Trans R Soc B*, 371: 20150337
- Li X, Zhao N, Jin R, Liu S M, Sun X M, Wen X F, Wu D X, Zhou Y, Guo J W, Chen S P. 2019. Internet of Things to network smart devices for ecosystem monitoring. *Chin Sci Bull*, 64: 1234–1245
- Piao S L, Fang J Y, Zhou L M, Ciais P, Zhu B. 2006. Variations in satellite-derived phenology in China's temperate vegetation. *Glob Change Biol*, 12: 672–685
- Piao S L, Tan J G, Chen A P, Fu Y S, Ciais P, Liu Q, Janssens I A, Vicca S, Zeng Z Z, Jeong S J. 2015. Leaf onset in the northern hemisphere triggered by daytime temperature. *Nat Commun*, 6: 6911
- Peng S S, Chen A P, Xu L, Cao C X, Fang J Y, Myneni R B, Pinzon J E, Tucker C J, Piao S L. 2011. Recent change of vegetation growth trend in China. *Environ Res Lett*, 6: 27–44
- Rousseeuw P J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*, 20: 53–65
- Ruml M, Vulić T. 2005. Importance of phenological observations and predictions in agriculture. *J Agric Sci BGD*, 50: 217–225
- Schwartz M D, Ahas R, Aasa A. 2006. Onset of spring starting earlier across the Northern Hemisphere. *Glob Change Biol*, 12: 343–351
- Schwartz M D, Ault T R, Betancourt J L. 2013. Spring onset variations and trends in the continental United States: Past and regional assessment using temperature-based indices. *Int J Climatol*, 33: 2917–2922
- Schwartz M D, Chen X Q. 2002. Examining the onset of spring in China. *Clim Res*, 21: 157–164
- Shen S, Cheng C X, Song C Q, Yang J, Yang S L, Su K, Yuan L H, Chen X Q. 2018. Spatial distribution patterns of global natural disasters based on biclustering. *Nat Hazards*, 92: 1809–1820
- Wang H J, Dai J H, Zheng J Y, Ge Q S. 2015a. Temperature sensitivity of plant phenology in temperate and subtropical regions of China from 1850 to 2009. *Int J Climatol*, 35: 913–922
- Wang H J, Ge Q S, Dai Y X, Dai J H. 2015b. Parameterization of temperature sensitivity of spring phenology and its application in explaining diverse phenological responses to temperature change. *Sci Rep*, 5: 8833
- Wang J F, Zhang T L, Fu B J. 2016. A measure of spatial stratified heterogeneity. *Ecol Indic*, 67: 250–256
- Wang X F, Xiao J F, Li X, Cheng G D, Ma M G, Zhu G F, Arain M A, Black T A, Jassal R S. 2019. No trends in spring and autumn phenology during the global warming hiatus. *Nat Commun*, 10: 2389
- White M A, de BEURS K M, Didan K, Inouye D W, Richardson A D, Jensen O P, O'keefe J, Zhang G, Nemani R R, van LEEUWEN W J D, Brown J F, de WIT A, Schaepman M, Lin X, Dettinger M, Bailey A S, Kimball J, Schwartz M D, Baldocchi D D, Lee J T, Lauenroth W K. 2009. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Glob Change Biol*, 15: 2335–2359
- White M A, Hoffman F, Hargrove W W, Nemani R R. 2005. A global framework for monitoring phenological responses to climate change. *Geophys Res Lett*, 32: L04705
- Wu X C, Liu H Y. 2013. Consistent shifts in spring vegetation green-up date across temperate biomes in China, 1982–2006. *Glob Change Biol*, 19: 870–880

- Wu X J, Zurita-Milla R, Kraak M J. 2015. Co-clustering geo-referenced time series: Exploring spatio-temporal patterns in Dutch temperature data. *Int J Geogr Inf Sci*, 29: 624–642
- Wu X J, Zurita-Milla R, Kraak M J. 2016. A novel analysis of spring phenological patterns over Europe based on co-clustering. *J Geophys Res-Biogeosci*, 121: 1434–1448
- Zhang Y, Hepner G F, Dennison P E. 2012. Delineation of phenoregions in geographically diverse regions using *k*-means++ clustering: A case study in the Upper Colorado River Basin. *GISci Remote Sens*, 49: 163–181
- Zhu D, Wang N H, Wu L, Liu Y. 2017. Street as a big geo-data assembly and analysis unit in urban studies: A case study using Beijing taxi data. *Appl Geogr*, 86: 152–164
- Zirlewagen D, von Wilpert K. 2010. Upscaling of environmental information: Support of land-use management decisions by spatio-temporal regionalization approaches. *Environ Manage*, 46: 878–893

(责任编辑: 李新)