

分类号_____

密级_____

UDC _____

编号_____

中国科学院研究生院

硕士学位论文

遥感植被指数时间序列数据滤波重建算法研究

李 儒

指导教师 张 兵 研究员 中国科学院对地观测与数字地
球科学中心

郑兰芬 研究员 中国科学院遥感应用研究所

童庆禧 院 士 中国科学院遥感应用研究所

申请学位级别 硕士 学科专业名称 地图学与地理信息系统

论文提交日期 2008 年 5 月 论文答辩日期 2008 年 5 月

培 养 单 位 中国科学院遥感应用研究所

学位授予单位 中国科学院研究生院

答辩委员会主席_____

M.S. Dissertation

**STUDY ON ALGORITHM FOR RECONSTRUCTION OF
THE VEGETATION INDEX TIME SERIES OF REMOTELY
SENSED DATA**

By

LI Ru

Directed by **Prof. Zhang Bing**

Prof. Zheng Lanfen

Prof. Tong Qingxi

Submitted in fulfillment of the requirements for the degree of

Master of Science

in

The State Key Laboratory of Remote Sensing Science,
Institute of Remote Sensing Applications, Chinese Academy of Sciences
Beijing

June 2008

中国科学院遥感应用研究所

学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明： 所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名： 日期： 年 月 日

学位论文使用授权说明

本人完全了解中国科学院遥感应用研究所关于收集、保存、使用学位论文的规定，即：
按照学校要求提交学位论文的印刷本和电子版本；
研究所有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务；
研究所可以采用影印、缩印、数字化或其它复制手段保存论文；
在不以赢利为目的的前提下，研究所可以公布论文的部分或全部内容。

（保密论文在解密后遵守此规定）

论文作者签名： 导师签名：

日期： 年 月 日

摘 要

源于卫星遥感数据的植被指数时间序列数据能够从不同尺度（从全球到区域）反映植被生长状况、地表覆盖变化等信息，因而有着十分广泛的应用，已经成为许多科学研究、工程项目的重要数据源之一，如应用于植被生长监测，土地覆盖类型解译及其监测，物候特征识别与信息提取，全球、大洲、区域等尺度的陆地生态系统建立，甚至植物物候变化对 GPP 和 NPP 的影响研究等。应用此类数据须重视数据序列的时空一致性问题，影响数据序列时空一致性的主要因素有卫星遥感器本身的原因、云、大气条件、视场角和几何校正等。针对这些问题，学者们已经发展出一些算法（如最大合成法），但处理后的数据产品仍然存在着比较严重的噪声残差，这制约了其深入应用，因此在应用前必须正确有效的去除残存噪声、重建时间序列。为了有效的解决此类问题，研究者发展了许多算法，包括最佳指数斜率提取法、基于非对称高斯函数拟合法、基于 Savitzky-Golay 滤波的拟合法、傅立叶谐波分析法等。但这些方法都存在着一些问题，最突出的就是关键参数需要人为通过实验、经验选择，因而引入后天误差、导致对原始数据的干扰。在众多方法中，傅立叶谐波分析法由于使用不同周期的谐波在拟合时间序列曲线时能通过谐波周期来模拟一定的物候规律，而非单纯的数学运算，因而更受关注。

本文即在此类方法基础上，通过引入新的参量或策略改进原方法，试图减少人为影响及其对原始数据的过度扰动，使算法运行结果更为客观。改进算法的主要思路为：

（一）.时间序列数据滤波重建原则的提出，即 a.尽可能减少对数据过度扰动，滤波重建算法对原始数据扰动越大，其结果可信度越低，某些重要规律甚至会因此而被抹杀；b.尽可能少的使用辅助数据，辅助数据的使用一方面为滤波重建提供了一定的指示，如 MODIS 云图像，但由于数据获取、生产等过程中的不确定性因素，使辅助数据本身就存在类型复杂的误差，它的使用一定程度上增加程序开销、可能引入新的误差（或不确定性）；c.尽可能多的考虑参量物理意义，植被指数时间序列的优势在于可以通过对其分析提取一定的季相物候等规律，因此滤波重建过程能考虑此问题将使数据处理更有意义，这也是傅立叶谐波分析系列算法的优势所在；d.算法效率与效果的兼顾。

（二）. 通过设置异常值检测增加数据拟合的真实性。由于残存误差的存在，

真实值和异常值难以通过表面得以区分。传统方法通过设置阈值处理突变值（疑似异常值）虽然也有效，但难免出现“误杀误放”的情况，因此有必要从数据内部入手，通过分析数据内在规律探测序列中异常值的存在。改进算法引入异常值检测算法，检测拟合迭代过程中的异常值，减少因传统方法根据距离定权剔除无效点而引入的人为误差和新的不确定性。

（三）. 频数的自动选择和拟合影响因子的自动迭代计算。频数控制拟合效果，也是傅立叶谐波分析方法体现序列数据物理意义的载体。但传统方法通过尝试人为设定、全局使用同一固定频数，即默认整个运算区域都是同一类植被生长模式（从物理意义上考虑）；改进方法通过迭代前预处理，动态估算出待处理序列点的峰值个数，即频数，参与下步拟合。另一个关键参数是本文借鉴基于 Savitzky-Golay 滤波的拟合法中拟合影响因子，计算每次拟合的影响因子，在该因子达到最小值（或局部最小值）时自动终止迭代，而放弃使用传统方法中人为设置的阈值限差。

通过与 HANTS 软件运算结果对比，本文分别从数据扰动、异常值检测、自动选频结果、迭代次数等方面分析，说明本文提出的改进方法取得了一定的试验效果。改进算法结果较 HANTS 软件结果在某些方面，如对真实数据扰动影响等，都有明显的改进。但是在算法普适性等方面，本研究仍需要进一步的试验验证。同时在分析过程中，本文还发现了 HANTS 软件运算结果呈负偏性趋向（拟合值小于原始值），拟合曲线呈低于原始曲线的趋向，这对以后应用 HANTS 软件时有一定的帮助作用。

最后，本文使用滤波重建后的植被指数时间序列，分析提取北京及周边地区两类裸地，即常年裸地和季节性裸地。经分析发现，北京西北部存在大量的常年裸地，季节性裸地也零星存在。该区处于北京西北风向风口，又是沙尘进入北京的必经之路，成为北京周边沙尘源地可能性极大。北京南部门头沟、房山、大兴存在着大量季节性裸地。该区冬春季节耕地裸露，又处在北京南部北风向风口，也极可能是北京本地沙尘源地之一。

关键词：植被指数时间序列，傅立叶谐波分析，HANTS，频数，异常值检测

ABSTRACT

Because vegetation index time series of remotely sensed data, such as Normalized Difference Vegetation Index (NDVI) products derived from NOAA/AVHRR, SPOT/VEGETATION, TERRA, or AQUA/MODIS, carry valuable information regarding land-surface properties in kinds of scales, they become more and more important and even one of main data sources for lots of applications including scientific researches and engineering projects. Take NDVI data set for example, various different scales of these products have been applied for detecting long-term land use/cover changes, modeling terrestrial ecosystems on global, continental and regional scales, extracting seasonal metrics of vegetation phenology to classify vegetation or land cover types, and even for estimating gross primary productivity (GPP) and net primary productivity (NPP).

However, since disturbed by cloud contamination, atmospheric variability and bi-directional effects, the vegetation index time series of remotely sensed data have serious noise. Although the most frequently-used data sets are the MVC products, such as MODIS 16-days NDVI/EVI data sets, they still include a lot of such serious residual noise. For this reason, many methods for reconstructing high-quality time-series data sets have been developed, including Best Index Slope Extraction Algorithm (BISE), Asymmetric Gaussian Function Fitting Approach (AGFF), Algorithm based on Savitzky-Golay Filtering(S-GF), Harmonic Analysis Algorithm based on Fourier transform (HAA), and so on. However, these methods also suffer several drawbacks that limit their further applications. The most serious problem for them is that the key parameters, almost for all of these algorithms, need to obtain through lots of trials which are easily influenced and brought new kinds of noise or uncertainties to the data sets by operators. Among these algorithms, because Harmonic Analysis Algorithm takes account of the physical meanings represented by the original data while fitting the curve, it associates well temporal changing regulation and spatial distribution characteristics showed on the time-series data sets. Therefore, based on this algorithm, a new method for reducing residual noise and

constructing high-quality time-series data sets for further application has been given in this paper. The new improved algorithm tries to bring less subjective noise in the data sets through the following ways:

1. Four rules have been given to guide the improvement of the algorithm: fewer disturbances to the original sets, less auxiliary data used; more consideration of the data physical meanings, good balance between effect and efficiency.

2. Outlier detection algorithm has been used to find the data points which are not proper to join the next curve fitting.

3. Key parameters automatically generating instead of obtaining by trials and experience. There are two important parameters, Numbers of Frequency (NOF) and Fitting-effect Index (F_k). The former is used to control the result of fitting, and it also reflects some phonological regulations implied in the data sets. The latter decides when to terminate the iteration.

Comparing to the HANTS result by analyzing disturbance to the data sets, detecting outlier and generating key parameters automatically, it can be concluded that the improved method has a good performance in some aspects such as fewer disturbance to the original data, while new problems, applicability of algorithm for instance, have been found which need to keep improving further.

Finally, the vegetation index time series of remotely sensed data have been applied to detect the source regions of dust weather. After analysis of the detected result and local meteorological data, two possible dust-sand sources regions of Beijing dust weather have been found. One is the junction of Inner Mongolia, Hebei and Shanxi in northwest Beijing. Another is in north Beijing, mainly is in Mengtougou country, Fangshan country and Daxing country. More attention should be paid here to find a balance between Economic and ecological benefits which will be helpful to improve Beijing dust weather.

Keywords: the vegetation index time series of remotely sensed data, Harmonic Analysis, HANTS, number of frequency, outlier detection

目 录

第一章 绪 论.....	1
1.1 研究目的及其意义.....	1
1.2 相关领域发展概述.....	2
1.2.1 遥感植被指数及其应用概述.....	2
1.2.2 遥感植被指数时间序列滤波重建算法概述.....	3
1.3 本文研究目标与研究内容.....	5
第二章 三类重要遥感植被指数时间序列滤波重建算法评述.....	7
2.1 基于 Savitzky-Golay 滤波的拟合算法.....	7
2.2 非对称高斯拟合法.....	8
2.3 傅立叶谐波分析法.....	9
2.4 重建算法分析评述.....	11
2.5 本章小节.....	11
第三章 基于傅立叶谐波分析算法的改进算法研究.....	13
3.1 傅立叶谐波分析算法原理及优缺点分析.....	13
3.2 傅立叶谐波分析算法改进设计.....	19
3.2.1 时间序列数据重建原则.....	20
3.2.2 数据异常值检测策略.....	21
3.2.3 算法关键参数的自动选择.....	22
3.2.4 算法终止策略的选择.....	23
3.3 改进算法技术流程设计.....	25
3.4 算法实现及对比试验分析——与 HANTS 软件结果对比.....	25
3.4.1 数据源介绍.....	25
3.4.2 相关参数确定.....	27
3.4.3 结果整体评价.....	29
3.4.4 算法对序列数据扰动分析.....	33
3.4.5 异常值处理结果分析.....	43
3.4.6 自动选频 (NOF) 结果分析.....	45
3.4.7 迭代终止条件分析.....	47
3.5 本章小节.....	48
第四章 基于植被指数时间序列的北京及周边沙尘源地提取.....	51
4.1 北京沙尘源地综述.....	51
4.2 数据源介绍.....	52
4.3 北京及周边地区两类裸地提取.....	53
4.3.1 土地分类规则确定.....	53
4.3.2 基于植被指数时间序列两类裸地提取.....	54
4.4 可能沙尘源地分析.....	55
4.4.1 可能沙尘源地分析.....	55
4.4.2 对策分析.....	56
4.5 本章小节.....	56
第五章 结论与展望.....	59
5.1 结论.....	59

5.2 展望	60
参考文献	61
致 谢	67

第一章 绪 论

1.1 研究目的及其意义

源于卫星遥感数据的植被指数时间序列数据 (MODIS TERRA/AQUA、NOAA/AVHRR、SPOT/VGT 植被指数时间序列等) 能够在不同尺度 (从全球到局地) 反映植被生长状态、地表覆盖变化等信息 (Stockli R. et al, 2002; Zhou L. et al, 2001; 张峰等, 2004), 因而有着十分广泛的应用, 已经成为许多科学研究、工程项目的重要数据源之一 (王正兴等, 2003; IGBP, 1992; Justice, C.O. et al, 1985; Myneni, R.B., et al., 1997; Reed, B.C., et al., 1994; Holben et al., 1986)。这类应用包括植被生长监测 (张霞, 2006; 范锦龙等, 2004), 长时间尺度内土地覆盖类型解译及监测 (Bethany et al., 2007; Swets et al., 1999), 物候特征识别与信息提取 (Jing F. et al., 2002; 张峰等, 2004), 全球、大洲、区域等尺度的陆地生态系统模型的建立 (Moulin et al., 1997; Chen X et al., 2000), 甚至植物物候变化对 GPP 和 NPP 的影响研究等 (White et al., 1999)。

应用此类数据之前, 必须重视数据序列的时空一致性问题。影响数据序列时空一致性的主要因素有云、大气条件、视场角和数据几何校正等 (Gutman et al., 1995; Gutman, 1991)。针对这些问题, 学者们已经开发出了一些方法, 并应用于实际生产。但由于卫星传感器自身性能、云层、大气条件等影响, 这些时间序列产品仍然存在着比较严重的噪声残差, 这制约了其深入应用, 因此应用前必须正确、有效的去除残余噪声、重建时间序列 (林忠辉等, 2006; Cihar, J., 1997)。

为了有效解决这个问题, 研究者发展了许多方法 (顾娟等, 2006), 但这些方法都存在相应的缺陷, 最突出的就是关键参数需人工通过实验、依靠经验选择, 因而引入后天人为主观误差、导致对原始数据的 (不当) 扰动。这一缺陷制约了遥感植被指数时间序列的深入应用。

1.2 相关领域发展概述

1.2.1 遥感植被指数及其应用概述

植被指数即应用多（高）光谱遥感数据时，经分析有目的的选择某些波段通过运算（加、减、乘、除等线性或非线性组合），产生某些对植被长势、生物量等有一定指示意义的数值。（赵英时，2003）。其理论依据是叶片的反射光谱特征。研究表明，植被指数有助于增强遥感影像中绿色植被的解译力，并已作为一种重要的遥感数据处理手段广泛应用于土地利用覆盖探测、作物物候关键期预测、作物分类识别和作物产量预报等方面（张霞，2006）。

由于植被光谱受到植被本身、环境条件、大气状况等多种因素影响，因此植被指数往往具有明显的地域性和时效性。到目前为止国内外学者已研究发展了几十种不同的植被指数模型，大致可以归纳为以下几类：比值植被指数（Ratio Vegetation Index, RVI）、归一化植被指数（Normalization Difference Vegetation Index, NDVI）、调整土壤亮度的植被指数（SAVI、TSAVI、MSAVI）、差值植被指数（Difference Vegetation Index, DVI）、穗帽变换中的绿度植被指数（GVI）、垂直植被指数（Perpendicular Vegetation Index, PVI）、其他的一些植被指数，如叶绿素吸收比值指数等。这些植被指数各有优势，适用于不同的研究领域中，解决特定的问题（赵英时等，2003）。

本文关注增强型植被指数（Enhanced Vegetation Index, EVI）：

$$\text{MODIS_EVI} = \frac{G(\rho_{\text{NIR}} - \rho_{\text{Red}})}{(L + \rho_{\text{NIR}} + C_1\rho_{\text{Red}} - C_2\rho_{\text{Blue}})} \quad (1-1)$$

其中 ρ_λ 是给定波长（波段）的反射率（ $\lambda = \text{Red}, \text{NIR}, \text{Blue}$ ），该增强型植被指数的目的是改善植被指数对大范围全球尺度植被状况的敏感度。公式中 L 是冠层背景的调整因子（或称土壤调节系数）； C_1 ， C_2 是权重系数，描述了用蓝光通道对红光通道进行大气气溶胶散射的修正， G 是增益因子（Huete, A et al., 2002）。

EVI 数据有十分广泛的应用：邹金秋等（2007）利用多时相 Terra/MODIS EVI 数据提取 2004 年冬小麦播种面积；徐永明等（2007）利用 250m 分辨率 MODIS EVI 数据进行土地覆盖制图研究；彭代亮等（2007）利用 2001-2004 年 MODIS EVI 数据和 52 个气象站点的日平均气温和日降水量，结合区域土地利用现状数

据,采用时滞互相关分析方法,分析耕地、林地、园地增强型植被指数(EVI)季节变化与气温、降水的相关性;张霞等(2005)、胡晓等(2006)也都分别使用EVI数据展开了多项研究。

1.2.2 遥感植被指数时间序列滤波重建算法概述

应用遥感植被时间序列首要的问题即数据集的重建工作,又称为时间序列的滤波重建。尽管在生成各种植被指数时间序列产品时(如MODIS NDVI、EVI时间序列产品),已经通过相应的算法处理噪声,但这些数据产品仍然存在比较严重的噪声,在实际应用中还需进一步校正(林忠辉等,2006)。研究者提出了一些方法解决此类问题,如最大合成法(Maximum Value Composite, MVC; Holben B N. et al., 1986)、最佳指数斜率提取法(Best Index Slope Extraction Algorithm, BISE; Viovy N et al.,1992; Lovel J L et al., 2001)、时间窗口的线性内插算法(Temporal Windows Operation, TWO; Park J et al., 1998)、基于非对称高斯拟合法(Asymmetric Gaussian Function Fitting Approach, AGFF; Per Jönsson et al., 2002)、基于Savitzky-Golay滤波拟合法(Savitzky-Golay Filtering, S-GF; Savitzky et al., 1964; 辜智慧, 2003; Chen J. et al., 2004)、局部最大值拟合法(Local Maximum Fitting, LMF; Nagatani, et al., 2002; Sajia Akhter et al., 2007)、傅立叶谐波分析法(Sellers P J et al., 1996; W. W. Immerzeel et al., 2005)等。

(1) 最大值合成法

时间序列数据首先需要处理的是去云和缺损值的拟合,最大值合成法是在生产时间序列过程中解决这一问题的主要方法。它遵循一个假设,即云层的位置是不断变化的,可以理解为在一段时间内任意一处都有可能存在一个没有云层覆盖的晴天,反映在图像上就是该位置在这段时间内的图像DN值(或植被指数)具有某种特性,如在NDVI图像中即表现为最大值。按照事先规定的时间间隔,将数据分类,新图像某点的值为在各组时间图像中该点最大值。这一方法在生产时间序列初级图像产品时去除云噪声和缺损点时很有效,但实际应用中得到的都已经是生成好的合成数据序列,再次使用最大值合成法效果不十分明显,且大幅降低原时间序列的时间分辨率。

(2) 最佳指数斜率提取法

该方法认为时间序列中的突变(突增突降)可能是有云或无云状态的转变,

也可能是传感器视场角发生变换造成的, 因此可能会出现虚假最大值的情况, 故采用滑动合成时段避免上述问题, 从而降低噪声干扰, 识别并替代序列中的噪声。这一方法有两个版本, 一是提出者 Viovy 的原始方法, 另外是 Lovell 和 Graetz 提出的改进算法。但不论是原算法还是改进算法, 都需要根据不同气候条件和地域植被特征, 根据经验尝试选择参数。

(3) 时间窗口的线性内插算法

时间窗口的线性内插算法, 是 Park 在 1998 年提出用来重建 NDVI 时间序列的修正算法。该算法可以在无辅助数据的情况下, 仅利用时间序列本身来检测并消除云和大气影响。它基于 3 个假设: a.真实的 NDVI 不低于观测值; b.在一个给定窗口内给定像元的真实 NDVI 值表现单调递增或者递减; c.在一定时期内, 给定像元的真实 NDVI 不低于同期观测 NDVI 的最大值。该方法的核心问题是时间窗口大小的选择。

(4) 基于非对称高斯拟合法

基于非对称的高斯曲线拟合法是由 Per 等人在 2002 年提出的, 这种方法首先在时间维选择一最大或最小值区间作为局部拟合区间, 然后使用局部拟合模型, 即高斯形拟合函数对这一区间曲线进行分段拟合, 最后再通过全局拟合模型, 将局部拟合结果合并。该方法分段拟合曲线, 保证了拟合结果更符合当前时段的真实特性, 较少受外界干扰, 仅就从这点而言, 优势明显。

(5) 基于 Savitzky-Golay 滤波拟合法

Savitzky-Golay 滤波, 该方法获得关注相对较多, 许多学者对它进行改进、或加入辅助数据以提高其运行精度。这种方法重建的 NDVI 时间序列能够清晰的描述时间序列的长期变化趋势以及局部的突变信息, 且不受数据时间空间尺度和传感器的限制。对该方法产生影响最大的因素是滑动窗口大小的选择, 由于参数也依靠经验获得, 自然也就存在人为扰动。

(6) 基于离散傅立叶分析的系列算法

基于离散傅立叶的分析方法, 也称谐波分析法。该方法将每个独立像素的时间曲线表达为一系列余弦波和一个加性项的和, 每个余弦波都由唯一的振幅和相位确定, 每一项都指明在给定时间内完成循环的个数。将这些连续的谐波项叠加起来就生成一条复杂的曲线, 每一项在原始时间序列曲线的总方差中占有一定百分比。该方法的一些输入参数不能够通过一定客观准则确定, 而是通过经验比较获得, 这使它处于不利地位。但尽管如此, 其应用仍然十分广泛,

是时间序列重建的主要算法。

(7) 局部最大值拟合法

局部最大值拟合方法先进之处在于在傅立叶谐波分析前先做数据滤波，用滤波数据作为谐波分析输入数据集。据称该方法可以有效的去云层、雾霭和其它一些大气影响，在作物种植模式识别应用中效果良好。

该方法分为两步：数据滤波和傅立叶曲线拟合（重建）。在两个相邻等大小（指数据点个数）的区间 A、B 中分别选出其中最大值，则两个区间共同拥有的那个数据点 t_{middle} 的值即为 $\min(\max(A), \max(B))$ ，该值作为 t_{middle} 滤波结果输出值。其滤波基本方程为：

$$d'_t = \text{Min} \begin{bmatrix} \text{Max}(d_{t-w+1}, d_{t-w+2}, \dots, d_t) & t-w+1 \sim t \\ \text{Max}(d_t, d_{t+1}, \dots, d_{t+w+1}) & t \sim w+1 \end{bmatrix} \quad (1-2)$$

d_t ：时间 t 处的观测值； d'_t ：时间 t 处的修正值； w 为滤波窗口。显然该窗口直接控制滤波效果。该方法与傅立叶等算法相比，把数据滤波与曲线拟合分开处理，如果该策略合理有效，将非常具有现实意义。因为在常规的傅立叶曲线拟合中，一大半的时间和精力都耗费在为了数据滤波而进行的迭代运算中，而且傅立叶系列方法最被诟病地方之一也在于滤波迭代终止条件的选择。

1.3 本文研究目标与研究内容

由于遥感植被指数时间序列有着十分重要和广泛的应用，因而对它有效的预处理也就显得至关重要。但遗憾的是目前的这些方法中，除了 MVC 在生成合成时间序列数据初级产品时得到大家基本认可外，其他方法学界尚没有达成共识，研究者都在努力改进或者研究新的滤波方法，希望新方法可以克服现有方法固有的缺陷，有效的滤去噪声、反映出植被真实的生长特性和物候规律等。

(一) 本文研究目标：

本文正是出于上述考虑，试图引入新策略，取众家之长，改进出一种能够最大程度克服目前滤波重建方法弊病，尤其参数非客观选择这一滤波重建方法通病，通过一些方法（或策略）实现重要参数自动生成而非人工实验选取，使时间序列的滤波重建过程更为客观，减少人为对数据扰动、对处理过程的干扰。

(二) 本文研究内容：

傅立叶谐波分析方法“能够充分利用遥感图像存在时间性和空间性的特点，

将其空间上的分布规律和时间上的变化规律联系起来.……它充分考虑了植物生长周期性和数据本身的双重特点”(张霞, 2006, p83), 即该类方法不仅是单纯的数据滤波, 更考虑了数据反映的物理意义, 因而较其他方法优势更为明显。本文以该类方法为蓝本, 尝试改进算法缺陷, 主要研究内容有:

(1) 植被指数时间序列滤波重建原则的提出: 即应尽可能减少对数据扰动, 滤波重建算法对原始数据扰动越大, 其结果可信度越低; 尽可能少的使用辅助数据, 辅助数据本身就存在类型复杂的误差, 它的使用一定程度上会增加程序开销, 并引入新的误差(或不确定性); 尽可能多的考虑参量物理意义, 植被指数时间序列的优势就在于可以通过对其分析提取一定的季相物候等规律, 因此在滤波重建过程时就能考虑此问题将更有意义; 算法效率与效果的兼顾。

(2) 通过设置异常值检测增加数据拟合的真实性及效果评价。由于残存误差存在的复杂性, 真实值和异常值难以通过表面得以区分。传统方法通过设置阈值滤去疑似异常值虽也有效, 但难免常出现“误杀误放”的情况, 因此有必要从数据内部入手, 通过分析数据内在规律探测序列异常值。因此改进算法引入异常值检测算法, 检测拟合迭代过程中的异常点, 减少因传统方法根据距离定权剔除无效点而引入的人为误差。

(3) 频数的自动选择和拟合影响因子的自动迭代计算及效果评价。频数控制着拟合效果也是傅立叶谐波分析方法体现其物理意义的载体, 传统方法通过尝试选择、人为设定全局使用同一固定频数, 即默认整个运算区域都是同一类植被生长模式(从物理意义上考虑)。改进方法通过迭代前预处理, 动态估算出待处理点的峰值个数, 即频数, 再参与下步拟合。另一个关键参数是借鉴基于 Savitzky-Golay 滤波拟合法中拟合影响因子, 放弃使用传统方法中人为设置的阈值限差, 而通过计算每次拟合的影响因子, 在该因子达到最小值(或局部最小值)时自动终止迭代计算。

(4) 在研究改进算法的同时, 希望能通过对比实验, 客观的评价现在广泛使用由 National Aerospace Laboratory 开发的 HANTS 软件谐波分析结果。

(5) 使用遥感植被指数时间序列, 分析提取北京及周边地区两类可能沙尘源地, 为北京沙尘天气治理提供一定的依据。

第二章 三类重要遥感植被指数时间序列滤波重建算法评述

2.1 基于 Savitzky-Golay 滤波的拟合算法

Savitzky-Golay 滤波基本公式：

$$Y_j^* = \frac{\sum_{i=-m}^{i=m} C_i Y_{j+i}}{N} \quad (2-1)$$

其中 Y 是指 NDVI 的原始值，Y* 是 NDVI 的拟合值，C_i 是第 i 个 NDVI 值滤波时的系数，N 是指卷积数目，也等于滑动数组的宽度 (2m+1)。系数 j 是指原始 NDVI 数组的系数。滑动数组包含有 (2m+1) 个点。该方法本质上是种平滑滤波，因此两个参数控制滤波效果，一个是 m，即滤波窗口大小；二是平滑多项式的次数，较低的次数可以得到更平滑的结果，但会保留异常值，高的次数可以去掉异常值但又过拟合，造成新噪声出现。由于使用窗口滑动进行数据平滑，会使序列前 m 个和后 m 个数据无法参与运算，如果将数据系列首尾相接成环，即可解决此问题。辜智慧 (2003)、Chen J. et al. (2004) 中都使用了这样策略。

需要特别指出的是上述二人使用该算法时引入了“拟合影响因子”，通过比较前后相邻拟合计算得到的拟合影响因子，确定拟合是否达到最佳而停止。这比傅立叶系列方法中通过经验确定拟合终止条件要客观先进的多。

S-GF 方法只需确定两组参数 (滑动窗口尺寸 m 和多项式次数) 即可获得较高质量的 NDVI 数据，并且参数敏感度低于 BISE。辜智慧 (2003) 给出了研究长期变化趋势时的最优滤波参数为 (7, 2)，拟合循环中的最优滤波参数为 (3, 4)。该组参数的选定还需要结合研究区域的实际情况确定。据称该方法可以有效地去除多时相 NDVI 遥感数据中由于云、气溶胶等大气影响造成的噪声，充分利用对应的云状态数据，可应用于不同时间尺度、空间尺度和传感器的数据以获取较高质量的时序数据，理论简单并易于实现。但辜智慧 (2003) 指出该方法在两种情况下会失效：一种是云和大气条件对 NDVI 的影响过于频繁或正好在植物生长高峰；另一种是由于收割造成 NDVI 正常低值可能会被该方法加以提高。另外该方法只能对等间距的时间序列数据滤波重建，这是由其方法本身造成的。

2.2 非对称高斯拟合法

该方法使用分段高斯函数（曲线）组合来模拟植被季相生长（物候）规律，一个组合代表一次植被盛衰过程，最后通过平滑连接各高斯拟合曲线，实现时间序列重建。其过程大致分为：区间提取、局部拟合（Local Model Function）、整体连接（Global Model Function）。

（1）区间提取

由于该方法是将整个时间序列中时间点对应的值按极大极小值分成多个区间，每个区间只包含一个极值，分别对该区间进行高斯拟合，故区间选择是关键一步，影响后期全局拟合（曲线连接）效果。

单纯从数据序列中选取这样的区间本身不是难点，但由于噪声影响，使得该区间内数据跳动很大。因此区间提取前，必须进行数据平滑处理。Per Jonsson 等（2002）中通过滑动均值窗口来实现数据平滑提取时间窗口。该方法区分左右半边窗口（ n_L, n_R ），亦即滑动窗口的大小为（ $n_L + n_R$ ）。窗口中心点左右分别选用不同数目的邻居点参与数据平滑，体现了对前后数据不同程度的信赖，近似可理解为一种加权方式。显然这两个值控制着数据平滑效果，该值愈大结果愈平滑，但损失的信息也就相对愈多。

（2）局部拟合

定义：高斯函数 $f(t)$

$$f(t) \equiv f(t; c_1, c_2, a_1, \dots, a_5) = c_1 + c_2 g(t; a_1, \dots, a_5) \quad (2-2)$$

$$\text{其中: } g(t; a_1, \dots, a_5) = \begin{cases} \exp\left[-\left(\frac{t-a_1}{a_2}\right)^{a_3}\right], & \text{if } t > a_1 \\ \exp\left[-\left(\frac{a_1-t}{a_4}\right)^{a_5}\right], & \text{if } t < a_1 \end{cases} \quad (2-3)$$

线性参数 c_1, c_2 决定曲线的基准和振幅； a_1 是与时间 t 相独立的参数，决定曲线最大值或最小值的位置； a_2, a_3, a_4, a_5 分别定义左右半边曲线的宽度和峭度（与曲线峰挺拔程度有关的量）。这四个参数控制拟合结果尽可能适应非对称条件下时间序列曲线拟合，其效果优于傅立叶拟合中使用绝对对称的 \sin 和 \cos 曲线。

（3）整体连接

定义全局拟合函数：使用 $f_L(t)$, $f_C(t)$, $f_R(t)$ 分别表示一个极大值区间的局部拟合函数，及其左右两边两个极小值区间的左极小值拟合函数和右极小值拟合函数，整个区间用 $[t_L, t_R]$ 表示，则全局拟合函数：

$$F(t) = \begin{cases} \alpha(t)f_L(t) + [1 - \alpha(t)]f_C(t) & , \quad t_L < t < t_C \\ \beta(t)f_C(t) + [1 - \beta(t)]f_R(t) & , \quad t_C < t < t_R \end{cases} \quad (2-4)$$

裁切函数 (cutoff function) $\alpha(t)$, $\beta(t)$ 是分别定义在 $(t_L+t_C)/2$ 和 $(t_C+t_R)/2$ 的函数。

通过全局拟合函数 $F(t)$ 将局部拟合曲线合并是该方法的关键之一。Per Jonsson 等人强调，采取这样分段拟合的策略避免全局数据对局部拟合的干扰，各个峰值相对独立使该方法在拟合复杂情况下的时间序列曲线具有良好的适用性和灵活性，拟合后的曲线更接近真实情况。但该方法存在区间提取问题。使用滑动均值窗口实现数据平滑效果暂且不论，仅对序列中数值的干扰，就足以对后期拟合产生相当影响。而且对于某些强噪声，算法也很容易将其误判为提取目标而获得虚假区间。Per Jonsson 指出有望通过傅立叶序列拟合确定平均季节数目来解决该问题。

2.3 傅立叶谐波分析方法

傅立叶方法在时间序列曲线拟合重建中有丰富的应用，以它为基础发展出了很多方法，如傅立叶拟合 (L. Olsson et al., 1994; M. Menenti et al., 1993; L. Andres et al., 1994; A. Moody et al., 2001)、Sellers 算法 (林忠辉等, 2006; Sellers P J et al. 1996)、HANTS 算法 (W. W. Immerzeel et al., 2005; Mark E. et al., 2001; Mark E. et al., 2002; Roerink, G. J. et al., 2000; 王丹等, 2000; Yves Julien et al., 2006) 等。这些算法使用最小二乘法拟合傅立叶系数矩阵，通过迭代保留或剔除“假点”实现滤波最终获得拟合曲线达到序列重建目的。该算法中引入最小二乘拟合求出合适的傅立叶曲线系数阵 C , C 为 $(2*m-1, 1)$ 的矢量。谐波波数 (即频数 m) 是该方法重要参数, m 的大小决定拟合曲线的形状: m 大, 拟合曲线不够平滑, 但保留细节多, 同时可能造成拟合结果偏离实际情况, 达不到滤波效果; m 值小, 则产生的拟合曲线平滑, 描述细节少。

Seller 算法是在傅立叶原理基础上发展起来的。该算法引入权重, 确立了相应的定权准则使傅立叶分析方法更加合理, 因为去除云和大气影响主要途径即

是比较植被指数观测值和拟合值，通过一定的准则，对不同值赋高低权重，根据权重来选择参与拟合的点。权重可以理解为待拟合点对曲线的贡献率，权重越大在拟合过程中贡献越大。不同学者的权重制定准则不同，运行结果自然有优有劣。Sellers (1996) 算法引入的定权准则：

$$W_i = \begin{cases} 0 & \text{if } U_i \leq -k \\ [1 + (U_i + r)/k]^4 & \text{if } -k < U_i < -r \\ 1 & \text{if } -r \leq U_i \leq r \\ [1 + (U_i - r)/k]^2 & \text{if } U_i > r \end{cases} \quad (2-5), \text{ 同时应有: } \begin{cases} 0 \leq W_{i=1} \leq 1 \\ 0 \leq W_{i=n} \leq 1 \end{cases}$$

式中 $U = (Y - \hat{Y}) / M$ ，M 是真值与拟合值之差绝对值的中数，显然，这本质

上就是根据观测值与估值间距离赋权，这一方法通常情况是适用的，但在拟合迭代过程中可能会出现某些噪声值更接近拟合值，造成非噪声点估计与真值之差反而变大。这种情况下如果按照距离定权拟合就有可能把非噪声点剔除而继续保留噪声参与曲线拟合。如果去除噪声点后，非噪声点观测值与拟合值的差值马上变小。林忠辉等 (2006) 在研究中国华北 10 天 NDVI 时间序列研究时，还指出 Sellers 定权准则的其他缺陷。该准则主要适用于热带阔叶林，而对短时间间隔合成的时间序列效果欠佳。这些问题也表现在 Sellers 定权未很好考虑地表植被的动态变化过程，如对动态异常高值仍赋高权重使冬季 NDVI 值整体高估等。

傅立叶谐波分析发展的另一种算法：HANTS (Harmonic Analysis of Time Series) 不像快速傅立叶变换算法 (FFT) 那样严格要求图像等时间间隔，可以处理不等时间间隔的图像序列 (张霞, 2006; Roerink, G. J. et al., 2000)。现在大家通常使用的 HANTS 方法多是由 National Aerospace Laboratory 编程实现的 HANTS 软件。该软件运行总共需要输入 8 个参数，其中尤为重要是频数和拟合允许误差 (FET) 两参数。但遗憾的是这两个参数都要靠经验、不断的尝试才能获得最佳值，这也直接引入了人为干扰因素，是目前基于傅立叶分析方法的通病之一。同时正如前所述，如何定权关系到该类方法性能好坏，而不同的地域通常在定权上都会有相应的侧重。HANTS 软件的定权准则是集成在软件里面的，属于黑箱操作，无法根据现实条件及时调整，从而影响其具体应用效果。

2.4 重建算法分析评述

从遥感时间序列数据被学者关注以来,很多时间序列滤波重建算法被提出,但并没有形成学者一致认同的方法。已有方法总有不能避免的理论或技术缺陷,主要表现在:

1) 现有方法为了使自己得以应用而附加假设或者限制条件、排除适用,表现在数据预处理中,通常通过设置阈值提前剔除数据,这对序列重建到底产生多大的影响并没有严密的论证。

2) 各类方法中的参数大多都要依靠经验选取,如谐波分析中的频数、迭代拟合终止条件的选择(张霞, 2006; Roerink, G. J. et al., 2000), Savitzky-Golay 滤波(辜智慧 2003; Chen J. et al. 2004)、LMF(Izumi NAGATANI et al., Sajia Akhter et al. 2007)、非对称高斯法(Per jonsson et al, 2002)中滑动窗口大小的选择等。这些都是应用中的关键参数,选择好坏直接影响算法应用的成功与否,人工选择难免引入个人主观干扰。同时由于是经验选择,整个测区使用同一参数值也多与实际不符,尤其是像中国这样土地利用类型复杂的地域。

3) 各类方法中都有尚未很好解决的技术难点,该问题影响算法应用。如在谐波分析中定权准则的确定:传统的按距离定权是否合理、滤波中的抛点方法是否得当;非高斯对称法中区间提取时数据平滑对序列拟合产生的影响也尚未有评估,这样会产生结果可靠性问题。

研究者往往有选择地针对不同应用开发适用于该领域的算法,如谐波分析因为其自身特点在植被生长季相特征、物候信息提取等领域使用频繁, Savitzky-Golay 滤波在植被类型提取中表现出色,而非对称高斯拟合在长期季相信息提取中效果看好。同时,对现有算法改进和应用新思路解决问题的步伐也从未停止过,如局部最大值拟合法的提出。如何妥善解决以上问题,是今后遥感时间序列数据滤波重建研究努力的方向。而目前,根据不同的应用领域选择合适的方法,通过实验选择参数组合使用各类方法也是应用趋势之一。

2.5 本章小节

本章重点介绍了三类重要遥感植被指数时间序列滤波重建算法的原理,分析其优势以及缺陷:

基于 Savitzky-Golay 滤波的算法引入了拟合影响因子,通过该参量控制拟

合达到最优后自动终止。另外本方法只需确定两组参数（滑动窗口尺寸 m 和多项式次数）即可获得较高质量的 NDVI 数据，并且参数敏感度低于 BISE，但参数需要通过实验人为比较选择。

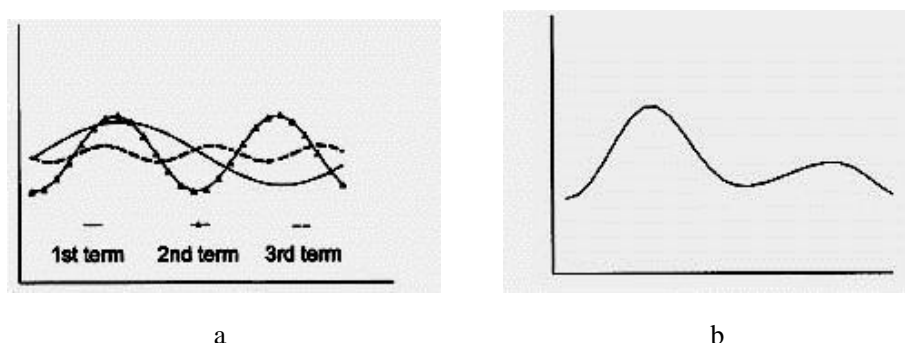
非对称高斯拟合法采取分段拟合的策略避免全局数据对局部拟合的干扰，各个峰值相对独立使该方法在拟合复杂情况下的时间序列曲线具有良好的适用性和灵活性，拟合后的曲线更接近真实情况。但同样存在着重要参数人为选择，另外，滑动窗口的选择也依然存在技术问题。

傅立叶谐波分析系列算法优势在于它既很好的去除噪声，又能够利用代表不同植被不同生长周期的谐波，最大可能的反映植被生长曲线内部蕴含的周期性规律，应用较为广泛。但与上述两种方法类似，重要参数人为实验比较选择。

第三章 基于傅立叶谐波分析算法的改进算法研究

3.1 傅立叶谐波分析算法原理及优缺点分析

傅立叶谐波分析方法“既很好的去除噪声，又能够利用代表不同植被不同生长周期的谐波，最大可能的反映植被生长曲线内部蕴含的周期性规律，利用遥感图像存在的时间性和空间性特点，将其空间上分布规律和时间上的变化规律联系起来”（张霞，2006，p83）。



a.1 级谐波项、2 级谐波项和 3 级谐波项示例 b.各级谐波项叠加产生曲线

图 3-1 谐波及其叠加曲线示意图（张霞，2006）

傅立叶谐波分析方法基本原理：

离散时间序列 y_1, y_2, \dots, y_n ，由于序列为离散时间点上的波形函数，根据采样原理，谐波波数最多只能取 $n/2$ ，即只能取有限个正弦波来逼近序列 $\{y_z\}$ 。若采用 y_i 表示 m 个谐波叠加后的序列估计值，则有：

设某要素时间序列 $Y=(y_1, y_2, \dots, y_n)$ ，将这 n 个资料表示成为有限个正弦波（谐波）的叠加形式，就称为谐波分析。由这些正弦波叠加构成的数据序列，即为傅立叶谐波序列，其数学表达式是：

$$y_i = A_0 + \sum_{j=1}^m A_j \sin(\varpi_j i + \theta_j) \quad (3-1)$$

式中： A_0 ——谐波的余项，等于序列的均值； A_j 各谐波的振幅； θ_j 各谐波的初相位， $i=1, 2, \dots, n$ ； n 表示时间序列中时间点个数， $\varpi_j=2j\pi/n$ ，为各谐波的频率； m 为谐波个数。

傅立叶序列还有另外一种表示方式，即表示为一个正弦波和余弦波函数和的形式：

$$y_i = a_0 + \sum_{j=1}^m [a_j \cos(\varpi_j i) + b_j \sin(\varpi_j i)] \quad (3-2)$$

式中， a_0 表示谐波余项，与 (3-1) 中的 A_0 大小相等； a_j ， b_j 傅立叶方程系数。

为了方便计算机编程，将 (3-1) (3-2) 作如下变形：

$$y_i = \sum_{j=1}^m A_j \sin((j-1)\phi_i + \theta_j) \quad (3-3)$$

$$y_i = \sum_{j=1}^m [a_j \cos((j-1)\phi_i) + b_j \sin((j-1)\phi_i)] \quad (3-4)$$

式中 $\phi_i = 2\pi(i-1)/n$ ，其它参数意义如上。

这样我们可以对将要处理的时间序列与上述方程作如下对应，即 Y_i 表示时间序列中某点值，它们的集合用 Y 表示，大小为 $(n, 1)$ ； i 表示该时间序列中第 i 个时间点的值。可将 (3-4) 式展开后写成：

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{n-1} \\ Y_n \end{bmatrix} \quad (3-6)$$

$$Y = F \cdot C \quad (3-5)$$

$$F_{(n, 2m-1)} = \begin{bmatrix} 1 & \cos \phi_1 & \sin \phi_1 & \cdots \cdots \cdots \cos(n-1)\phi_1 & \sin(n-1)\phi_1 \\ 1 & \cos \phi_2 & \sin \phi_2 & \cdots \cdots \cdots \cos(n-1)\phi_2 & \sin(n-1)\phi_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos \phi_n & \sin \phi_n & \cdots \cdots \cdots \cos(n-1)\phi_n & \sin(n-1)\phi_n \end{bmatrix} \quad (3-7)$$

$$C_{(2m-1, 1)} = \begin{bmatrix} c_1 \\ \vdots \\ c_{2m-1} \end{bmatrix} \quad (3-8)$$

F 为傅立叶阵， C 为傅立叶系数。傅立叶最小二乘拟合过程中，就是要求出合适傅立叶曲线的系数阵 C 。选择谐波波数为 m ， m 的大小决定拟合曲线的形状： m 大，拟合曲线不够平滑，但保留细节多，同时产生的问题是造成拟合结果偏离实际情况，达不到滤波效果； m 值小，则产生的拟合曲线平滑，描述细节少。 F 是大小为 $(n, 2 \cdot m - 1)$ 的矩阵， C 为 $(2 \cdot m - 1, 1)$ 的矢量。

傅立叶序列拟合采用最小二乘法：

$$(F^T \bullet F) \bullet C = F^T \bullet Y \quad (3-9) \quad C = (F^T \bullet F)^{-1} \bullet F^T \bullet Y \quad (3-10)$$

即 (3-10) 为通过最小二乘拟合后得到的傅立叶方程系数。其与方程 (3-4) 系数对应，有 $C_1=A_1=a_1$; C_2, C_3, \dots 分别于 a_2, b_2, \dots

在方程 (3-4) 中，当 $j=1$ 时， $\sin((j-1)\phi)=0$ ， $\cos((j-1)\phi)=1$ ，故 F 、 C 阵列数为 $2 \times m-1$ ，即在矩阵生产过程中自动去除了零值，且 F 阵第一列全为 1。

方程 (3-4) 与方程 (3-3) 在系数上有这样的对应：

$$A_1=a_1, \quad A_1 = \sqrt{a_2^2 + b_2^2}, \quad A_2 = \sqrt{a_3^2 + b_3^2} \dots\dots$$

$$\phi_1 = \arctan\left(\frac{a_2}{b_2}\right), \quad \phi_2 = \arctan\left(\frac{a_3}{b_3}\right), \dots\dots$$

使用这样的换算公式是用来通过公式 (3-4) 根据拟合曲线，计算对应时间点上的拟合值（估值）。

如果在最小二乘中，加入权值控制不同点对曲线拟合的贡献度（也可以理解为对不同点值有不同的信任度），这样更容易控制拟合曲线，利于噪声的剔除。

加权后的傅立叶最小二乘拟合模型变为：

$$WF = WFC \quad \text{亦即} \quad (F_w^T \bullet F_w) \bullet C = F_w^T \bullet Y_w \quad (3-11)$$

$$F_w = \begin{bmatrix} W_1 & \dots & \dots & W_1 \cos(n-1)\phi_1 & W_1 \sin(n-1)\phi_1 \\ \vdots & \dots & \dots & \vdots & \vdots \\ \vdots & \dots & \dots & \vdots & \vdots \\ W_n & \dots & \dots & W_n \cos(n-1)2\phi_n & W_n \sin(n-1)2\phi_n \end{bmatrix} \quad (3-12)$$

$$Y_w = \begin{bmatrix} W_1 Y_1 \\ \vdots \\ \vdots \\ W_n Y_n \end{bmatrix} \quad (3-13)$$

Y_w 是赋值的观测序列； F_w 赋权值的傅立叶矩阵。

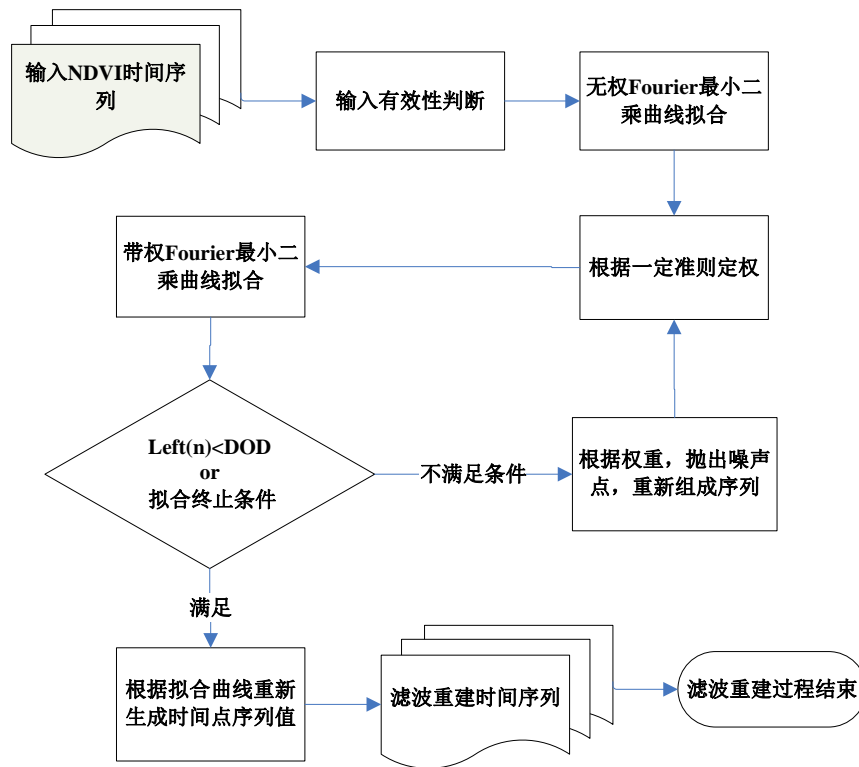


图 3-2 NDVI 时间序列 Fourier 谐波分析算法基本流程

遥感植被指数时间序列傅立叶谐波校正算法：

在上述傅立叶原理的基础上，不同的学者经过改进，产生了不同的谐波分析方法。

傅立叶谐波分析算法之一：HANTS (Harmonic ANalysis of Time Series)，它可以不向 FFT 那样严格要求时间序列，而是可以以不等时间间隔。现在大家通常使用的 HANTS 都是采用 National Aerospace Laboratory 编程实现的封装软件包，它需要输入的参数有：

有效值范围：只有落在用户给定范围内的数据才被 HANTS 用于计算；

周期长度：待分析的周期长度；

频数个数 (NOF)：即拟合曲线选用的谐波个数，须根据物候周期确定；

周期：以天为单位，大小等于频数；

方向性界外值切换：即选定是高值抛出还是低值抛出；

拟合允许误差 (FET)：曲线迭代过程中，当所有点的估值与观测值绝对差若小于 FET，曲线拟合迭代终止；

超定度 (DOD)：确定参加拟合的最少点数，少于这个值，拟合结果不可信；

缩放比例：振幅输出结果乘以该值，使得最终输出值达到最大值范围。

HANTS 算法总共需要输入 8 个参数，这其中尤为重要是 NOF 和 FET，但遗憾的是这两个参数都要靠经验，不断的尝试获得最佳参考值，这直接增加了人为因素的干扰，其实这也是目前基于傅立叶分析方法的通病。同时，它的定权准则是集成在软件里面的，是种黑箱子操作，这使得在具体的应用领域使用时会产生一些麻烦。

Sellers 等（1996）提出一种改进的傅立叶谐波分析方法，他们在使用该方法处理 NDVI 时间序列时，提出两条假设：

- （1）植被状态的变化具有时间连续性，并呈现季节性规律；
- （2）云和大气对 NDVI 的影响总是表现出削弱 NDVI 信号的趋势；

第一条假设存在一个问题，即在作物收割或者植被突然被铲除等地表突变的情况下，会出现 NDVI 值的突变，时间序列点中时间点的间隔越小，如 10 天、5 天，这一问题就越突出。但如果不作这样的假设，就不能使用傅立叶分析方法，因为该方法适用的条件就是曲线连续。使用这种方法在遇到间断点时候，如果突变处的 NDVI 相对上点值变化很大，算法就有可能把该点当作噪声滤除，此时就会出现错误拟合，导致与地面实际情况不符。至于第二条假设通常都是成立的。对于要处理的数据，就整体情况而言，这两条假设还是能够接受的。

在这种情况下，算法性能是否优越就表现在两处：（1）权重的确定方法；（2）判别算法拟合终止条件的选择。在去除云和大气影响时，主要就是通过比较 NDVI 观测值和拟合值，通过一定的准则，对不同的值赋高低权重，通过权重来选择参与拟合的点，这些权重可以理解为带拟合点对曲线的贡献率，权重越大，在拟合过程中贡献越大，也可以理解为可信度。不同的学者对权重制定准则不同，运行结果自然有优有劣。同时，终止条件的选定影响拟合次数和曲线和曲线拟合的效果。

Sellers 等学者提出的定权原则本质上根据距离定权：

$$W_i = \begin{cases} 0 & \text{if } U_i \leq -k \\ [1 + (U_i + r)/k]^4 & \text{if } -k < U_i < -r \\ 1 & \text{if } -r \leq U_i \leq r \\ [1 + (U_i - r)/k]^2 & \text{if } U_i > r \end{cases} \quad (3-14) \quad \text{同时应有：} \begin{cases} 0 \leq W_{i=1} \leq 1 \\ 0 \leq W_{i=n} \leq 1 \end{cases}$$

(3-14) 中 $U = \left(Y - \hat{Y} \right) / M$ ，M 是真值与拟合值之差的绝对值的中数，显

然这本质上就是根据观测值与估值间距离赋权，这一方法通常情况是适用的，但是在拟合迭代过程中会出现一严重问题，即某些噪声值有可能更接近拟合值，造成某些非噪声点的估计与真值的差反倒变大。这种情况下，如果按照距离赋权拟合，就有可能把非噪声点剔除，不能下次拟合，而继续保留噪声参与曲线拟合。如果正确去除噪声点后，非噪声点观测值与拟合值的差值马上变小。这种现象在遥感影像几何校正过程中经常可以碰到。

Sellers 定权规则中， k ， r 为经验参数，Sellers 没有给出具体含义，在与林忠辉教授讨论时，他认为 k ， r 的取值决定权重曲线的形状。 k 取值与 $U = \left(Y - \hat{Y} \right) / M$ 有很大关系， k 的取值 4，可以有效降低高值的权重、提升低值的权重。这是因为 Sellers 的权重算法原来考虑的主要是热带地区的常绿阔叶林，因此他们认为应当相信高值，即给高值赋予高权重，这样就导致在作物生长季初期和末期的 NDVI 低值也是不可信的，给与的权重偏小。将 k 值提升，则会增高低值的权重。

林忠辉等（2006）在研究中国华北 10 天 NDVI 时间序列的研究论文中指出按照 Sellers 等学者的定权法定权存在问题，尤其是对短时间间隔合成的时间序列，如 10 天合成序列。这些问题突出表现在 Sellers 定权未很好考虑地表植被的动态变化过程，如对动态异常高值仍赋高权重，会整体高估冬季 NDVI 值；对冬季 NDVI 整体上赋予的权重较小，导致冬季 NDVI 校正结果偏低。

虽然存在研究区域差异而造成结果不够理想的问题，但更主要的也是反映了这种定权方法的局限。针对定权问题，林等人提出改进意见，即通过进一步设置阈值提高可信赖的冬季 NDVI 低值的权重：裸地 NDVI 多 0.11~0.17 之间，考虑到混合像元的问题，设定阈值为 0.20；对冬季（ $i \leq 9$ 或 $i \geq 33$ ）NDVI 在 0.2 以下的点，认为这些点 NDVI 值可信，提高权重；对生长季节中低于 0.2 的点则认为是云干扰或作物收获所得。据此定权：

$$W_i = \begin{cases} 0 & \text{if } U_i \leq -k \\ [1 + (U_i + r)/k]^4 & \text{if } -k < U_i < -r, NDVI \geq 0.2 \\ 1 - U_i/k & \text{if } -k < U_i < -r, NDVI < 0.2 \\ 1 & \text{if } -r < U_i < r \\ [1 + (U_i - r)/k]^2 & \text{if } U_i > r \end{cases} \quad (3-15)$$

Sellers 方法中关于拟合迭代终止条件的认定是：将序列中的每一点与它的

原始值和它最近邻的四个点的值进行比较 ($i-2$, $i-1$, $i+1$, $i+2$), 如果新拟合值不超出这个 5 个点中最大值, 且不小于其原始值, 则认为该拟合达到效果。其实, 在这个条件之上还有一个当然迭代终止条件, 就是: 剩余拟合点个数不能小于 $(2*NOF+1)$ 。张霞博士在其论文中是这样解释这一条件的, 即: 拟合中必须保证剩余拟合点的个数大于或等于 $(2*NOF+1)$ 。这样才能保证拟合结果的可信性, 拟合点剩余个数如果小于这个值, 就会造成方程组欠定, 不可能对矩阵求逆。所以只要是在使用最小二乘方法拟合曲线时, 都是要考虑这个隐含的拟合终止条件。

同时这些方法都存在一个普遍的严重问题。根据前面论述频数是根据物候周期确定的, 比如某地作物是一年两熟, 那么在时间谱上植被指数就大体反映出两个峰, 那么此时就该选用 2 个谐波, 即频数为 3 (0 频率, 即频率均值也算为一个频数, 它图像上反映是条平行 x 轴的直线, 其他谐波值在其上下波动), 来拟合时间谱曲线。之后该频数被应用到了整个作业区域内, 实际上这就暗含一条假设: 在作业区域内, 所有地方作物都是一年两熟。显然这是不符合事实。所以静态的频数值存在严重的缺陷, 它代表了傅立叶谐波分析的特色却又抹杀了该方法的优势。

3.2 傅立叶谐波分析算法改进设计

以上是傅立叶谐波分析的基本原理。通常对该方法的主要异议集中在两点:

- (1) 使用绝对对称的正余弦曲线参数拟合无规律或者非对称数据效果不佳;
- (2) 拟合曲线会产生虚假的抖动。

虽然该方法使用对称的正余弦曲线参与拟合, 但最终曲线是各谐波的叠加结果, 控制好参与叠加的谐波级数, 在一定程度上可以缓解第一个问题。造成 2 号问题的原因是使用过高级数的谐波, 造成了虚假的峰值或者使用过少谐波, 导致拟合曲线过度平滑。因此解决这两个问题的最佳方案是使用合理级数的谐波参与曲线拟合。

如上文分析在实际应用中, 碰到了另外两个问题:

- (1) 全局使用同一个频数不合理
- (2) 参数靠经验获得, 引入新干扰。

这两个问题也是现有谐波分析的通病，而参数经验获得甚至是绝大部分时间序列滤波重建的通病。所以本文试图针对上述问题，在傅立叶谐波分析 Sellers 算法的基础上进行改进。

Sellers 算法基本形式为：

$$WY = WFC \quad (3-16)$$

式中 W 为权阵，其它参数同上。

同时，引入两个概念：

(1) 异常值：Hawkins (1980) 对异常的本质进行了科学地定义：异常是在数据集中与众不同的数据，使人怀疑这些数据并非随机偏差，而是产生于不同的机制。使用判断准则和异常 (outliers) 检测，则可以有效的去除序列中无效值。对于判定为无效值的点，不参与后面曲线拟合。

(2) “拟合影响因子 F_k ” (详细定义参见 3.2.4)。

3.2.1 时间序列数据重建原则

植被指数时间序列图像重建应该遵守的 4 条原则：

- (1) 应尽可能减少对数据扰动
- (2) 尽可能少的使用辅助数据；
- (3) 尽可能多的考虑参量物理意义；
- (4) 算法效率与效果的兼顾。

Sellers 算法改进最主要的地方应该是增强它的普适性，尽可能的减少人为选择参数带来的不确定性，针对这些问题进行改进：

- (1) 通过设置异常值检测增加数据拟合的真实性。包括两方面内容，输入值有效性判断和定权抛点策略。输入值有效性确定：传统的方法是时间序列立方体的逐波段（也即逐时相）的统计曲线得到，比如根据最大值、最小值和平均值曲线，结合图像确定（张霞，2006）。这本质上也是根据经验获得。抛点和定权准则：尝试抛点后动态定权而非简单根据距离定权。
- (2) 动态选定谐波个数，即动态选频：由于谐波个数与像元对应地面的物候规律有关，因此对不同的像元，动态选定参与下步拟合的谐波个数。
- (3) 迭代拟合终止条件的改进。

3.2.2 数据异常值检测策略

(一) 输入值有效性判定

确定观测值的有效性就是为了去除一些粗差点(错误或异常值)。这类点有:错误的赋值、丢失的点、云等严重污染的值。通常无效值有两种类型:(1)不同植被指数时间序列,这类无效值有所不同。在 NDVI 时间序列中,如果只考虑植被,那么凡是 $NDVI < 0$ 的点都可以认为是无效值,在拟合前先行剔除。除此之外,还有另一类型的无效值,(2)它们不满足第一类型,看似都正常,但与这个序列点间确有差别,可以称它为:异常点(outliers)。尽管对每个序列进行检测一定程度上增加了算法时间开销,但检测算法简单有效,消耗的时间并不多。

本文使用的异常检测算法是根据 Grubbs Test 实现的。

Grubbs Test¹:

一组重复观测值(残差),按大小顺序排列,得到 $V_{(1)} \leq V_{(2)} \leq \dots \leq V_{(n)}$. Grubbs

$$\text{统计量: } g_n = \frac{V_{(n)} - \bar{V}}{\sigma} \quad g'_n = \frac{\bar{V} - V_{(1)}}{\sigma} \quad (3-17)$$

式中 $\bar{V} = \frac{1}{n} \sum_{i=1}^n V_{(i)}$, σ 为样本标准差, g_n 用来判别最大值是否为异常值, g'_n

可用来判断最小值是否为异常值, g_n 和 g'_n 有相同的概率分布。当 g_n 和 g'_n 大于显著水平 α 下的临界值 $g(n, \alpha)$ 时,就可以判断最大值(或最小值)是异常值, $g_0(n, \alpha)$ 已有学者作出,查表即可获得。

查临界值 $g_0(n, \alpha)$ 表,当 $|g_n| \geq g_0(n, \alpha)$ 或 $|g'_n| \geq g_0(n, \alpha)$ 时,可认为该残差相应的观测值含有粗差, α 为显著度,一般取 0.05 或 0.01。关于 Grubbs test 详细的介绍请参考相关文献。

针对本文中数据,采用观测值与拟合值之差 $\Delta = |Y - \hat{Y}|$ 做异常检测。由于 Grubbs Test 要求数据服从正态分布,为了避免麻烦、减少算法开销,将 Δ 对数化,即使用 $\log(\Delta)$ 替代 Δ ,作近似处理以适应监测算法。

本文使用的是根据上述算法,由 Brett Shoelson 博士编写的 matlab 程序参与数据预处理²。

使用异常检测(outliers)对数据有效性进行判断是种非常好的思路,它避

¹Engineering Statistics: <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm>

²Delete outliers:

<http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=3961&objectType=file>

免处理人员被数据表面迷惑，得到的结果更加客观准确。但是具体选用什么样的异常检测算法，哪种算法更优，对哪种运算结果检测之类问题还有待于探讨。

（二）.抛点和定权准则

Sellers 算法是根据观测值与估值间距离赋权，使用该方法的其他学者大体上也遵从这一规则，这一方法通常情况是适用的。但是在拟合迭代过程中会出现这样一个悖论，即某些噪声值更接近拟合值，造成一些非噪声点的估计与真值的差反倒变大。这种情况下，如果按照距离赋权拟合，就有可能把非噪声点剔除下次拟合，而继续保留噪声。如果正确去除噪声点后，非噪声点观测值与拟合值的差值马上变小。这种现象在遥感影像几何校正过程中经常可以碰到。故应避免直接使用距离定权和抛点。克服这一问题的首要方法就是拟合算法的选择。比如最小二乘法，该方法最大的优势就是使拟合曲线尽可能的接近真实值，其次就要考虑一些策略诱出噪声点。所谓“诱出策略”是指通过异常检测算法，或者其他尝试方法，尽可能找出造成整体观测值与拟合值偏差很大的点，将它以赋 0 权重的方法抛出。单纯异常检测，最好的情况是直接检测出真实噪声点，赋 0 权抛出，最差的情况是检测不出异常点，此时，如果补充定义了 Sellers 定权准则，则剩余点根据距离大小继续拟合。

实验证明，只使用异常检测，很有可能不能诱出真实误差点，虽然采信 Sellers 算法定权准则予以补充，但最初诱出目的并未达到，所以很有必要使用多种方法联合作业，达到诱出真实噪声点的目的。有种思路是先对拟合值与估值的绝对差进行异常值检测，异常值当然赋 0 权重抛出，如果未检测出异常值，则使用补充策略，即依距离依次尝试抛出拟合点，对剩余点再次拟合后拟合值与估值的绝对差方差最小的点赋 0 权重抛出。其它点则按照 Sellers 或者林辉忠改进规则定权。这样的策略至少比 Sellers 算法时间复杂度高出了 $(n-1)$ 倍。而造成迭代次数猛增的原因就是对异常检测的补充策略。在考虑一定精度要求下，可以不选用这种补充策略，而直接使用 Seller 算法定权准则对异常检测不足进行补充。至于其它的更高效的异常检测算法或者补充策略还有待于进一步实验研究。

3.2.3 算法关键参数的自动选择

傅立叶谐波分析最大的优势就是考虑了植被物候规律，这种优势靠谐波数

(频数 NOF) 来体现。以前的方法对时间序列上所有点都使用同一频数, 这显然暗含一条假设, 即: 在作业区域内, 所有地方物候规律都是相同的。这个假设并不是普遍成立的。即便是同一地区, 耕地也可以一年数茬, 即在曲线上反映出数个波峰, 但是林地则不会有这样的现象, 显然代表耕地和林地的谐波数是不同的。所以不加区分, 使用同一谐波数模拟植被物候明显不合理, 它的存在足以影响曲线拟合精度和可靠性。

针对这个问题, 本文设计了前置动态估算方法, 估算待处理地区应该采用的谐波个数。谐波最终是用来叠加拟合研究点的时间谱曲线, 时间谱曲线峰值的个数与谐波个数有对应的关系, 即时间谱曲线有几个峰, 就应该选用几个谐波 (此处谐波不含 0 谐波)。显然, 如果对进行了“输入值有效性”判定后的时间序列估算其局部最大值个数, 就能解决这一问题。考虑到时间序列噪声影响, 如果直接进行局部最大值统计, 其结果非常不可靠。这个问题可以转化为对该序列的粗拟合曲线的波峰的统计来解决。所谓“粗拟合”是指不苛求拟合曲线与真实曲线符合, 只要拟合曲线的波峰与实际的相同就可以了。这一目标, 通过多项式拟合或者曲线滤波可以很容易达到。一种算法思想是: 将进行“输入值有效性”判断后的序列首尾相连, 或将序列第一个值加到尾部, 形成含 $n+1$ 个值的新序列, 这是为了保证不因首尾截断而丢失波峰。对新序列采用 3 次多项式拟合生产新曲线。通常使用 3 次就足够了, 低于 3 次, 可能会造成丢失波峰, 高于 3 次, 拟合时间长, 容易产生虚假波峰。对新曲线求一次导数, 获得零点, 再求二次导数, 判断零点正负性, 最终获得极大值个数, 亦即曲线波峰个数。

3.2.4 算法终止策略的选择

必要多的剩余点参与曲线拟合, 保证拟合结果可信度是拟合迭代终止的当然条件。此外, 还有最优迭代终止条件。在 Sellers 算法中采用与原始值的最近 4 领域进行比较, 来判断拟合是否终止。本文借鉴 Savitzky-Golay 滤波中拟合终止条件, 使用“拟合影响因子 F_k ”, 其定义为: 第 k 次拟合的影响因子 F_k ,

$$F_k = \frac{\sum_{j=1}^L (N_j^{k+1} - N_j^0)^2}{L} \quad (3-18)$$

N_j^{k+1} 是指 NDVI 经过 k 次拟合后得到的数列, N_j^0 原始的 NDVI 值, 辜智慧 (2003) 介绍 L 值是经过“输入值有效性”判别后, 被认定为有效输入值、参加下步拟合的点的个数。当 F_k 符合 $F_{k-1} \geq F_k \leq F_{k+1}$ 条件时, 拟合终止。在 Jin Chen (2004) 论文中使用了加权版的 F_k , 其公式形式为:

$$F_k = \sum_{j=1}^L \left(\left| N_j^{k+1} - N_j^0 \right| \times W_j \right) \quad (3-19)$$

Jin Chen 还给出了 F_k 值随拟合次数改变的曲线, 说明了使用拟合影响因子的可行性:

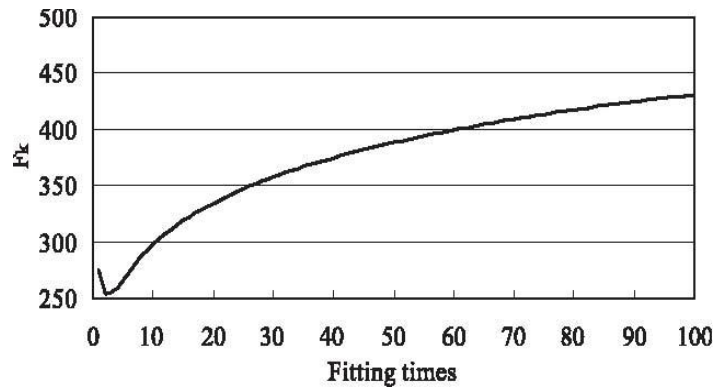


图 3-3 F_k 与拟合次数关系曲线 (Jin Chen et al., 2004)

显然, F_k 是对观测值与真值间接近程度的刻画, 当它达到最小时, 标志着这一序列点的拟合值与真值达到了最接近的程度。当然, 非常有可能在在计算过程中根本达不到这一迭代终止条件时, 剩余点个数已经等于 $2 \times \text{NOF} + 1$ 了, 此时使用迭代停止的当然条件, 迭代自动终止。这样的拟合终止判别条件显然要比 Sellers 算法的最近邻域比较 2% 控制和 HANTS 软件设置 FET 要客观的多, 同时从时间复杂度上讲, 也未必增加多少。

3.3 改进算法技术流程设计

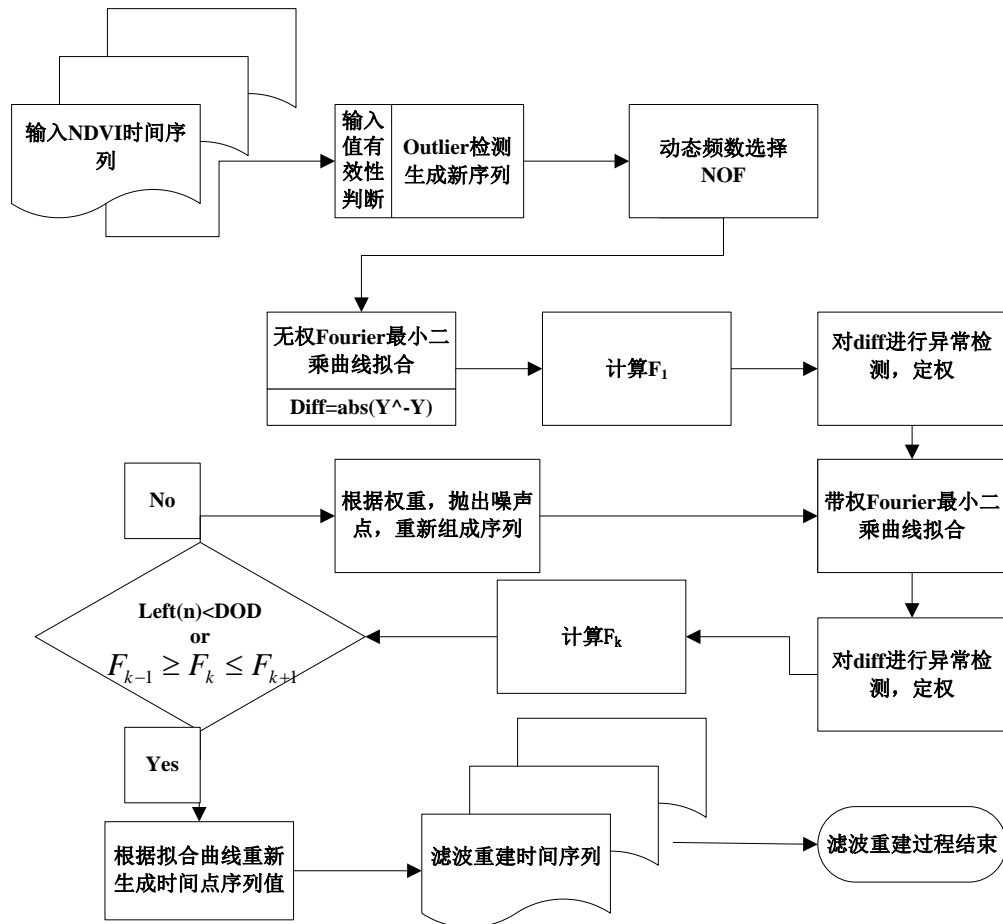


图 3-4 改进型的植被指数时间序列 Fourier 谐波分析算法流程

根据上文分析, 图 3-4 设计出改进算法流程。对于每一个时间点, 进行一次选频频数 (NOF) 计算, 而异常值检测则可以根据需要, 在输入值有效性判断、迭代计算中分别使用。输入值有效性判断时也可以不选用异常值检测, 而只给出一个宽泛的有效值范围, 将异常检测留到拟合迭代过程中。

3.4 算法实现及对比试验分析——与 HANTS 软件结果对比

3.4.1 数据源介绍

实验使用 MODIS EVI 时间序列数据和地面监测站的观测物候资料进行测试, 其中影像为 2003 年华北某地区全年 16 天合成所数据。

植被指数数据:

表 3-1 MODIS 16 天合成植被指数产品 (MOD13Q) 时间定义

产品时间编号	时间区间	时间序号
001:	01Jan ~ 16Jan	1
017:	17Jan ~ 01Feb	2
033:	02Feb ~ 07Feb	3
049:	18Feb ~ 05Mar	4
065:	06Mar ~ 21Mar	5
081:	22Mar ~ 06Apr	6
097:	07Apr ~ 22Apr	7
113:	23Apr ~ 08May	8
129:	09May ~ 24May	9
145:	25May ~ 09Jun	10
161:	10Jun ~ 25Jun	11
177:	26Jun ~ 11July	12
193:	12July ~ 27July	13
209:	28July ~ 12Aug	14
225:	13Aug ~ 28Aug	15
241:	29Aug ~ 13Sep	16
257:	14Sep ~ 29Sep	17
273:	30Sep ~ 15Oct	18
289:	16Oct ~ 31Oct	19
305:	01Nov ~ 16Nov	20
321:	17Nov ~ 02Dec	21
337:	03Dec ~ 16Dec	22
353:	17dec ~ 03Jan	23

数据主要来源于 NASA USGS 的植被指数产品—MOD13Q (Huete, 2002), 为 V004 版数据, 属于科学研究数据。该产品同时提供两种植被指数 NDVI 和增强型植被指数 EVI, 本论文选用 EVI 参与分析应用。该产品的空间分辨率为 250 米, 为 MODIS 的最佳空间分辨率, 时间分辨率为 16 天。数据选择的时段为 2003.1.1~2003.12.30, 共计 23 个时相的图像。研究区为经纬度范围为: 34.1°~41°N, 113°~120°E, 该区属于华北平原一部分。表 3-1 定义了 MOD13Q 合成植被指数产品时相, 每个时相合成对应应有固定且明确的时间定义

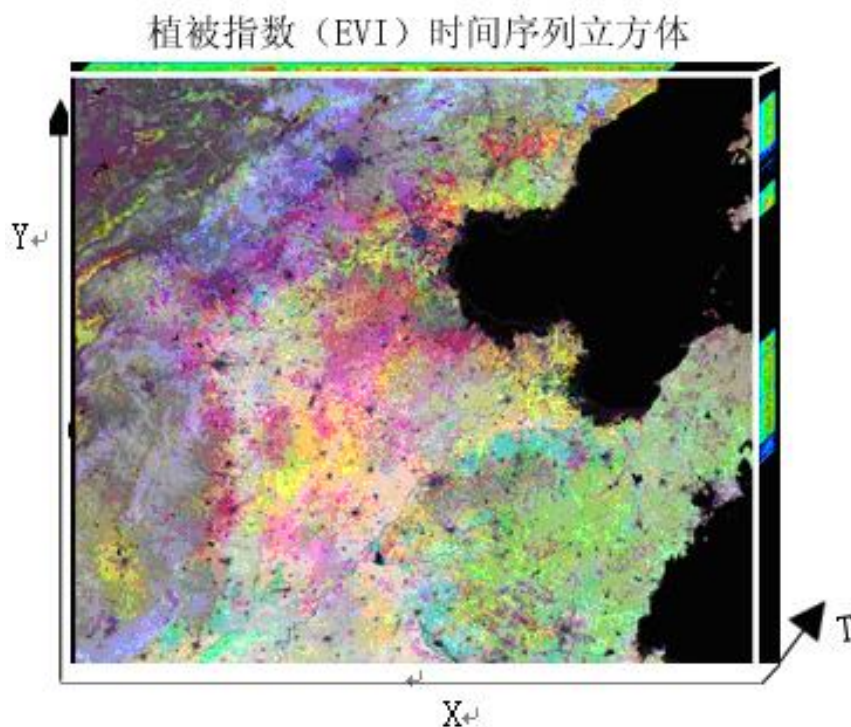


图 3-5 2003 年华北部分区域 250 米 EVI 时间立方体

RGB 分别对应波段 17, 13, 9 (相应儒历日为 257, 193, 129)

物候资料:

为便于结合时相分析, 论文使用部分张霞博士 (2006) 收集整理的华北平原标准气象台站的农气观测报表资料数据。这些台站包括河北 11 个站, 山东 6 个。据称资料都是通过调查大田作物而非调查试验作物获得的, 因此对大面积的作物物候具有很好的面上代表性。本文选取数个地点的数据进行对比分析, 同时对内陆水体、草地等类型指数序列也做相应的对比分析。

3.4.2 相关参数确定

(一) .改进算法参数:

改进算法需要提供的输入参数较少, 主要是输入值的有效范围、最大最小频数 (NOF) 的确定, 以及处理异常点的临界点个数。

输入值的有效范围是根据研究内容确定, 不同的研究内容会有不同的结果, 有时还会通过这样的预处理来剔除一些粗差 (错误)。本文更倾向于处理前不对

数据限定范围或是给定一个宽泛的有效值范围（本文选择-3000-10000），而通过后续的异常值检测剔除粗差，以减少后天扰动对数据结果产生的二次污染，这与本文中提到的滤波重建原则是一致的。

最大最小频数确定：由于改进方法使用粗拟合，通过拟合曲线峰值确定 NOF。在某些情况下，由于数据噪声过大，导致粗拟合曲线峰值个数与真实 NOF 偏差过大，如果使用该值作为 NOF 值参与后续运算，则会产生拟合曲线细节保留过多，产生虚假噪声。同时傅立叶方法的优势在于使用谐波（由 NOF 控制）来模拟植被物候规律，理论上上一周期内植被生长峰值个数与频数（NOF）相等。如果放任粗拟合获得 NOF 而不加干涉就会影响数据结果。因此，控制动态选频结果是有必要的。对于最小值频数，如果使用谐波个数过少（2 以下），则会导致拟合曲线过于平滑，因而丢失有用信息，所以本文将最小谐波个数（NOF）设定为 2（不包括零频），这是可行的。最大频数设定为 5，对于中国华北地区而言，一年内植被生长峰值不会超过 5。不同地区，可以根据当地的特点适当调整最大最小频数，实现曲线拟合效果和 NOF 表征物理意义间的调和。这是一种妥协策略。

处理异常点的临界点个数可以与 HANTS 方法中的超定度（DOD）对比理解。傅立叶方法使用最小二乘拟合傅立叶系数，当参与点个数多于 $2*NOF-1$ 时就可以获得结果。但允许 NOF 最大值才 5，如果只用 9 个点参与拟合显然是很难准确的描述 23 时相序列，使得处理结果不可信。因此有必要控制异常点处理临界点个数。同时如果将临界点设置过低，又会出现序列无法处理的情况。所以处理临界点个数通过妥协策略确定。本文中确定为 13。

改进方法数据处理结果：

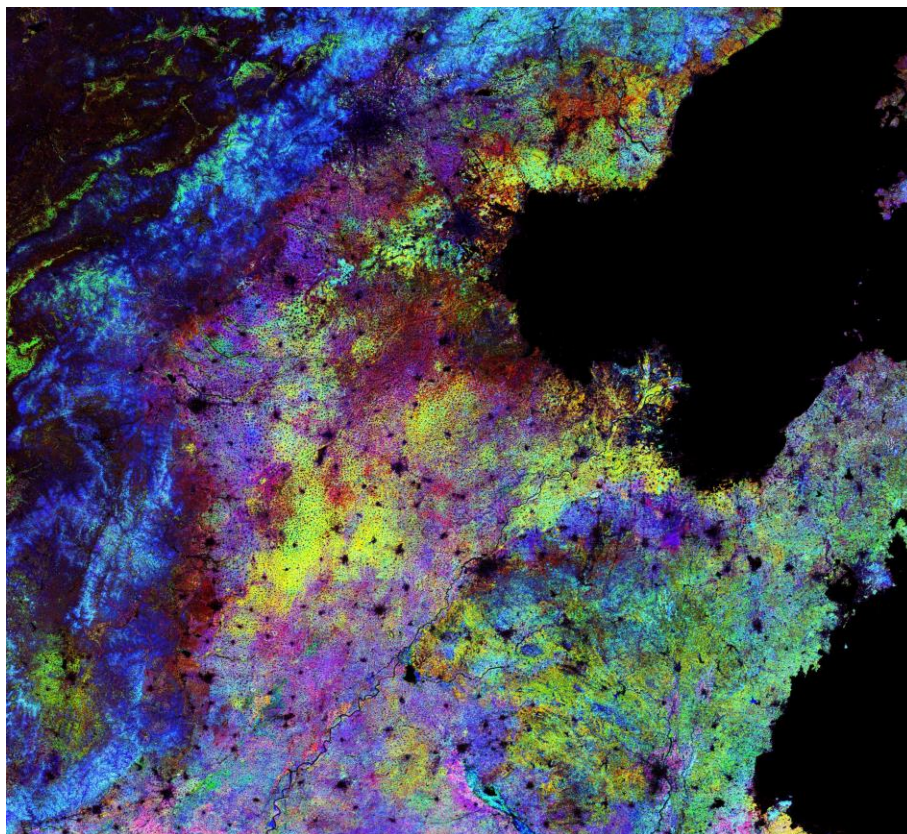


图 3-6 经改进方法处理后得到的 2003 年华北部分区域 250 米 EVI 影像（进过均衡化拉伸）
RGB 分别对应波段 17、13、9，相应儒历日为 257（9 月中下旬），193（7 月中下旬），129
（5 月中下旬）

（二）.HANTS 软件参数：

HANTS 软件参数需要通过反复实验确定最佳参数，本文与张霞（2006）使用同样的数据，故直接使用其确立的最佳参数，获得 HANTS 软件处理结果进行对比分析，具体参数如下：

- 1) 有效值范围（VDR）：-1000~8000；
- 2) 频率数：3；
- 3) 周期长度：23；
- 4) 周期：23 12 8；
- 5) FET：500；
- 6) DOD：4。

3.4.3 结果整体评价

HANTS 方法结果：

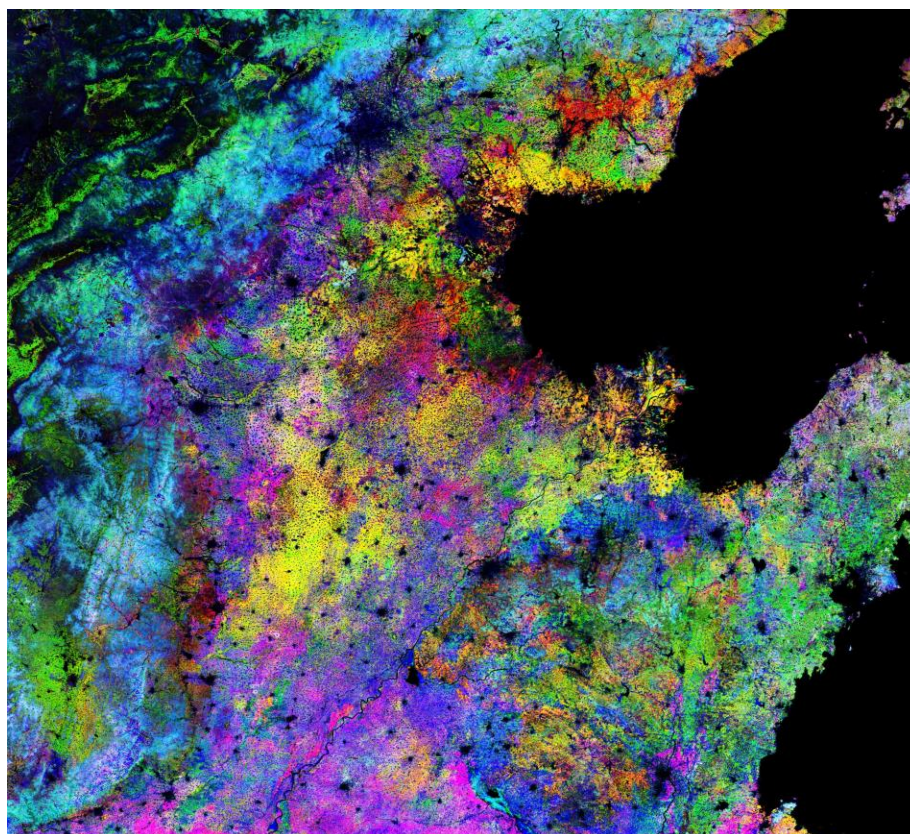


图 3-7 经 HANTS 软件处理得到的 2003 年华北部分区域 250 米 EVI 影像(经过均衡化拉伸) RGB 分别对应波段 17、13、9，相应儒历日为 257（9 月中下旬），193（7 月中下旬），129（5 月中下旬）

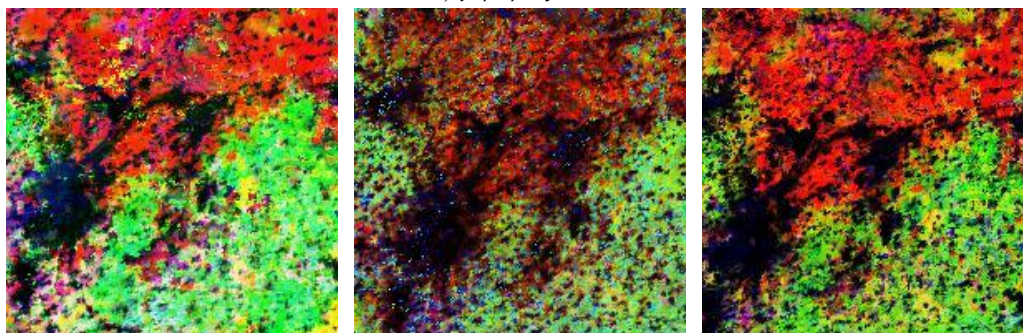
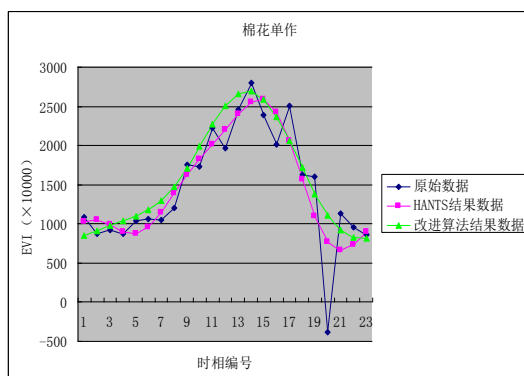
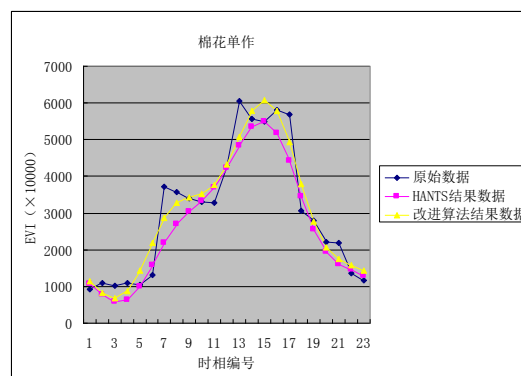


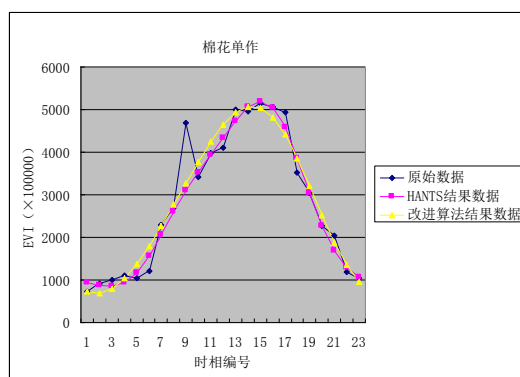
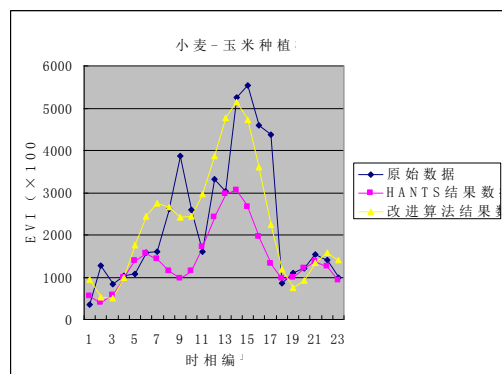
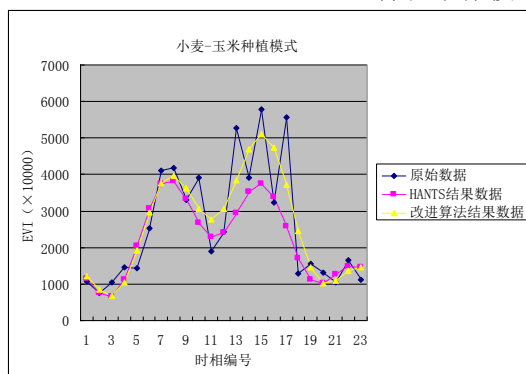
图 3-8 局部对比图（经过均衡化拉伸），依次为：原始影像、改进方法处理影像、HANTS 软件处理影像



a. 棉花单作模式（112° 59'E, 39° 9'N）



b. 棉花单作（115° 17'E, 38° 2'N）

c. 棉花单作模式 ($115^{\circ} 30'E$, $38^{\circ} N$)

d.

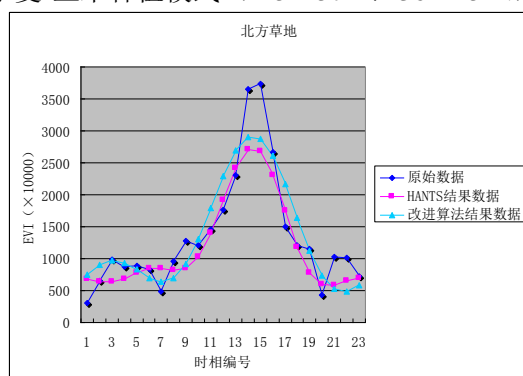
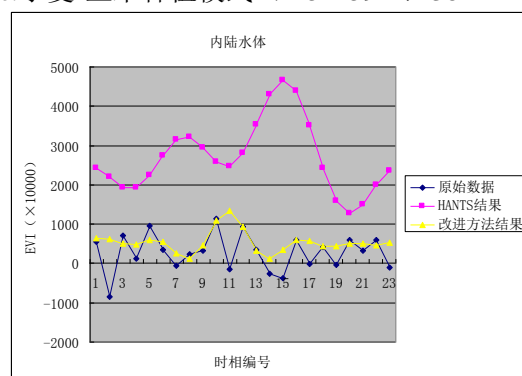
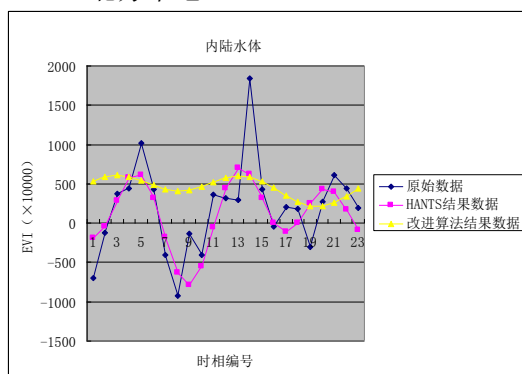
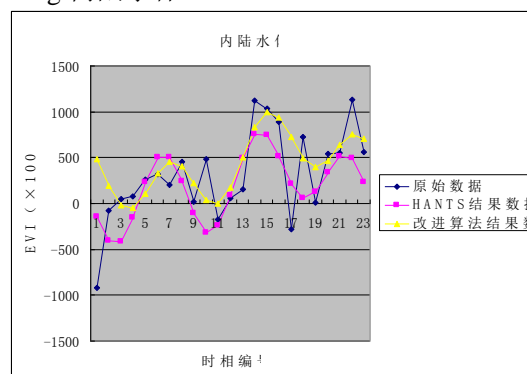
小麦-玉米种植模式 ($115^{\circ} 57'E$, $36^{\circ} 25'N$) e. 小麦-玉米种植模式 ($115^{\circ} 59'E$, $38^{\circ} 2'N$)f. 北方草地 ($114^{\circ} 9'E$, $40^{\circ} 31'N$)g. 内陆水体 ($119^{\circ} 19'E$, $32^{\circ} 55'N$)h. 内陆水体 ($116^{\circ} 11'E$, $35^{\circ} 57'N$)i. 内陆水体 ($117^{\circ} E$, $36^{\circ} 45'N$)

图 3-9 不同种植模式或土地覆盖类型样点对照曲线

对比分析提取的各类土地覆盖类型样点的时间序列曲线，改进方法结果与 HANTS 软件结果都忠实的反映了一定的物候规律。

棉花单作种植模式下，通常冬春土地空闲，一直延续到开春后开始种植棉花（即 3 月以后，不同地方时间略有区别，对应为第 6 时相以后）；四、五月份后棉花逐步进入生长旺季（物候资料显示河北深州出苗期为 5.2 日，下同，对应为 6-7 时相）；在 8 月份，棉花生长达到旺盛期（对应棉花裂铃期为 8.16 日，对应为 15-16 时相），之后棉花生长逐步减缓、直至成熟拔杆收获（通常为 10 月中下旬，物候资料显示为 10.23 日，对应为 20-21 时相）。文中 a、b、c 图为该类种植模式时间谱曲线。在 1-4 时相内，根据分析及物候资料，该时段地表覆盖变化小，从 EVI 上反映即为一段平直曲线，或者略有波动，但在 a 上 HANTS 结果出现了走低在回升的情况，而 b 图上不论 HANTS 还是改进方法，均出现了这样的问题。出现这样情况一方面固然有程序实现原因，本质上还是傅立叶谐波方法本身的问题，即使用绝对对称的正余弦曲线，该问题无法避免，只能通过合适的波数尽量减小影响。图中 HANTS 与改进方法之所以出现差异，也即各自使用的频数（参与拟合的谐波个数）不同造成的，HANTS 使用 3 谐波，而改进方法根据曲线峰值个数自动选频。从第五时相起，两类方法对数据的拟合都非常理想，也为物候资料相印证。总体来说，两种方法拟合曲线都清晰的反映了该种植模式。

d、e 反映小麦-玉米种植模式，样点中数据波动较大，但总体上反映出了一年两峰值的特征。冬小麦在 3-4 时相开始返青（观测时间为 2.20，落在第 4 时相）；第 7-8 时相内开始抽穗（观测时间为 5.3，落入第 8 时相），EVI 迎来第一个峰值；第 10 时相成熟（观测时间为 6.8 日，落入第 10 时相），EVI 值出现一极低值；由于后茬玉米的种植，在第 14-15 时相，又出现 EVI 峰值，此时为玉米吐丝期（观测时间为 8.18，落入第 15 时相），第 19 时相，玉米开始收割，EVI 值再次将为极低（观测时间为 9.30，接近第 19 时相）。两方法拟合曲线都正确的反映了该种植模式。

需要注意的是，HANTS 拟合曲线呈现出低于真实曲线的倾向，后文中将有分析。

内陆水体中由于浮游植物的生长，EVI 在一定程度上也出现了波动。但是理想水体的 EVI 曲线是一水平直线，其值也不会很高，而水中浮游植物对 EVI 值影响能力有限，因而在内陆水体真实曲线上很难有大起大落的波动（出现赤潮则另当别论）。g、h、i 上的内陆水体样点来在北方内陆湖泊，出现大规模赤

潮可能性相对小，从拟合曲线上看，改进方法表现更好，更接近实际情况下的时间曲线。

3.4.4 算法对序列数据扰动分析

通常情况下，在滤波重建过程中，会有三种原因对原数据产生扰动：

(1) .源于数据有效性判断导致的数据扰动；(2) .算法迭代过程剔除偏离值过大数据造成的数据扰动；(3) .使用拟合函数生产序列新数据导致的数据扰动；

这三种干扰是不可避免的，只能尽可能的使拟合值接近原始值，进而减小算法对原序列数据的过度扰动。本节将就原始序列样点数据分别与 HANTS 软件结果、改进方法结果、HANTS 软件处理结果与改进方法结果展开定量分析。本节共选 14 个样本点，其中棉花单作模式 3 个采样点、小麦-玉米模式 2 个采样点、北方草场 2 个采样点，内陆水体 3 个采样点、季节性裸地（冬春低 EVI 值）2 个采样点、三峰采样点 2 个，均来在华北平原地区。

表 3-2 原始数据—HANTS 结果较差对照表

Δ 时相	A	B	C	D	E	F	G	H	I	J	K	L	M	N	时相内均 方差	时相内较 差正值数
1	64	752	436	-129	241	-205	-134	-51	317	-212	-504	-779	31	66	383.08	7
2	-174	898	-591	315	1851	48	1058	-4	597	867	-81	329	-141	1838	741.17	9
3	-67	967	1251	453	2011	150	1260	378	-22	266	89	466	208	1826	684.30	12
4	-17	937	234	464	1409	172	986	340	-65	54	-138	234	100	1842	599.26	15
5	166	828	3657	32	-57	-131	248	-596	656	-328	409	36	-194	6173	1858.2	9
6	104	701	273	-264	197	-369	-302	-543	1506	10	112	-183	148	-62	513.01	11
7	-102	632	-116	1513	-203	210	642	376	10	172	-225	-299	-396	-3	506.01	7
8	-182	693	-483	890	-162	90	850	369	2302	1472	-298	219	1373	16	797.55	10
9	129	921	14	368	212	1596	-80	-26	2043	2913	648	126	1231	548	895.27	12
10	-98	1307	400	-15	342	-104	182	1231	1395	1457	150	807	1172	388	589.72	11
11	220	1795	161	-425	-452	54	618	-410	38	-101	424	71	-134	-61	569.84	8
12	-230	2292	-286	44	53	-227	-197	24	-90	902	-118	-34	-39	36	679.31	6
13	58	2691	-87	1194	1722	257	22	2325	2926	68	-404	-345	2734	3353	1392.3	11
14	251	2901	-28	218	0	-110	1356	395	2693	2196	1223	364	236	6365	1786.8	12
15	-192	2873	68	-1	145	-55	898	2051	471	2893	113	285	1270	788	1060.77	11
16	-410	2609	122	647	732	16	1683	-155	194	2626	-51	384	-50	1797	1029.05	10
17	448	2165	14	1283	1905	351	241	2982	1964	3061	325	-498	2703	2921	1254.80	13
18	57	1636	162	-392	-284	-344	-183	-413	-507	-111	180	668	220	1	556.43	7
19	499	1131	-272	239	193	-4	641	416	479	106	-559	-117	2178	-14	671.67	9
20	-1159	743	-91	275	686	0	999	274	687	-16	-164	207	3796	102	1095.96	9
21	479	528	728	573	141	358	1039	-202	575	173	207	37	2894	-34	754.60	12
22	221	491	342	-101	25	-100	163	151	-38	148	273	646	1657	274	445.63	11

23	-35	590	-246	-141	-369	-22	39	-347	-278	70	283	335	-306	-24	283.17	5
时相 间均 方差	343.5	1610	878.2	610.1	901.7	391.7	784.25	1004	1277	1417	404.075	404.202	1536.707	2265.366		
时相 内较 差正 值数	12	23	14	15	16	11	18	13	17	18	13	16	16	17	较差标准 差	2282.99

注：△表示序列中原始值与 HANTS 方法重建后对应值之差。红色为绝对值最大或最小的较差。

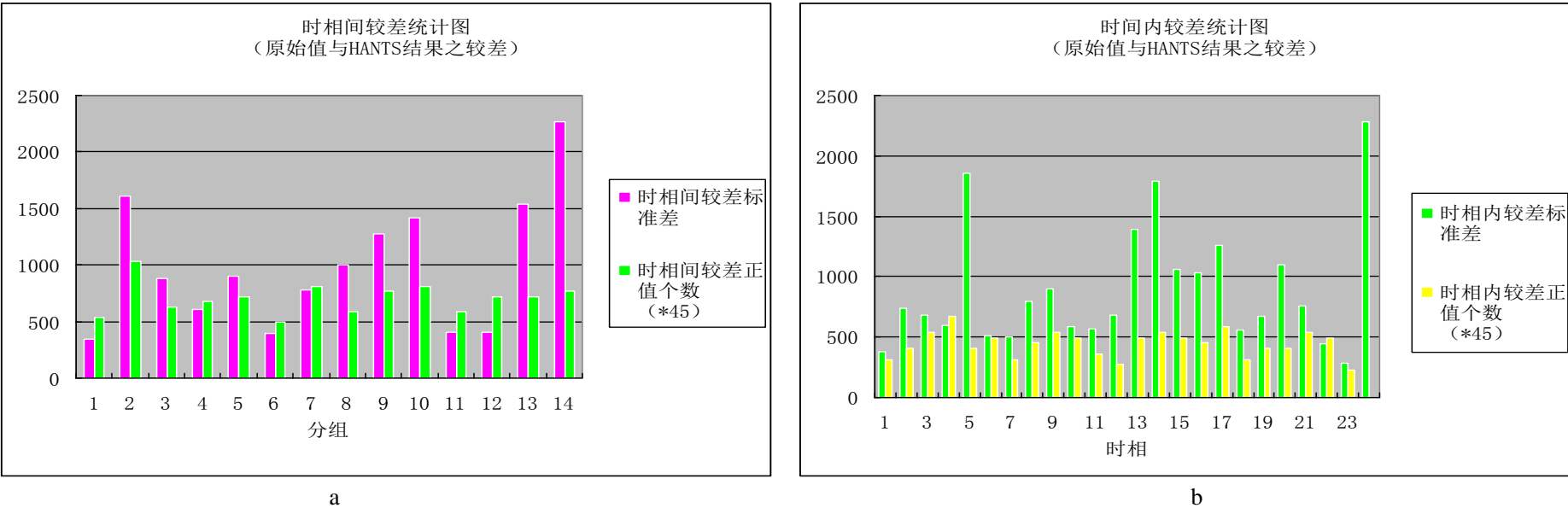


图 3-7 原始数据—HANTS 结果较差标准差及较差正负值个数统计图，a 为时相间统计图，b 为时相内统计图

表 3-3 原始数据—改进算法结果较差对照表

Δ 时相	A	B	C	D	E	F	G	H	I	J	K	L	M	N	时相内标准差	时相内较差正值个数
1	172	-449	889	-217	-424	11	-683	-137	805	-609	-1225	-1403	454	-545	674.7051	5
2	137	-245	-741	249	707	234	437	-100	162	733	-712	-271	479	308	462.2882	9
3	7	15	-233	342	568	206	593	357	-1235	330	-236	70	122	-564	483.1811	10
4	-141	-62	-2328	220	35	72	313	419	-1491	55	-153	127	-975	-994	797.0548	7
5	-227	57	1172	-405	-978	-335	-379	-467	-392	-691	482	155	-1800	3514	1224.504	5
6	-220	128	-1026	-872	-72	-595	-828	-454	1056	-871	-47	-11	-1165	-1967	722.5462	2
7	-150	-142	-63	840	81	19	251	350	-128	-1152	-836	-254	-829	-885	549.2748	5
8	-88	259	132	299	333	-70	598	206	1885	-41	-1339	50	1740	8	781.9383	10
9	-90	352	277	-29	515	1418	-234	-303	865	1463	-561	-207	1684	924	730.26	8
10	-163	-102	78	-201	203	-348	47	863	-565	164	-872	451	833	553	499.1612	8
11	-258	-337	-260	-487	-1011	-253	405	-885	-2207	-1357	-160	-170	-1696	-547	700.8086	1
12	-304	-524	-285	-43	-668	-528	-570	-623	-1885	-551	-255	-118	-2468	-1223	690.7824	0
13	-258	-386	297	952	1168	66	-550	1430	2105	-1739	-309	-352	408	1508	1026.409	8
14	-136	747	247	-225	-180	-120	604	-770	2828	101	1255	292	-987	4287	1419.825	8
15	-4	864	-100	-587	327	106	36	699	1004	818	-107	36	1535	-1185	696.8742	9
16	54	52	-274	40	1085	252	806	-1513	364	978	-499	-49	1184	134	708.9512	10
17	-2	-665	-125	776	2196	524	-564	1840	1275	2129	-146	-1007	3743	1636	1361.924	8
18	-143	-434	433	-744	-179	-335	-863	-1169	-1982	-318	-85	231	-71	-901	622.5551	2
19	-285	19	16	25	146	-171	93	88	-1181	351	-524	-386	115	-531	395.4084	8
20	-341	-313	-272	133	611	-253	549	277	-407	273	52	78	477	-23	345.0626	8
21	-268	501	106	442	119	153	632	-50	488	190	347	-82	-501	166	314.9472	10
22	-99	529	-146	-239	-8	-151	-257	278	739	-178	105	384	-659	584	391.2833	6
23	79	136	-124	-266	-613	99	-436	-336	726	-427	-245	-141	-1052	30	413.6672	5
时相	145	406	668	480	723	409	533	780	1355	895	556	414	1400	1463		

间标准差																
时相间较差正值数	5	12	10	11	14	12	13	10	13	12	5	10	12	11	较差标准差	1520.072

注：△表示序列中原始—改进算法结果对应值之差。红色为绝对值最大或最小的较差。

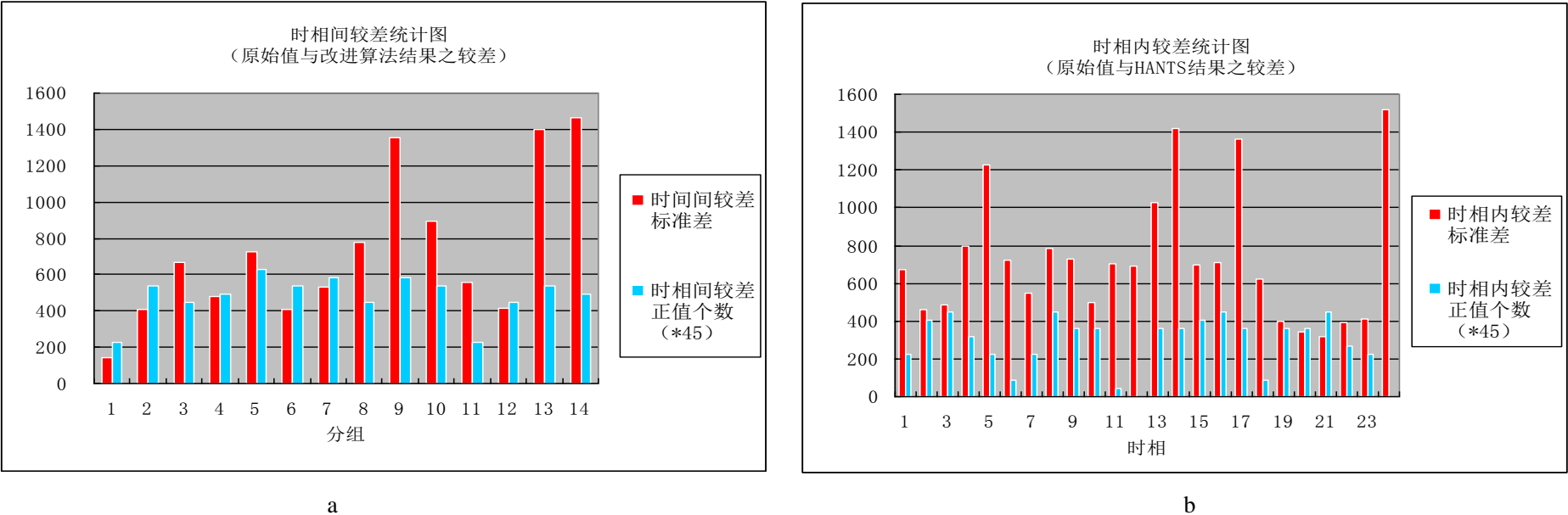


图 3-8 原始数据—改进算法结果较差标准差及较差正负值个数统计图，a 为时相间统计图，b 为时相内统计图

表 3-4 HANTS 结果—改进算法结果较差对照表

Δ 时相	A	B	C	D	E	F	G	H	I	J	K	L	M	N	时相 内均 方差	时相 内负 值数
1	172	-69	453	-88	-665	216	-549	-86	488	-397	-721	-624	423	-611	447	9
2	137	-256	-150	-66	-1144	186	-621	-96	-435	-134	-631	-600	620	-1530	551	11
3	7	-333	-1484	-111	-1443	56	-667	-21	-1213	64	-325	-396	-86	-2390	752	11
4	-141	-250	-2562	-244	-1374	-100	-673	79	-1426	1	-15	-107	-1075	-2836	965	12
5	-227	-53	-2485	-437	-921	-204	-627	129	-1048	-363	73	119	-1606	-2659	921	11
6	-220	141	-1299	-608	-269	-226	-526	89	-450	-881	-159	172	-1313	-1905	617	11
7	-150	217	53	-673	284	-191	-391	-26	-138	-1324	-611	45	-433	-882	450	11
8	-88	128	615	-591	495	-160	-252	-163	-417	-1513	-1041	-169	367	-8	573	10
9	-90	-75	263	-397	303	-178	-154	-277	-1178	-1450	-1209	-333	453	376	605	10
10	-163	-283	-322	-186	-139	-244	-135	-368	-1960	-1293	-1022	-356	-339	165	565	13
11	-258	-393	-421	-62	-559	-307	-213	-475	-2245	-1256	-584	-241	-1562	-486	614	14
12	-304	-373	1	-87	-721	-301	-373	-647	-1795	-1453	-137	-84	-2429	-1259	742	13
13	-258	-274	384	-242	-554	-191	-572	-895	-821	-1807	95	-7	-2326	-1845	803	12
14	-136	-190	275	-443	-180	-10	-752	-1165	135	-2095	32	-72	-1223	-2078	787	11
15	-4	-199	-168	-586	182	161	-862	-1352	533	-2075	-220	-249	265	-1973	815	10
16	54	-300	-396	-607	353	236	-877	-1358	170	-1648	-448	-433	1234	-1663	810	9
17	-2	-418	-139	-507	291	173	-805	-1142	-689	-932	-471	-509	1040	-1285	620	11
18	-143	-455	271	-352	105	9	-680	-756	-1475	-207	-265	-437	-291	-902	449	11
19	-285	-356	288	-214	-47	-167	-548	-328	-1660	245	35	-269	-2063	-517	663	11
20	-341	-150	-181	-142	-75	-253	-450	3	-1094	289	216	-129	-3319	-125	898	11
21	-268	61	-622	-131	-22	-205	-407	152	-87	17	140	-119	-3395	200	909	9

22	-99	165	-488	-138	-33	-51	-420	127	777	-326	-168	-262	-2316	310	688	10
23	79	108	122	-125	-244	121	-475	11	1004	-497	-528	-476	-746	54	437	7
时相 间均 方差	145	209	855	209	556	180	212	504	913	766	401	220	1345	1011		
时相 间负 值数	18	17	13	23	16	15	23	16	17	18	17	20	16	18	较差 标准 差	405.4

注：△表示序列中 HANTS 结果—改进算法结果对应值之差。红色为绝对值最大或最小的较差。

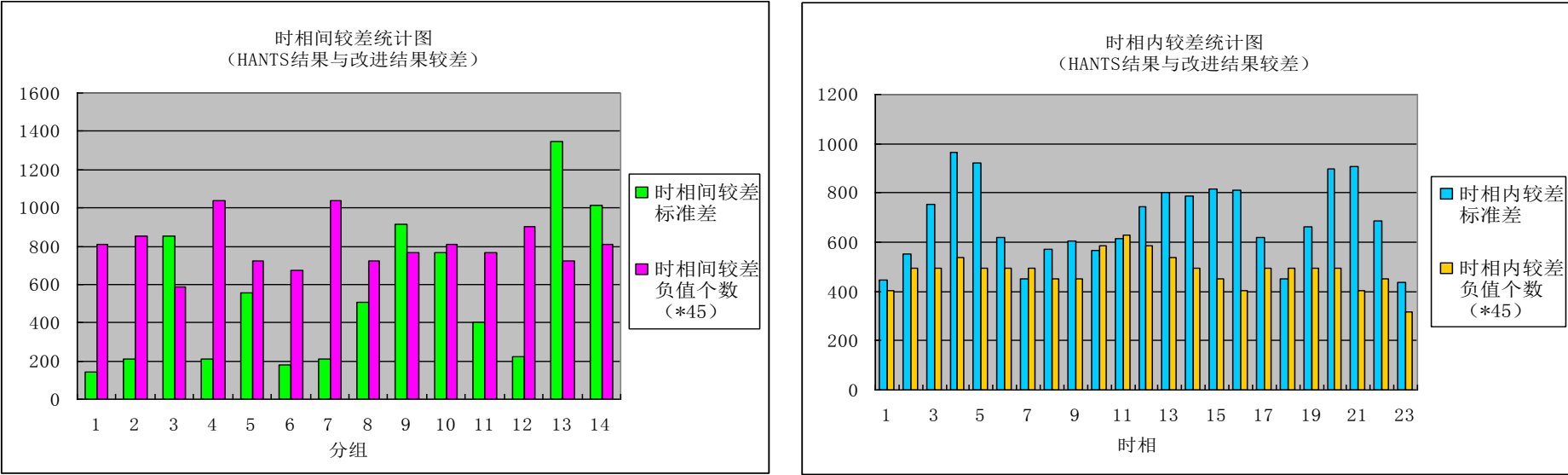
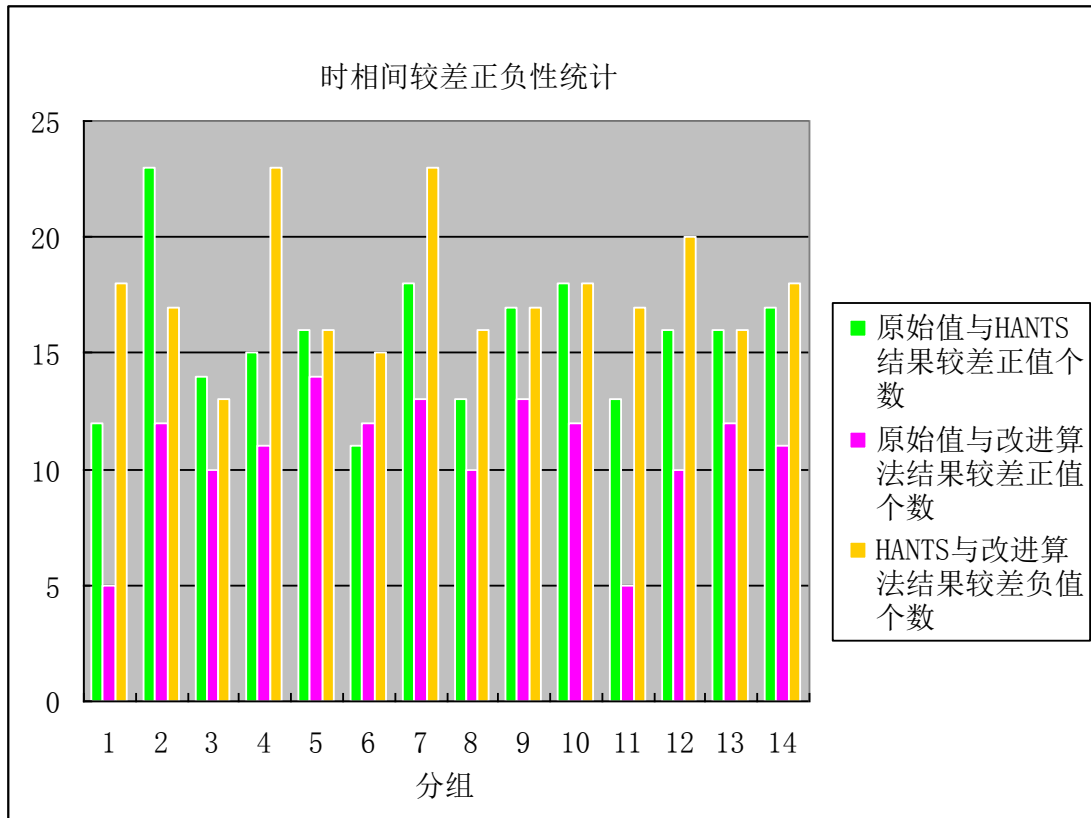
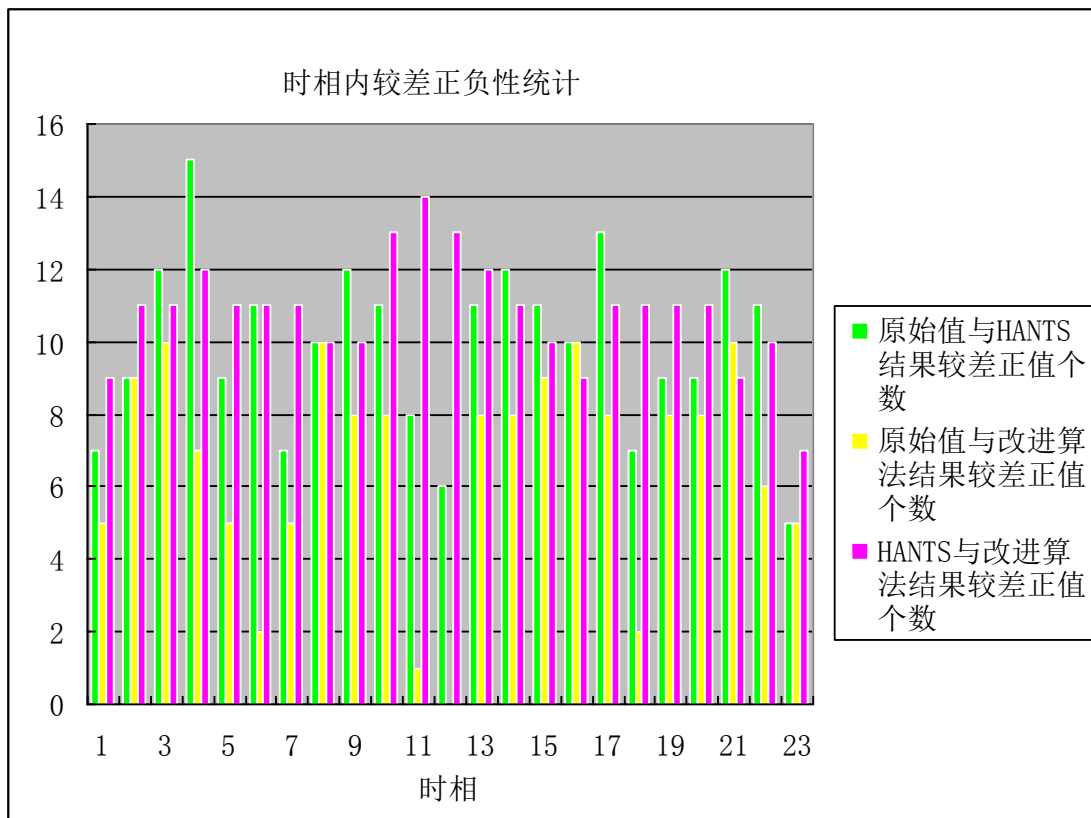


图 3-8 HANTS 结果—改进算法结果较差标准差及较差正负值个数统计图，a 为时相间统计图，b 为时相内统计图



a



b

图 3-9 较差正负性统计图：a 为时相间较差，b 为时相内较差

从“原始数据—HANTS 结果较差对照表”可以得知,在选取的样点中,原始与 HANTS 较差绝对值最大为 6365,最小为 0,较差均方差为 2282.99。时相内的均方差最大值为 1858,最小 283.17,时相内均方差为 417.045;时相间的均方差最大为 2265.366,最小 343.5,时相间均方差为 571.38。

从“原始数据—改进算计结果较差对照表”中可以得到如下信息:在选取的样点中,原始与改进方法结果较差绝对值最大为 4287,最小为 8,较差均方差为 1520.72。时相内的均方差最大值为 1419.82,最小 314.95,时相内均方差为 344.099;时相间的均方差最大为 1463,最小 145,时相间均方差为 410.668。

依据上面的统计,不论 HANTS 还是改进方法,对原始数据的扰动,时相间都小于时相内的扰动,这是由于在谐波分析过程中,对于每个时相的值,总是试图找到一个合适的拟合值代替,以达到最佳滤波的效果,保证了部分点接近曲线,必然损失另一些点值,因而造成时相内较差离散,均方差偏大,对原始数据产生较大的扰动,因而时相内较差大。而对于时相间,因为最终拟合曲线是要尽可能的接近原始曲线,因此拟合曲线,在形态、振幅上都接近原始曲线,因而较差标准差相对聚集,从时相间较差方差上显示均方差值相对小,即扰动小。

但这两个参数只反映了数据的整体扰动情况,再结合较差绝对值最大最小值、均方差最大最小值,即可全面评价滤波重建方法对原始数据的扰动。HANTS 方法较差均方差达到 2282.99,较改进方法超过 762,比其时相内或组内均方差最大值都大;较差绝对值最大值与最小值相差 6365,悬殊颇为诡异。尽管这有可能是因为异常点造成的误会,但偏差也过于夸张。

从“HANTS 结果—改进算法结果较差对照表”中,可以得到如下信息: HANTS 与改进算法结果较差绝对值最大值 2836 (-2836),最小值 1。较差均方差为 405.4。时相内均方差最大值为 965,最小值为 437,时相内均方差为 163.47;时相间均方差最大为 1345,最小为 145,时相间均方差为 381.73。

分析发现,两类方法在时相内表现较为一致,均方差聚集,其值集中在数百以内,但在时相间,均方差较为离散。这反映了两种方法,在处理上的不同,如输入值有效性的判定、对异常值的取舍等,不同的策略都会对拟合曲线在形态、曲线振幅上有所反映,因而在时相间均方差上表现出离散程度相对较大。

上述三张数据对照表中,还给出了:原始数据-HANTS 结果、原始数据-改进算法值、HANTS-改进算法值较差正负值个数。通过这个指标,一定程度上

反映两类方法在滤波重建时数据拟合的接近程度。在应用中，我们时常需要把握拟合值的偏向性：即拟合结果与原始值相比，是正偏性还是负偏性。HANTS 软件中，尽管有“方向性界外值切换”参数，但该值并不能表明在实际滤波重建中拟合值的偏向性，其用于表明高于还是低于拟合值的点被剔除，相当于是异常值判断准则之一。

从“原始数据—HANTS 算法结果较差标准差及较差正负个数统计图” a、b 图中可知，样点中 HANTS 结果比原始值普遍较小。时相内同一个时相至少有一半的拟合值都小于原始值，甚至有 HANTS 值全部小于原始值的情况。而在时相间，虽未有 23 个时相全部小于原始值的情况，但每个时相内 HANTS 结果普遍小于原始值的情况更为突出，大部分都在 17 个以上。这反映了 HANTS 拟合结果具有负偏性，即拟合值比原始偏小，曲线偏低。

与 HANTS 相比，“原始数据—改进算法结果较差标准差及较差正负个数统计图” a、b 图上反映改进结果无论是时相内还是时相间，都未出现大量小于原始的情况：时相内，除个别时相外，其余正值个数均小于 10（即原始值>改进结果）；时相间这一结果更明显，通常都是正负对半。

进一步对比分析 HANTS 结果和改进算法结果，如 HANTS 结果—改进算法结果较差标准差及较差正负个数统计图，HANTS 结果值普遍小于改进结果值：时相内，负值个数保持在 9 以上(样本点为 14 个)；时相间，绝大时相改进方法值都大于 HANTS 方法结果。

从统计图上还反映出了一个有趣的现象，即不论是 HANTS 方法结果和改进方法结果，其较差正（负）值个数，都是按照某个周期波动，波形更像是三角曲线。而对于较差标准差，如果略去突起值，其大体也是按照某个周期波动，波形也似三角曲线。这或许与傅立叶谐波分析方法使用正余弦曲线存在某种关系。

通过上面分析对比，总体看来，改进方法对原始数据扰动要小于 HANTS 结果，并且 HANTS 结果成负偏性，拟合曲线呈低于原始曲线趋势，改进方法较 HANTS，结果更为接近原始曲线。

3.4.5 异常值处理结果分析

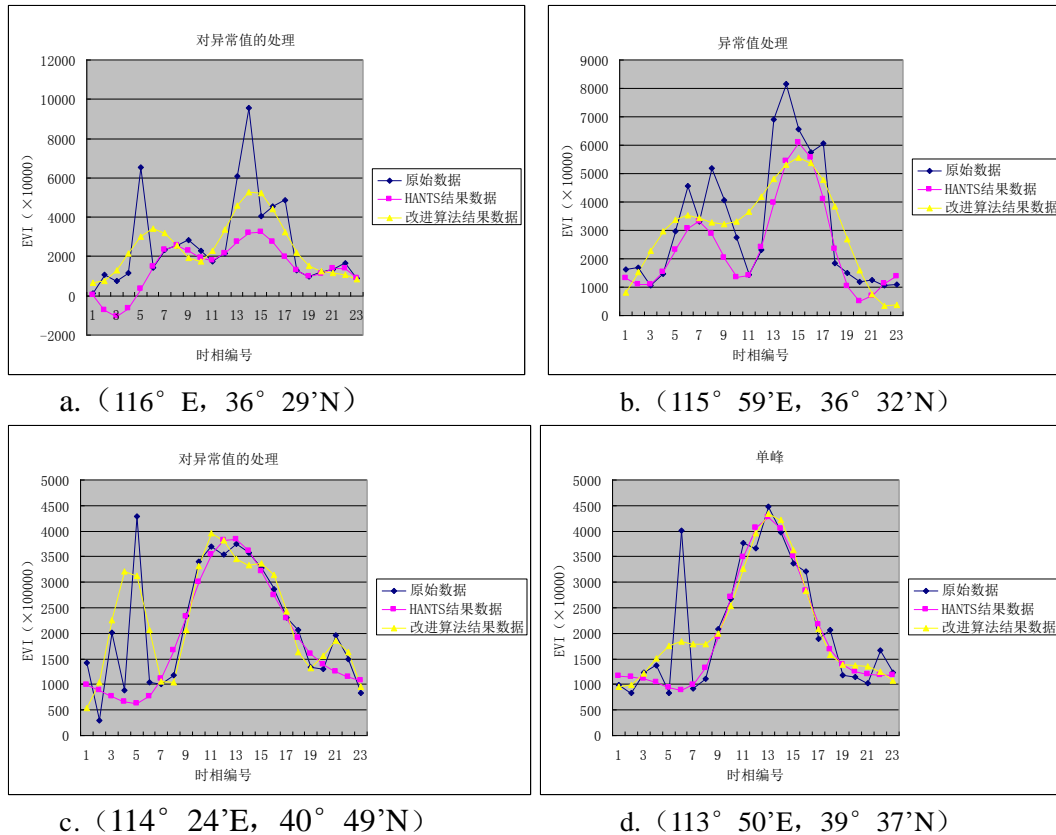


图 3-10 异常样点对照曲线

对于突变值，本文提出要区别对待，即区分是异常值还是正常突变，进而采取对策处理。改进方法在两处进行突变值的检测判别，一是在输入值有效性判别时，将过分大或小的值滤去（人为设定阈值，主观性大），起辅助左右；二是在程序运行过程中，通过判别拟合值与原始值较差，使用异常监测算法动态检测提出异常值（检测算法请参见上文）。

文中选取两类异常点，一为正常生长情况下的高值突变点（参见图 3-10 a、b）它的出现，有可能是因为某种天气或人为因素造成的植被生长期内的某种突变（如 b 图中第 7 时相点），该类异常某种程度上讲依然反映一定的真实情况；另一种则为纯粹噪声点（参见图 3-10 c、d），它的存在严重影响了时间曲线反映植被真实生长规律。

本节样点曲线 a 中，在第 5、13、14、15、17 时相，发生了 EVI 值的异常的突变。从第 4 时相到第 5 时相，EVI 出现陡升，第 6 时相又恢复到第 4 时相水平；第 13 时相又跃升，第 14 时相时跃升至最高点。到第 15 时相时出现突降，但第 16、17 时相内小幅上升，第 18 时相又跌至谷底。其中又以第 5、12、13

为异常。

该样点采自河北南部地区，是典型的一年两熟制种植区。该区第一茬作物通常是第 4 时相开始返青（表现为 EVI 值上升），第 6 时相内抽穗（表现为 EVI 值跃升至第一稿顶峰），在第 10 时相内收获（表现为 EVI 值下落至局部最小值）。在接下来的时相里是第二茬作物生长达到高峰，然后收获（即在第 15 时相达到生长峰值，第 18 时相收获）。之后是越冬作物的种植，并且在越冬前出现一个小的长势高峰（即第 22 时相左右），然后进入越冬期，生长减缓或停滞（表现为 EVI 值下降或保持低水平状态）。

分析清楚该区作物生长模式，就可以分析两种方法对突变点（异常点）的应对策略。HANTS 方法在 1~6 时相内较少采信原始数据，导致样点曲线在该时相内出现负值，严重与实际情况不符。因此从整体上讲，就该区域改进方法的策略更为优胜一些。它既兼顾了对异常值的回避，又保证能客观的反映作物真实的生长模式。b 图曲线中异常值（突变点）与 a 图中情况类似。

本节样点曲线 c 中该样点异常值主要出现在 1~6 时相内，时相内数据 EVI 值出现大的波动，对该波动不同的处理策略导致了明显不一致的结论。HANTS 方法放弃突变值，因而获得一条单峰曲线，而改进方法部分采信突变值，最终获得一条双峰曲线。本样点采自北京西北部地区，该区主要为一年一作种植模式，因而单峰曲线较为可信，即便个别地区存在双作情况，也不可能第 2 时相就返青，5 时相就开始抽穗，第 8 时相即行收获。d 图中异常值（突变点）情况与 c 图中情况类似。

改进方法的异常检测算法在 c、d 图中分别失效。失效的原因很多，如算法中使用 Grubbs Test。该算法要求数据服从正态分布，由于为了避免麻烦，增加算法开销，算法直接将拟合值与真值较差对数化，参与检测。这样做是否正确或者在一定的精度范围内是否合理就需要论证。在某些情况下，即便不是异常值，但该值对于拟合整条曲线利小于弊时，通过简单的阈值剔除反到更有助于拟合。此外，某些情况下，突变值是异常值，但该值的确是地表植被覆盖状况的真实反映，使用算法辛苦的剔除，反到不好。但是，通过异常检测算法剔除异常值（突变值）而非单纯根据某个阈值剔除能够尽可能的减小操作者主观误差的引入，减少方法对数据的扰动。因而使用异常值监测算法尽管有问题存在，但更为合理，所需要的只是找到一个合适的方法或者策略，有效的避免上述问题的出现。

3.4.6 自动选频（NOF）结果分析

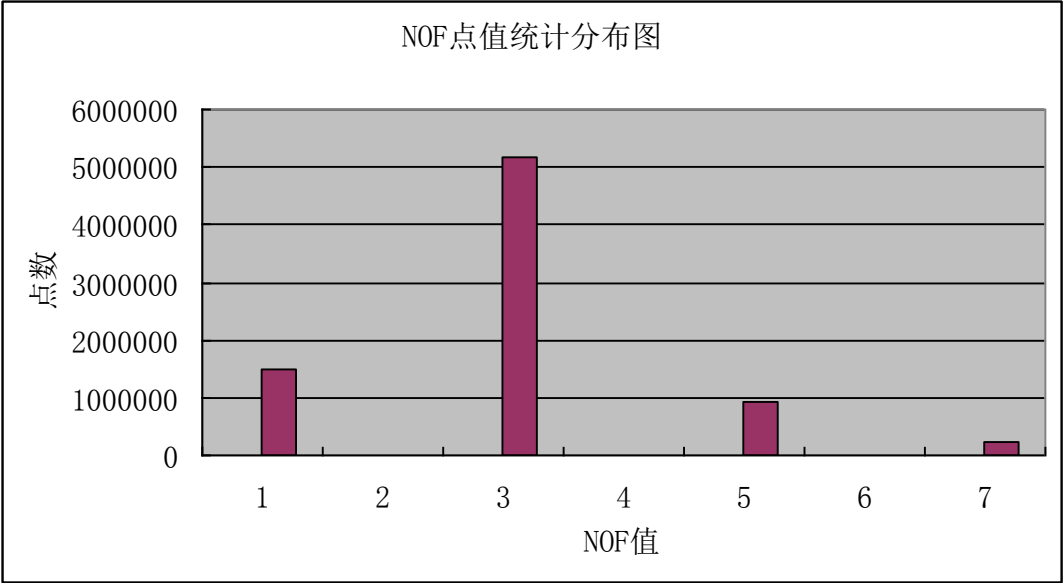
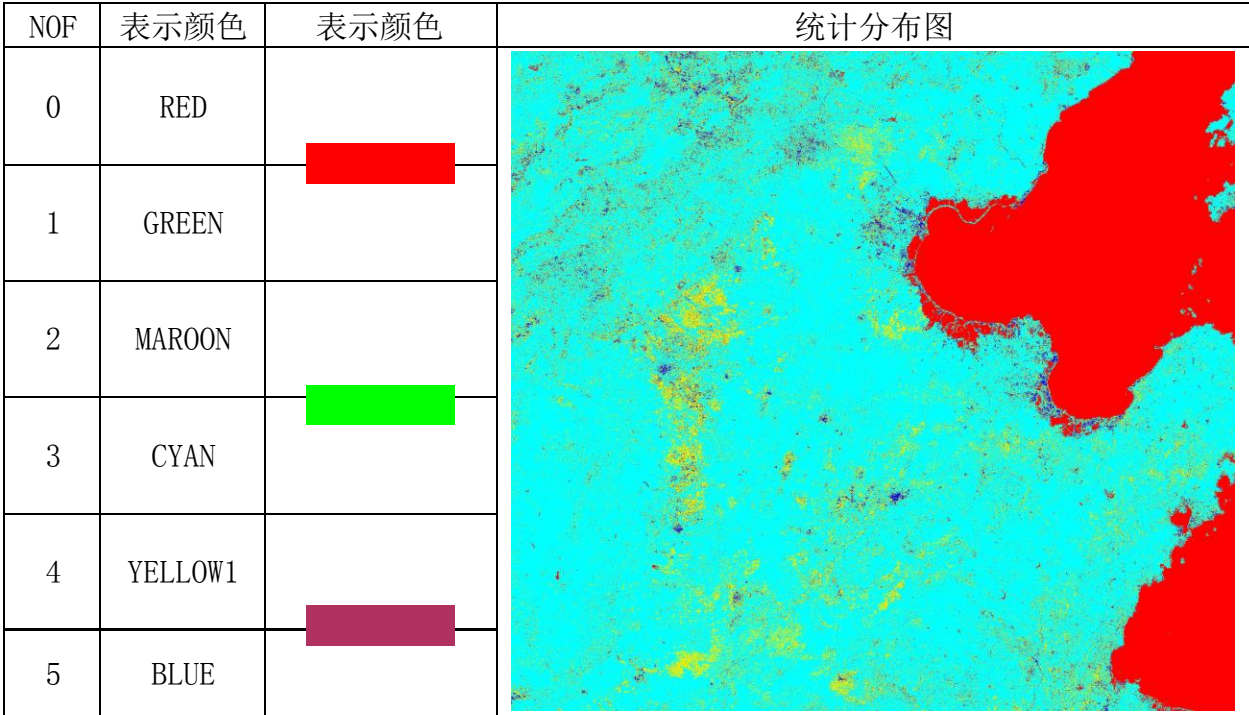


图 3-11 NOF 统计分布图

表 3-5NOF 统计分布表

NOF 值	0	3	4	5
点个数	1481562	5184839	939204	228175
所占比例 (%)	(无值点)	81.623	15.258	3.593



注：红色代表无数据区。

图 3-12 自动选频结果图

自动选频策略的提出主要是因为在传统的谐波分析中，对全局使用同一个频数（参与拟合的谐波个数）。尽管这样也能满足要求，但有可能谐波分析的特色因此被抹杀。谐波分析正是借助不同周期的谐波，来模拟不同种植模式曲线（或植物生长曲线）。如果全局使用同一个频数，则不可能体现这一优势。因而本文提出动态选频，根据前期预测不同的生长模式使用不同的频数。

从 NOF 统计分布表上可知，使用 3 谐波（NOF）参与拟合的点占总点数 81%，另有使用 4 频数的点占 15.258%，使用 5 及其以上频数的点有 3.593%。在自动选频结果图中，可以看出这些点有分布。显然，HANTS 方法使用全局同一频数存在牵强（有 18.851% 的点不认同全局同一 NOF）。另外，本文使用数据为中国华北部分地区，范围相对较小，区内种植模式相对简单，倘若是应用于全国范围，或者更大范围，使用同一 NOF 引起的问题更为突出。

两点需要说明：一是频数值（NOF）。本文中为了保证拟合效果，在选用谐波个数与表征物理意义上有所折中，即 NOF 值限定在 2-5。这只是为了控制程序运行而设定的，有主观认识因素在其间；而将 NOF 最小值设置为了保证拟合曲线接近真实值，因为如果选用谐波个数过少，会造成拟合曲线过分平滑，结果失真。二是本文中选择出的 NOF 都在自动选择值上加一，目的也是确保足够的谐波参与拟合，是拟合曲线更接近真实情况。

3.4.7 迭代终止条件分析

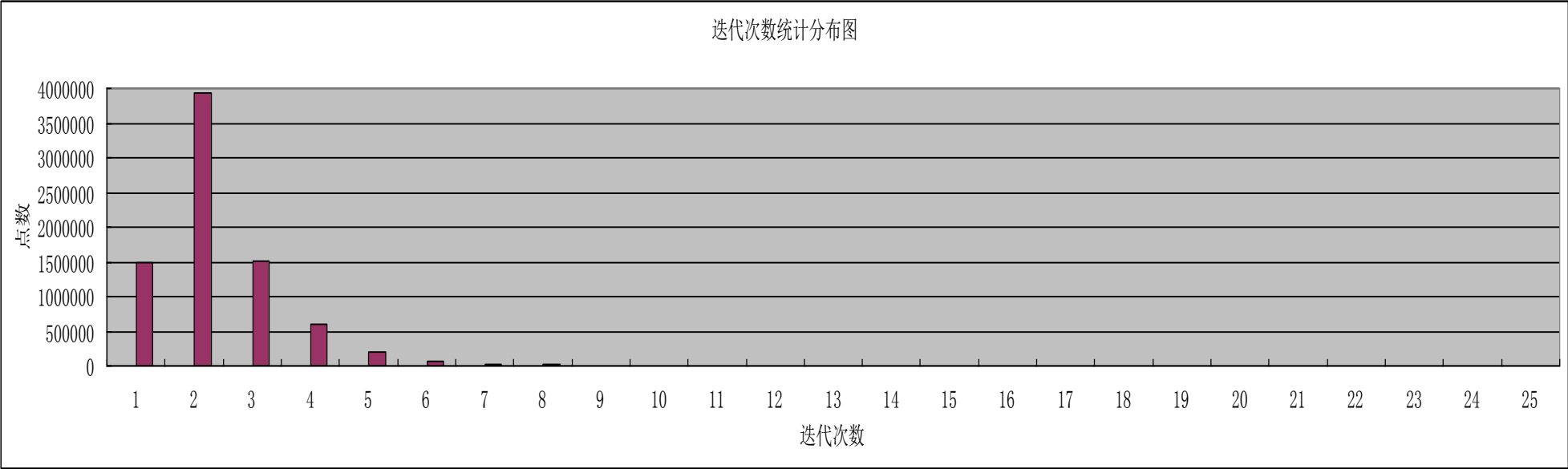
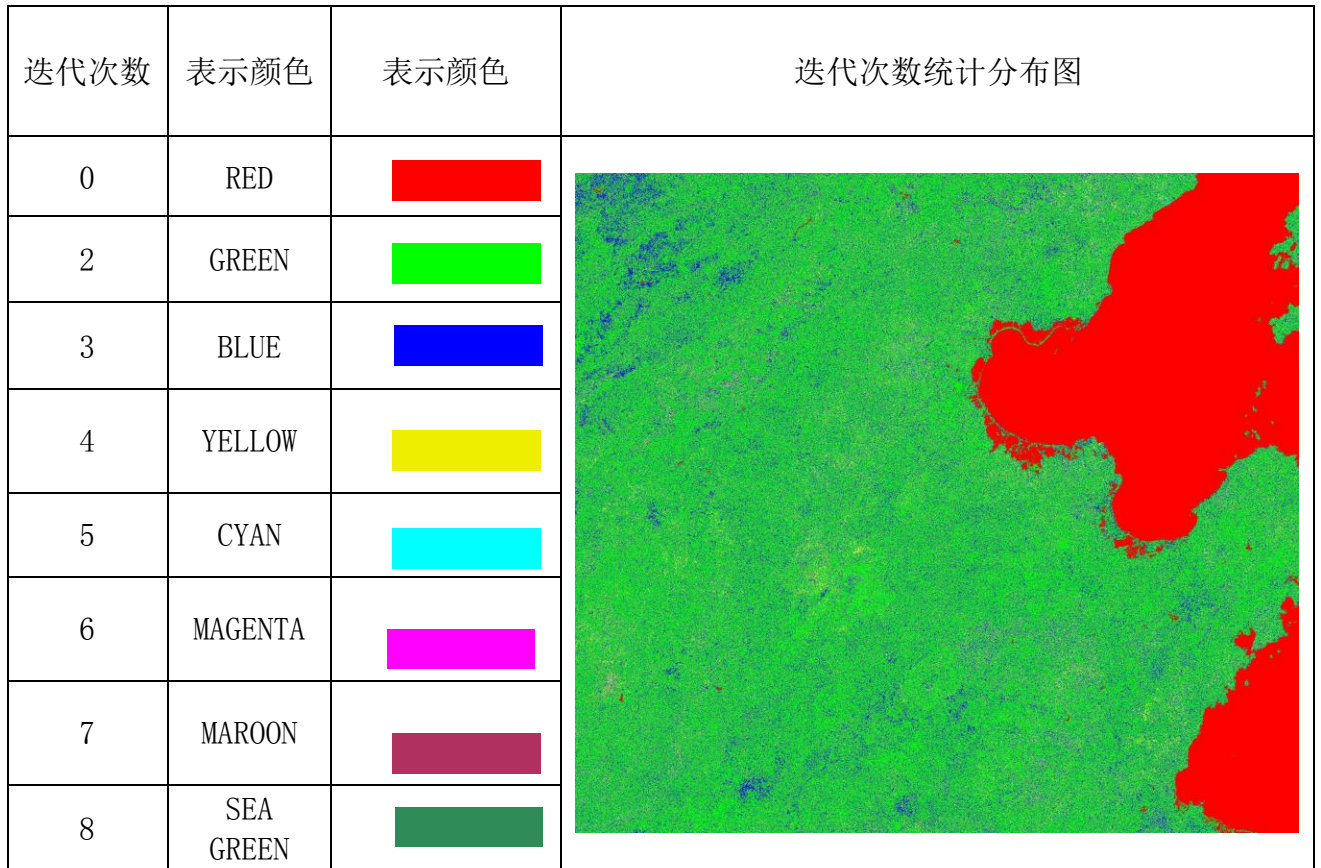


图 3-13 迭代统计分布图

表 3-6 迭代次数统计分布表

迭代次数	0	2	3	4	5	6	7	8	9	10	11	12	13
点个数	1481562	3927698	1505479	599345	201575	70275	27445	11792	5092	2123	809	341	其他
所占比例	18.9125	50.1380	19.2178	7.6508	2.5732	0.8971	0.3503	0.1505	0.0650	0.0271	0.0103	0.0044	0.003



注：红色代表无数据区。

图 3-14 迭代次数统计分布图

本文除了使用超定度 (DOD) 来控制拟合迭代外, 还引入了拟合影响因子, 放弃使用人为设置的阈值 (FET) 控制迭代终止, 试图减小后天人为主观干扰。从数据结果分析图表中可以发现大多数点可以在 5 次迭代以内完成拟合, 占到了总点数的 98.49%, 这说明使用拟合影响因子是有效的, 在控制了拟合迭代正常终止的情况下亦未大量增加程序时间开销。

3.5 本章小节

本章详细介绍了傅立叶谐波分析方法的原理, 总结其在滤波重建的运行流程和相关技术问题以及存在的重大缺陷, 在通过分析后, 针对这些缺陷提出了改进弥补方案。四条改进原则的提出, 即: 尽可能少的扰动原始数据、尽可能少的使用辅助数据、非单纯的数据运算、效率效果的兼顾。提出使用异常检测算法选择参与拟合数据, 对剔除异常值; 提出使用粗拟合的方法, 动态估计频数 (NOF); 引入拟合影响因子, 控制拟合迭代的自动终止。最后通过与 HANTS

软件结果对比分析, 定量的给出算法对原始数据扰动、异常值处理、自动选频、迭代终止策略等分析评价。结果显示改进之处都运行良好, 但是也存在瑕疵。同时分析发现, HANTS 软件结果存在负偏性, 即拟合结果大部分小于实际值, 拟合曲线大部分低于真实曲线。

第四章 基于植被指数时间序列的北京及周边沙尘源地提取

4.1 北京沙尘源地综述

一)、北京沙尘概述

沙尘一直是北京环境治理的一块心病,近年来北京地区上空更是不时出现遮天蔽日的风沙,市区的风力达5至8级,扬沙天气、局部地区沙尘暴一再肆虐,严重时能见度不足100米(苏海波,2001)。

根据沙尘造成大气水平能见度的差异,气象观测中将沙尘天气划分为3类,即浮尘、扬沙和沙尘暴(方翔等,2002),分别定义如下:

(1) 浮尘悬浮在大气中的沙或土壤粒子使大气水平能见度小于10km的天气现象。

(2) 扬沙风将地面尘沙吹起,使空气相当混浊,且水平能见度在1~10km以内的天气现象。

(3) 沙尘暴强风将地面尘沙吹起,使空气很混浊,大气水平能见度小于1km的天气现象。根据水平能见度大小,又将水平能见度小于500m的沙尘暴称为强沙尘暴。

北京地区同时遭受上述三种大气污染。

二)、北京沙尘天气的时间分布:

扬沙、浮尘和沙尘暴等三种天气现象,在一年中都是以春季最多,都是4月份最多,8、9月份最少。扬沙天气主要集中在5月。浮尘日数4月份占全年总数39%。尽管北京发生沙尘暴次数较少,但年变化中仍表现出春季最多,主要发生在4月份(周冰冰,2001;吕新苗等,2004)。

三)、北京沙尘天气的空间分布

沙尘源是风沙产生的必要条件,北京的沙尘源问题也是近几年来众多专家学者争论的焦点之一。根据各学术观点,可以肯定影响北京风沙的沙源可分为2部分:京外沙尘源,京内沙尘源,即就地起沙(贾丹等,2006;)。

关于京外沙尘源,中国环境科学研究院任阵海院士指出沙尘移动的路径主要有三条:1)北路:源区(蒙古国东南部)内蒙古乌兰察布盟-锡林郭勒盟西部的二连浩特、阿巴嘎旗-内蒙浑善达克沙漠西部-朱日和-四子王旗-张家口至北京;2)西北路:源区(蒙古国中、南部)-内蒙古阿拉善盟的中蒙边境-乌拉特

中、后旗-河西走廊-从贺兰山南、北两侧分别经毛乌素沙地和乌兰不和沙漠-呼和浩特-张家口至北京；3) 西路：源区（新疆塔里木盆地塔克拉玛干沙漠边缘）-敦煌-酒泉-张掖-民勤-盐池-鄂托克旗-大同至北京。（苏海波，2001；任阵海等，2003；陈广庭，2002；吕新苗等，2004）

关于京内及其周边地区沙尘源。1985-1986 年，中国科学院兰州沙漠研究所的科研人员，在北京设点连续观测研究，认为北京扬沙物质的来源是就地起沙。据北京市环境保护监测中心测定，2000 年春季影响北京的风沙，本地扬尘高空污染占 20%，低空污染占到 80%。（周冰冰，2001，朱祖希，2000）尽管北京地区沙尘源地相对较弱，起沙量小，污染往往低于外来沙尘输送，但由于爆发突然，更难预报，往往造成措手不及，影响也较严重。（王英，等，2005；贾丹等，2006；杨维西，2002）

四)、北京沙尘天气的治理

要消灭沙尘，最直接的办法就是绿化造林。就京内沙尘源治理，2000 年北京启动了首都绿化隔离带和四环路绿化带及二、三环路绿化改造工程。周边省份，如河北，也加大对环北京地区防沙治沙工程任务投资力度，建设好首都的生态屏障（朱祖希，2000）。

但治理北京沙尘之前就需要理清本地及其周边地区沙尘源地状况。本章就尝试使用遥感植被指数时间序列，分析、提取北京本地及周边可能沙尘源（主要包括北京本地及周边附近常年裸地和季节性裸地沙尘源），希望对沙尘天气的治理有所帮助。

4.2 数据源介绍

本章仍然采用 MODIS 植被指数（EVI）时间序列数据展开研究，数据具体介绍请参见前章。本章通过对目标区域全年 23 个季相（时间分辨率为 16 天）的 EVI 值，来判别区域植被覆盖状况、区分高 EVI 区域和低 EVI 区域。根据 EVI 特性，当区域 EVI 值低于一定程度时，即可判定该区为裸地区域，使用时间序列也方便提取季节性裸地区域。由于使用数据空间分辨率为 250 米过低，故无法监测过小地块，只关注与北京市郊及周边山地、部分耕地等。

4.3 北京及周边地区两类裸地提取

本节主要研究提取北京本地及周边附近的裸地。将裸地区分为两类：常年裸地和季节性裸地。根据前文综述有必要也将研究沙尘源区分为两类，即常年可能沙尘源和季节性可能沙尘源。

常年裸地即指地表常年裸露的区域，从而形成潜在沙尘源的称为常年可能沙尘源。

季节裸地指由于耕作，造成的冬春季土地裸露的区域，从而形成潜在沙尘源地的称为季节性可能沙尘源。

由于沙尘天气形成复杂，不仅需要沙尘源，还需要别的外力共同作用，所以下文提取出来的两类裸地只能以可能沙尘源标示，确认尚需其他资料。

4.3.1 土地分类规则确定

由于本章只关心裸地与非裸地，因此对土地覆盖类型分类要求也低，只需分析裸地或者非裸地 EVI 值可能的临界点就可以了。本章使用 `parallelepiped` 监督分类，选取四季 EVI 低值、冬春 EVI 低值样本，展开分类。

经过处理提取出的两类地块虽并不能立即判断它是否是沙尘源，但这对沙尘源地治理仍是有帮助的，通过这样的处理缩小了治理范围，可以集中力量有针对性的治理，缓解沙尘危害。

4.3.2 基于植被指数时间序列两类裸地提取

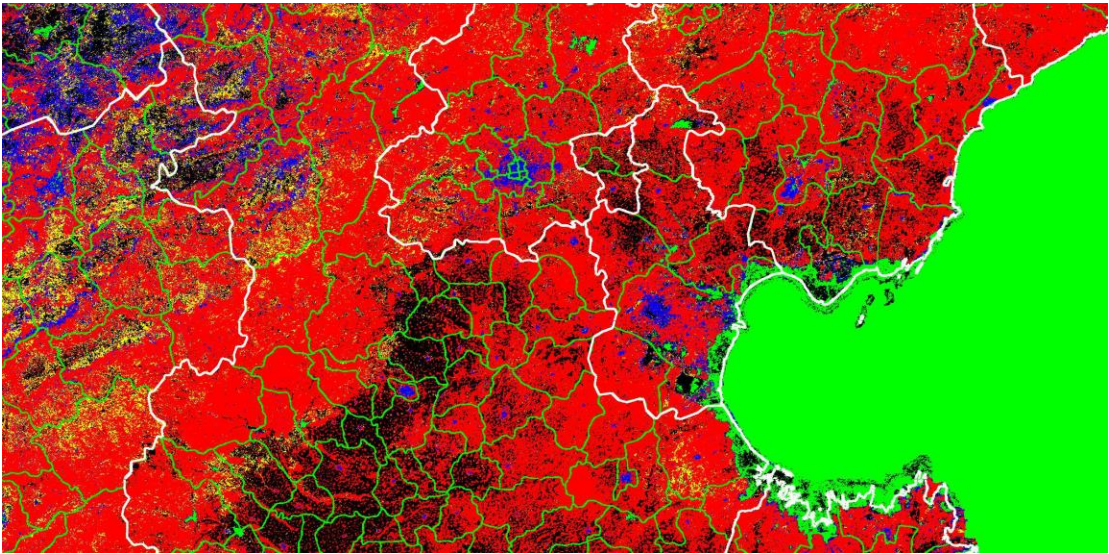


图 4-1 土地覆盖分类图（白线为省界、绿线为县界）

表 4-1 分类表

地类	表示颜色	表示颜色
四季 EVI 值 偏低区	BLUE	
冬春 EVI 值 偏低区	YELLOW	
其他地类 1	RED	
其他地类 2	BLACK	
其他地类 3	GREEN	

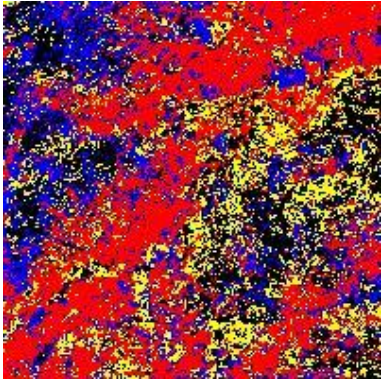


图 4-2 细部

蓝、黄区域正是本文需要提取的两类裸地。经过统计，四季 EVI 值偏低区域（蓝色，可能为常年性裸露地区）约为 11500 平方公里（包括城镇地类），主要分布在北京西北部地区，即河北、山西、内蒙交界处，尤其以内蒙境内居多，北京市郊各区县内也有零星分布。但由于文中未对城镇用地和裸地加以区分，则图中表现出北京市郊各县的该地类也有可能是当地居民聚集区。冬春季 EVI 值偏低区域（黄色，可能为季节性裸露区域）约 16820 平方公里，主要分布在北京西北和西部等区域，集中在河北西北部、山西东部地区，北京的大兴、房

山、门头沟、昌平、延庆、密云、顺义、通州等郊县也有零星分布。

4.4 可能沙尘源分析

4.4.1 可能沙尘源地分析

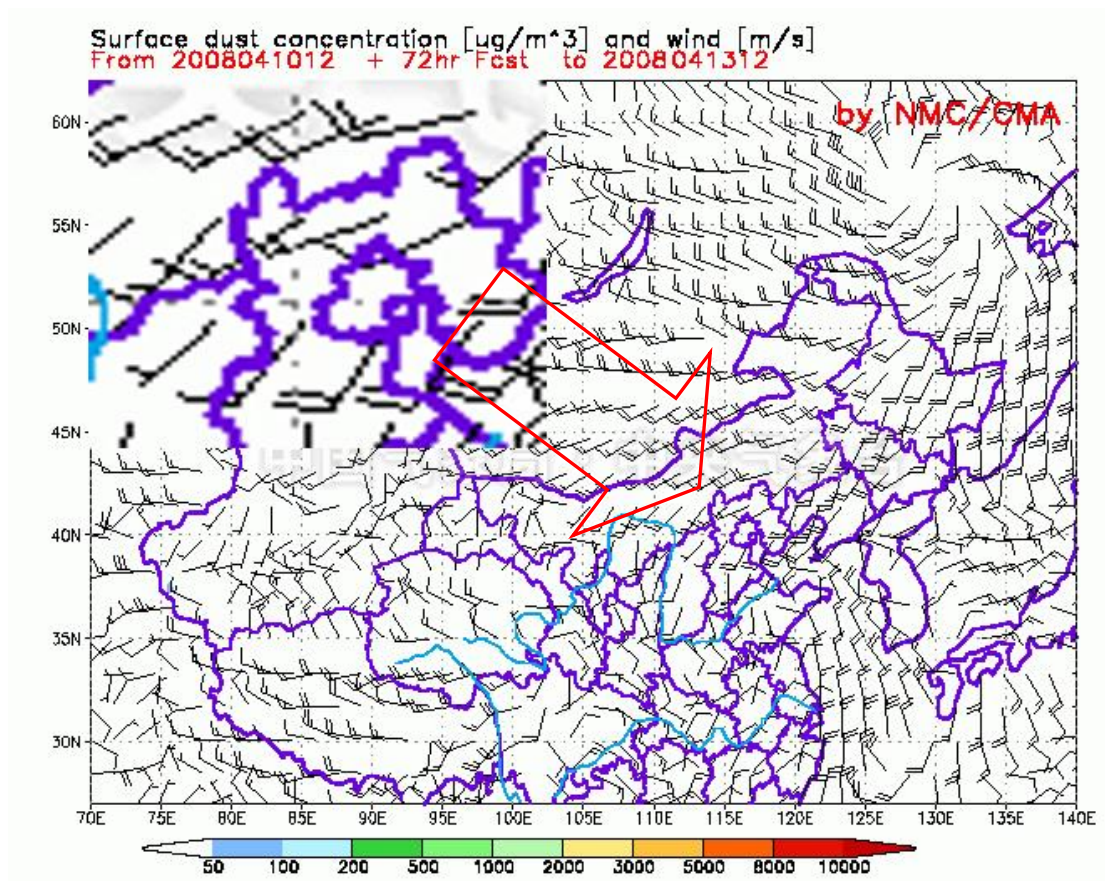


图 4-3 72 小时预测沙尘浓度和风场合成图

注：此图从国家气象中心网站获得，链接地址：

http://www.nmc.gov.cn/product_mode.php?v1=MS&v2=SCMS_SCNDHFCHCT_3H&v3=1。

形成沙尘源地除存在提供沙尘的裸地（或其他）外，另一个主要条件是大风和风向。北京气候为典型的暖温带半湿润大陆性季风气候。冬春季强大的蒙古高压散发出来的西北季风容易构成对本地的影响。北京位于华北平原的西北边缘，太行山余脉和燕山山脉分立其北部、西部，其南面是华北大平原。北部、西部由于山脉的横亘，大风受到阻挡因此风向发生变化，以西北风向从北京西北部（内蒙古、河北、山西交界地区），穿过大山，进入北京市郊区到达

北京城区，这是外地沙尘进入北京的必经之路。而南部为平坦的平原地区，无阻无当，南风一路横扫北京地区。这些点可以从风场合成图上得到印证。

对比分析可以发现，北京西北部土地覆盖分类图中蓝、黄地区正是处在沙尘进入北京的必经之路上，该区干旱，植被稀疏，又有沙漠存在。值得注意的是，该区多山脉，内部又有河流途径，滩涂沙化。大风进入该区后携当地沙尘向北京挺进，这也与实际情况相符。因此可以推断，分类图中北京西北地区的蓝、黄区域成为北京沙尘区域可能性很大。北京南部地区大兴、房山、门头沟等存在冬春季节 EVI 值偏低区（黄色区域）。据前分析，由于地势影响该区冬春盛行南风，大风无阻挡很容易从该地区携带沙尘，进入北京城区。周吉红（2007）指出该区（尤其大兴）后茬种植花生和红薯等作物，在作物收获后耕地完全裸露，直到来年五月才再次栽种。因此冬春季见风起沙，形成沙尘源地，因此该区也极可能成为北京本地沙尘源。

4.4.2 对策分析

北京西北部为多省交界处，该区气候多干旱、山脉纵横，且有沙漠存在，又是西北地区进京的主要通道，治理意义重大同时也难度很大，必须多方协调，风沙治理与生态恢复并行。除大量资金投入防护林、草场建设上外，还需减小畜牧量甚至是居民迁移、退耕还林还草，以减少人为对山地植被破坏，形成沙尘进入北京的西北绿色屏障。

由于不合理耕作模式形成的北京北部地区冬春沙尘源地，该类沙尘源地对北京沙尘天气有极大的贡献。但此区治理相对难度小，主要是如何处理经济效益和生态效益的问题，更容易见效。因此设计、推广合理的种植模式是主要的解决手段。

治理上述两区，虽然不能彻底解决北京地区受到沙尘天气危害，但对于减少内部沙尘、缓解北京地区中低空大气污染是十分有意义的。

4.5 本章小节

本章将遥感植被指数时间序列应用于北京及周边两类裸地提取中，试图通

过这样的方法，发现可能沙尘源地。通过结合本地区气候资料分析、发现：

北京西北部，即内蒙古、山西、河北交界处存在这大量常年性裸地，季节性裸地也有零星分布。该区处在多条沙尘进入北京必经之路，区内干旱且有沙漠存在，又在北京西北风向风口，成为沙尘源的可能性极大。

北京南部，即门头沟、房山、大兴等区县存在大量的季节性裸地。由于特定的种植模式，导致该区冬春耕地裸露，同时该区处在北京南风向风口，故大风携带沙土进入北京可能性也极大。该发现得到周吉红（2007）文献支持。

第五章 结论与展望

5.1 结论

本文对植被指数时间序列滤波重建算法在傅立叶谐波分析系列方法的基础上做出改进,提出四条改进原则,并遵从该原则,重点对谐波分析方法中重要参数的获取提出改进,试图减少操作者主观因素对数据序列的影响。通过与 HANTS 软件结果对比分析,验证了改进方法的优势,也暴露了改进方法的不足:

(一).自动选频。该策略起到了很好的作用,从 NOF 统计分布表上也充分体现了不同区域使用不同频数的必要性。对于大区域,如整个中国地区、整个亚洲地区,经纬度跨度大,种植模式多,地面覆盖类型复杂,自动选频就更能体现出其优势。

(二).算法终止策略。本文在使用传统超定度(DOD)参与极端情况下终止拟合迭代的同时,从 Savitzky-Golay 滤波重建方法中引入了拟合影响因子,希望通过计算不同迭代时的拟合影响因子,自动终止迭代,减少人为设置阈值的主观干扰。从对比实验分析可以得到结论,即使用该终止策略,有效的终止了拟合迭代,大部分点(98.49%)都可以在5次迭代内结束拟合。这也证明并未因为计算新的参量导致算法时间的大量消耗。

(三).异常值检测策略。引入异常值检测主要出于两个目的:一希望通过更为客观的方法判断输入值是否应该参与下步拟合;二不希望通过简单的距离定权,使用人为设置的阈值剔除某些疑似异常点,试图通过这样的策略来减少拟合迭代对原始数据的扰动。该策略也起到了一定的左右,从样本较差均方差分析发现,改进方法与 HANTS 结果相比,结果更接近真实情况,较少出现 HANTS 方法出现的拟合结果负偏性。

最后本文使用 MODIS EVI 植被指数 23 时相时间序列分析提取了北京地区及周边可能沙尘源地,希望能对北京地区治理沙尘有所帮助。在分析过程中,本文第四章将北京周边沙尘源地明确区分为两类,即常年地表裸露形成的沙尘源地和因为不合理的种植模式导致冬春耕地裸露而形成的沙尘源地。在分析基础上,利用 EVI 值的特性和前人研究成果,设计分类策略提取两类裸地。接合北京本地物候数据,分析北京本地及周边两块主要的可能沙尘源,即北京西北部,内蒙古、河北、山西交界处的可能外来沙尘源地,该沙尘源大部分属于第

一类沙尘源，也是外来沙尘进入北京的必经之路。另一处为北京南部门头沟、房山、大兴等区域在冬春形成的可能沙尘源地，该类主要因为不合理的种植模式导致冬春耕地裸露，形成北京南部最主要的可能沙尘源地。该区域问题已经有专家学者开始关注，试图寻找新的途径，解决农民经济效益和生态效益间的矛盾。

5.2 展望

本文主要研究目的是在时间序列滤波重建过程中尽可能的减少人为主观干扰对数据产生的过度扰动、避免引入操作者的主观情绪，而这恰恰是傅立叶谐波分析系列方法甚至大部分滤波重建方法的通病。本着上述思想改进滤波重建算法。

在与 HANTS 软件结果对比分析时，大部分改进都起到了良好的作用，但这些改进策略对原始数据会产生何种影响尚需分析。另外，出于拟合效果的考虑，在谐波波数（NOF）选择上并未完全按照谐波代表的实际物理意义确定，而是使用了妥协策略。算法中其他几处出于同样的目的或算法实现需要，也使用了妥协策略，这些妥协对原始数据的扰动到底有多大仍需要进一步研究。

此外，本文得承认尽管费了很大脑筋，采取了很多策略保证数据的合理处理，但与 HANTS 软件相比，还是有很大差距，如算法普适性方面，HANTS 软件更具有优势。另外，在一些处理技巧上，改进算法与 HANTS 软件差距也很大，显得过于僵硬，对于特殊情况的考虑欠周全。这些在以后的研究中，将逐步得到完善。

本文在第四章提取了北京周边地区两类可能沙尘源，因为缺乏实地调研，尚不能确认是否为真实沙尘源。但区分不同类型沙尘源，对症下药、有效治理，对于北京地区治理沙尘危害是十分有意义的。

参考文献

1. 陈广庭. 北京强沙尘暴史和周围生态环境变化[J]. 中国沙漠, 2002, 22(3): 210—213.
2. 范锦龙, 吴炳方. 复种指数遥感监测方法[J]. 遥感学报, 2004.11, 8(6): 628-636.
3. 方翔, 郑新江, 陆均天, 孙家民, 屠志方, 李梦先, 王国胜, 周卫东. 2002 年春季北京沙尘天气成因及源地分析[J]. 国土资源遥感, 2002.12, 4: 17-22.
4. 辜智慧. 中国农业复种指数的遥感估算方法研究[D]. 2003, 北京师范大学硕士论文.
5. 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.
6. 胡晓, 马耀明, 田辉, 王介民. 4~10 月藏北地区地表植被参数的卫星遥感研究[J]. 高原气象, 2006, 25(06): 1020-1027.
7. 贾丹, 岳得鹏. 北京地区风沙的现状、成因及防治对策[J]. 延边大学农学学报, 2006.6, 28 (2): 140-144.
8. 彭代亮, 黄敬峰, 王秀珍. 基于 MODIS-EVI 区域植被季节变化与气象因子的关系[J]. 应用生态学报, 2007, 18(05): 983-989.
9. 林忠辉, 莫兴国. NDVI 时间序列谐波分析与地表物候信息获取[J]. 农业工程学报, 2006.12, 22(12): 138-144.
10. 吕新苗, 刘惠清, 王文杰, 吴绍洪. 北京周边地区沙尘暴时空分布特征及其环境背景[J]. 地理研究, 2004.1, 23 (1): 38-44.
11. 任阵海, 高庆先, 苏福庆, 王耀庭, 张志刚, 杨新兴. 北京大气环境的区域特征与沙尘影响[J]. 中国工程科学, 2003.2, 5 (2): 49-56.
12. 苏海波. 北京鏖战风沙[J]. 科技潮, 2001, 10: 30-31.
13. 王丹, 姜小光, 唐伶俐, 习晓环. 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, 2005.6, 14: 29-32.
14. 王英, 李令军. 北京沙尘天气污染分析[J]. 中央民族大学学报 (自然科学版). 2005.8, 14 (3): 269-276.
15. 王正兴, 刘闯, Huete A. 等. 植被指数研究进展[J]. 生态学报, 2003, 23 (5): 979-987.
16. 徐永明, 刘勇洪, 魏鸣, 吕晶晶. 基于 MODIS 数据的长江三角洲地区土地覆盖分类[J]. 地理学报, 2007, 62 (06): 640-648.
17. 杨维西. 北京沙尘天气的沙尘来源及其治理[J]. 林业经济, 2002, 7: 19-22.

18. 赵英时等. 遥感应应用分析原理与方法[M]. 北京: 科学出版社, 2003.6, 166~208.
19. 张峰, 吴炳放, 刘成林等. 利用时序植被指数监测作物物候的方法研究[J]. 农业工程学报, 2004, 20 (1): 155-159.
20. 张峰, 吴炳放, 刘成林等. 区域在作物生长过程的遥感提取方法[J]. 遥感学报, 2004, 8 (6): 515-528.
21. 张霞. 光谱指数时间谱特性研究及其在种植模式信息提取中的应用[D]. 中国科学院遥感应用研究所博士论文, 2006: 83-106.
22. 张霞, 张兵, 卫征, 陈正超, 郑兰芬. MODIS 光谱指数监测小麦长势变化研究[J]. 中国图象图形学报, 2005, 10(04): 420-424.
23. 周冰冰. 北京的风沙灾害及其控制技术[J]. 林业科技管理, 2001 (4): 55-58.
24. 周吉红. 京郊花生、红薯裸露地块冬春季扬尘综合控制技术探讨[J]. 北京农业, 2007, 6: 49-51.
25. 朱祖希. 构筑北京绿色屏障防治风沙肆虐危害[J]. 城市环境, 2000, 4: 36-38, 5: 46-48.
26. 邹金秋, 陈佑启, Satoshi Uchida, 吴文斌, 许文波. 利用 Terra/MODIS 数据提取冬小麦面积及精度分析[J]. 农业工程学报, 2007, 23(11): 195-200, 封 3.
27. A. Moody and D. Johnson. Land-surface phenologies from AVHRR using the discrete Fourier transform[J]. Remote Sensing of Environment, 2001:305-323.
28. Bethany A. Bradley, Robert W. Jacob, John F. Hermance, John F. Mustard. A Curve Fitting Procedure to Derive Inter-annual Phonologies from Time-series of Noisy Satellite NDVI Data[J]. Remote Sensing of Environment, 2007, 106: 137-145.
29. Chen J., Per Jönsson, Masayuki Tamura, Gu Z., Bunkei Matsushita, Lars Eklundh. A Simple Method for Reconstructing a High-quality NDVI Time-series Data Set Based on the Savitzky-Golay Filter[J]. Remote Sensing of Environment, 2004:332-344.
30. Chen X, Tan Z, Schwartz M D, et al. Determining the growing season of land vegetation on the basis of plant phenology and satellite data in Northern China[J]. International Journal of Biometeorology, 2000, 44(2):97-101.
31. Cihar, J., Ly, H., Li, Z. Q., Chen, J., Porkrant, H., & Huang, F. T.. Multi-temporal, multi-channel AVHRR data sets for land biosphere studies-artifacts and corrections. Remote Sensing of Enviroment, 1997, 60:35-57.
32. Goward S N, Turner D G, Dye D G, et al. The University of Maryland improved global

- vegetation index product[J]. *International Journal of Remote Sensing*, 1994, 15:3365-3395.
33. Gutman G. G. Vegetation indices from AVHRR: An update and future prospects. *Remote Sensing of Environment*, 1991, 35:121-136.
 34. Gutman G. G., A. Ignatov. Global Land Monitoring from AVHRR: Potential and Limitations. *International Journal of Remote Sensing*. 1995, 16: 2301-2309.
 35. Holben B N. Characteristics of Maximum-value Composite Images for Temporal AVHRR Data[J]. *International Journal of Remote Sensing*. 1986,7: 1435-1445.
 36. Huete, A., Didan, K., Miura, T., Todriguez, E. P., Gao, X., Ferreira, L. G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.*, 83: 195~213.
 37. Hawkins D. Identification of outliers. Chapman and Hall, London, 1980.
 38. IGBP(1992). J.R.G Townshend(Ed.), Improved global data for land applications. IGBP Global Change Report, vol. 20. Stockholm, Sweden: International Geosphere–Biosphere Programme.
 39. Jing F., X Louise L. Mapping Crop Key Phonological Stages in the North China Plain Using NOAA Time Series Images[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2002.4:109-117.
 40. Justice, C.O., Townshend, J.R.G, Holben, B.N., & Tucker, C.J. (1985). Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, 6(8), 1271–1318.
 41. L. Andres, W. Salas, and D. Skole. Fourier analysis of multi-temporal AVHRR data applied to a land cover classification[J]. *International Journal of Remote Sensing*, 1994, 15:1115-1121.
 42. L. Olsson, L. Eklundh. Fourier Series for Analysis of Temporal Sequences of Satellite of Sensor Imagery[J]. *International Journal of Remote Sensing*, 1994, 15(18):3735-3741.
 43. Lovel J L, Graetz R D. Filtering Pathfinder AVHRR Land NDVI Data for Australia[J]. *International Journal of Remote Sensing*, 2001, 22(13):2649-2654.
 44. M. Menenti, S. Azzali, W. Verhoef, and R. van Swol. Mapping agroecological zones and time lag in vegetation growth by means of Fourier analysis of time series of NDVI images[J]. *Adv. Space Res.*, 1993, 13:233-237.
 45. Mark E. Jakubauskas, David R. Legates, Jude H. Kastens. Crop Identification Using

- Harmonic Analysis of Time-series AVHRR NDVI Data[J]. Computers and Electronics, 2002, 37:127-139.
46. Mark E. Jakubauskas, David R. Legates, Jude H. Kastens. Harmonic Analysis of Time-series AVHRR NDVI data[J]. Photogrammetric Engineering & Remote Sensing, 2001, 67(4):461-470.
47. Moulin S, L. K, Viovy N, et al. Global-scale assessment of vegetation phenology using NOAA/AVHRR Satellite Measurements[J]. Journal of Climate, 1997, 10:1154-1170.
48. Myneni, R.B., Keeling, C.D., Tucker, C.J., Asrar, G., & Nemani, R.R. (1997). Increased plant growth in the northern high latitudes from 1981 to 1991. Nature, 386(6626), 695–702.
49. Nagatani, I., Saito, G., Toritani, H. and Sawada, H. Agricultural Map of Asian Region Using Time series AVHRR NDVI Data[C]. In Proceedings Online of the 23rd Asian Conference on Remote Sensing, 25–29 November 2002, Birendra International Convention Centre in Kathmandu, Nepal. Available online at: <http://www.gisdevelopment.net/aars/acrs/2002/pos2/184.pdf>
50. Park J, Tateishi R. Correction of Time Series NDVI by the Method of Temporal Window Operation (TWO)[C]. Proceedings of the 1998 Asian Conference on Remote Sensing (OL), 1998. <http://www.gisdevelopment.net/aars/acrs/1998/ps2/ps2004.shtml>.
51. Per Jönsson, Lars Eklundh. Seasonality Extraction by Function Fitting to Time-Series of Satellite Sensor Data[J]. IEEE Transactions on Geosciences and Remote sensing, 2002, 40(8):1824-1832.
52. Reed, B.C., Brown, J.F., VanderZee, D., Loveland, T.R., Merchant, J.W., & Ohlen, D.O. (1994). Measuring phenological variability from satellite imagery. Journal of Vegetation Science, 5(5), 703–714.
53. Roerink, G. J., Menen, M., Verhoef, W. Reconstructing Cloudfree NDVI Composites Using Fourier Analysis of Time-series. International Journal of Remote Sensing, 2000, 21(9):1911-1917.
54. Sajia Akhter, Ipshita Sarkar, Kazi Golam Rabbany, Nahid Akter, Shamim Akhter, Yann Chemin, Honda Kiyshi. Adapting the LMF Temporal Splining Procedure From Serial to MPI/Linux Clusters[J]. Journal of Computer Science, 2007, 3(3):130-133.
55. Savitzky, A., Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least

- Squares Procedures[J]. Analytical Chemistry, 1964,36:1627-1639.
56. Sellers P J, Los S O, Tucker C J, et al. A Reversed Land Surface Parameterization (SiB2) for Atmospheric GCMs, Part 2: The Generation of Global Fields of Terrestrial Biophysical Parameters from Satellite data[J]. Journal of Climate, 1996,9:706-737.
57. Stockli R., Vidale P L. European plant phenology and climate as seen in a 20-year AVHRR land-surface parameter dataset[J]. International Journal of Remote Sensing, 2002, 25(17):3303-3330.
58. Swets, D.L., Reed, B.C., Rowland, J.R., & Marko, S.E. (1999). A weighted least-squares approach to temporal smoothing of NDVI. 1999 ASPRS Annual Conference, From Image to Information, Portland, Oregon, May 17–21, 1999, Proceedings: Bethesda, Maryland, American Society for Photogrammetry and Remote Sensing.
59. Viovy N, Arino O, Belward A S. The Best Index Slope Extraction (BISE): A Method for Reducing Noise in NDVI Time Series(J). International Journal of Remote Sensing, 1992,13:1585-1590.
60. W. W. Immerzeel, R. A. Quiroz S. M. De Jong. Understanding Precipitation Patterns and Land Use Interaction in Tibet Using Harmonic Analysis of SPOT VGT-S10 NDVI Time series[J]. International Journal of Remote Sensing, 2005, 26(11):2281-2296.
61. White M. Z., Running S. W., Thornton P. E. The Impact of Growing-season Length Variability on Carbon Assimilation and Evapotranspiration Over 88 Years in the Eastern US Deciduous Forest [J]. International Journal of Biometeorology, 1999,42(3): 139-145.
62. Yves Julien, José A. Sobrino, Wout Verhoef. Changes in Land Surface Temperatures and NDVI Values over Europe Between 1982 and 1999[J]. Remote Sensing of Environment, 2006,103:43-55.
63. Zhou L., Tucker C J., Kaufmann R K, et al. Variations in northern vegetation index during 1981 to 1999[J]. Journal of Geophysical Research. 2001,106(D17):20069-20083.

致 谢

三年的硕士研究生生活转眼就结束了。当年决定报考遥感所，接着参加考试、复试、面试的场景还历历在目，又怎能轻易忘怀呢！那段是到目前为止我人生当中最为潦倒的一段光阴。好在都挺了过来，这就是希望和曙光。

这里我感谢父母对我这些年来的支持和鼓励。现在他们仍然坚强有力的支持着我对人生的探求，将来他们也还会继续默默的支持我的选择与人生，他们是我最强后盾！除父母外，还要感谢我其他的亲人们，他们共同为我创造了平稳的环境让我能够安心。每次看到他们自豪的、充满期望的眼神，动力都由心底而升。

这里我要郑重的感谢我的导师：郑兰芬、张兵、童庆禧三位研究员。三年前在我最无助、最潦倒的时候郑老师、张老师选择录取了我，在我心里这等同于再造；在遥感所的日子，童老师睿智大度，对我们高屋建瓴的指导使我们受益匪浅！这三年里，尽管我也有过不听老师话、有过埋怨的时候，但从来没有忘记他们的帮助，对他们的敬意是一辈子的事！感谢他们在这三年对我的悉心照顾和指导！

这里我要郑重的感谢王晋年研究员。虽然和王老师相处时日不长，但王老师在工作、生活、以至未来选择等各个方面的关怀和指导都对我有着深远的影响，谢谢您一直以来对我的信任、照顾和鼓舞，学生不会忘记！

这里我非常感谢张霞师姐、陈正超师兄、卫征师兄。在我正式进所第一天见到的第一位师兄就是张霞大师姐。大师姐这些年来对我学业的完成多有帮助，从选题、研究内容确定到论文撰写全面指导，感谢大师姐！我刚进所，被安排跟陈师兄一起做事，陈师兄帮着我安排住宿、安排学习。尽管后来一起工作的时间少了，但是我们间的交流没有断过，学业上的，生活上的，人生奋斗上的，我受益匪浅。感谢陈师兄一直以来的照顾和信任！进所之后，没多久我被安排和卫师兄一起工作，他也是这三年半来一起合作最多的人：从沙河机场实验，到山东飞行实验，再到3号临建、二十一世纪公司；从学业安排、生活安排到人生之路。卫师兄对我有很深的影响，从他身上，我学到了踏实，也将会学到勤奋。也感谢方俊永师兄在各次实验时的帮助和张立福老师的指导！

这里我感谢施永胜、孙鸿睿、邓飏、魏小兰、黄群英。他们是我大学的师兄、师姐和同学，在我复习考研以及后来复试录取过程中，他们鼎力支持，多

方帮助，非常感谢他们！

这里我感谢实验室的其他兄弟姐妹们，他们是高连如、和海霞、胡方超、胡兴堂、焦全军、李俊生、李庆亭、李兴、刘波、刘翔、刘学、罗文斐、申茜、孙旭、温建亭、吴頔、张浩、张靓、张文娟、周丽萍等。大家在一起相互帮助，一起玩得开心。感谢你们的笑声（笑声是可以传染的，呵呵）。记否，上凤凰岭，下北戴河？要特别提出北戴河之旅时，张靓兄的全力支持和帮助，以及大厨总在管刘学、张文娟，娱乐总在线刘翔，差旅总在帮李俊生，谢谢大家共同努力，换来了全程的笑声！

也感谢研究生处的吴晓清、刘戈平两位老师，在我平时学习以及学生会工作中给予的支持！