

东南亚地区 2010 年 MODIS-NDVI 再分析数据集

王正兴^{1*}, 曹云锋²

1. 中国科学院地理科学与资源研究所资源环境信息系统国家重点实验室, 北京 100101;

2. 北京林业大学林学院, 北京 100083

摘要: MODIS-NDVI 是广泛使用的植被指数之一, 但是基于全球算法开发的植被指数产品依然有许多噪音。在多云多雨的东南亚地区, 这些噪音严重影响了数据质量和数据的可用性。根据 2010 年 Terra-MODIS 和 Aqua-MODIS 两个 16 d 植被指数产品的数据噪音分布情况, 研发了二颗卫星 MODIS-NDVI 再分析(加密、去噪音、数据重建)方法, 即首先根据象元真实时间(Day of Year, DOY)和数据质量(QA)把两个 16 d 数据序列合并为一个 8 d 数据序列, 然后利用象元质量信息(QA)和 Timesat 软件提供的时间序列重建模型(S-G), 对低质量的象元进行了重建。最后获得东南亚地区 2010 年时间分辨率 8 d、月度空间分辨率为 250 m 的 MODIS-NDVI 时间序列重建数据集。数据集由 46 个每 8 天重建数据文件包和 12 个每月重建数据文件包组成。数据采用.tif 格式存储, 数据量为 47.68 GB(压缩为 9.77 GB)。

关键词: MODIS-NDVI; 时间序列; 数据质量; 重建; 东南亚

DOI: 10.3974/geodp.2017.03.10

1 前言

东南亚地区是世界三个重要的热带雨林分布区之一, 也是人口压力大, 土地覆盖变化快的地区。但是, 由于多云多雨的天气条件, 由光学遥感观测开发的植被指数往往包含很多噪音, 使得植被指数的时间、空间覆盖不完整, 影响了植被指数在本地区资源环境研究中的应用。本数据集在 2010 年 Terra-MODIS 和 Aqua-MODIS 两个 16 d 植被指数产品的基础上, 首先利用象元真实时间(Day of Year, DOY)把两个数据集合并为一个 8 d 数据序列, 然后利用象元质量评估信息(QA)和 TimeSAT 软件提供的时间序列重建模型(S-G), 对低质量的象元进行了重建。最后, 利用高质量象元检查重建模型, 获得一个时间分辨率 8 d、空间分辨率为 250 m、覆盖东南亚地区的 MODIS-NDVI 时间序列。

2 数据集的元数据简介

东南亚地区 2010 年 MODIS-NDVI 再分析(加密、去噪、重建)数据集(SEA_Rea_MODIS_NDVI_2010)^[1]的名称、短名、通讯作者、作者、地理区域、数据年代、空间分辨率、时间分辨率、数据集组成、数据开发资助基金、数据出版与共享服务平

收稿日期: 2016-02-12; 修订日期: 2016-08-15; 出版日期: 2017-09-25

基金项目: 中华人民共和国科学技术部(2016YFA0600201, 2015DFA11360); 中国科学院(KZZD-EW-08-01-02-01, TSYJS04)

*通讯作者: 王正兴 L-5255-2016, 中国科学院地理科学与资源研究所, E-mail: wangzx@igsrr.ac.cn

论文引用格式: 王正兴, 曹云锋. 东南亚地区 2010 年 MODIS-NDVI 再分析数据集[J]. 全球变化数据学报, 2017, 1(3): 317-323. DOI: 10.3974/geodp.2017.03.10.

数据集引用格式: 王正兴, 曹云锋. 东南亚地区 2010 年 MODIS-NDVI 再分析(加密、去噪、重建)数据集[DB/OL]. 全球变化科学研究数据出版系统, 2016. DOI: 10.3974/geodb.2016.03.16.V1.

台和数据政策等信息见表 1。该数据时间分辨率 8 d，2010 年全年共 46 个合成期。地图投影为 WGS84/Geographic。数据类型采用 MODIS 陆地产品标准的 16-bit Signed，本数据集数值换算为 NDVI 的尺度系数为 0.000,1。

3 数据采集与研发方法

本数据集是在前期相关研究的基础进一步改进算法而开发的。输入的两个基础数据来自 MODIS 第五代植被指数产品^[3]，有关 Timesat 软件及其 3 种滤波算法的比较研究见文献[4-5]；有关时间序列从 16-day 加密到 8-day 的概念和方法，见文献[6]；有关象元尺度的质量信息提取，对植被指数产品，目前可以直接使用 Reliability 波段，如果需要提取更详细的质量评估信息，可以参考文献[7]。本文仅对本次在东南亚地区 MODIS-NDVI 时间序列重建中的改进部分作简要介绍。

表 1 东南亚地区 2010 年 MODIS-NDVI 再分析数据集元数据简表

条目	描述
数据库（集）名称	东南亚 2010 年 MODIS-NDVI 再分析（加密、去噪、重建）数据集
数据库（集）短名	SEA_Rea_MODIS_NDVI_2010
数据作者	王正兴 L-5255-2016, 中国科学院地理科学与资源研究所, wangzx@igsnrr.ac.cn 曹云锋 S-1593-2016, 北京林业大学林学院, yfcao@bjfu.edu.cn
地理区域	10.95°S-28.57°N, 92.18°E-141.03°E 行政范围：东南亚 11 国；中国广东、广西、福建、云南、贵州全部或大部
数据年代	2010 年
数据时间分辨率	8 d
数据空间分辨率	250 m
地图投影	WGS84/Geographic
数据格式	.tif, 压缩为.rar
数据集组成	数据集由 46 个.tif 文件组成
数据量	47.68 GB（压缩为 9.77 GB）
基金项目	中华人民共和国科学技术部（2016YFA0600201, 2015DFA11360）；中国科学院（KZZD-EW-08-01-02-01, TSYJS04）
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101，中国科学院地理科学与资源研究所
数据共享政策	全球变化科学研究数据出版系统的“数据”包括元数据（中英文）、实体数据（中英文）和通过《全球变化数据学报》（中英文）发表的数据论文。其共享政策如下：（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报》（中英文）编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源 ^[2]

本数据集使用了两个输入数据集(MOD13Q1 和 MYD13Q1)中的 3 个波段：归一化植被指数 (NDVI)，象元时间（Day of Year, DOY），象元质量评估信息（QA）。在进行数据预

处理（数据完整性检查，格式转换，影像裁剪，数据类型调整）后，主要数据处理分 3 步：

（1）时间序列加密：以实际 DOY 为序，把 16-day 的 MOD 与 MYD 合并为 8-day 时间序列。该处理使用 IDL 8.1 完成。

（2）时间序列重建：利用 Timesat GUI 进行参数训练，结合训练所得参数进行重建。该处理在 Timesat 2.3 完整。

（3）数据后处理：重建后数据分解，质量检查，奇异值处理。用 ENVI4.8 软件完成。

3.1 数据时间序列加密

根据文献^[6]，两个 16-day 的 MOD 和 MYD，在文件尺度是“错位 8-day”。但是在实际的象元尺度，如果把这“错位 8-day”理解为新的“8-day 合成期”，那么会出现 3 种情况：（1）MOD 和 MYD 都出现。这是最佳情况，但也意味着其他某个 8-day 合成期就可能没有数据了。（2）MOD 或 MYD 其中之一出现。（3）MOD 和 MYD 都未出现（图 1）。

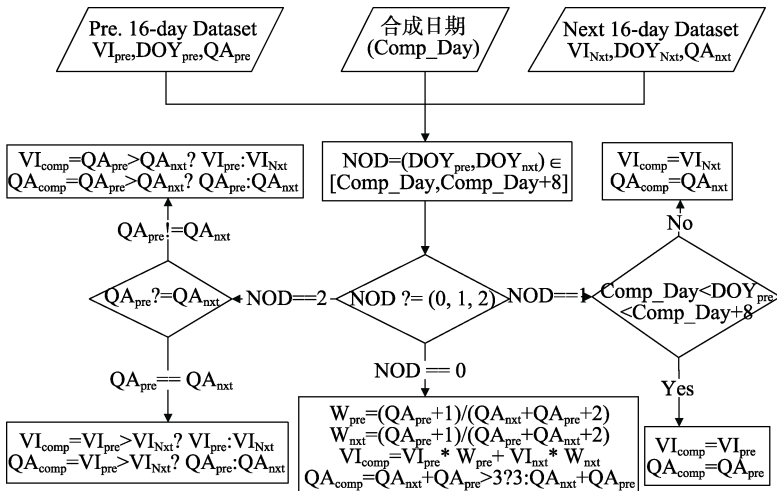


图 1 MODIS-NDVI 时间序列加密流程图

对应以上 3 种情况，时间序列加密的方法分别介绍如下：

（1）MOD 和 MYD 都出现（NOD=2）：如果二者质量不同，选择质量较好的一个（QA 小）；如果二者质量相等，选 NDVI 较大的一个。

IF No. of DOY = 2 THEN

如果 QA 不相等，取 Min_QA 对应数据

如果 QA 相等，取 NDVI 较大者(最大值合成原则)

（2）MOD 或 MYD 其中之一出现（NOD=1）：直接使用对应原值。

IF No. of DOY = 1 THEN 直接使用在合成期内的 MOD 或 MYD NDVI 值

（3）MOD 和 MYD 都未出现（NOD=0）：该合成期没有数据，则需要借用前、后 2 个合成期的数据。同时，由于在该合成期“MOD 和 MYD 都未出现”，意味着其前后 2 个合成期“数据一定出现”，因此，需要先按照“数据质量好坏”，质量相同时再按照“NDVI 大小”的优选顺序进行处理。

3.2 数据时间序列重建

（1）算法选取：考虑东南亚热带雨林植被生长的非高斯特征，选择 S-G。

(2) QA 权重分级: 东南亚地区雨季时数据质量较差, 如将 3 级数据权重设 0, 导致原始数据太少无法完成重建过程, 因此采用宽松权重。QA0 权重为 1; QA1 和 QA2 权重为 0.5; QA3 权重为 0.1。图 2 为 2010 年 1-5 月各级质量分布。在 DOY153 (6 月 2-9 日), 大部分地区已经进入雨季, 质量最差的 3 级象元比例高达 40%。

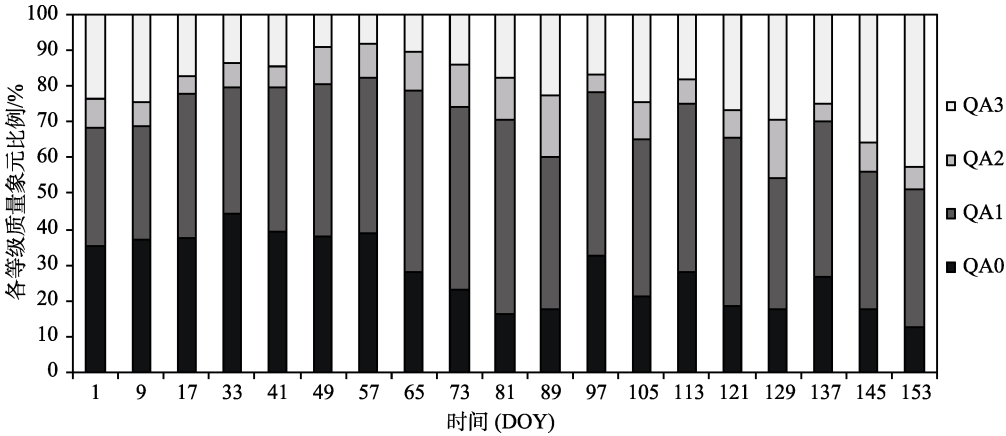
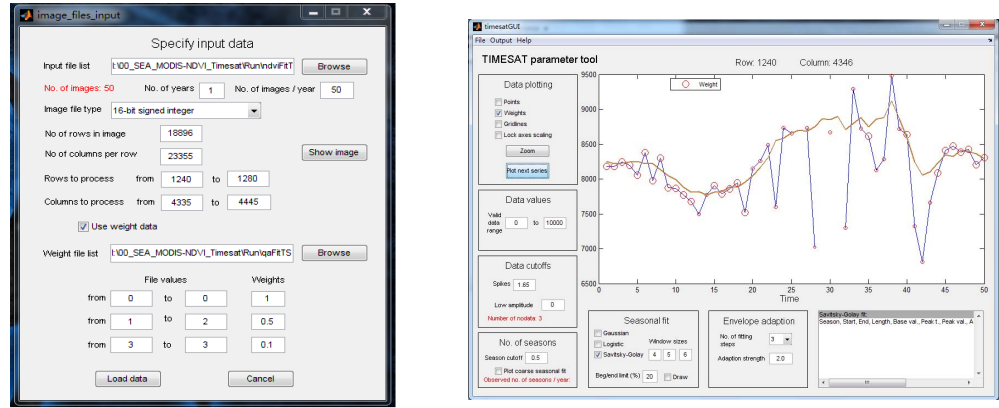


图 2 东南亚地区 2010 年 1-5 月 MODIS-NDVI 各级质量分布

- (3) Spike 参数: 1.65, 实际利用 Z 指数, 当 Z 指数为 1.65 时, 达到 95% 显著性水平。
- (4) Season para: 0.5; 季节性参数。主要考虑区域内部分地区农作物为 2-3 季;
- (5) Adaptation strength: 2; 数据与拟合模型适应性强度值, 采用默认值。
- (6) SG window: 4, 5, 6; 三次迭代, 窗口逐渐增大。



Timesat S-G 算法参数设置

参数设置对应某一点重建前后对比

图 3 Timesat 软件中 S-G 方法进行滤波重建的参数设置与效果

3.3 数据后处理

首先, 检查所有文件是否在正常值域范围。然后, 利用重建前的高质量数据检查重建前后数值变化, 如果后者偏离超过 2,000 (对应实际 NDVI=0.2), 需要恢复原来的高质量数据。最后, 通过目视检查, 对照已知地物, 检查重建后数据的空间连续性以及对地物的解译能力。

3.4 数据共享方式

8-day/250 m: 这是本次再分析直接输出的结果。2010 年全年共 46 个文件。

4 数据重建结果与数据质量验证

4.1 数据集组成

本数据集由 58 个.tif 文件组成，包括 12 个月的数据文件（表 2）和 46 个 8 天合成的 250 m 分辨率的数据文件（表 3）。

表 2 SEA_Rea_MODIS-NDVI_2010 的每月的数据文件

数据文件名	数据量 (KB)
1_SEA_Rea_MODIS-NDVI_2010_JAN.rar	176,096.10
2_SEA_Rea_MODIS-NDVI_2010_FEB.rar	175,212.03
3_SEA_Rea_MODIS-NDVI_2010_MAR.rar	175,327.30
4_SEA_Rea_MODIS-NDVI_2010_APR.rar	175,900.24
5_SEA_Rea_MODIS-NDVI_2010_MAY.rar	176,199.49
6_SEA_Rea_MODIS-NDVI_2010_JUN.rar	178,383.50
7_SEA_Rea_MODIS-NDVI_2010_JUL.rar	179,857.91
8_SEA_Rea_MODIS-NDVI_2010_AUG.rar	178,859.01
9_SEA_Rea_MODIS-NDVI_2010_SEP.rar	177,037.20
10_SEA_Rea_MODIS-NDVI_2010_OCT.rar	174,397.40
11_SEA_Rea_MODIS-NDVI_2010_NOV.rar	175,371.10
12_SEA_Rea_MODIS-NDVI_2010_DEC.rar	175,905.04

表 3 SEA_Rea_MODIS-NDVI_2010 的每 8 天合成的数据文件

文件名	数据量 (KB)	文件名	数据量 (KB)
SEA_Rea_MODIS-NDVI_2010-001.rar	176,967.56	SEA_Rea_MODIS-NDVI_2010-185.rar	179,419.49
SEA_Rea_MODIS-NDVI_2010-009.rar	176,989.77	SEA_Rea_MODIS-NDVI_2010-193.rar	180,214.66
SEA_Rea_MODIS-NDVI_2010-017.rar	176,020.70	SEA_Rea_MODIS-NDVI_2010-201.rar	180,736.62
SEA_Rea_MODIS-NDVI_2010-025.rar	175,513.47	SEA_Rea_MODIS-NDVI_2010-209.rar	180,454.42
SEA_Rea_MODIS-NDVI_2010-033.rar	175,098.94	SEA_Rea_MODIS-NDVI_2010-217.rar	179,693.48
SEA_Rea_MODIS-NDVI_2010-041.rar	175,137.40	SEA_Rea_MODIS-NDVI_2010-225.rar	179,117.06
SEA_Rea_MODIS-NDVI_2010-049.rar	175,010.29	SEA_Rea_MODIS-NDVI_2010-233.rar	178,720.78
SEA_Rea_MODIS-NDVI_2010-057.rar	175,040.92	SEA_Rea_MODIS-NDVI_2010-241.rar	178,210.19
SEA_Rea_MODIS-NDVI_2010-065.rar	175,189.89	SEA_Rea_MODIS-NDVI_2010-249.rar	177,686.87
SEA_Rea_MODIS-NDVI_2010-073.rar	175,419.63	SEA_Rea_MODIS-NDVI_2010-257.rar	177,237.44
SEA_Rea_MODIS-NDVI_2010-081.rar	175,610.37	SEA_Rea_MODIS-NDVI_2010-265.rar	176,841.12
SEA_Rea_MODIS-NDVI_2010-089.rar	175,757.41	SEA_Rea_MODIS-NDVI_2010-273.rar	176,198.74
SEA_Rea_MODIS-NDVI_2010-097.rar	175,830.47	SEA_Rea_MODIS-NDVI_2010-281.rar	175,003.33
SEA_Rea_MODIS-NDVI_2010-105.rar	176,116.94	SEA_Rea_MODIS-NDVI_2010-289.rar	174,313.30
SEA_Rea_MODIS-NDVI_2010-113.rar	176,405.35	SEA_Rea_MODIS-NDVI_2010-297.rar	174,543.72
SEA_Rea_MODIS-NDVI_2010-121.rar	176,668.54	SEA_Rea_MODIS-NDVI_2010-305.rar	175,081.09
SEA_Rea_MODIS-NDVI_2010-129.rar	176,672.63	SEA_Rea_MODIS-NDVI_2010-313.rar	175,607.18
SEA_Rea_MODIS-NDVI_2010-137.rar	176,617.79	SEA_Rea_MODIS-NDVI_2010-321.rar	175,824.29
SEA_Rea_MODIS-NDVI_2010-145.rar	176,758.38	SEA_Rea_MODIS-NDVI_2010-329.rar	176,116.84

续表

文件名	数据量 (KB)	文件名	数据量 (KB)
SEA_Rea_MODIS-NDVI_2010-153.rar	178,245.56	SEA_Rea_MODIS-NDVI_2010-337.rar	176,065.76
SEA_Rea_MODIS-NDVI_2010-161.rar	179,343.16	SEA_Rea_MODIS-NDVI_2010-345.rar	175,925.25
SEA_Rea_MODIS-NDVI_2010-169.rar	179,752.97	SEA_Rea_MODIS-NDVI_2010-353.rar	176,782.02
SEA_Rea_MODIS-NDVI_2010-177.rar	179,495.96	SEA_Rea_MODIS-NDVI_2010-361.rar	176,590.19

4.2 重建象元比例

表 4 为 2010 年 1-6 月各月第一个合成期重建前后象元变化抽样统计。其中“不变象元”为重建前后没有变化的象元，而“改变象元”为重建后数值变化了的象元，可以近似描述重建的贡献。6 个抽样数据显示，重建对 MODIS-NDVI 的贡献率为 27% 以上，主要对应图 2 中质量最差的 QA3 和 QA2 象元重建。

表 4 东南亚地区 2010 年部分合成期 MODIS-NDVI 重建前后数值变化象元比例

合成时段	对应日期	全部象元(个数)	不变象元(个数)	不变象元(%)	改变象元(%)
DOY001	1月:1-8日	441,316,080	321,891,754	72.94	27.06
DOY033	2月:2-9日	441,316,080	321,821,304	72.92	27.08
DOY065	3月:6-13日	441,316,080	321,906,903	72.94	27.06
DOY097	4月:7-14日	441,316,080	321,928,179	72.95	27.05
DOY121	5月:1-8日	441,316,080	321,846,869	72.93	27.07
DOY153	6月:2-9日	441,316,080	321,602,839	72.87	27.13

4.3 重建改进了时间序列连续性。

重建前，云雾造成的噪音影响模式识别，重建后的时间序列可以清晰识别植被物候特征。图 4 显示一个二季植被和一个常绿植被的 NDVI 季节特征。

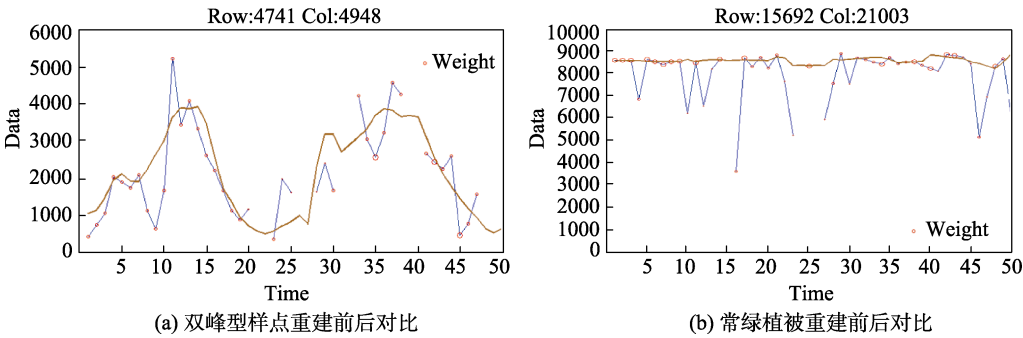


图 4 数据重建前后时间序列对比

4.4 重建改进了空间连续性。

图 5 为 DOY153 合成期重建前后对比。此时，中南半岛进入雨季，持续云雾导致数据质量以 3 级为主。重建前图 5 (a) 与重建后图 5 (b) 后对比发现，东南亚地区数据质量都有不同程度提升，其中最集中地区为缅甸北部、老挝、中国南部沿海地区。中南半岛盆地

虽也受云雾覆盖影响,但重建前后变化较小,可能与该季节农作物覆盖度较低有关。

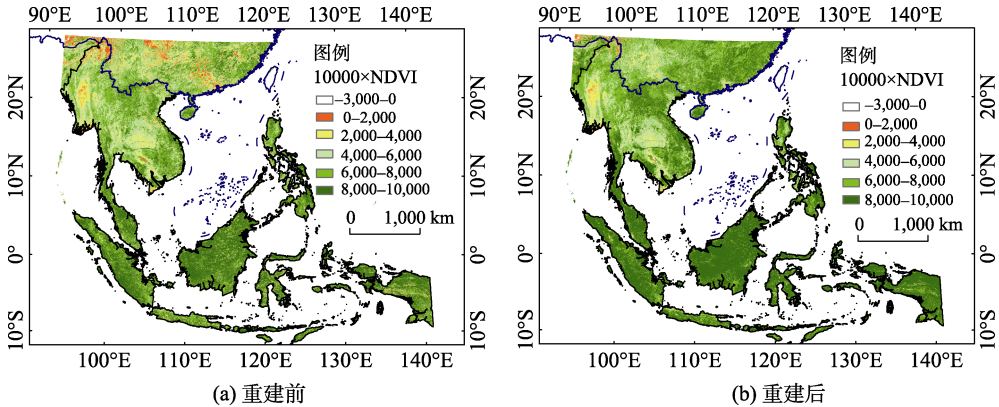


图5 MODIS-NDVI(DOY153)数据质量与重建前后对比

5 结论

东南亚地区多云多雨的气象条件导致光学遥感数据形成的 MODIS-NDVI 时间序列数据存在很多噪音,影响了数据的正常使用。本文以数据质量评估信息和数据实际时间为基础,首先把 Terra 和 Aqua 两个平台的 16-day 标准产品合成为一个 8-day 的时间序列,然后利用时间序列重建根据 Timesat 提供的 S-G 算法,对新的 8-day 时间序列进行重建。对比重建前后时间序列发现,27%以上的象元质量得到提高,数据整体的时空完整性有明显提高,可以用于进一步的科学研究。

作者分工:王正兴负责数据库开发的整体设计、科学算法、数据质量控制和论文撰写。曹云锋负责数据采集、软件开发、数据分析和数据验证。

参考文献

- [1] 王正兴,曹云锋. 东南亚地区 2010 年 MODIS-NDVI 再分析(加密、去噪、重建)数据集[DB/OL]. 全球变化科学研究数据出版系统, 2016. DOI: 10.3974/geodb.2016.03.16.V1.
- [2] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策[OL]. DOI: 10.3974/dp.policy. 2014.05 (2017 年更新).
- [3] Didan, K., Huete, A. MODIS vegetation index product series collection 5 change summary [Z/OL]. <http://www.ctahr.hawaii.edu/grem/mod13ug/index.html>. TBRs Lab.,The University of Arizona, 2006-06-29.
- [4] Jönsson, P., Eklundh, L. TIMESAT-A program for analyzing time series of satellite sensor data [J]. *Computers and Geo-science*, 2004, 30 (8): 833-845.
- [5] 曹云锋,王正兴,邓芳萍. 3 种滤波算法对 NDVI 高质量数据保真性研究[J]. *遥感技术与应用*, 2010, 25 (1): 118-125.
- [6] 王正兴,柯灵红,邓芳萍. MODIS-NDVI 时间分辨率加倍算法: 从 16-Day 到 8-Day[J]. *遥感技术与应用*, 2011, 26(4): 437-443.
- [7] 王正兴,王亚琴. 利用 LDOPE 工具解码 MODIS 陆地产品质量信息[J]. *遥感技术与应用*, 2013, 28(3): 459-466.