

MODIS NDVI 时间序列数据的去云算法比较

梁守真^{1,2}, 施平¹, 邢前国¹

(1 中国科学院烟台海岸带研究所, 烟台 264003; 2 中国科学院研究生院, 北京 100039)

摘要: 受多重因素的影响, MODIS NDVI 数据产品中存在着大量的噪声, 需要进行去噪重建。针对目前几种常用的 NDVI 时间序列数据去云方法, 如 HANTS 法、SPLNE 插值法以及 Savitzky-Golay 法, 以山东省 MODIS NDVI 时间序列数据 (一年的) 作为检验数据, 从不同角度比较几种算法的去云能力和使用范围。结果表明: SPLNE 插值法的去噪效果取决于云掩模数据的质量, 但有时会产生异常值; HANTS 算法和 Savitzky-Golay 算法会改变几乎所有像元的值, 得到一个比较平滑的时间序列曲线, 但这两种算法的输入参数没有统一标准, 需多次试验才能确定最优参数。

关键词: NDVI 时间序列; HANTS; SPLNE; Savitzky-Golay (S-G)

中图分类号: TP 75 **文献标识码:** A **文章编号:** 1001-070X (2011) 01-0033-04

0 引言

NDVI 是监测地表植被活动简单、有效的度量指标。目前基于 NOAA/AVHRR、MODIS、SPOT/VEGETATION 等卫星遥感数据得到的 NDVI 时序数据已经在植被动态变化监测、宏观植被覆盖分类和植物生物物理参数反演方面得到了广泛应用^[1]。应用中低分辨率传感器对地面进行观测时, 太阳光照角度、观测角度以及云的覆盖状况随时间变化很大, 因此得到的观测值包含了大量的噪声。尽管观测数据经过严格的预处理, 并采用最大值合成法 (Maximum Value Composite, MVC) 或者限制视角的最大值合成法 (Constrained-View Angle Maximum Value Composite, CVMVC) 将多天的 NDVI 数据进行合成, 但是得到的产品仍然受云、气溶胶及水汽等的影响。尤其是当合成期内一直有云存在的情况下, 云成为对 NDVI 产品影响最大的噪声。这些噪声的存在使得合成后的 NDVI 随时间变化呈无规律状态, 相邻值高低变化没有规律, 难以进行应用。因此, 这些 NDVI 产品需要进行进一步去云处理, 重建高质量的 NDVI 数据。

在图像处理中, 滤波去除噪声方法是一种二维的空域处理; 时间序列数据的噪声去除是将同一个目标像元的时间序列作为处理单元, 属于一维时间域信号处理。目前已有多种时间序列去噪 (云) 方法被提出, 主要包括 HANTS 算法^[2,3]、SPLNE 插值法^[4]、Savitzky-Golay 法^[5]、滑动平均法^[6]以及中值

滤波法^[7]等。利用 HANTS 算法、SPLNE 插值法和 Savitzky-Golay 法进行植被指数时间序列去噪的技术近些年才出现, 目前还未见对它们的除噪效果进行比较的文章出现。为了更好地应用这些算法, 有必要对它们的除噪能力进行比较。

1 实验数据和方法

采用的实验数据为山东省一年内的 MODIS A2 产品。该产品共有 23 幅图像, 缩放尺度为 10 000 有效值范围为 -2 000~10 000 填充值为 -3 000。

1.1 HANTS 算法

HANTS (Harmonic Analysis of Time Series) 算法是以傅立叶变换为基础的谐波分析法, 它将 NDVI 时间序列表示为不同相位、频率和幅度的正弦函数组合。其中, 植被的生长过程可用几个低频正弦函数描述, 而 NDVI 图像中以斑点形式出现的云被认为是高频噪声。该算法的具体实现过程如下:

首先, 针对每个像元点的时间序列进行傅立叶变换; 然后, 选择几个低频分量进行反傅立叶变换, 得到一个新的序列; 计算原始时间序列和新序列的差值, 如果差值大于设定的阈值, 那么该点将被认为是受到了污染, 便从原始序列中去掉, 用新序列中对应的值来填充; 对改变的原始序列重复上述过程, 直到没有受云污染的点被找到或者达到设定的迭代结束条件为止。该过程的输出结果为一平滑曲线。

HANTS 算法在进行 NDVI 时间序列处理时, 需要设置频率个数、误差阈值、最大删除点个数及有效

收稿日期: 2010-05-20; 修订日期: 2010-07-14

基金项目: 中国科学院国际合作项目 (编号: GJHZ200811); 山东省科技发展计划项目 (编号: 2007GG2QT06019); 中国科学院信息化项目 (编号: NFO-115-D01-Z005); 国家自然科学基金项目 (编号: 30800149)。

数据范围等参数,这些参数的设置没有客观标准,只能根据经验或多次试验来确定。本研究采用的各参数值(或数据范围)分别为 4、1500、8及 0~10 000。

1.2 SPLNE 插值法

SPLNE 是一种常用的时间和空间插值方法。它采用 SPLNE 函数对节点范围内所有区间进行插值,能使用低阶多项式样条而产生和高阶多项式插值类似的效果,并且避免了高阶多项式插值出现的龙格现象。目前应用较多的是三次样条插值,具体定义如下:

设在区间 $[a, b]$ 上取 $n+1$ 个节点 $a = x_0 < x_1 < x_2 < \dots < x_n = b$ 函数 $y = f(x)$ 在各个节点处的函数值为 $y_i = f(x_i) (i = 0, 1, \dots, n)$ 。若 $S(x)$ 满足:

- (1) $S(x_i) = y_i$;
- (2) 在区间 $[a, b]$ 上, $S(x)$ 有连续的二阶导数;
- (3) 在区间 $[x_i, x_{i+1}] (i = 0, 1, \dots, n-1)$ 上, $S(x)$ 是 x 的三次多项式;

则称 $S(x)$ 是函数 $y = f(x)$ 在 $[a, b]$ 上的三次 SPLNE 插值函数。

SPLNE 插值算法只能对标识为受云影响的像元进行插值,因此,首先要确定受云影响的像元。MODIS NDVI 数据集中包含有 Pixel Reliability 和 V I Quality 数据,可以用来确定被云覆盖的像元。

1.3 Savizky-Golay 滤波法

Savizky-Golay(简称 S-G)滤波法是由 Savizky 和 Golay 提出的一种简便的、基于最小二乘法的卷积算法^[8]。该算法主要基于以下两个前提假设: ① 序列的变化趋势应服从植被年际动态变化的渐进特征; ② 由于云和大气对 NDVI 的影响一般为负偏移,所以多数噪声应低于 NDVI 序列数据的平均值。S-G 法可以简单地理解为一种权重滑动平均滤波,其权重取决于在一个滤波窗口范围内做最小二乘拟

合的多项式。在拟合时,一些远离大多数点的边沿点不会参与拟合。这样,拟合时过于偏离正常生长趋势线的噪声部分会被丢弃。

$$Y_j^* = \sum_{i=-m}^m C_i Y_{j+i} / N \quad (1)$$

式中, Y 是原始 NDVI 值; Y^* 是平滑后 NDVI 值; j 表示原始 NDVI 时间序列的序号; C_i 是从滤波窗口首部开始第 i 个 NDVI 值的权值; N 是滤波窗口的大小(为 $2m+1$), m 为滤波窗口长度的一半。

S-G 法需要人为设定滤波窗口宽度和多项式拟合阶数两个参数,对滤波器窗口大小敏感,如果滤波窗口的宽度设置偏小,容易产生大量冗余数据,不易获取数据的长期变化趋势;反之,又容易遗漏一些细节所描述的信息。一般来讲,阶数越小,处理后的结果越平滑,但拟和效果差;阶数越高,拟和效果越好,但有可能出现过拟合现象,而且在数据点会有震荡。通常情况下阶数取 2~4,滤波窗口宽度取 7~15(m 为 3~7)就能满足要求^[5]。这样就有 15 种组合,需要分别进行处理比较,以便选择一组处理效果较好的参数。实验发现,当 $m > 5$ 以后,时间序列就很难体现一年两熟农作物的季节变化;而当 $m = 4$ 时,能体现一年两熟的季节变化,但是波动太小,不符合实际情况。最终,本研究取的滤波窗口大小为 7,即 $m = 3$ 阶数为 2。

2 结果比较

2.1 视觉比较

选择云覆盖量最大的第 14 个(即第 209~224 d 的合成数据)MODIS NDVI 合成图像。图 1(a)、(b) 分别为合成期的云覆盖图和 NDVI 图。

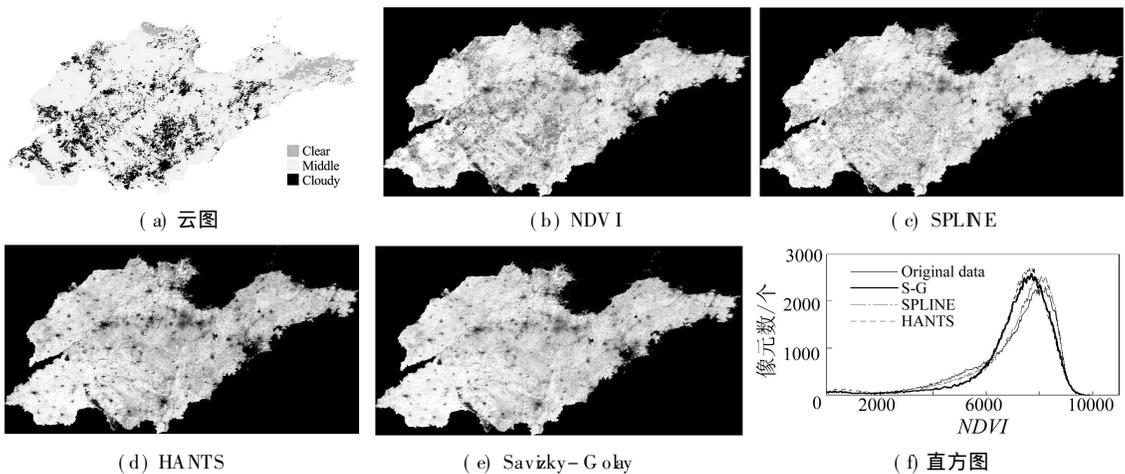


图 1 实验数据

Fig. 1 Test data

由图 1 可见, HANTS 和 S-G 算法处理得到的影像(图 1(d)、(e))整体显得较为平滑,原始图像中被云覆盖而导致 $NDVI$ 值偏低的像元已经得到恢复;经过 SPLINE 插值以后的图像(图 1(c))仍然存在一些暗斑点,且还有零星的亮斑出现。但是,仅凭这一点并不能说明 HANTS 算法和 S-G 算法比 SPLINE 插值法在去噪处理上更有竞争性。其实, SPLINE 插值效果较差的部分原因是由于云掩模数据的不准确造成的,因为 SPLINE 插值只能对设定的点进行插值处理,其本身无法识别哪些像元受到云的干扰。此外,本文还对比了不同算法处理后图像的直方图(图 1(f))。可以看出, SPLINE 插值得到的图像直方图和原始图像的直方图最为接近,但

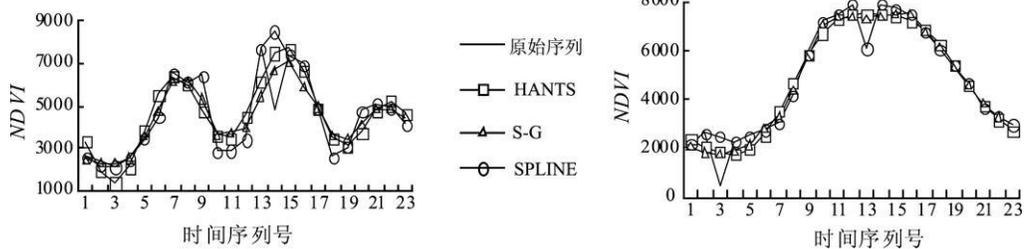


图 2 农田(左)和林地(右)的原始时间序列和不同算法结果序列

Fig 2 Original time series and resulting time series of cropland (left) and (right) forest

列值变得较为平滑,尤其是 HANTS 算法,而 SPLINE 插值处理所得的时间序列值相对比较粗糙。但前两种算法处理后的时间序列值已和原始序列完全不同,这在几个波谷处表现得最明显。其中,第一个波谷偏低,而其他两个波谷较原始数据偏高。也就是说, S-G 和 HANTS 算法会改变几乎所有原始序列的值,只是保持整体趋势和原始序列一致。SPLINE 插值所得结果比较接近原始序列,除了受云影响的像元外,其他像元值保持不变。此外,在农作物序列的第 8 个合成期 $NDVI$ 也有一个小的下降,这不太符合健康作物正常的生长规律,有可能该时期作物生长受到了环境的胁迫,也有可能是受薄云的影响。HANTS 和 S-G 算法认为它是受云影响而对其校正,但在 SPLINE 插值结果中并没有校正,原因是在云掩模数据中该点并没有被标识为云,同样的现象也出现在林地的时间序列中(图 2 右)。林地序列中有两个点的 $NDVI$ 值明显低于相邻点的值, HANTS 算法和 S-G 算法对这两点都进行了校正,而 SPLINE 插值只处理了其中的一个点,同样是因为在云掩模数据中有一个点没有被标识为云,而无法对其进行校正。

从图 1(c) 可以看到, SPLINE 插值后的图像有亮斑出现,其像元值超出了 $NDVI$ 的有效范围(图

最大值已经超过 10 000,说明部分插值存在错误; S-G 算法和 HANTS 算法的图像直方图相似,说明两者所得的结果比较接近,但它们偏离原始数据较大。

2.2 像元比较

更为详细地分析和比较不同算法的处理效果,需要观察典型地物像元的时间序列在处理前后的变化情况。山东省典型土地利用类型为农田和林地,在选择典型地物的纯净像元时,参考了 1:25 万土地覆盖遥感调查与监测数据库。

从图 2(左)可以看出,农作物在第 14 个合成图像中明显受云影响, $NDVI$ 值较 2 个临近时期值明显低, 3 种方法都能较好地去除云的影响。其中经 S-G 和 HANTS 算法处理以后地物的 $NDVI$ 时间序

3)。经过对像元时间序列分析对比,发现这种情况一般出现在时间序列的两端(冬季)有连续 2 个以上的像元都受到云或冰雪影响的时候。可能是因为在这样的条件下, SPLINE 算法缺乏足够的有效数据进行插值运算,受云影响的像元只能通过外推的方式来计算,从而导致了异常值的出现。因此,这种情况下的图像应避免使用 SPLINE 算法进行去云或重建,或者对 SPLINE 插值后的时序图像再进行二次处理,用其他方法校正这些异常值。

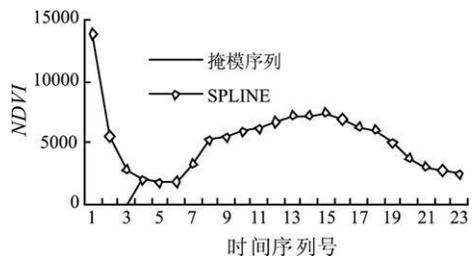


图 3 SPLINE 算法插值异常

Fig 3 Abnormity of SPLINE interpolation

3 结论

(1) SPLINE 插值结果强烈依赖于云掩模数据,只有在高质量的云掩模数据存在的前提下,该方法才能获取较好的效果,当缺乏足够的有效数据进行

插值时,会产生异常值。HANTS和 S- G 算法则不需要云掩模数据。

(2) SPLNE 算法不会改变无云覆盖的像元值,而其他两种方法处理之后图像的所有像元值都会发生改变,新的像元时间序列值较原始序列值平滑,尤其是 HANTS 算法。

(3)HANTS 算法和 S- G 算法对所需参数的设定没有统一标准,需多次试验对比才能确定合适参数,而 SPLNE 算法则不涉及参数设置问题。

(4)HANTS 和 S- G 算法处理后得到的结果相似,因此很难判断哪种方法更好。但根据目前出版的文献,前者应用得更多。这可能是因为 HANTS 算法不仅能产生时间间隔的无云图像,同时还产生频率组分图像(振幅和相位图),这些组分可用于各种分析。

(5)每种算法都有自己的优点和缺点,应该根据应用的目的来选择时间序列的去噪(云)算法。比如平滑的时间序列曲线有利于提取植被的物候信息,如果要提取物候信息,HANTS 算法是一个较好的选择。但是,如果要定量分析和提取植被的生物学参数,最好选择 SPLNE 算法(要排除异常值),因为该算法不会改变序列的正常值。

致谢: 特别感谢国家科学数据共享工程——地

球系统科学数据共享网 (www. geodata. cn) 提供土地覆盖数据。

参考文献:

- [1] 顾娟,李新,黄春林. NDVI 时间序列数据集重建方法述评 [J]. 遥感技术与应用, 2006 21(4): 391-395
- [2] Roerink G J, Menenti M, Verhoef W. Reconstructing Cloud Free NDVI Composites Using Fourier Analysis of Time Series [J]. International Journal of Remote Sensing 2000, 21(9): 1911-1917.
- [3] 王丹,姜小光,唐伶俐,等. 利用时间序列傅立叶分析重构无云 NDVI 图像 [J]. 国土资源遥感, 2005(2): 29-32
- [4] Mucher C A, de Baets E P J, Gobal Land Cover 2000 Evaluation of the SPOT VEGTATION Sensor for Land Use Mapping [R]. Wageningen Alterra Green World Research, 2002
- [5] Chen J, Jönsson P, Tanura M, et al. A Simple Method for Reconstructing a High-quality NDVI Time-series Dataset Based on the Savitzky-Golay Filter [J]. Remote Sensing of Environment 2004, 91: 332-344
- [6] Malinreau J P. Global Vegetation Dynamics Satellite Observations over Asia [J]. International Journal of Remote Sensing 1986, 7: 1121-1146
- [7] Kogan F, Sullivan J, Cary R, et al. Post-fire Pinatubo Vegetation Index in Central Africa [J]. Geocarto International 1994, 9: 63-66
- [8] Savitzky A, Golay M J E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures [J]. Analytical Chemistry 1964, 36: 1627-1639

A Comparison Between the Algorithms for Removing Cloud Pixel from MODIS NDVI Time Series Data

LIANG Shou-zhen^{1,2}, SHI Ping¹, XING Qian-guo¹

(1. Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003, China;

2. Graduate University of Chinese Academy of Sciences, Beijing 100039, China)

Abstract Although composite data present lower atmospheric contamination than raw time series, MODIS NDVI products are still contaminated by clouds, especially when cloud cover lasts longer than the composite period (e.g., in the rainy season). The long-time cloud cover will weaken the application of MODIS NDVI time series data. To remove the effect of these clouds from NDVI data and reconstruct high-quality NDVI data, the authors propose three algorithms for cloud removal, namely SPLNE function, HANTS and Savitzky-Golay. The capabilities of the three algorithms in cloud removal was compared with each other in this study, with the MODIS NDVI time series data in Shandong province serving as the test data. The results show that the three algorithms can remove the effect of cloud from NDVI time series data effectively, with each algorithm having its own advantages and disadvantages. For the algorithm of SPLNE function, the result of cloud removal mainly depends on the quality of cloud data and sometimes extreme values will occur; this algorithm fails to change the values of pixels which have not been contaminated atmospherically. When HANTS and Savitzky-Golay algorithms are used, most of the pixels will lose their original values, and the parameters have to be determined after conducting many experiments because there is no objective rule to set them.

Key words NDVI Time series; HANTS; SPLNE; Savitzky-Golay(S-G)

第一作者简介: 梁守真 (1979-), 男, 博士研究生, 主要从事遥感图像处理和环境遥感研究。

(责任编辑: 刁淑娟)