

站点尺度的青藏高原时序NDVI重构方法比较与应用

刘建文^{1,2}, 周玉科^{3*}

(1. 福州大学空间信息工程研究中心, 福州 350003; 2. 数据挖掘与信息共享教育部重点实验室, 福州 350003;
3. 中国科学院地理科学与资源研究所 生态系统网络观测与模拟院重点实验室, 北京 100101)

摘要: 基于卫星遥感的植被指数时序数据广泛应用于植被覆盖监测、生物量反演等多个研究领域, 但由于传感器本身、大气条件、环境特征等因素引起的噪声会影响数据的应用效果, 因此开展植被指数时序数据重构研究具有实际意义。本文基于2000-2015年MODIS归一化差异植被指数(NDVI)数据, 采用三次样条函数法、双逻辑斯蒂函数法和奇异谱分析法3种常用方法, 对青藏高原106个气象站点所在的典型覆被NDVI时序数据进行重构, 并以植被物候信息提取作为应用, 比较分析了3种算法的保真性、细节拟合能力及物候特征提取效果。研究表明, D-L及Spline函数分别对受冰雪及云层影响较大(荒漠、灌木、林地)及较小的覆被类型(草原、农作物)表现出较好的细节拟合能力; SSA方法拟合能力较差, 易出现NDVI重构曲线整体“下移”的现象, 造成峰值拟合结果偏低, 并且表现出NDVI绝对值越小拟合效果越差的现象。从保持原始数据真值的能力来看, 受噪声点影响较大的覆被类型(林地、灌木、草原)Spline函数略优于D-L函数法; 而林地类型中SSA方法表现优于D-L函数法。从物候信息提取结果来看, D-L函数法所提取的生长季稍有提前, Spline函数及SSA方法分别表现出生长季开始点及结束点滞后的现象, 灌木、林地类型表现出明显的年际波动变化的特征, 荒漠类型由于NDVI绝对值偏低, 3种方法物候提取结果一致性表现出锯齿状不规则波动。此外, D-L方法生长季开始期(SOS)和生长季结束期(EOS)在各覆被区均小于其他方法, 波动较大; SSA方法提取的EOS在大部分覆被地区大于其他方法; Spline提取结果的年际波动与SSA高度相似。该研究可为高原植被不同覆被类型下NDVI时序数据噪声去除的方法选择提供借鉴。

关键词: 时序数据重构; 植被物候; 双逻辑斯蒂函数; 三次样条函数; 奇异谱分析法

1 引言

基于卫星遥感获取的植被指数能够较好地反映植被的生长态势, 已经广泛应用于植被变化监测、全球陆地表面建模、生物量计算等多个领域(Yu et al, 2004; Julien et al, 2006; 李斌等, 2017)。气象观测数据与植被指数的关系也得到广泛的研究, 基于多个站点的区域平均植被指数时序数据可有效表征相应覆被类型区内的植被生长状况(Chen et

al, 2002; 聂勇等, 2012)。理想的植被指数能增强对植物信息的敏感性, 最大限度地反映植被真实的生长状况及物候规律。但由于传感器性能各异、光照与观测角度不同、数据传输误差、地形与大气条件因时而异, 观测值通常带有不可预测的噪声, 从而影响了植被指数的应用。因此, 研究NDVI时间序列数据的重构方法, 尽量真实地刻画植被生长情况, 是植被生态遥感领域的重要研究课题, 也是植被综合对比研究的热点。

收稿日期: 2017-04-12; 修订日期: 2017-11-08。

基金项目: 国家自然科学基金项目(41601478); 国家重点研发计划项目(2016YFC0500103); 中科院STS项目(KFJ-SW-STS-167) [Foundation: National Natural Science Foundation of China, No.41601478; National Key R&D Program of China, No. 2016YFC0500103; Science Technology Service Project of Chinese Academy of Sciences, No. KFJ-SW-STS-167]。

作者简介: 刘建文(1989-), 男, 硕士研究生, 研究方向为地图学与地理信息系统, E-mail: liujw@lreis.ac.cn。

通讯作者: 周玉科(1984-), 男, 博士, 助理研究员, 研究方向为生态遥感, E-mail: zhoyk@igsrr.ac.cn。

引用格式: 刘建文, 周玉科. 2018. 站点尺度的青藏高原时序NDVI重构方法比较与应用[J]. 地理科学进展, 37(3): 427-437. [Liu J W, Zhou Y K. 2018. Comparison and application of NDVI time-series reconstruction methods at site scale on the Tibetan Plateau[J]. Progress in Geography, 37(3): 427-437.]. DOI: 10.18306/dlkxjz.2018.03.013

当前NDVI时间序列重构的主要方法包括:阈值法、延时滑动平均法、函数拟合法等。阈值法出现较早,经历了固定阈值法、动态阈值法、多参量阈值法的发展过程易与其他方法相结合,应用较为广泛(Jönsson et al, 2004)。延迟滑动平均方法(Delayed Moving Average, DMA)最先是Reed等(2009)提出的,该方法基于自回归滑动平均模型,通过NDVI时间序列曲线与滑动平均曲线之间的比较结果来反映实际NDVI值偏离既定趋势的程度。函数拟合法则是利用平滑模型函数拟和时间序列遥感数据,进而提取物候信息的方法,包括逻辑斯蒂拟合法、谐波函数法、非对称高斯函数法及Savitzky-Golay滤波法等(Jonsson et al, 2002; Chen et al, 2004; Sakamoto et al, 2005; Beck et al, 2006)。

近年来植被指数时序数据重构的研究较多,其中非对称性高斯函数拟合法(AG)、双逻辑斯蒂函数拟合法(D-L)、Savitzky-Golay (S-G)滤波法是目前较为常用的函数拟合方法。宋春桥等(2011)对藏北地区的研究发现,AG及D-L函数拟合法的拟合效果优于S-G滤波法。周惠慧等(2016)基于不同时间分辨率的研究表明,S-G与D-L拟合法分别适用于较大及较小时间间隔的时序数据重构。刘亚南等(2016)认为,在噪声点优化及保留原始数据质量方面,AG与D-L整体效果相接近,但部分像元处理中AG方法表现出更好的拟合效果。此外,李杭燕等(2010)在常用方法的基础上,探索了基于标准

差权重和噪声点性质的时序数据综合重构方法;Geng等(2014)通过回归估计标准差(RMSE)、Akaike信息准则(AIC)、贝叶斯信息准则(BIC)对8种NDVI重构方法的性能进行了比较研究。

尽管国内外已有针对不同的植被指数时间序列重构方法的比较研究,但多数基于常用方法的对比,有关三次样条函数(Spline)、双逻辑斯蒂函数法(D-L)及奇异谱分析法(SSA)在多种覆被类型下方法适宜的对比研究仍然较少,尤其对于植被类型复杂的青藏高原区域,仍有必要对不同的方法和区域的适用性进行深入探索。因此本文以青藏高原几种主要的植被覆被区域的NDVI时序数据为研究对象,分别采用Spline、D-L、SSA 3种方法对不同覆被区域气象站点的NDVI时间序列的重构效果进行研究,该研究内容可为进一步探索NDVI与气象因子的关系奠定基础。

2 研究区域及数据集

2.1 研究区域

青藏高原平均海拔在4000 m以上,属于全球变化的敏感区和脆弱区,也是陆地生态系统研究的热点区域(张懿锂等, 2002)。在大气环流及高原地势格局的制约下,青藏高原形成了不同的水热地域状况配置,气候呈现从东南温暖湿润向西北寒冷干燥变化的特征。复杂的地形和气候条件造就了丰富的植被类型(图1),水平带谱上依次出现了森林、草

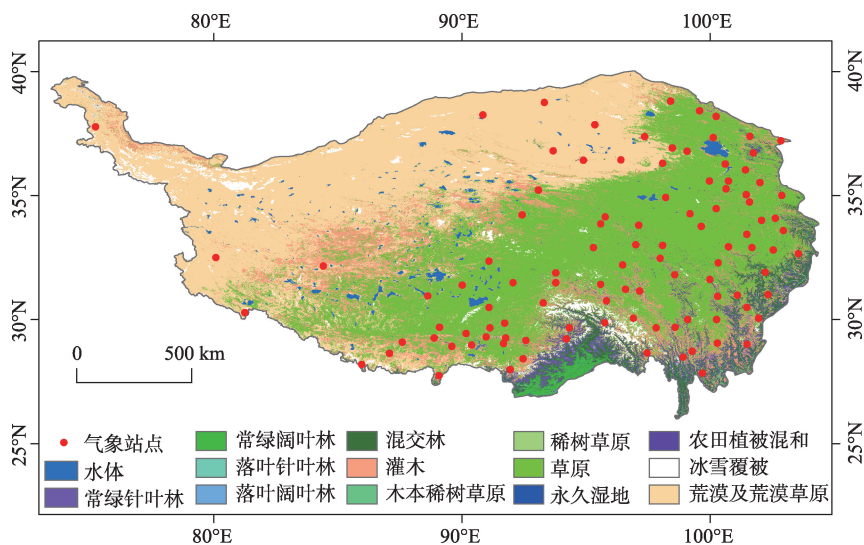


图1 青藏高原气象观测站点及土地覆被类型分布图

Fig.1 Spatial distribution of the meteorological stations and land cover types on the Tibetan Plateau

原、草甸、荒漠等植被,垂直自然带也由喜马拉雅山南侧的热带雨林开始,依次向常绿阔叶林、针叶林、灌丛、草甸和高山永久冰雪带过渡,成为中国最完整的地山自然景观垂直带谱(于海英等, 2009)。研究区域范围和主要植被覆盖类型如图1所示。

2.2 数据集

本文NDVI时间序列数据采用MODIS的MOD13Q1产品(<https://daac.ornl.gov/MODIS/modis.shtml>),其空间分辨率为250 m,时间分辨率为16天。考虑研究区域巨大的空间异质性,以青藏高原106个气象站点为中心获取2 km×2 km范围内平均的NDVI时序数据(图1)。首先通过MODIS官网获取MODIS多种产品的子数据集,并根据坐标点提取对应的像素、一定距离缓存范围或整个对应瓦片的时间序列数据,然后利用R语言,通过SOAP网络服务的接口下载MOD13Q1数据。该数据时间覆盖范围为2000年2月-2015年12月,共365期。MODIS NDVI数据请求的关键代码如下(基于R语言MODIS package):

```
MODISSubsets(LoadDat=SubsetExample,Products="MOD13Q1", Bands=c("250 m_16_days_NDVI", "250 m_16_days_pixel_reliability"), Size=c(2,2), StartDate=TRUE)
```

为保证数据的精度,同时从MODIS官网请求了像素质量控制信息,在数据处理的过程中,将像素质量可信度(pixel reliability)高于75%的像素作为可用数据,否则设置为空值。MOD13Q1产品一年可获取23期的NDVI时序数据,若一年中空值期数大于或等于5期,则本年度的NDVI时序数据不可用。土地覆盖类型的数据源于2007年MODIS影像土地覆盖分类产品MCD12Q1。该分类产品依据国际地圈生物圈计划(IGBP)定义的分类方案数据,分类类别及代号见表1。

研究区域内主要土地覆盖类型包括:荒漠草原、草原、林地、冰雪、灌木、水体等,其中冰雪覆被、农田与自然植被混合、水体、永久湿地、城镇与建设

用地不在研究范围之内。研究中根据需要将类型1-5合并为林地、6-7合并为灌木、8-10合并为草原,鉴于农田及荒漠与其他覆被类型的特征差异,将12和16分别作为一类,最终将研究区域分成荒漠、草原、灌木、农作物、林地5类。

3 研究方法

本文NDVI时序数据重构方法的对比研究流程如图2所示:首先利用R语言通过SOAP网络服务获取MOD13Q1产品中的NDVI子数据集,构建原始NDVI时间序列。然后基于R语言实现双逻辑斯蒂函数法、三次样条函数法、奇异谱分析法的NDVI时间序列数据重构,并对数据进行平滑去噪处理。最后依据MODIS土地覆盖类型及气象站的分布数据,选取不同覆被类型的典型站点,分析不同覆被类型情形下,3种方法在NDVI时间序列的重构效果及物候特征提取方面的差异。

3.1 双逻辑斯蒂函数法

双逻辑斯蒂函数法(Double logistic function method, D-L)是由Beck等(2006)提出的一种半局部拟合的经典方法,该方法无需确定阈值,仅依赖NDVI曲线的变化特征,即可客观地反映植被生长过程的细节特征,可用于区域乃至全球尺度上不同植被物候监测。D-L方法首先将整个时间序列中时间点对应值按极大、极小分成多个区间,然后分别对该区间进行D-L函数局部拟合,其局部拟合方式与非对称高斯拟合方法类似,公式为:

$$g(t; a_1, \dots, a_4) = \frac{1}{1 + \exp\left(\frac{a_1 - t}{a_2}\right)} - \frac{1}{1 + \exp\left(\frac{a_3 - t}{a_4}\right)} \quad (1)$$

式中: a_1 决定峰值和谷值的位置, a_2 、 a_3 、 a_4 控制着左半部分的宽度和陡峭度。然后基于式(1),将局部拟合函数的特征加以综合,重构新的NDVI时间序列曲线。整体拟合的函数为:

$$F(t) = \begin{cases} a(t)f_L(t) + (1 - a(t))f_C(t) \cdots t_L < t < t_C \\ \beta(t)f_C(t) + (1 - \beta(t))f_R(t) \cdots t_C < t < t_R \end{cases} \quad (2)$$

表1 MCD12Q1产品土地分类方案及其代号

Tab.1 IGCP land cover classification scheme and codes in the MCD12Q1 product

代码	0	1	2	3	4	5	6	7	8
类别	水体	常绿针叶林	常绿阔叶林	落叶针叶林	落叶阔叶林	混交林	封闭灌丛	开放灌丛	木本稀树草原
代码	9	10	11	12	13	14	15	16	254
类别	稀树草原	草原	永久湿地	农田	城镇与建设用地	农田/自然植被混合	冰雪覆被	荒漠及荒漠草原	未分类

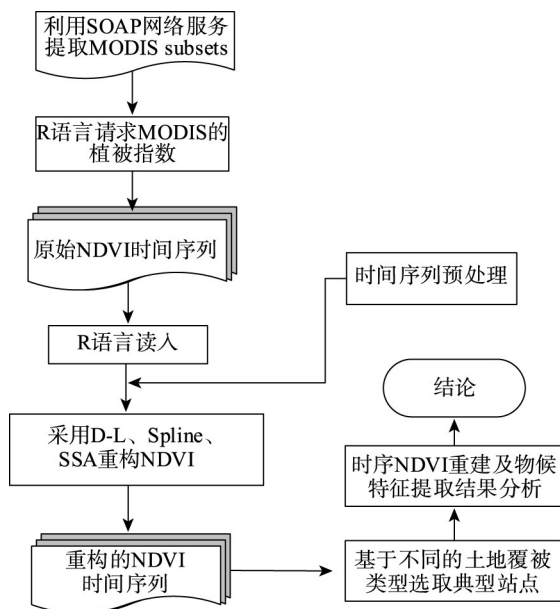


图2 NDVI时序重构方法对比研究流程图

Fig.2 Flowchart for the comparison of time-series NDVI reconstruction methods

式中: $[t_L, t_R]$ 区间是 NDVI 整个变化区间, $f_L(t)$ 、 $f_C(t)$ 、 $f_R(t)$ 分别代表 $[t_L, t_R]$ 区间内左边谷值、中间峰值及右边谷值所对应的局部函数, $\alpha(t)$ 和 $\beta(t)$ 为 $[0, 1]$ 之间的剪切系数。值得指出的是, D-L 函数法通过时间序列的季节性确定函数内部的参数, 若时间序列无季节性则 D-L 函数的拟合没有意义, 所以时间序列预处理时应考虑季节性。

3.2 奇异谱分析法

奇异谱分析法 (Singular Spectrum Analysis, SSA) 是在 Karhunen-Loeve 分解理论的基础上发展起来的, 是一种基于时间序列的动力重构并与经验正交函数相联系的统计技术 (Vautard et al, 1992)。由于该方法具有稳定识别和强化信号、时频域相结合、不受正弦波假设约束的特征, 广泛应用于信号处理、测量与空间插值、语言识别、气象监测等多个领域。NDVI 时间序列存在多尺度的震荡特征, 其中月际尺度的 NDVI 时序数据周期震荡微弱, 常规手段不易度量, SSA 可从包含噪声的数据序列中尽可能多地提取可靠信息, 把最可预报的分量聚集到若干个时间序列, 并提取有意义的分量进行序列重构降低噪声, 对于 NDVI 时序数据的分析具有较好的适用性 (王解先等, 2013)。SSA 中窗口长度为实验的关键参数, 长度过大会包含过多的噪声, 过小则会丢失数据的有效信息。本文中的时间序列长

度为 365, 为有效滤除噪声并保留数据的有效信息, 多次实验最终确定序列长度的一半 (182) 作为最终的窗口长度参数。

3.3 三次样条函数法

三次样条函数法 (Cubic spline function method, Spline) 是时间序列重构的经典方法, 在处理任意有限间隔的时间序列的数据时, 可有效地滤除噪声并具有较强的健壮性及数据处理能力, 在时序数据平滑、噪声处理等方面应用较为广泛 (Musial et al, 2011)。Spline 方法强调产生变量的原因及时间序列的数学架构, 首先架构了一种连续的三次多项式的分段曲线, 然后在节点组合在一起, 通过这种方式, 完成局部到整体的拟合。其高度的灵活性及适应性可由单个平滑参数 λ 进行控制, 此值越小, 样条曲线与原数据点越接近。当 λ 值从大到小变化时, 函数灵活性增加; 当值从小到大增加时, 曲线的僵硬程度相应增大, λ 可以到无穷大, 样条变成一个线性平方值。该方法具有较好的健壮性, 计算开销也相对较小, 适用于较大数据集处理。在构建模型过程中, 为保证方程组具有唯一解, 在区间的端点需补充边界条件, 本文使用 Lagrange 三次样条插值函数作为边界条件, 即通过数据的两端各取 4 个点拟合一个精确的立方的方式确定边界。

3.4 导数法提取植被物候参数

NDVI 时间序列重构以后, 本文采用了经典的导数法求取植被生长轨迹中的关键物候参数: 生长季开始点 (SOS) 和生长季结束点 (EOS)。导数法能够通过全年的植被指数时序数据, 采用散点图组成的边界建立生长轨迹模型, 具有较强的适宜性 (Gu et al, 2009)。生长轨迹曲线可通过如下公式描述:

$$NDVI(t) = y_0 + \frac{a_1}{\left[1 + \exp\left(-\frac{t-t_1}{b_1}\right)\right]^{c_1}} - \frac{a_2}{\left[1 + \exp\left(-\frac{t-t_2}{b_2}\right)\right]^{c_2}} \quad (3)$$

式中: $NDVI(t)$ 是第 t 天的归一化植被指数, y_0 、 a_1 、 a_2 、 b_1 、 b_2 、 c_1 、 t_1 和 t_2 分别为经验估计参数。该模型具有较强的灵活性, 对于植被群落不同季节周期的拟合能力具有较好的适应性, 能精确模拟生长季开始到结束整个过程。与创建断点分段考虑的方法不同, 此方法具有更少的限制条件, 并用迭代算法进行参数的估计, 可以有效提取 SOS、EOS 等物候参数。

4 结果与分析

4.1 不同覆被NDVI时序曲线重构结果比较

由于青藏高原面积广阔且站点分布并不均匀,为减少海拔、温度等空间异质性的影响,本文中时序数据重构及物候提取的过程中均采用各覆被类型下单个典型站点时序数据作为输入。数据输入完成后,首先对原始数据进行预处理剔除无效数据,然后采用Spline函数法、D-L函数法、SSA方法对初始NDVI时间序列数据进行重构,NDVI时序曲线重构结果见图3。不同覆被NDVI时序数据的初始值如图3a所示:荒漠、草原、灌木、农作物、林地的NDVI值依次升高;荒漠、草原NDVI绝对值较小,生长季NDVI值相对于冷季仅有轻微抬升;灌木、农作物与林地生长季NDVI值较大,季节性较为明显。另外由于林地大部分位于森林线以上,易受冰雪、云层的影响,所以受连续噪声的影响较大。

荒漠及荒漠草原主要分布于柴达木盆地、昆仑山北翼和阿里山地等高原温带区及昆仑山高山高原亚寒带区,气候严寒干旱、风大、植被组成单调。

在荒漠覆被区,NDVI范围仅在0.05~0.085范围内小幅波动,6-9月份NDVI值相对冷季有轻微抬升并伴随连续噪声点,以局部拟合为主的Spline函数及D-L函数法可较为精确地描述NDVI时序数据的细微变化,有效降低噪音。其中D-L函数法对上包络线的拟合能力较强,在返青期至成熟期的过渡阶段,D-L函数法的陡峭程度明显大于另外2种方法。SSA由于NDVI的振幅(NDVI最大值与NDVI最小值的差值)较小,使得该方法的NDVI重构曲线特征(峰值、坡度)较真实值偏低。Spline函数由于其局部拟合“高低填平”的效果,重构后的曲线平滑效果良好(图3b)。

草原主要分布在羌塘高原湖盆、青南高原宽谷及藏南高山谷地等半干旱高寒区域,植被以高山嵩草、紫花针茅和披碱草等为主。草原区域全年的NDVI值普遍偏低,从2000-2015的统计数据来看,全年NDVI最大值通常不超过0.25,生长季大约 120 ± 10 d,部分区域受地面冰雪、冻融影响波动较大。Spline函数法与D-L函数法对草原的细节拟合能力较强。SSA重构结果与荒漠区类似,对上包络

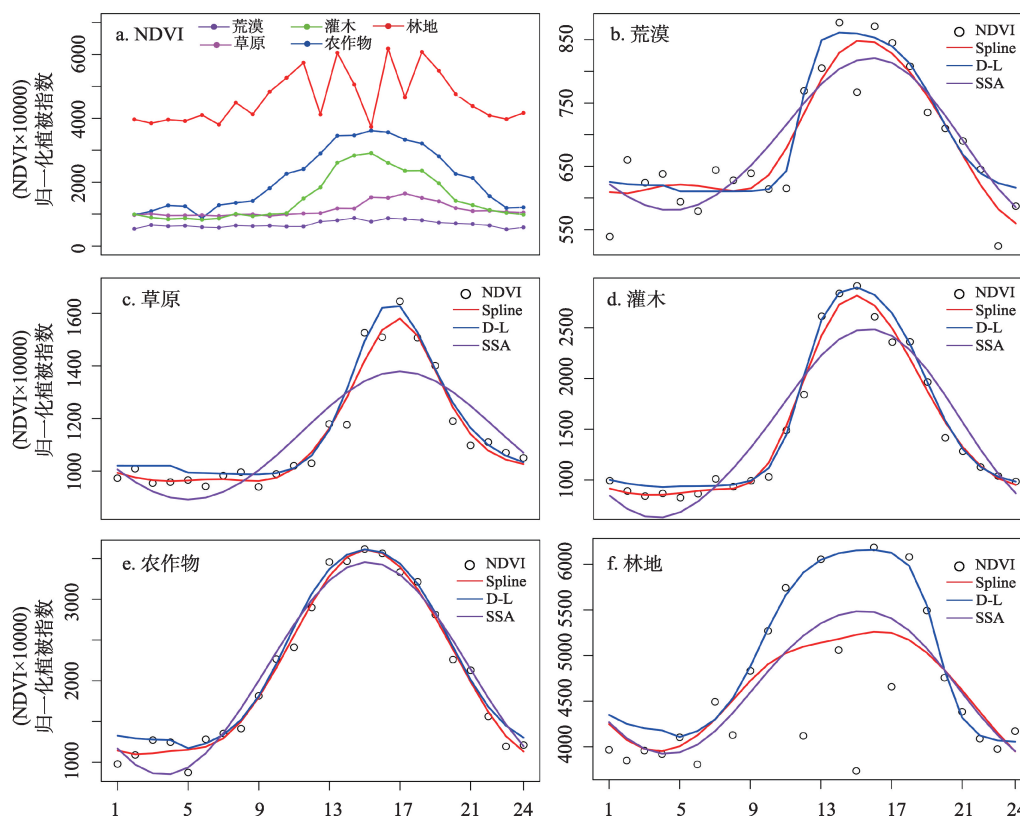


图3 不同土地覆被类型NDVI时间序列重构曲线对比

Fig.3 Comparison of the original and the reconstructed MODIS time-series NDVI

线的拟合能力较弱,生长季曲线出现明显“左移”现象,易造成物候特征提取误差(图3c)。

灌木主要分布在藏南高山谷地,果洛那曲高原等半干旱或半湿润地区,NDVI振幅相对较大,成熟期峰值NDVI可达0.6左右。由于该覆被类型主要分布在高原森林线以上,冬季受降雪影响较大。整体拟合过程中,由于NDVI绝对值较大并且噪声点较少,3种方法都一致表现出平滑过渡的效果,另外D-L及Spline函数表现出极强的细节拟合能力,在返青期至成熟期过渡阶段,NDVI曲线峰值及曲线的陡峭程度都表现出极大的相似性,SSA拟合效果则稍有欠缺,重构后NDVI曲线特征较另外2种方法偏低(图3d)。

青藏高原的农业用地包括耕地和旱地,分布范围极小。农作物NDVI振幅较大,具有明显的生长季特征;由于人为因素的干预,较灌丛、草原、荒漠覆被类型,有返青期提前、休眠期滞后、生长季跨度大的现象。由于农作物类型受噪声影响较少且NDVI绝对值较大,3种方法对农作物NDVI时间序列的拟合存在高度一致性,基本上都能保持NDVI的真实值,有较好的拟合效果(图3e)。

青藏高原的林地主要分布在四川西部西藏东部的高山深谷、祁连山青藏高原东部高山盆地,气候湿润或半湿润地区以及中亚热带部分湿润地区,少量的针叶林阔叶林呈条带分布。全年的NDVI值较高,但由于森林线以上的冰雪覆盖及云量相对较多,使得NDVI时间序列曲线易出现不规则锯齿状波动。由于D-L函数法对上包络线较好的拟合能力,能有效地滤除噪声,其NDVI重构曲线特征(峰值,生长季)明显优于另外2种方法;Spline函数及SSA则受噪声的影响较大,峰值拟合明显偏低,易出现部分月份或整个年份无法拟合的现象(图3f)。

综上所述,3种NDVI时序数据重构方法都不同程度地消除了噪声的影响,重建后的NDVI曲线更为平滑。但是由于青藏高原各区域生态地理条件各异、方法的原理不同,3种方法表现出不同的特点:D-L方法保持曲线峰值的能力较优,对于存在连续低值噪声的覆被类型,仍能有效地保持曲线的整体特征,但对于NDVI低值区域拟合能力较弱,休眠期阶段D-L方法易出现拟合值偏大的现象,该结论与宋春桥等(2011)研究的结论一致。Spline方法对于受噪声影响较小的区域能有效地表现NDVI时序曲线的细节特征,且可较好地平衡拟合曲线的整体

趋势,但由于Spline方法分段拟合的原理,对于易出现连续低值噪声的区域,拟合效果不佳。SSA方法重构曲线最为平滑,对于NDVI时序曲线下包络线的拟合能力较强,但由于其基本为原理为信息的分解与重构,窗口大小的设定是影响结果重构的关键因素,同时受研究区域的影响,在植被返青期及枯黄期部分区域所拟合的曲线过于“平缓”。

4.2 3种重构方法的保真性比较

保真性即保持原始数据真值的能力,能有效地表示不同覆被类型下各类方法的稳定性。本文通过相关系数及回归估计标准差来衡量各方法的保真性,相关系数越高、回归估计标准差越小则保真性越强,其值与原始NDVI的绝对值及各方法的去噪能力有关。图4所示为各类覆被类型2000-2015年365期原始NDVI时间序列曲线,由于各类覆被类型强烈的空间异质性,不同覆被类型受噪声影响程度、NDVI绝对值及年际变化趋势均有明显差异。荒漠的NDVI有上升趋势,草原、灌木则较为稳定,但农作物则有轻微下降趋势,而林地由于受冰雪及云层的影响易出现连续的噪声和缺失值。

相关系数是反映变量间线性相关关系的统计指标,可用于表示重构后NDVI时序曲线保持原始NDVI曲线特征的能力。从覆被类型对比结果来看,农作物由于人为因素的影响较大,受随机噪声影响较小,相关性最强;而易受冰雪、大气等随机噪声影响的荒漠、草原、灌木类型则相关程度较低;林地类型由于易受云层、冰雪等影响而出现连续的低值噪声,相关程度最低。从算法对比结果来看,对荒漠及农作物类型,D-L函数表现出较好的相关性;对灌丛和林地类型,由于D-L函数对于上包络线的拟合能力较强,当低值噪声点较多时,相关系数相对较低。Spline函数由于局部拟合的特性,NDVI时序重构时具有“高低填平”效果,在易受噪声点影响的草原、灌木、林地地区,NDVI重构结果具有较强的相关性。SSA方法由于易出现“欠拟合”现象,整体拟合效果及相关性最低。

回归估计标准差也称作均方根误差(Root Mean Square Error, RMSE, 见式(4)),用于描述重构后NDVI时间序列数据与原始数据之间的平均差异程度,值越小拟合代表性越强。如图5所示,荒漠、草原、灌木、农作物、林地RMSE值依次呈上升趋势,其原因一方面受到重构算法拟合能力的影响,

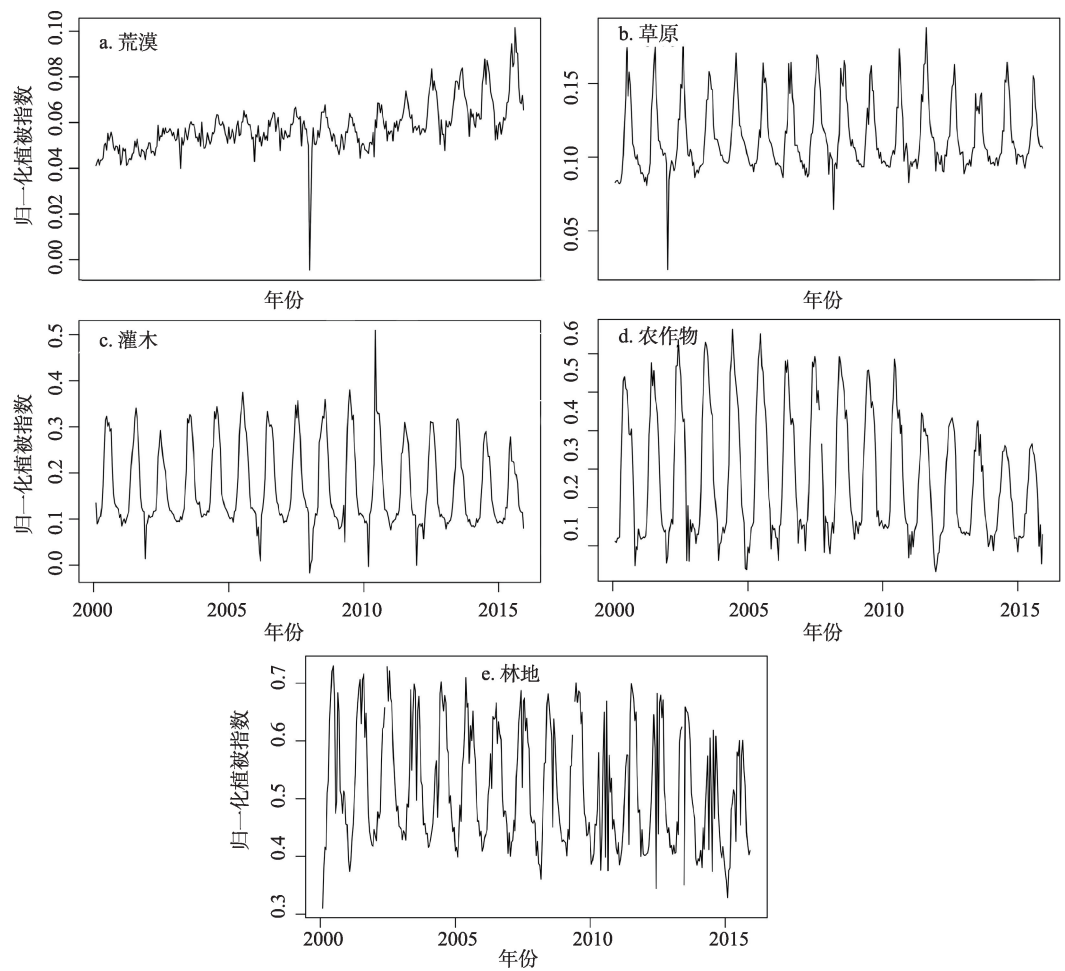


图4 2000-2015不同覆被类型的归一化植被指数(NDVI)时序曲线
Fig.4 NDVI time-series curve of different land cover types, 2000-2015

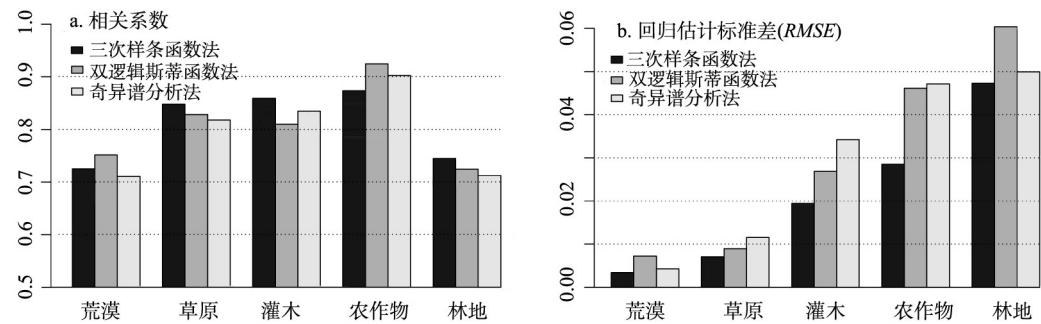


图5 不同覆被类型NDVI重构保真能力比较
Fig.5 Comparison of the ability of three time-series reconstruction methods on keeping the main characters of NDVI for different land cover types

另一方面与该植被类型本身NDVI绝对值大小有关。 $RMSE$ 统计结果与NDVI重构效果比较吻合,Spline函数由于局部拟合的特性, $RMSE$ 值较小;D-L函数对于上包络线的拟合能力较强,当受冰雪或大气影响而产生较多噪声点时, $RMSE$ 值相对较高;

SSA由于NDVI重构过程中易于出现的“左移”现象,在草原、灌木类型的重构时 $RMSE$ 值偏高。

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (NDVI_{p_i} - NDVI_{o_i})^2}{N}}$$

(4)

4.3 3种方法在植被物候参数提取上的应用与比较

植被物候是研究植被与气候、环境变化间关系的重要参量,它反映了生物圈与环境因子(地形、地貌、土壤类型等)年内与年际间的变化响应(昝佳星等, 2016)。由于气候及地理条件的差异,植被物候监测结果呈现区域性规律变化的特征,不同植被类型之间差异较大。基于卫星遥感的物候参数提取能大范围监测空间连续的植被物候情况,其中NDVI时间序列重构是物候参数提取的关键环节。

上述3种方法的物候关键参数的提取结果如图6所示,每个子图下半部分表示生长季开始点(Start of Growing Season, SOS),上半部分表示生长季结束点(End of Growing Season, EOS)。从SOS及EOS提取结果的年际变化规律来看,Spline及D-L函数由于局部拟合的特性年际变化波动明显;SSA由于较强的滤除噪声的能力,表现出较强稳定性。另外从SOS及EOS结果的时间先后排序,可以看出由于D-L函数上包络线良好的拟合能力,D-L函数生

长季表现出整体前移现象;Spline与SSA则分别表现出生长季开始点与结束点滞后的现象。

荒漠类型由于其特殊的生态环境,生长季不明显且NDVI绝对值较小,物候提取结果均表现出较强的波动性,与其他覆被类型相比,该覆被类型SOS指标明显滞后(图6a)。草原、灌木由于受降水、海拔等自然因素的影响较大,Spline函数及D-L函数波动变化明显,波动范围20 d左右;另外草原生长季稍有滞后,灌木生长季起点稍有提前(图6b-6c)。农作物类型受人为因素的影响较大,物候提取结果均表现出较强的稳定性,该覆被类型下D-L函数生长季提前、Spline函数的SOS指标、SSA的EOS指标滞后的现象表现突出(图6d)。

林地类型中,Spline与SSA方法表现出较强的相似性,二者生长季开始时间与结束时间几乎一致;不同的是由于林地大部分在森林线以上,易受冰雪、云层等随机噪声的影响,D-L函数的物候特征提取结果波动性较大,振幅可达40 d;该覆被类型

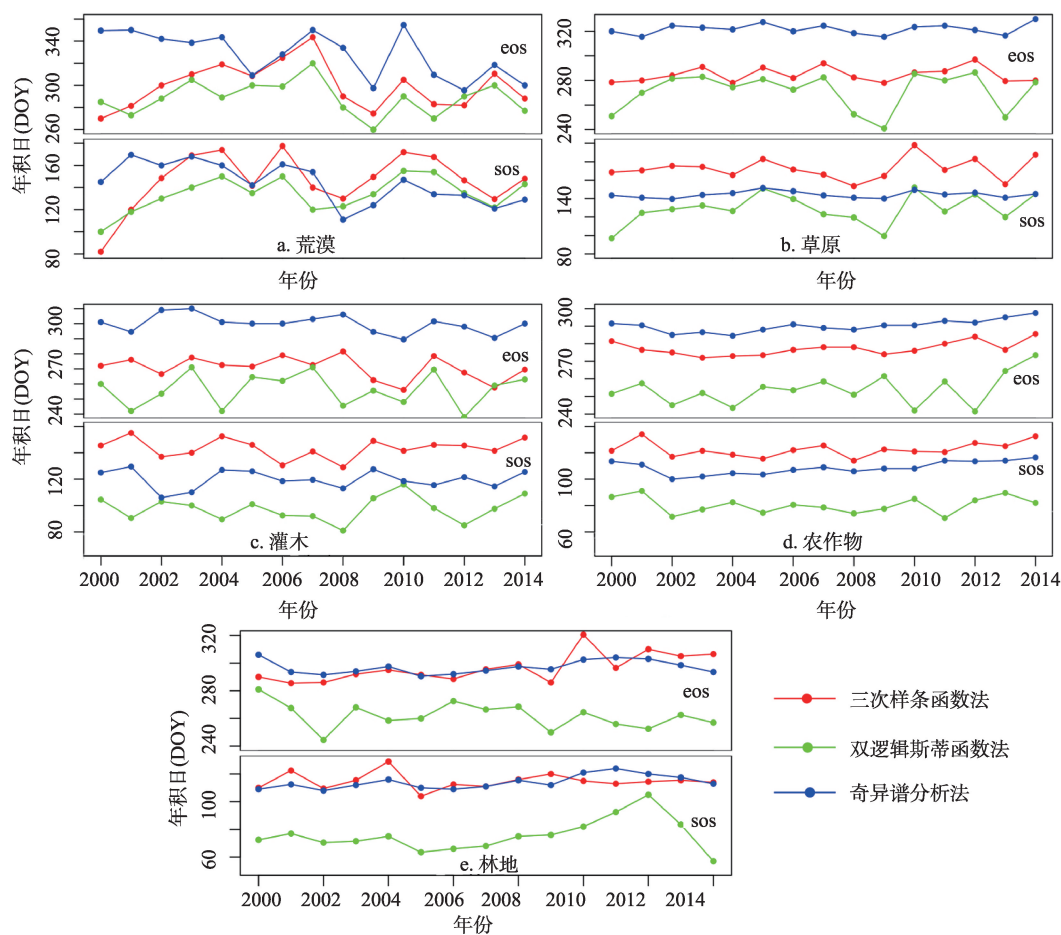


图6 3种方法不同植被类型的物候参数提取结果

Fig.6 The phenology parameters derived from three time-series reconstruction methods for different land cover types

中SOS较其他覆被类型提前约15 d左右,EOS滞后约10 d左右。此外2009年草原及2014年林地覆被,D-L函数受冰雪等噪声影响较大,SOS与EOS比往年明显偏低(图6e)。

从物候指标的提取结果来看,D-L方法由于上包络线较优的拟合能力,植被返青期及枯黄期NDVI曲线“坡度”较大,导致SOS、EOS指标均有提前且年际波动明显大于Spline及SSA方法。SSA方法由于其对于信息先分解然后重构的原理,稳定性优于D-L及Spline方法;另外该方法SOS指标稍有提前,原因为:该方法对于下包络线较强的拟合能力导致所拟合的休眠期值低于实际值,加之返青期及枯黄期整体“坡度”较小,导致SOS指标稍有提前、EOS指标稍有滞后。与D-L及SSA方法不同,Spline方法由于分段拟合的特征易造成“上下填平”的拟合效果,从而使得所拟合的曲线返青期及枯黄期的“坡度”小于D-L方法,休眠期拟合结果大于SSA方法,从而造成SOS指标稍有滞后,而EOS指标介于D-L及SSA方法之间的现象。荒漠类型中各方法的SOS、EOS指标波动较大、农作物类型中SOS、EOS波动较小,反映了不同覆被类型下各方法物候提取结果波动性的差异。

综上所述,基于遥感的植被物候提取结果精度取决于研究区域的植被类型、生态地理条件、时序数据噪声的处理过程、植被物候的提取方法等多个环节。由于青藏高原温度、海拔等因素强烈的空间异质性,不同覆被类型下地理条件存在显著差异。NDVI时序数据重构虽然能部分消除噪声带来的误差,但各方法去除噪声的能力不尽相同且噪声的影响存在一定延续性,从而对拟合后的NDVI时序曲线在极值、幅宽等多个方面造成影响。此外,植物物候遥感识别方法各具优点,不同方法有时仅对特定研究区域有效。为准确提取各项物候指标,应针对特定研究区域选择合适的方法并加以参数化和本地化。由此可见,研究区域的植被类型、地理环境、时序数据去噪过程、物候提取方法等各个环节共同影响了植被生长曲线拟合及物候信息提取的准确性。上述不同覆被类型下各方法的相关系数及回归估计标准差的分析亦有类似结论。

5 结论

本文在综合考虑时间尺度及以往研究的基础

上,以青藏高原106个气象站点为中心,利用R语言获取MOD13Q1产品中的NDVI时间序列。分别采用D-L、Spline及SSA方法对NDVI时间序列数据进行重构,并选取不同覆被类型的典型站点,对比分析不同覆被类型情形下3种方法在NDVI时间序列的重构及物候特征提取方面的差异。

从NDVI时间序列的重构结果来看,以局部拟合为主的Spline函数法及D-L函数法可较为精确地描述NDVI时序数据的细微变化,有效降低噪声。但由于方法原理不同,D-L函数法对上包络线拟合较好,适合用于灌木、林地等易受连续噪声点影响的覆被类型;Spline函数法由于其局部拟合的特性,对于草原、农作物等受噪声点影响较小的覆被类型的拟合效果较好;SSA重构结果则易出现“欠拟合”的现象,造成峰值拟合结果偏低,并且表现出NDVI绝对值越小拟合效果越差现象。

从相关系数及RMSE结果来看,不同覆被类型的原始NDVI值越大,噪声点越多,保真性越差。从3种重构方法的比较结果来看,受噪声点较大的覆被类型(林地、灌木、草原)Spline函数保真性较强;荒漠及农作物类型,D-L函数法保真性较优;而林地类型由于易受连续噪声点的影响,SSA表现最佳。从物候提取结果来看,D-L函数的SOS、EOS指标,较另外2种方法提前约20 d左右,并且受噪声影响时波动较大。Spline函数及SSA方法物候特征提取结果稍有滞后,但二者稳定性均优于D-L函数法。此外,草原的生长季较灌木、农作物稍有滞后,且年际波动较小;易受噪声影响的灌木、林地类型,物候结果表现出明显的年际波动特征;荒漠类型由于NDVI绝对值偏低,物候指标的年际变化最为剧烈。总结上文的研究结果发现,物候遥感监测结果的精度不仅取决于具体的植被覆盖类型及地理位置环境条件,还与去除噪声方法的选取有关;不存在具有广泛适应性的算法,实际应用中仍需依据植被类型与研究目的选择合适的方法。

本文基于不同覆被类型的多方法的比较研究,可进一步了解各方法在不同的研究区域及覆被类型的适应性及自身的优缺点,便于研究人员依据研究区域及研究目的选择更适宜的方法。在进一步探索NDVI时间序列与气象因子关系研究中,该研究成果可作为数据处理工作的关键部分。

参考文献(References)

李斌,王慧敏,秦明周,等. 2017. NDVI、NDMI与地表温度关

- 系的对比研究[J]. 地理科学进展, 36(5): 585-596. [Li B, Wang H M, Qin M Z, et al. 2017. Comparative study on the correlations between NDVI, NDMI and LST[J]. Progress in Geography, 36(5): 585-596.]
- 李杭燕, 马明国, 谭俊磊. 2010. 时序NDVI数据集重建综合方法研究[J]. 遥感技术与应用, 25(6): 891-896. [Li H Y, Ma M G, Tan J L. 2010. Integrated reconstruction methods of time-series NDVI dataset[J]. Remote Sensing Technology and Application, 25(6): 891-896.]
- 刘亚南, 肖飞, 杜耘. 2016. 基于秦岭样区的四种时序EVI函数拟合方法对比研究[J]. 生态学报, 36(15): 4672-4679. [Liu Y N, Xiao F, Du Y. 2016. Analysis of four time series EVI data reconstruction methods[J]. Acta Ecologica Sinica, 36(15): 4672-4679.]
- 聂勇, 刘林山, 张懿锂, 等. 2012. 1982-2009年珠穆朗玛峰自然保护区植被指数变化[J]. 地理科学进展, 31(7): 895-903. [Nie Y, Liu L S, Zhang Y L, et al. 2012. NDVI change analysis in the Mount Qomolangma (Everest) national nature preserve during 1982-2009[J]. Progress in Geography, 31(7): 895-903.]
- 宋春桥, 柯灵红, 游松财, 等. 2011. 基于TIMESAT的3种时序NDVI拟合方法比较研究: 以藏北草地为例[J]. 遥感技术与应用, 26(2): 147-155. [Song C Q, Ke L H, You S C, et al. 2011. Comparison of three NDVI time-series fitting methods based on TIMESAT: Taking the grassland in northern Tibet as case[J]. Remote Sensing Technology and Application, 26(2): 147-155.]
- 王解先, 连丽珍, 沈云中. 2013. 奇异谱分析在GPS站坐标监测序列分析中的应用[J]. 同济大学学报: 自然科学版, 41(2): 282-288. [Wang J X, Lian L Z, Shen Y Z. 2013. Application of singular spectral analysis to GPS station coordinate monitoring series[J]. Journal of Tongji University: Natural Science, 41(2): 282-288.]
- 于海英, 许建初. 2009. 气候变化对青藏高原植被影响研究综述[J]. 生态学杂志, 28(4): 747-754. [Yu H Y, Xu J C. 2009. Effects of climate change on vegetations on Qinghai-Tibet Plateau: A review[J]. Chinese Journal of Ecology, 28(4): 747-754.]
- 张懿锂, 李炳元, 郑度. 2002. 论青藏高原范围与面积[J]. 地理研究, 21(1): 1-8. [Zhang Y L, Li B Y, Zheng D. 2002. A discussion on the boundary and area of the Tibetan Plateau in China[J]. Geographical Research, 21(1): 1-8.]
- 周惠慧, 王楠, 黄瑶, 等. 2016. 不同时间间隔下的遥感时间序列重构模型比较分析[J]. 地球信息科学学报, 18(10): 1410-1417. [Zhou H H, Wang N, Huang Y, et al. 2016. Comparison and analysis of remotely sensed time series of reconstruction models at various intervals[J]. Journal of Geo-Information Science, 18(10): 1410-1417.]
- 祖佳星, 杨健. 2016. 东北地区植被物候时序变化[J]. 生态学报, 36(7): 2015-2023. [Zu J X, Yang J. 2016. Temporal variation of vegetation phenology in northeastern China[J]. Acta Ecologica Sinica, 36(7): 2015-2023.]
- Beck P S A, Atzberger C, Høgda K A, et al. 2006. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI[J]. Remote Sensing of Environment, 100(3): 321-334.
- Chen J, Jönsson P, Tamura M, et al. 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter[J]. Remote Sensing of Environment, 91(3-4): 332-344.
- Chen J M, Pavlic G, Brown L, et al. 2002. Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements[J]. Remote Sensing of Environment, 80(1): 165-184.
- Geng L Y, Ma M G, Wang X F, et al. 2014. Comparison of eight techniques for reconstructing multi-satellite sensor time-series NDVI data sets in the Heihe River Basin, China [J]. Remote Sensing, 6(3): 2024-2049.
- Gu L H, Post W M, Baldocchi D D, et al. 2009. Characterizing the seasonal dynamics of plant community photosynthesis across a range of vegetation types[M]//Noormets A. Phenology of ecosystem processes. New York, NY: Springer: 35-58.
- Jönsson P, Eklundh L. 2002. Seasonality extraction by function fitting to time-series of satellite sensor data[J]. IEEE Transactions on Geoscience and Remote Sensing, 40(8): 1824-1832.
- Jönsson P, Eklundh L. 2004. TIMESAT: A program for analyzing time-series of satellite sensor data[J]. Computers & Geosciences, 30(8): 833-845.
- Julien Y, Sobrino J A, Verhoef W. 2006. Changes in land surface temperatures and NDVI values over Europe between 1982 and 1999[J]. Remote Sensing of Environment, 103(1): 43-55.
- Musial J P, Verstraete M M, Gobron N. 2011. Technical Note: Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series[J]. Atmospheric Chemistry and Physics, 11(15): 7905-7923.
- Reed B C, Schwartz M D, Xiao X M. 2009. Remote sensing phenology[M]//Noormets A. Phenology of ecosystem processes. New York, NY: Springer: 365-381.
- Sakamoto T, Yokozawa M, Toritani H, et al. 2005. A crop phenology detection method using time-series MODIS data[J]. Remote Sensing of Environment, 96(3-4): 366-374.
- Vautard R, Yiou P, Ghil M. 1992. Singular-spectrum analysis:

A toolkit for short, noisy chaotic signals[J]. *Physica D: Nonlinear Phenomena*, 58(1-4): 95-126.
Yu F F, Price K P, Ellis J, et al. 2004. Satellite observations of

the seasonal vegetation growth in Central Asia: 1982-1990 [J]. *Photogrammetric Engineering & Remote Sensing*, 70 (4): 461-469.

Comparison and application of NDVI time-series reconstruction methods at site scale on the Tibetan Plateau

LIU Jianwen^{1,2}, ZHOU Yuke^{3*}

(1. Key Laboratory of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fuzhou 350003, China; 2. Spatial Information Research Center of Fujian, Fuzhou 350003, China; 3. Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China)

Abstract: Vegetation index data based on satellite remote sensing have been widely applied in many fields such as vegetation monitoring and biomass estimation. However, the noise caused by the sensors, atmospheric conditions, and environmental factors affect the application of such data. Therefore, it is of practical significance to carry out research on the reconstruction of vegetation index. In this article, MODIS normalized differential vegetation index (NDVI) datasets of 106 meteorological stations on the Qinghai-Tibet Plateau with typical vegetation types, from 2000 to 2015, were reconstructed using three methods: cubic spline function (Spline), double logistic function (D-L), and singular spectrum analysis (SSA). Based on the results of phenological parameter extraction and NDVI time-series reconstruction, the ability of preserving the authenticity of the original data, detail fitting, and phenological character extraction of the three algorithms are compared and analyzed. The result suggests that there is no single method that performed the best for all types of vegetation, owing to the spatial heterogeneity of vegetation cover types and the varying denoising ability of the algorithms. D-L showed better performance for desert, shrubs, and woodland, which are heavily affected by random noise from snow and clouds, whereas for grassland and cropland, Spline had better performance. The fitting curves of SSA are below the D-L and Spline curves and the lower the NDVI value, the worse the reconstruction performance. With regard to maintaining the true values of the original data, Spline is superior to the D-L method for woodland, shrubs, and grassland, while the SSA is superior to the D-L function for woodland. The D-L method resulted in an earlier phenological period, Spline and SSA had a lagged outcome of the start of the growing season (SOS) and end of the growing season (EOS) respectively. Phenological index is unstable for shrubs and woodland and fluctuate irregularly for desert with all three methods, due to the low absolute value of NDVI. In addition, the phenology curve of spline is similar to that of SSA, and the SOS and EOS index derived from the D-L method is ahead of other methods for all vegetation cover types. The EOS extracted by the SSA method is larger than that by other methods for most of the vegetation cover types. This research could provide a reference for the selection of noise reduction methods for NDVI time-series data with different vegetation cover types in plateau vegetation.

Key words: time-series data reconstruction; vegetation phenology; double logistic fitting; cubic spline; singular spectrum analysis