文献 DOI:
10.11922/csdata.170.2017.0132
数据 DOI:
10.11922/sciencedb.397
10.11922/sciencedb.398
10.11922/sciencedb.399

文献分类：地球科学
收稿日期：2017-03-29
开放同评：2017-06-19
发表日期：2017-09-27

青藏高原地区物候数据集、地表覆盖种类图和青海湖环湖地区草地生物量数据集

张丽[1*]，王翠珍[2]，杨昊翔[1]，张炳华[1]，郑艺[1]

摘　要：青藏高原高山植被类型丰富、覆盖范围大，因此成为了研究高山植被空间分布和物候变化的热点区域。以MODIS地表反射率产品为主要数据源，辅以Landsat数据、气象数据、野外实测数据等补充数据，得到了高原植被物候数据集、地表覆盖种类图与青海湖环湖地区草地生物量数据集。主要研究方法和成果如下：①基于由MODIS地表反射率产品计算得到的NDVI时间序列，结合非对称高斯函数，利用阈值法提取物候参数，得到2000～2010年高原植被物候数据集，较从AVHRR GIMMS提取的物候参数具有更丰富的时空信息。②基于从样本数据中提取的植被物候和生物物理参数，利用支持向量机方法对MODIS影像实现青藏高原地区地表覆盖分类，得到2010年青藏高原地表覆盖种类图，分类精度为93%，较已出版的全国植被地图精度更高，且高山草甸与高山草原两种地类的分界线也更清晰和平滑。③基于MODIS和Landsat TM影像，利用STARFM数据融合算法及支持向量机非参数草地生物量估算模型，生成2000～2015年青海湖环湖地区草地生物量数据集，估算精度为82%。3组数据集可作为青藏高原地区植被研究的基础数据集，为相关科研工作提供参考依据。

关键词：青藏高原；植被；物候；地表覆盖种类；草地生物量

数据库（集）基本信息简介

数据集中文名称	2000～2010年青藏高原植被物候数据集	2010年青藏高原地表覆盖种类图	2000～2015年青海湖环湖地区草地生物量数据集
数据集英文名称	Phenological metrics dataset of alpine grasslands over the Tibetan Plateau (2000 – 2010)	Land cover map for Tibetan Plateau (2010)	Alpine aboveground biomass dataset for Qinghai Lake Basin (2000 – 2015)
数据作者	王翠珍、张丽、刘双俞、邱玉宝	王翠珍、张丽、邱玉宝、张镱锂	张炳华、张丽
通讯作者	王翠珍（cwang@mailbox.sc.edu）、张丽（zhangli@radi.ac.cn）	王翠珍（cwang@mailbox.sc.edu）、张丽 (zhangli@radi.ac.cn)	张丽（zhangli@radi.ac.cn）
数据时间范围	2000～2010年	2010年	2000～2015年
地理区域	青藏高原地理范围包括26°00′12″N～39°46′50″N，73°18′52″E～104°46′59″E，由昆仑山脉、喀喇昆仑山脉、唐古拉山脉、喜马拉雅山脉等主体山脉组成。	青藏高原地理范围包括26°00′12″N～39°46′50″N，73°18′52″E～104°46′59″E，由昆仑山脉、喀喇昆仑山脉、唐古拉山脉、喜马拉雅山脉等主体山脉组成。	青海湖环湖地区地理范围包括37°17′N～38°19′N，90°48′E～101°11′E，由大通山、日月山与青海南山之间的断层陷落形成。

1. 中国科学院遥感与数字地球研究所，数字地球重点实验室，北京　100094；
2. 美国南卡罗来纳大学，地理系，哥伦比亚　S.C. 29208
*通讯作者（Email: zhangli@radi.ac.cn）

（续表）

数据格式	*.Geotiff	*Geotiff	*.Geotiff
数据量	4.08 GB	70.6 MB	93.7 GB
数据服务系统网址	http://www.sciencedb.cn/dataSet/handle/397	http://www.sciencedb.cn/dataSet/handle/398	http://www.sciencedb.cn/dataSet/handle/399
基金项目	中国科学院遥感与数字地球研究所数字地球重点实验室开放研究基金（2013LDE002），国家自然科学基金国际合作交流项目(41120114001)，国家自然科学基金（41271372）	国家自然科学基金国际合作交流项目（41120114001），国家重点基础研究发展计划（973计划，2009CB723906）	国家自然科学基金（41271372），中国科学院国际合作局对外合作重点项目（131C11KYSB20160061）
数据集组成	2000 ~ 2010年青藏高原植被物候数据集包括植被生长季开始时间、生长季结束时间、生长季持续时间、NDVI峰值日期（pheno_change文件夹），空间分辨率为500 m，时间分辨率为年；及植被物候参数变化趋势数据（pheno_trend文件夹），空间分辨率为500 m。	2010年青藏高原地区地表覆盖种类图（LandCoverTypesMap文件夹），空间分辨率为500 m。	2000 ~ 2015年青海湖环湖地区草地生物量数据集（biomass result文件夹），空间分辨率为30 m，时间分辨率为8天（每年5 ~ 9月植被生长季）。

引　言

青藏高原是世界上海拔最高的高原，被称为“世界屋脊”“第三极”。青藏高原南至喜马拉雅山脉，北抵昆仑山脉、阿尔金山脉和祁连山脉，西起帕米尔高原，东到横断山脉，平均海拔4000米以上，总面积约250万平方千米，拥有丰富的自然资源和高山高寒动植物资源，其中高寒植被是青藏高原生物多样性的重要组成成分。然而，由于青藏高原特殊的地理位置、寒冷干旱的气候特点、高原地区的积雪和融雪效应[1]以及高寒植被的脆弱性，该区域植被对气候比较敏感，植被物候期、生长状况和地域分布易受到气候变化的影响[2-4]，草地也存在一定程度的退化[5]。随着近年来全球气候的不断变化，研究植被物候参数、理化生化参数及其与气候因子的关系也因此成为了热点。

由于青藏高原地区覆盖范围大、气候复杂多样、环境恶劣，实地考察和观测研究受到了极大的限制，导致科研人员难以开展相关研究。遥感技术具有观测范围广、时效性强、信息量大等众多优势，为青藏高原地区的科研工作开展带来了新的突破。

中分辨率成像光谱仪（Moderate resolution imaging spectroradiometer，简称MODIS）的地表反射率产品是重要的基础数据产品，广泛应用于植被指数计算、地物光谱获取、变化监测等方面的研究。通过建立数学模型，代入数据进行反演和分析，可以得到目标区域的空间分布、变化趋势、参数之间相关性等信息，为青藏高原地区的相关研究提供数据支持[6]。

本文介绍了3组青藏高原地区高山植被遥感数据集，包括2000 ~ 2010年高原植被物候数据集、2010年高原地表覆盖种类图和2000 ~ 2015年青海湖环湖地

区草地生物量数据集，并阐述了数据集反演方法、精度，旨在为相关科研人员提供基础地理信息数据。

1 数据采集和处理方法

1.1 数据采集

3 组青藏高原地区高山植被遥感数据集产品的原始数据主要采用 MODIS 产品、Landsat 遥感影像数据、野外实测数据及其他补充数据。相关数据介绍如下：

1.1.1 MODIS 产品和 Landsat 遥感影像数据

生产数据集用到的 MODIS 产品为 MOD09A1 和 MCD43A4。MOD09A1 为 2000 ~ 2010 年每 8 天合成的 500 m 青藏高原地区地表反射率产品，每年 46 期影像，由 MODIS 产品分幅 h25v05、h25v06、h26v05、h26v06 和部分 h23v05、h24v05 组成，作为青藏高原植被物候参数数据集的原始数据，其中 2010 年的 MOD09A1 数据还用作青藏高原地表覆盖种类图的原始数据；MCD43A4 为 2000 ~ 2015 年植被生长季期间（5 ~ 9 月）每 16 天合成的 500 m 地表反射率产品，由 MODIS 产品分幅 h26v05、h25v05 组成，作为青海湖流域草地生物量数据集的原始数据。MODIS 产品的下载地址为 https://ladsweb.modaps.eosdis.nasa.gov/search/。

Landsat 遥感影像数据主要为 Landsat5 TM 数据，数据的下载地址为 http://espa.cr.usgs.gov/index/。

1.1.2 野外实测数据

在生成 2010 年青藏高原地区地表覆盖种类图和 2000 ~ 2015 年青海湖流域草地生物量数据集的过程中，分别使用了不同的野外实测数据。前者使用的野外实测数据为 2011 年在青藏高原地区实地观测得到的数据，记录了包括 28 个高山草甸、54 个高山草原和 2 个高山荒漠在内共 84 个站点数据，每个站点数据覆盖范围为 1.5 km × 1.5 km，作为数据集生成过程中的训练样本。后者则是 2013 ~ 2015 年植被生长季期间（5 ~ 9 月）的生物量野外实测数据，获取方式为定点及不定点野外观测：定点观测通过选取样带作为定点观测实验区，在每年生长季（5 ~ 9 月）逐月进行 1 ~ 2 次的定点覆盖度观测；不定点观测通过采集样方的方式对样带进行加密，每个样方记录信息包括坐标、主要物种、草地高度、草地覆盖度、地上生物量、地表温度以及地物波谱信息等资料，其中生物量采用收获法获取，并记录取得的草样的鲜重。截至 2015 年底，三年共获取 409 个样方的草地生物量数据（每年 7 月底至 8 月初），1125 个样方的草地覆盖度数据（每年 5 ~ 9 月），其中青海湖环湖流域共采集有效生物量数据 287 个。

1.1.3 其他数据

生成 2010 年青藏高原地区地表覆盖种类图过程中，使用的样本点数据集含：

（1）基于 2011 年 8 月 11 ~ 13 日在西藏中部地区进行的航空高光谱实验（本实验共设计了 62 条飞行轨道，其中有 28 条与图 1 所示的四条航线平行）所拍摄的 2 m 空间分辨率影像，通过目视解译随机选取 33 个高山草甸和 35 个高山草原样本点，用于检验分类精度；

（2）基于 2001 年中国科学院发布的 1 ∶ 100 万中国植被地图集（ACAS）及其他出版资源 [7]，随机选取 22 个高山荒漠训练样本点，用于生成地表覆盖种类图；

（3）基于张镱锂等人提供的实地考察数据，选取的 2012 ~ 2013 年西藏西部尼玛、改则 4 个站点的高山荒漠样本点，用于检验分类精度；

（4）基于 MODIS 地表反射率产品（2010 年，第 217 天），选取 30 个常年积雪、33 个湖泊和 38 个裸地训练样本点，用于生成地表覆盖种类图；采用目视解译独立选取的 50 个湖泊和 50 个积雪样本点，用于检验分类精度 [8]。

部分样本数据点分布如图 1 所示。

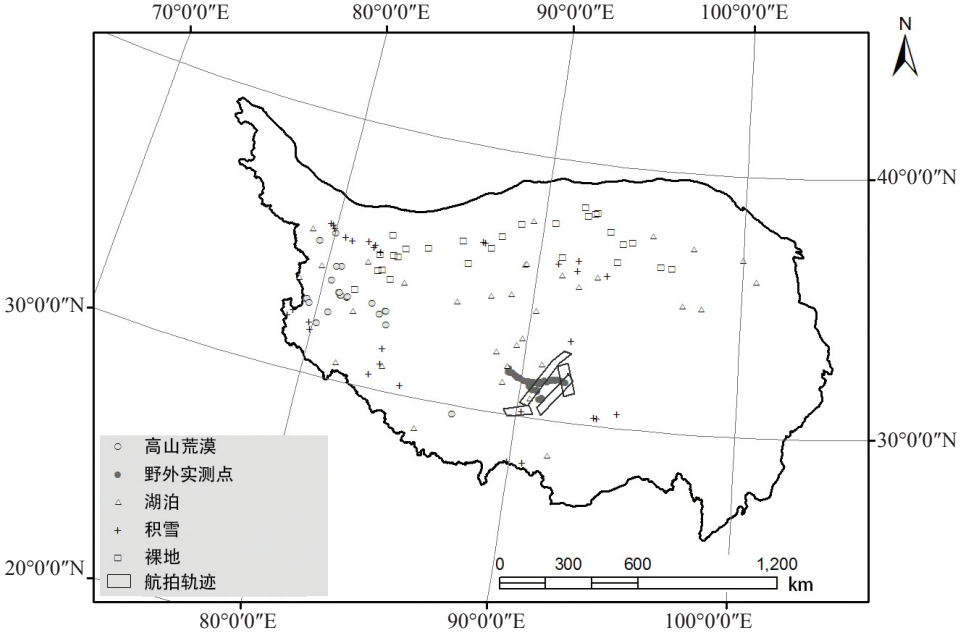


图 1　2011 年野外航拍轨迹与野外实测数据及部分地表覆盖类型样本数据分布示意图

生成 2000 ~ 2015 年青海湖流域草地生物量数据集的过程中，使用的补充数据包括：

（1）中国科学院“八五”重大应用项目“全国资源环境遥感宏观调查与动态研究”中基于 Landsat 数据提取的青海省 2000 年 1 ： 10 万土地利用数据，主要包括水体、建设用地、林地、耕地等，是目前全国精度最高的土地利用类型产品 [9]，在生物量数据集生产中用于非草地区域的掩膜；

（2）来源于美国太空总署（NASA）和国防部国家测绘局联合执行的 SRTM（Shuttle Radar Topography Mission）计划中测量得到的全球 30 m 空间分辨率的 DEM，用于计算研究区海拔信息，数据下载网址为 http://glcf.umd.edu/data/；

（3）由中国气象科学数据共享中心提供的气象数据，为中国地面气候资料日值数据集产品（V3.0），主要获取自国家级地面站观测数据。

1.2　数据处理

1.2.1　植被物候数据处理

青藏地区植被物候研究的数据处理可分为 3 个部分：数据预处理、植被物候参数提取和植被物候变化趋势分析 [10-11]。

（1）数据预处理：基于 MODIS 的地表反射率产品，计算得到 2000 ~ 2010 年的 NDVI 时间序列影像和曲线。舍弃 NDVI 明显偏小的像素后，利用 Savitzky-Golay 滤波器对存在云污染的 NDVI 时间序列曲线进行平滑处理。

（2）植被物候参数提取：将平滑后的曲线与非对称高斯函数进行拟合，从而

提取出植被的生长季开始时间、生长季结束时间、生长季持续时间和 NDVI 峰值日期 4 个物候参数，得到了青藏高原地区 2000 ~ 2010 年植被物候数据集。4 种物候参数分别对应 pheno_change 文件中的 4 景物候参数影像。该影像由 11 个波段组成，分别对应 2000 ~ 2010 年每年的植被物候参数。

（3）植被物候变化趋势分析：虽然通过一般的线性拟合法，可以得到植被物候参数的变化趋势，但植被物候参数存在自相关性，仅利用线性拟合法进行趋势分析时，会导致包括零均值、常数方差在内的这些残差不满足要求。因此，采用 Mann-Kendall 趋势分析方法对物候参数的变化趋势进行分析，可以得到该地区 2000 ~ 2010 年更加准确的植被物候变化趋势数据集。4 种物候参数分别对应 pheno_trend 文件中的 4 景物候参数变化趋势影像。该数据集反映了各物候参数在 2000 ~ 2010 年间的总体变化动向。

1.2.2 高原地表覆盖种类数据处理

该数据处理流程主要由数据预处理、参数提取、地表覆盖分类 3 部分组成 [8]：

（1）数据预处理：基于 MODIS 的地表反射率产品（MOD09A1），计算了 2010 年青藏高原地区的 NDVI 时间序列影像，共计 46 期。参照 1.2.1 节中的数据预处理方法，对 NDVI 曲线进行了平滑去噪处理。

（2）物候和生物物理参数提取：针对平滑后的 NDVI 时间序列曲线，除了用 1.2.1 节物候参数提取的方法提取了 4 个物候参数之外，还从 NDVI 曲线中提取了 NDVI 峰值与生长季 NDVI 总和 2 个生物物理参数。这两类参数反映了草地的生长状况和生长周期，可用于区分地表覆盖类型中的不同草地类型。

（3）地表覆盖分类：基于野外实测数据和补充数据，将提取的各草地类型生物物理参数和物候参数作为训练样本，利用图 2 所示的支持向量机（SVM）和多阈值判断综合起来的分类方法，将青藏高原地区分成了高山荒漠、高山草原、高山草甸、裸地、湖泊、积雪 6 个地表覆盖类型，得到了 2010 年青藏高原地区地表覆盖种类图。

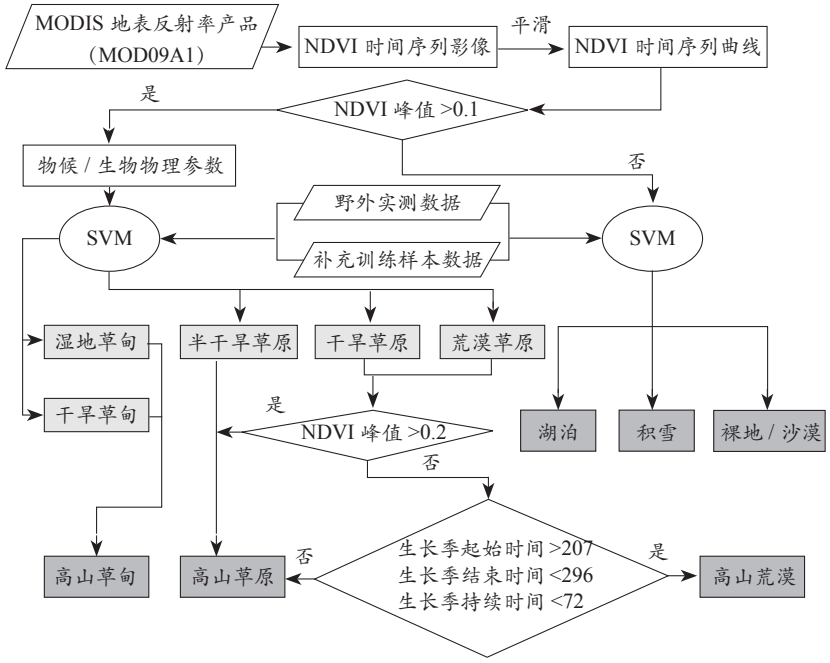


图 2　青藏高原地区地表覆盖分类流程图

1.2.3 草地生物量数据处理

该数据处理过程主要包括基于 STARFM 算法的数据融合与草地生物量估算模型的建立 [12]，数据集生成的流程如图 3 所示：

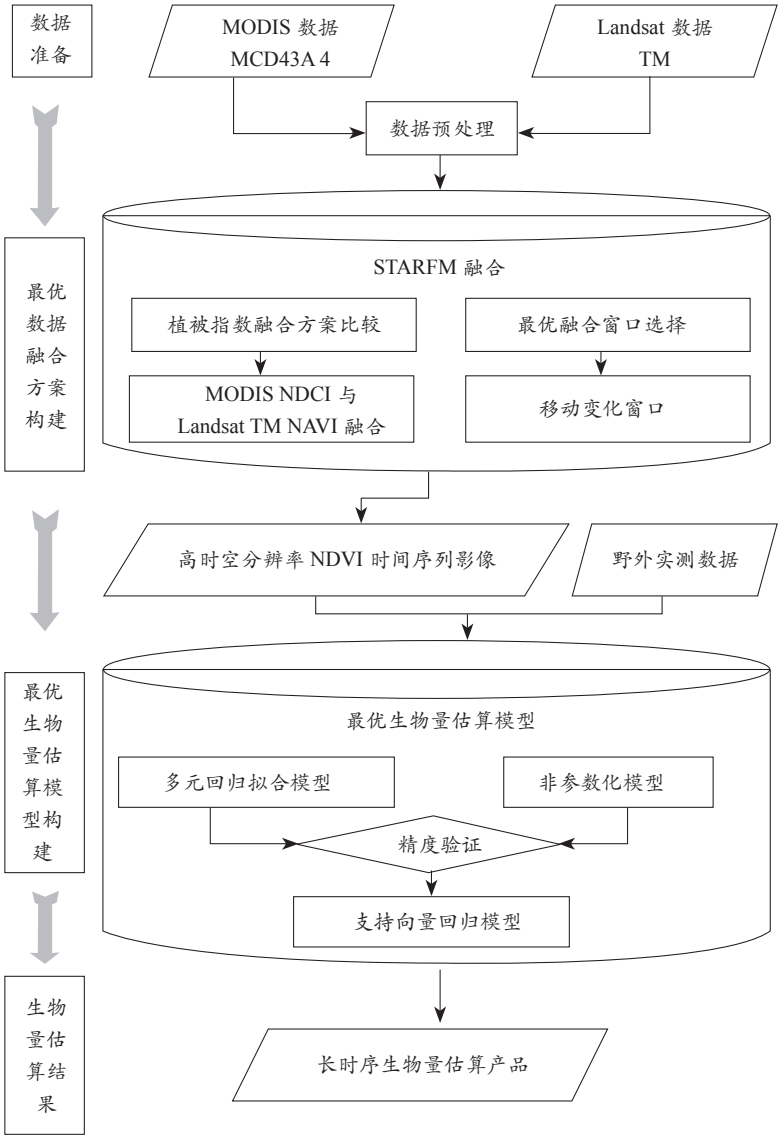


图 3 青海湖环湖地区草地生物量数据集获取流程图

（1）基于 STARFM 算法的数据融合：在综合比较了多种融合方案后，本数据集生成过程中以 MODIS MCD43A4 数据和根据获取情况而选择的一年内、间隔一年、间隔两年的 TM 数据为遥感输入数据，利用 STARFM 融合算法，将 MODIS 和 Landsat 影像分别计算得到的植被指数进行融合，生成新的植被指数。在忽略大气误差和影像配准误差前提下，STARFM 算法假设单一地物类型、同时相的 Landsat 影像像元可以由 MODIS 影像像元加权平均表示。然而由于混合像元、地表覆盖类型、物候等因素的影响，单个 Landsat 像元很难由 MODIS 影像表示。因此实际操作过程中利用邻近相似像元对 Landsat 像元值进行预测，即以预测像元为中心，依据像元值相似程度、空间距离以及时间距离等，在某个固定大小的窗口内寻找预测像元的固定点，并对所有相似像元附以一定的权值，最终获得中间

预测像元的值，以此方法对整幅影像进行卷积运算，获得整幅影像预测值。其中耕地采用 350 m 窗口，林地采用 950 m 窗口，草地及其他用地类型采用 750 m 窗口，最后生成了兼有 MODIS 的高时间分辨率和 Landsat TM 中空间分辨率优势的 NDVI 时间序列影像。

（2）草地生物量估算模型的建立：结合数据融合生成的 NDVI 时间序列和野外实测数据，分别建立参数化和非参数化的生物量估算模型。基于精度验证分析，最终选择支持向量回归模型为最优模型并进行草地生物量的估算，生成了 2000 ~ 2015 年植被生长季期间（5 ~ 9 月）每 8 天 30 m 的草地生物量数据，并对该期间的草地生物量求均值，得到 2000 ~ 2015 年间 16 景植被生长季期间青海湖流域的年草地生物量均值影像。

2 数据样本描述

2.1 2000 ~ 2010 年青藏高原植被物候数据集

数据集中 2010 年青藏高原植被物候参数的空间分布情况如图 4 所示，2000~2010 年物候参数变化趋势空间分布如图 5 所示。图 5 中数字 1 ~ 6 表示各种典型的生态和气候区，其中 1 区为青海 – 祁连东部大草原，寒冷半干旱气候；2 区为那曲 – 玉树草甸，寒冷半湿润气候；3 区为西藏南部草原，温暖半干旱气候；4 区为阿里高山荒漠，寒冷干燥气候；5 区为羌塘草原，寒冷半干旱气候；6 区为青海南部草原，寒冷半干旱气候。

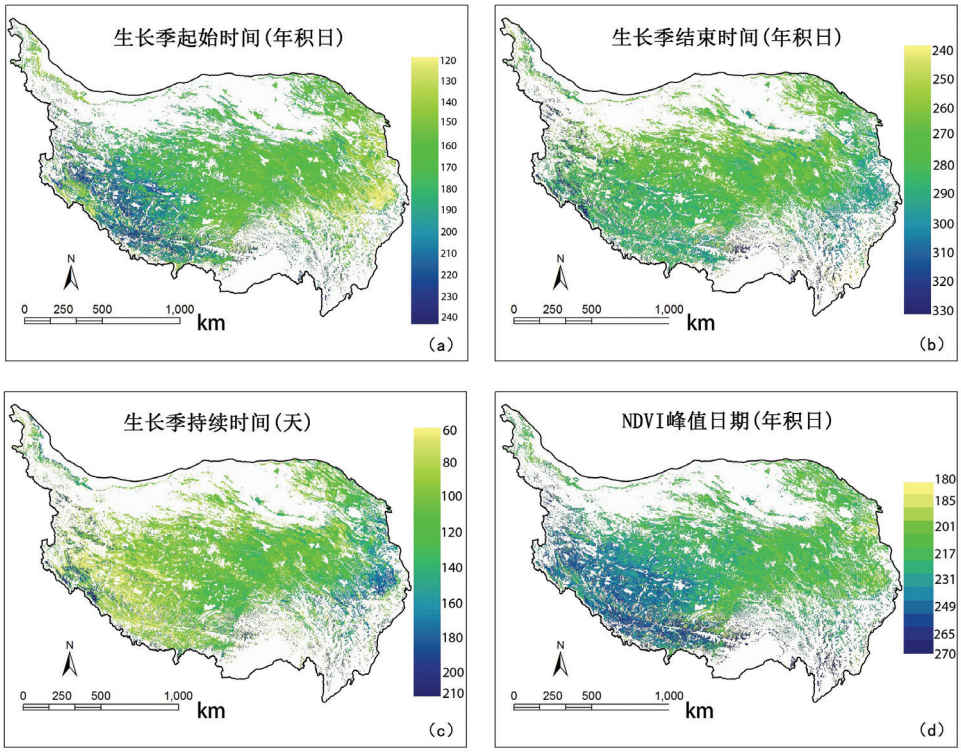


图 4 2010 年青藏高原植被物候参数

(a) 生长季起始时间；(b) 生长季结束时间；(c) 生长季持续时间；(d) NDVI 峰值日期

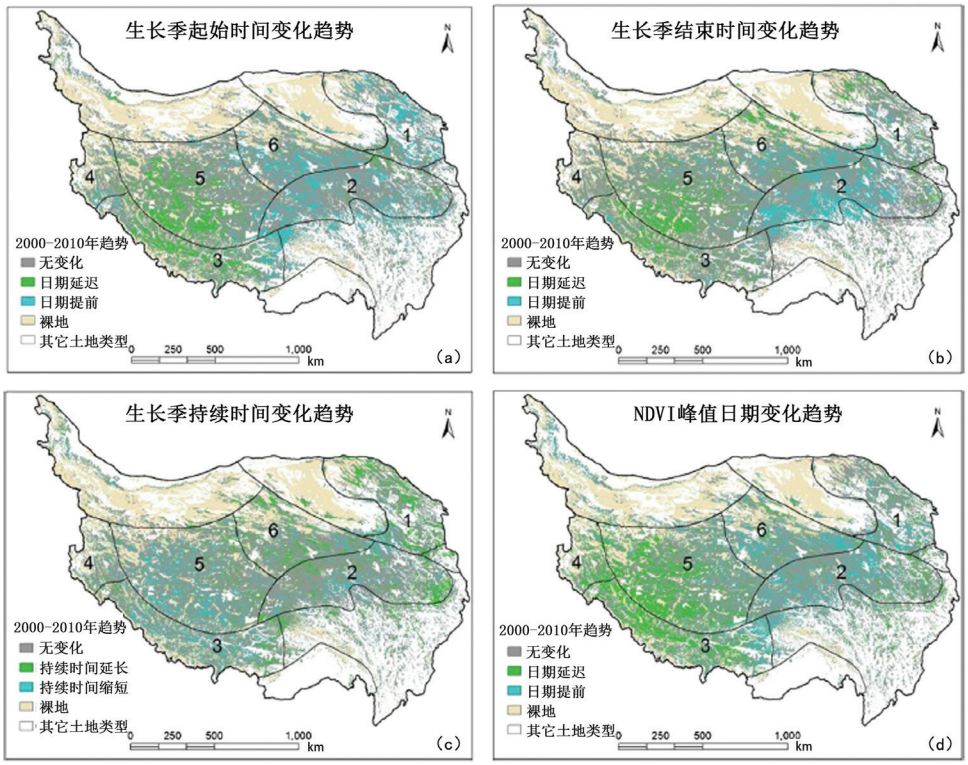


图 5　2000 ～ 2010 年青藏高原植被物候参数变化趋势图

(a) 生长季起始时间；(b) 生长季结束时间；(c) 生长季持续时间；(d) NDVI 峰值日期 [11]

图 4 中 a ～ d 分别表示 2010 年青藏高原地区植被生长季起始时间（年积日）、生长季结束时间（年积日）、生长季持续时间（天）和 NDVI 峰值日期（年积日），年积日即从当年 1 月 1 日起开始计算的天数。图 5 中 a ～ d 分别表示植被生长季起始时间、生长季结束时间、生长季持续时间和 NDVI 峰值日期的年积日变化趋势；无变化表示对应的物候参数年积日没有明显的提前和推迟；日期延迟是指生长季起始和结束时间、NDVI 峰值日期在这 11 年期间，到来的日期逐渐推后，而日期提前则指这些物候参数到来的日期逐渐提前；生长季持续时间则分别表现为持续时间延长和缩短的趋势。

从物候观点分析看，2000 ～ 2010 年间，青藏高原有将近 40%的地区存在明显的物候变化，其西部地区生长季起始、结束时期和 NDVI 峰值日期均有所延迟，而东部地区均有所提前；西部地区生长季持续时间缩短，东部则延长。该变化状况与生长季期间青藏高原西部降水的减少和东部降水增加的气候变化表现出较好的一致性。而随着青藏高原地区整体气温升高，生长季期间的降水变化趋势将可能成为高山植被类型变化和分布的主要驱动因素。

2.2　2010 年青藏高原地表覆盖种类图

2010 年青藏高原地区地表覆盖种类如图 6 所示，可以看出整个青藏高原地区从东部的高山草甸覆盖类型逐渐过渡到西部高山荒漠覆盖类型。高山草甸占整个青藏高原地区的 1/3，主要分布于因生长季季风影响而具有温暖半湿润和寒冷半干旱气候的青藏高原东部与东北部；高山草原覆盖类型分布于以半干旱气候为主的青藏高原中部和西南部。地表覆盖种类自青藏高原西部地区开始，逐渐转变为耐

热耐旱的高山荒漠类型。青藏高原地区分布着许多湖泊，其北部主要是沙漠和裸地，西部的喀喇昆仑山脉和南部的喜马拉雅山脉常年覆盖积雪和冰川，一些诸如森林、灌木、农耕用地等非高山植被覆盖土地类型则分布于青藏高原东南部。

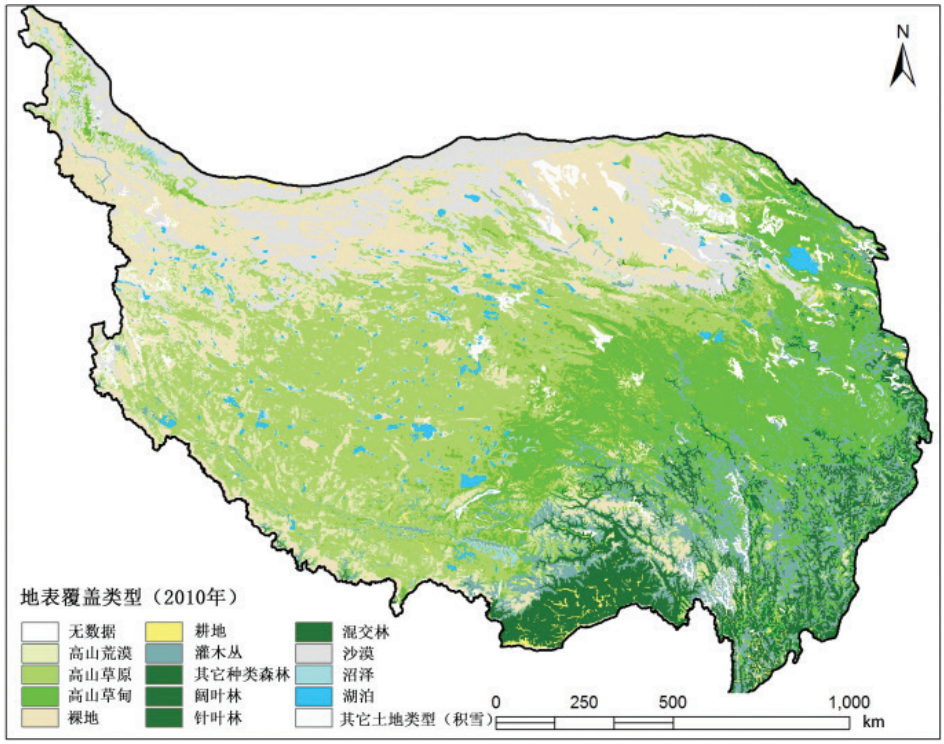


图 6　2010 年青藏高原地区地表覆盖种类图

2.3　2000 ~ 2015 年青海湖环湖地区草地生物量数据集

将基于 STARFM 算法得到的融合 NDVI 时间序列，代入到支持向量机的非参数模型，对青海湖环湖地区进行草地生物量估算，获得 2000 ~ 2015 年青海湖环湖地区草地生物量数据集。如图 7 所示为 2015 年植被生长季期间，8 天间隔的青海湖环湖地区草地生物量空间变化图，从图中可以看出融合植被指数生成的生物量在空间上显示了较大的异质性。2015 年，青海湖草地植被 5 月生物量较小，6 月植被生长较为旺盛，在 7 月初到 8 月初为生长旺季。7 月底整个流域生物量区域平均值为 250 g/m^2 左右，7 月底部分地区的植被生物量可达到 500 g/m^2 以上，主要集中在青海湖湖北、湖南以及流域中部的高寒草甸地区。8 月底生物量开始逐渐减少，至 9 月底，整个流域生物量与 5 月初的生物量在数值上持平。

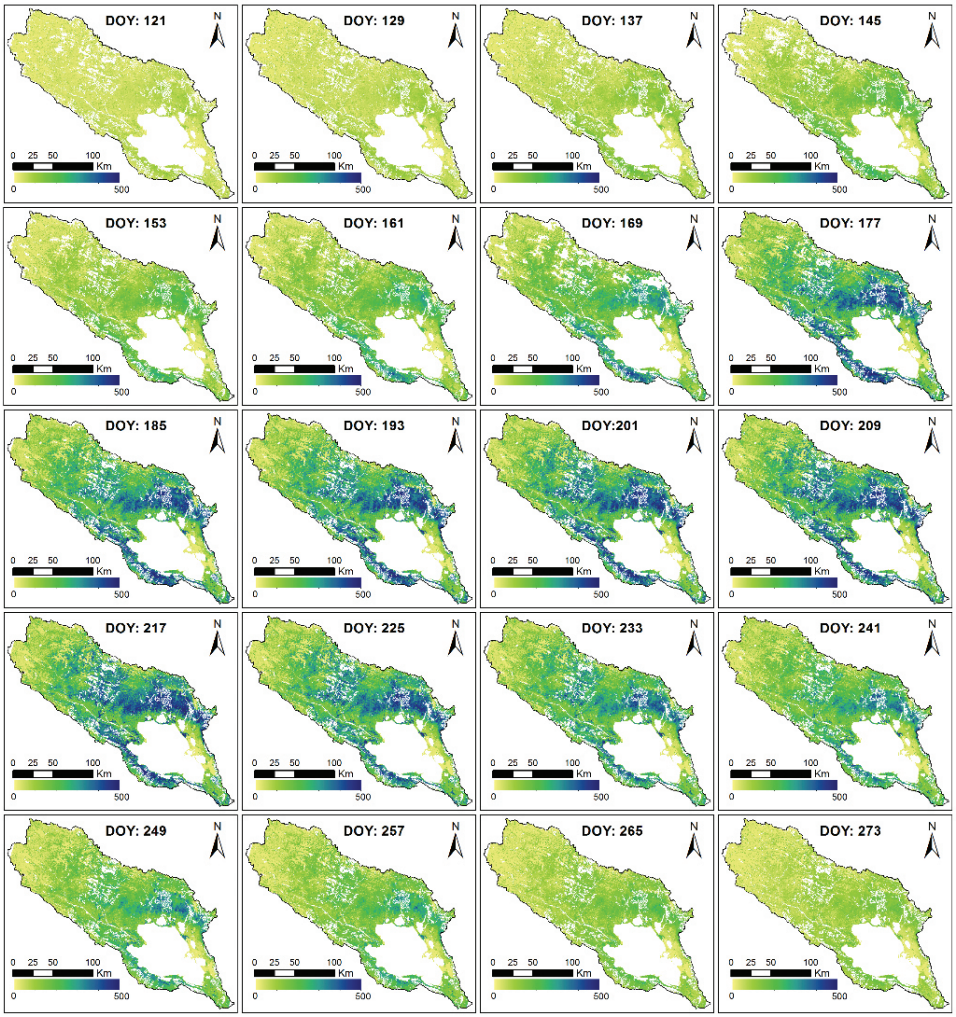


图 7　2015 年植被生长季青海湖地区 8 天间隔草地生物量空间分布图

3　数据质量控制和评估

3.1　植被物候数据集质量控制与评估

3.1.1　质量控制

为了提高植被物候数据集的精度和质量，采取以下数据质量控制方法：

（1）为了保证提取的物候数据的准确性，需要对由地表反射率产品计算得到的 NDVI 时间序列影像进行筛选，将其中 NDVI 最大值小于 0.1 的像元视为裸地，不参与物候参数的提取。

（2）对原始的 NDVI 时间曲线进行的平滑处理，可以去除原始曲线中的云污染和噪声，从而高效、正确地提取物候参数。

3.1.2　质量评估

本数据集中物候参数及其变化趋势与青藏高原 109 个气象站点记录的气象变化结果具有一致性。根据气象站点记录的青藏高原地区降水数据，分析得出降水的变化趋势与物候参数的变化趋势在地理空间分布上具有同质性，符合植被对气

候的响应，故验证本数据集提取的物候参数的可靠性。

3.2 地表覆盖种类图质量控制与评估

3.2.1 质量控制

为了提高地表覆盖种类图的精度和质量，采取以下数据质量控制方法：

（1）由于生成地表覆盖种类图时需要植被物候参数作为分类指标，因此数据处理过程中同样需要 3.1.1 小节的数据质量控制方法保证提取的物候参数的准确性。

（2）采用结合植被物候参数和生物物理参数的多级综合分类方法，可以减少由于各种草地覆盖类型之间空间和光谱的同质性所引起的错误分类现象，从而提高数据质量。

3.2.2 质量评估

为评估本数据集的数据质量，需要对地表覆盖种类图的分类结果进行精度验证。通过选取 1.1.3 小节中提到的若干样本点作为验证样点，从而生成混淆矩阵如表 1 所示。在不考虑高山荒漠草原类型和裸地的情况下，本数据集的分类精度可达 93%。对于高山植被覆盖类型，33 个高山草甸样本点中，仅有 1 个被错分为高山草原，用户精度达到 97%，35 个高山草原样本点中有 5 个被错分为高山草甸，用户精度达 86%。积雪和湖泊分类精度高是因为它们的训练样本点和验证样本点来自于同一景 MODIS 影像。

表 1 地表覆盖种类图样本验证混淆矩阵 [8]

分类结果 \ 参考样本点	高山草甸	高山草原	积雪	湖泊
高山草甸	32	5	0	0
高山草原	1	30	0	0
积雪	0	0	50	0
湖泊	0	0	0	44
裸地 / 沙漠	0	0	0	6
总计	33	35	50	50

总体精度：92.86%

3.3 草地生物量数据集质量控制与评估

3.3.1 质量控制

为了提高草地生物量数据集的精度和质量，采取以下数据质量控制方法：

（1）为保证数据融合的质量，在生成数据集的过程中选用了 MOD09A1、MOD09Q1、MCD43A4 三种地表反射率数据，采用融合反射率数据计算 NDVI 和对反射率数据计算得到的 NDVI 融合两种方案进行比较试验，并以均方根误差（*RMSE*）、相对均方根误差（*RMSEr*）等参数作为精度评价指标，获取最佳数据融合方案。相对均方根误差（*RMSEr*）的计算公式如式（1）所示：

$$RMSE_r = \frac{\sqrt{\frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}(p_{(i,j)}-q_{(i,j)})^2}}{\bar{p}} \times 100\% \quad (1)$$

其中 M 和 N 分别表示影像的行数和列数，p 表示融合前的原影像，q 表示融合后影像，表示原影像的像素均值。相对均方根误差表征了融合影像与观测值的相对误差程度，差异越小，说明融合效果越好。

（2）在建立生物量估算模型的过程中，通过比较线性、幂函数、指数、对数、多项式和 SVM 等多种参数 / 非参数模型的精度，选优最佳估算模型。

3.3.2 质量评估

在产生草地生物量数据集之前，需要将 287 个实验区有效的实地采样生物量数据分为两组，一组 215 个点，用来做生物量估算模型训练样本，另外一组 72 个点，用来作为模型精度验证数据，质量评估过程选用的精度验证指标包括相关系数（r）、均方根误差（*RMSE*）、相对均方根误差（*RMSEr*）。其中，r 越大，RMSE 越小，RMSEr 越小，说明模型的估算值与实测值越接近，当 RMSE 或者 RMSEr 等于 0 时，模型为理想状态，无误差。用生成反演模型的训练样本计算得到的相关系数 r 可达 85%，均方根误差（*RMSE*）为 74.45 g/m^2，RMSEr 为 34.5%；而将验证样本代入模型计算得到的相关系数 r 可达 82%，RMSE 为 78.28 g/m^2，*RMSEr* 为 37.05%。研究成果与之前在内蒙锡林郭勒地区的研究结果一致 [13]，可以看出生物量估算模型精度高，数据集成果具有可信度。

4 数据价值

将本文涉及的青藏高原物候数据集、地表覆盖种类图及生物量数据集与国内外其他同类数据集进行对比分析，结果表明本文提出的 3 组数据集在时空分辨率、产品精度等方面具有优势：

（1）不同研究学者利用 AVHRR GIMMS 的 NDVI 产品对青藏高原地区植被生长状况进行监测时得出了不同结论，例如一些学者研究得出近几十年来青藏高原地区植被绿度增加，植被返青期提前 [14-17]，其它研究却表明植被绿度没有明显变化 [18]，一些研究甚至得出青藏高原地区植被退化，高山草原生长开始日期推迟的结论 [19-21]，反映出 AVHRR GIMMS 的 NDVI 产品存在较大的不确定性；而本文生产 2000 ~ 2010 年青藏高原地区植被物候数据集使用的是 MODIS NDVI 产品，经过本文第 3.1 节的检验，证明了数据集与产品的可靠性。且本文提取物候时利用的 MODIS NDVI 产品较 AVHRR GIMMS NDVI 产品有更高的时空分辨率。

（2）将基于 MODIS 影像得出的地表覆盖种类图与 ACAS 在青藏高原地区的地表植被覆盖情况进行了对比。利用表 1 中的验证样本点对 ACAS 进行验证，得出如表 2 所示的混淆矩阵。由表 1 可知基于 MODIS 影像进行分类的总体精度达到了 93%，比总体分类精度为 79% 的 ACAS 要高（表 2）；同时，本研究的分类图在高山草甸和高山草原地表种类的分界线上较 ACAS 地图更为清晰，过渡更平滑，分界区域更完整；除此之外，MODIS 分类图在高山草原和高山草甸、高山荒漠、小型湖泊等地物种类上，较 ACAS 地图分类更加准确 [8]。这主要是因为本数据集使用的 MODIS 影像数据获取于 2010 年，而用来进行验证的野外数据点来源于 2011 年（地面调查及高光谱航片），在这两年间植被变化忽略不计的前提下，本数据集分类总体精度较高，而 ACAS 图集的原始影像来源于 20 世纪八九十年代的 Landsat TM 影像，随着近几十年来全球气候变化，地表植被变化会对 ACAS 图集的验证产生影响。但由于在生产本数据集时，当时的实地样本已经无从获取，

不能进行同期的ACAS分类验证，因此在选取验证点时，尽量采用了大的植被地块，以尽量减少验证误差。

表2　ACAS地图样本验证混淆矩阵[8]

参考样本点 / ACAS结果	高山草甸	高山草原	积雪	湖泊
高山草甸	24	12	0	3
高山草原	9	23	3	2
积雪	0	0	43	0
湖泊	0	0	0	43
裸地 / 沙漠	0	0	4	2
总计	33	35	50	50

总体精度：79.17%

（3）高原草地生物量是评价高原草地生长状况的一项重要指标，因此对该参数进行估算可以有效管理和保护高原草地资源。但由于单一的遥感数据源无法同时满足草地轮牧观测的中高时间分辨率和高空间分辨率的需要[12]，而通过融合中空间分辨率的TM影像和高时间分辨率的MODIS影像所具有的优势则凸显出来。将青海湖环湖地区样地融合影像NDVI与500 m的MODIS NDVI数据集对比，结果如图8所示。通过比较两种NDVI影像在同一区域的平均值与标准差，可以看出两者没有明显差异，从而说明融合后的NDVI影像序列不仅能够有效反映500 m的MODIS NDVI的信息，同时还兼顾了MODIS高时间分辨率和Landsat TM影像中空间分辨率的优点。除此之外，融合后的NDVI影像较500 m的MODIS NDVI能更好地观察植被的时空变化状况，从而更好地描述河流两岸的植被分布及生长状况。

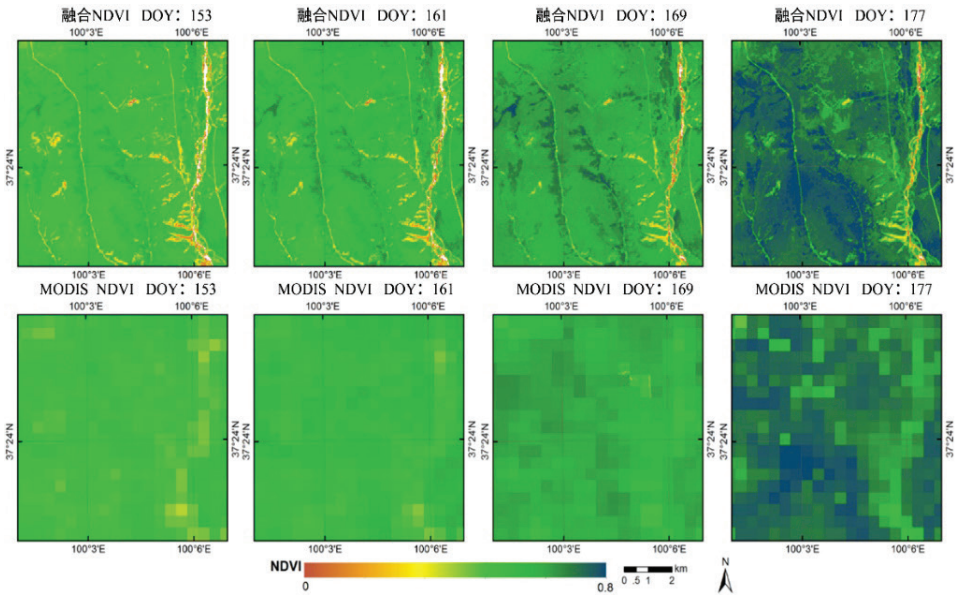


图8　2015年6月青海湖流域样地融合NDVI与MODIS NDVI变化比较图[12]

5 数据使用方法和建议

本文介绍的 3 组研究成果从植被物候参数提取、遥感数据融合和估算模型等角度出发，对青藏高原地区植被进行了研究，最终生成该地区 2000 ~ 2010 高原植被物候数据集、2010 年高原地表覆盖种类图和 2000 ~ 2015 青海湖环湖地区草地生物量数据集。

2000 ~ 2010 高原植被物候数据集分为 pheno_trend 和 pheno_change 两个文件。其中 pheno_change 文件中为 2000 ~ 2010 年青藏高原地区植被物候专题图，包括植被生长季开始时间，生长季结束时间，NDVI 峰值日期和生长季持续时间，共 4 期影像，每期影像 11 个波段，对应每年的植被物候信息，前三个物候参数的单位为年积日，生长季持续时间的单位为天，数据集空间分辨率为 500 m，时间分辨率为 1 年；pheno_trend 文件中为植被物候参数变化趋势数据，共 4 期影像，影像信息见图 5，空间分辨率为 500 m。本数据集中 Geotiff 格式的影像可用 Envi 软件或 ArcGIS 软件打开。

2010 年青藏高原地区地表覆盖种类图为 LandCoverTypesMap 文件夹中的 SVM_2010_3x3major_Plateau 文件，空间分辨率为 500 m，数据中的 class 11 为背景类，属于研究区之外。本数据集中 Geotiff 格式的影像可用 Envi 软件或 ArcGIS 软件打开。

2000 ~ 2015 青海湖环湖地区草地生物量数据集为 biomass result 文件夹。包括 2000 ~ 2015 年生长季（5 ~ 9 月）青海湖环湖地区草地生物量数据，共 16 期影像，每期影像 20 个波段，为每年生长季期间的草地生物量，单位为 g/m^2，空间分辨率为 30 m，时间分辨率为 8 天（每年 5 ~ 9 月植被生长季）。本数据集中 Geotiff 格式的影像可用 Envi 软件或 ArcGIS 软件打开。

以上数据集可为相关学者研究青藏高原地区植被的物候及其与气候的相关性、植被分布和变化趋势等方面提供可参考的资料，用以更全面地评价全球变暖及青藏高原变暖下高山植被类型的生态响应 [22]，辅助管理青藏高原地区草地资源和维护草原生态系统稳定性，并根据植被物候变化研究青藏高原地区，甚至全球范围的气候环境变化。

致　谢

感谢国家自然科学基金“基于多源遥感协同的草原生物量动态预测模型研究”（41271372）和中国科学院国际合作局对外合作重点项目（131C11KYSB20160061）对本数据集提供的资助。感谢 USGS 提供的 MODIS 和 Landsat 影像数据，感谢中国气象数据分享服务系统提供的青藏高原地区气温和降水站点实测数据。

数据作者分工职责

张　丽（1975—），女，新疆人，博士，研究员，研究方向为植被生态遥感。主要承担工作：草地生物量估算方案设计和建模工作，联合参与高原植被物候数据集和地表覆盖种类图制作。

王翠珍（1971—），女，山东人，博士，副教授。研究方向为植被环境遥感及环境监测。主要承担工作：高原植被物候数据集和地表覆盖种类图制作。

杨昊翔（1995—），男，湖北人，硕士研究生，研究方向为植被生态遥感。主要承担工作：数据处理和分析。

张炳华（1991—），女，山东人，硕士研究生，研究方向为植被生态遥感。主要承担工作：青海湖环湖地区草地生物量估算模型建立和反演。

郑　艺（1990 —），女，河北人，硕士研究生，研究方向为植被生态遥感。主要承担承担工作：数据处理和分析。

参考文献

[1] Wang K, Zhang L, Qiu Y, et al. Snow effects on alpine vegetation in the Qinghai-Tibetan Plateau[J]. International Journal of Digital Earth, 2015, 8(1): 56–73.

[2] 侯小丽, 张丽, 张炳华, 等. 青藏高原植被对气候变化响应的研究进展 [J]. 安徽农业科学, 2016, 44(17): 230–235, 244.

[3] Zhang L, Guo H, Wang C, et al. The long-term trends (1982–2006) in vegetation greenness of the alpine ecosystem in the Qinghai-Tibetan Plateau[J]. Environmental Earth Sciences, 2014, 72(6): 1827–1841.

[4] Zhang L, Guo H, Ji L, et al. Vegetation greenness trend (2000 to 2009) and the climate controls in the Qinghai-Tibetan Plateau[J]. Journal of Applied Remote Sensing, 2013, 7(1): 1–17.

[5] 李亚楠, 张丽, 廖静娟, 等. 藏北中部地区草地退化遥感监测 [J]. 遥感技术与应用, 2013, 28(6): 1069–1075.

[6] Li B, Zhang L, Yan Q, et al. Application of piecewise linear regression in the detection of vegetation greenness trends in the Tibetan Plateau[J]. International Journal of Remote Sensing, 2014, 35(4): 1526–1539.

[7] Zhang Y, Bai W, Liu L, et al. Land use and land cover change in the Tibetan Plateau[C]. Workshop on the Third Pole Environment, LUCC and Climate Adaption in Tibetan Plateau. Beijing, 2010.

[8] Wang C, Guo H, Zhang L, et al. Improved alpine grassland mapping in the Tibetan Plateau with MODIS time series: A phenology perspective[J]. International Journal of Digital Earth, 2013, 8(2): 131–150.

[9] Liu J, Liu M, Zhuang D, et al. Study on spatial pattern of land-use change in China during 1995–2000[J]. Science in China, 2003, 46(4): 373–384.

[10] 刘双俞, 张丽, 王翠珍, 等. 基于 MODIS 数据的青藏高原植被物候变化趋势研究 (2000 年 –2010 年) [J]. 遥感信息, 2014, 29(6): 25–30.

[11] Wang C, Guo H, Zhang L, et al. Assessing phenological change and climatic control of alpine grasslands in the Tibetan Plateau with MODIS time series[J]. International journal of biometeorology, 2015, 59(1): 11–23.

[12] 张炳华. 基于融合算法和 SVM 模型的青海湖流域草地生物量遥感估算研究 [D]. 北京 : 中国科学院大学, 2016.

[13] Zhang B, Zhang L, Xie D, et al. Application of synthetic NDVI time series blended from Landsat and MODIS data for grassland biomass estimation[J]. Remote Sensing. 2016, 8(1): 10.

[14] Yang Y, Piao S. Variations in grassland vegetation cover in relation to climatic factors on the Tibetan Plateau[J]. Journal of Plant Ecology, 2006, 30: 1–8.

[15] Xu X, Chen H, Levy J. Spatiotemporal vegetation cover variations in the Qinghai-Tibet Plateau under global climate change[J]. Chinese Science Bulletin, 2008, 53: 915–922.

[16] Zhong L, Ma Y, Salama M, et al. Assessment of vegetation dynamics and their response to variations in precipitation and temperature in the Tibetan Plateau[J]. Climatic Chang, 2010, 103: 519–535.

[17] Zhang G, Zhang Y, Dong J, et al. Green-up dates in the Tibetan Plateau have continuously advanced from 1982 to 2011[J]. PNAS, 2013, 110: 4309–4314.

[18] Zhang J, Yao F, Zheng L, et al. Evaluation of grassland dynamics in the Northern-Tibet Plateau of China using remote sensing and climate data[J]. Sensors, 2007, 7: 3312–3328.

[19] Liang S, Chen J, Jin X, et al. Regularity of vegetation coverage changes in the Tibetan Plateau in the past 21 years[J]. Advances in Earth Science, 2007, 22: 33–40.

[20] Mao F, Zhang Y, Hou Y, et al. Evaluation of grassland degradation in Naqu Northern Tibet[J]. Chinese Journal of Applied Ecology, 2008, 19: 278–284.

[21] Yu H, Luedeling E, Xu J. Winter and spring warming result in delayed spring phenology on the Tibetan Plateau[J]. PNAS Early Edition,2010: 1–6.

[22] Wang C. A remote sensing perspective of alpine grasslands on the Tibetan Plateau: Better or worse under "Tibet Warming"?[J]. Remote Sensing Applications: Society and Environment, 2016, 3:36–44.

数据引用格式

(1) 王翠珍, 张丽, 刘双俞, 等. 2000 ～ 2010 年青藏高原植被物候数据集 [DB/OL].Science Data Bank, 2017. DOI: 10.11922/sciencedb.397.

(2) 王翠珍, 张丽, 邱玉宝, 等. 2010 年青藏高原地表覆盖种类图 [DB/OL].Science Data Bank, 2017. DOI: 10.11922/sciencedb.398.

(3) 张炳华, 张丽. 2000 ～ 2015 年青海湖环湖地区草地生物量数据集 [DB/OL].Science Data Bank, 2017. DOI: 10.11922/sciencedb.399.

Phenological metrics dataset, land cover types map for the Tibetan Plateau and grassland biomass dataset for Qinghai Lake Basin

Zhang Li, Wang Cuizhen, Yang Haoxiang, Zhang Binghua, Zheng Yi

ABSTRACT With abundant types and large coverage of alpine vegetation, the Tibetan Plateau has been a hot region for researches on the spatial distribution and phenological changes of alpine vegetation. Using the MODIS Surface Reflectance Product as the main data source, supplemented by Landsat images, meteorological data and field measurements, we obtained a phenological metrics dataset of alpine grasslands, a land cover types map for Tibetan Plateau and an alpine aboveground grassland biomass dataset for Qinghai Lake Basin. Major research methods and outcomes of this study include: ① Based on NDVI time series derived from the MODIS Surface Reflectance Product, threshold method was used together with a symmetric Gaussian function to extract phenological metrics. The phenological metrics dataset of alpine grasslands (2000 – 2010) generated therefrom showed more spatial and temporal details than phenological products derived from AVHRR GIMMS. ② With phenological and biophysical metrics extracted from sample data, we applied the support vector machine (SVM) on MODIS images to get the classification of land cover types over the Tibetan Plateau, thereby creating a land cover types map for Tibetan Plateau (2010). Results showed that the map had a classification accuracy of 93%, higher than the published national vegetation altas. Besides, the newly-created land cover types map had a clearer and smoother transition from alpine meadow to alpine steppe. ③ Based on MODIS and Landsat TM images, the alpine aboveground grassland biomass dataset (2000 – 2015) for Qinghai Lake Basin was obtained by using STARFM data fusion algorithm and non-parameterized SVM biomass estimation model. The dataset had an estimated precision of 82%. The three datasets can serve as fundamental data for studies on vegetation over Tibetan Plateau, providing references for related scientific researches.

KEYWORDS Tibetan Plateau; vegetation; phenology; land cover types; grassland biomass

论文引用格式: 张丽, 王翠珍, 杨昊翔, 等. 青藏高原地区物候数据集、地表覆盖种类图和青海湖环湖地区草地生物量数据集 [J/OL]. 中国科学数据, 2017, 2(2). DOI: 10.11922/csdata.170.2017.0132.