

结合多源专题数据和目视解译的大区域密集湿地样本数据生产

彭凯锋^{1,2}, 蒋卫国¹, 侯鹏³, 凌子燕^{1,4}, 牛振国⁵, 毛德华⁶, 黄卓¹

1. 北京师范大学 地理科学学部 遥感科学国家重点实验室, 北京 100875;

2. 天津师范大学 地理与环境科学学院, 天津 300387;

3. 生态环境部 卫星环境应用中心, 北京 100094;

4. 南宁师范大学 地理科学与规划学院, 南宁 530001;

5. 中国科学院空天信息创新研究院, 北京 100094;

6. 中国科学院东北地理与农业生态研究所 湿地生态与环境重点实验室, 长春 130102

摘要: 样本数据是开展湿地制图的研究基础之一, 对于数据生产和精度验证具有重要作用。针对湿地生态系统类型多样, 大区域的全湿地类型样本生产困难的问题, 本研究提出了一种准确、高效的大区域密集湿地样本生产框架。该框架主要包括两部分: 首先, 基于已有的湿地数据产品, 使用规则筛选的方法直接生产稳定的湿地样本点, 能够得到河流、湖泊、水库、滨海木本沼泽(红树林)、滩涂的5种湿地类型样本点; 其次, 基于多源专题数据进行规则筛选, 生产潜在湿地样本点, 并利用Google Earth Engine大数据云平台、Google Earth平台和Collect Earth平台进行目视解译, 以确定潜在湿地样本点的类型属性。本文开展大洲尺度的全湿地类型样本生产, 结果表明: 本研究共生产了150688个湿地样本点, 其中内陆湿地样本点为121412个, 滨海湿地样本点为11563个, 人工湿地样本点为17693个。13种湿地类型中, 湖泊样本点占比最大(39.22%), 潟湖样本点占比最小(0.19%)。本文生产了稳定、高质量的湿地样本点, 样本数量充足, 空间分布合理, 能够为湿地分类器的训练和分类结果精度验证提供可靠的数据基础。

关键词: 遥感, 湿地, 样本生产, 多源专题数据, 目视解译, Google Earth Engine, Google Earth, Collect Earth
中图分类号: P2

引用格式: 彭凯锋, 蒋卫国, 侯鹏, 凌子燕, 牛振国, 毛德华, 黄卓. 2024. 结合多源专题数据和目视解译的大区域密集湿地样本数据生产. 遥感学报, 28(2): 334-345

Peng K F, Jiang W G, Hou P, Ling Z Y, Niu Z G, Mao D H and Huang Z. 2024. Dense wetland sample production at large scale by combining multi-source thematic datasets and visual interpretation. National Remote Sensing Bulletin, 28(2): 334-345 [DOI: 10.11834/jrs.20211152]

1 引言

基于遥感卫星的地表覆被制图是开展陆表景观变化和生态资源调查的重要技术手段, 具有成本低、效率高、覆盖范围大等优势(刘涵和宫鹏, 2021)。目前, 已有大量学者利用遥感卫星数据开展了水体制图(Pekel等, 2016; Deng等, 2019; Pickens等, 2020)、农田制图(薛冰等, 2019; Zhang等, 2020)、森林制图(Hansen等, 2013; 颜伟等, 2019)、土

地利用制图(Li等, 2017; Calderón-Loor等, 2021)等研究。根据人工干预的程度, 遥感制图方法大致分为人工解译法、监督分类法、自动分类法等, 其中监督分类方法的分类精度好、制图效率高, 应用最为广泛, 但该方法需要一定数量的样本数据训练分类器。无论采用何种方法, 均需要样本数据对制图结果进行验证, 以确保制图成果能够满足生产应用和科学研究要求(Huang等, 2020)。因此, 精确、高质量的样本数据生产是开展地表

收稿日期: 2021-03-26; 预印本: 2021-10-13

基金项目: 国家自然科学基金(编号: 42071393, 42301413, 41571077); 国家自然科学基金联合基金重点项目(编号: U1901219)

第一作者简介: 彭凯锋, 研究方向为植被生态与湿地遥感。E-mail: pengkaifeng@mail.bnu.edu.cn

通信作者简介: 蒋卫国, 研究方向为生态水文遥感。E-mail: jiangweigu@bnu.edu.cn

覆被制图的基础工作之一。

现阶段，样本数据生产方法大致包括野外实测、无人机调研、高分影像解译和自动生产等方法。野外实测法是指利用野外测量设备（譬如GPS、手持PAD测量仪），到实地进行勘察，在一定大小的样方内确定主导土地覆被类型，记录该地类的地理坐标。该方法大多应用于区域尺度的中、高分影像制图研究（陈劲松等，2014；Jia等，2020）。无人机调研法是指利用低空无人机设备对指定区域进行航空摄影，使用布设的地面控制点对获取图像进行几何校正，通过对其进行目视判读获取样本点。随着无人机设备的技术进步和应用普及，越来越多的研究和生产部门使用无人机进行样本数据生产（田富有等，2019；Sun等，2021）。高分影像解译是指利用米级或亚米级的高空间分辨率遥感影像，譬如GF-2影像（Wu等，2021）、Google Earth的高分影像（张磊等，2019）等，通过目视解译生产样本点。该方法大多用于中等空间分辨率的遥感制图研究（何昭欣等，2019；Xu等，2019）。自动生产法是指基于已有的制图数据集，使用一系列的约束条件筛选稳定、高质量的样本点。譬如，Xie等（2019）使用MCD12Q1的土地利用数据生产样本点，开展全球30 m的土地利用制图。Gong等（2019）基于30 m的土地利用数据样本点，对Sentinel-2影像开展10 m分辨率的全球土地利用制图。

湿地生态系统是位于水陆过渡区的生态系统，其丰富的生态系统服务功能和物质生产力对于人类福祉和生产生活具有重要作用（彭凯锋等，2019）。为此，基于遥感影像的湿地制图成为了当前生态资源管理与保护以及科学研究的热点之一，其中样本数据生产是其制图研究和精度验证的必备基础工作之一。湿地生态系统空间分布较为分散，且大都位于偏远、泥泞的地区（Mahdavi等，2018），野外实测或无人机调研的样本生产成本较高、效率较低，难以满足大尺度的湿地遥感制图研究需求。湿地的类型多样，光谱特征复杂（王鹏等，2017），基于高分影像解译方法难以辨识具有相似光谱和纹理特征的湿地类型，譬如某些草本沼泽和木本沼泽、草本沼泽和洪泛湿地等类型。此外，由于湿地生态系统的复杂性，湿地的定义及其子类型的分类系统尚未形成统一的标准，且现阶段大区域尺度、精细化类别的湿地产品数据较为缺乏

（Hu等，2017；Xu等，2020），基于自动生产法的湿地样本生产难以满足湿地精细化制图的需求。因此，耦合多种样本生产方法，构建一种准确、高效的湿地样本生产方法，对于开展大区域尺度、精细化类别的湿地遥感制图研究具有重要意义。

目前，已有很多学者针对某一湿地类别生产了一系列湿地专题数据集，譬如水体数据集（Pekel等，2016；Han和Niu，2020）、潮间带数据（Murray等，2019）、红树林数据集（Giri等，2011）等，这些数据能够辅助大区域湿地样本生产，极大减少工作量。同时，随着长时序遥感数据集的积累和Google Earth Engine大数据云平台的普及，为样本数据生产提供了良好的数据基础和工具平台。在此背景下，考虑某些湿地类别的易混淆性和湿地专题数据的局限性，本研究基于多平台和多源湿地类型数据，提出一种结合专题数据和目视解译的大区域密集湿地样本生产技术框架，以期开展大尺度湿地遥感制图提供样本数据基础。

2 研究区域与数据

2.1 研究区概况

本文将地理范围为44.04°E—138.25°E、7.44°S—60.0°N的区域定义为研究区，覆盖亚洲的北部、中部和南部区域。根据GLWD（Global Lake and Wetland Database）湿地数据的初步统计（Lehner和Döll，2004），研究区湿地资源丰富，包含湖泊、河流、水库、沼泽地等多种湿地类型，总面积为149.24万km²。该地区经济发展相对落后，当地政府和人们为了获得更多的经济利益，过度开发和使用湿地资源，加之内陆干旱区的湿地受气候变化的影响明显，研究区的湿地退化严重。因此，开展该区域的湿地研究对于全球自然资源保护和可持续发展目标的实现具有重要意义。

2.2 数据收集

为了直接生产和辅助生产湿地样本数据，本研究收集了全湿地类型数据、单一湿地类型数据、土地利用数据及其其他辅助数据等，各数据名称及其介绍如表1所示。JRC-GWS数据、GMW（Global Mangrove Watch）数据、Global Intertidal Change数据为单一湿地类型数据，其中JRC-GWS为全球水体数据，错分误差和漏分误差分别小于1%和5%

(Pekel 等, 2016)。使用 1984 年—2019 年的 JRC-GWS 积水频率产品, 能够可靠的筛选水体样本点和非水体湿地样本点。GMW 为全球红树林数据集, 总体精度为 94%, 本研究选择 2007 年—2016 年间共有 5 个时相的 GMW 产品数据 (Bunting 等, 2018), 将多时相的 GMW 数据进行空间相交, 选

择稳定的红树林区域生产红树林样点。Global Intertidal Change 为全球滨海滩涂数据, 总体精度在 82% 以上 (Murray 等, 2019), 用于生产滨海滩涂样本点。由于滨海滩涂数据将内陆区的部分地物, 错分为了滨海滩涂, 因此, 本研究的滩涂样本点生产仅限于滨海区。

表 1 湿地数据及其他辅助数据的列表
Table 1 List of wetland datasets and other corresponding auxiliary datasets

名称	数据描述	空间分辨率	时间范围	数据用途
JRC-GSW	全球水体数据	30 m	1984 年—2019 年	作为湿地样本区域, 通过积水频率生产水体和非水体样本点
GMW	全球红树林	30 m	2007 年—2016 年	用于生产红树林样本点, 并归类为滨海木本沼泽
Global Intertidal change	全球潮间带	30 m	2014 年—2016 年	用于生产潮间带数据集
GLWD	多湿地类型数据	1 km	1980s	用于筛选草本沼泽和木本沼泽
GlobeLand30	全球土地利用	30 m	2020 年	用于筛选有植被湿地样点、其他内陆湿地样点
HydroLAKES	湖泊与水库数据	1:24000	—	用于筛选湖泊和水库样本点
GRWL	全球河流矢量数据	30 m	—	用于筛选河流样本点
GSHHG-WVS	全球海岸线数据	1:250000	—	用于区分内陆湿地和滨海湿地

注: 表中 1:240000 和 1:250000 表示制图比例尺。

GLWD 为全湿地类型数据 (Lehner 和 Döll, 2004), 该数据的精度和空间分辨率较粗, 主要用于辅助目视解译。GlobeLand30 为土地利用数据, 2020 年的数据总体精度为 85.72% (Jun 等, 2014)。该数据包含了湿地图层, 主要用于筛选有植被湿地和其他内陆湿地样本点。HydroLAKES 是全球湖泊与水库矢量数据, 该数据是耦合多源专题数据和遥感解译产品, 生产了湖泊数据集, 包含了 10 ha 以上的湖泊、水库和其他湖泊类型等 (Messenger 等, 2016), 可用于生产湖泊样本和水库样本。GRWL (Global River Widths from Landsat Database) 是基于 Landsat 影像生产的河网数据集, 记录了河流中心线及其宽度 (Allen 和 Pavelsky, 2018), 可用于筛选河流样本点。GSHHG-WVS 为全球海岸线数据集 (Wessel 和 Smith, 1996), 可用于区分滨海湿地样本和内陆湿地样本。

3 研究方法 with 基础

3.1 研究方法

本文探究国家或大洲尺度的全湿地类型样本点生产技术方法, 基于已有的湿地专题数据集、土地利用数据集等, 并使用 GEE (Google Earth Engine)、Google Earth 和 Collect Earth 等平台工具, 提出一种结合多源数据和目视解译的大尺度密集

湿地样本生产方法, 总体思路如图 1 所示。首先, 使用专题数据生产初始湿地样本点, 即使用全球 JRC 水体数据生产水体样本点和非水湿地体样本点, 使用全球滩涂数据和 GMW 红树林数据分别生产滨海滩涂样本点和红树林样本点。其次, 基于湿地专题数据和 GEE 平台统计的样本点 NDVI 序列, 筛选潜在其他内陆湿地样本和有植被湿地样本。再次, 基于 HydroLAKES 和 GRWL 河流数据等筛选水体样本点, 生产河流、湖泊、水库、内陆水体和滨海水体。最后, 基于 Google Earth 和 Collect Earth 平台对未确认属性的样本点进行人工目视解译。

考虑研究区范围大, 通常的湿地分类研究采用分块迭代的方式开展。本研究按 5°×5° 格网将研究区分为 102 个单元, 设计每个单元的样本数约为 1500 个, 共需要生成约 15 万个湿地类型样本点。对于水体样本点, 本研究使用 80% 积水频率筛选 1984 年—2019 年稳定水体样点, 这些水体样点在 2020 年大都保持属性不变。对于非水体样本点, 大都使用 Google Earth 平台中 2020 年附近的高分影像解译。对于滨海滩涂样点, 使用 2014 年—2016 年的数据集近似表征 2020 年的滨海滩涂样点。对于红树林样本点, 使用 1996 年、2007 年、2008 年、2008 年、2010 年、2015 年、2016 年的数据集进行相交, 选择稳定的红树林区域作为样点区, 因此

红树林样本点时相也可以为2020年。综上所述，本研究生产的全湿地类型样本点的时相为2020年，

生成的样本点为稳定、高质量的样本点，可用于开展中等空间分辨率（10—30 m）的湿地制图研究。

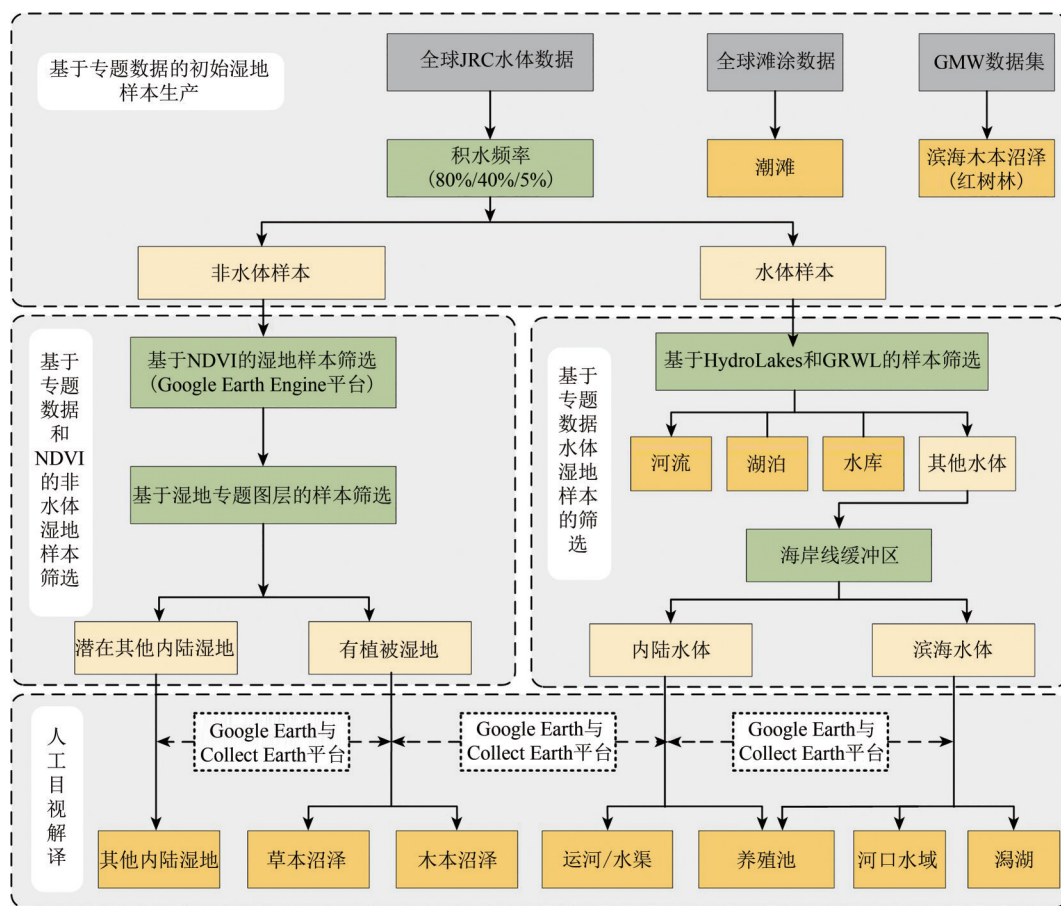


图1 湿地样本数据生产的技术流程图

Fig. 1 The flowchart of wetland samples production

3.2 湿地分类体系

由于湿地群落的复杂性和季节变化特征，湿地分类系统及其类型定义尚未形成统一的规范。为了使得生产的湿地数据具有较高的一致性，相关学者积极开展标准湿地分类体系的研究，譬如CLC (CORINE Land Cover) 分类系统、LCCS FAO (FAO Land Cover Classification System)、MAES (Mapping and Assessment of Ecosystem Service classification system)、基于LCCS的湿地分类系统等 (Finlayson, 2018; Weise 等, 2020; Xu 等, 2020)。本研究考虑湿地类型在遥感影像的可分性和湿地资源管理的需求，采用Mao等 (2020) 提出的湿地分类体系。同时，考虑其他内陆湿地（譬如河漫滩、藓类湿地、河心洲等）的重要性以及与其他湿地类型的可分性，本研究将该类型加入湿地分类体系中。本研究的湿地分类体系如表2所示。

3.3 工具平台介绍

本研究使用GEE (Google Earth Engine) 云平台辅助湿地样本数据生产。GEE是由美国谷歌公司开发的遥感大数据平台，不仅存储了Landsat、MODIS、Sentinel等遥感影像以及其他地理空间数据，同时也具有强大的数据分析和计算能力。本研究基于GEE平台，统计样本点在MOD13Q1的NDVI时间序列影像的像元值，时间范围为2001年—2020年，每个样本点对应457个NDVI值。分析样本点的NDVI时间序列变化，可以可靠的筛选出有植被湿地和潜在其他内陆湿地。

本研究使用Google Earth和Collect Earth软件开展样本点目视解译工作。Google Earth软件是开展样本数据生产的常用工具平台，可以为用户提供高、中、低等多种分辨率影像，但该平台样本管理功能不够友好。为此，联合国粮农组织 (FAO)

开发了 Collect Earth 软件，可以方便的对样本数据进行管理和分析，同时该软件能够轻松访问 Bing Maps、Google Earth Engine 的影像，获取每个样本点的空间信息和时间序列信息。结合 Google Earth 和 Collect Earth 平台的样本采集，能够提升样本生产的精度与效率。生产过程示意图如图 2 所示。

表 2 湿地分类体系及其类型定义

Table 2 Wetland classification system and its sub-type definition		
一级	二级	定义
内陆湿地	内陆木本沼泽	以林木植被为主的内陆自然沼泽,包括森林沼泽、灌丛沼泽
	内陆草本沼泽	以草本植被为主的内陆自然沼泽,包括沼泽地草甸
	河流	内陆地区的流动水体,呈线性分布,常年有水体淹没
	湖泊	内陆地区的汇水区,呈面状分布,常年有水体淹没
	其他内陆湿地	植被覆盖度低于 10%,指河流和湖泊周围漫滩、河流冲积扇、藓类湿地等
滨海湿地	滨海木本沼泽	以林木植被为主的滨海自然湿地,包括滨海森林沼泽、滨海灌木沼泽、红树林等
	滨海草本沼泽	以草本植被为主的滨海自然沼泽
	滨海滩涂	潮间带区域的植被覆盖度低于 10% 的区域,包括沙滩、岸岩等
	潟湖	由岛屿或礁石包围的滨海浅水水体,但至少有一个出水口与滨海区相连
人工湿地	河口水域	由内陆边界流向滨海水域的自然河流水体
	水库	有人工修筑边界的汇水区,常常有大坝分布
	运河/水渠	有明显大坝或堤岸的流动水体,呈线性分布
	盐田/养殖池	有规则形状的小斑块水体,用于水产养殖,大多位于河流或滨海区附近

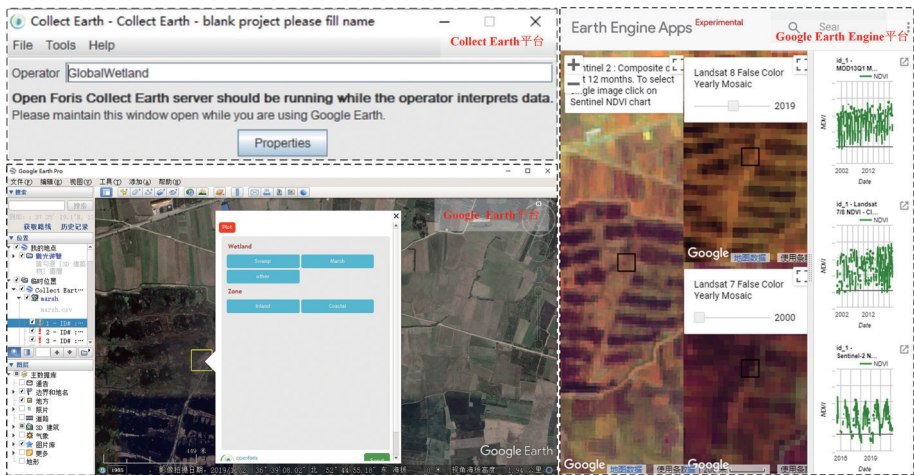


图 2 结合 Google Earth 和 Collect Earth 平台的样本采集示意图
Fig. 2 Sample collection diagram combining Google Earth and Collect Earth software

4 样本生产过程与结果

4.1 水体类型的湿地样本数据生产

分析湿地类型的定义发现，河流、湖泊、水库、河口水域、潟湖、运河/水渠、人工养殖池等均属于水体类型。本小节基于 JRC 水体数据、HydroLAKES 数据、GRWL 数据和 GSHHG-WVS 数据，通过规则筛选和目视解译的方式生产水体类型样本点。首先，基于 JRC 水体数据生产水体样本点：将 JRC 积水频率大于 80% 的区域作为稳定水体区域，并在此区域内生成随机点，样本之间的最小

采样距离为 300 m，共生成了 124839 个水体样本点，如图 3 (a) 所示。其次，基于 HydroLAKES 数据生产湖泊、水库样本点，该数据包含了 10 ha 以上湖泊、水库和其他湖泊类型。将 80% 的积水频率边界与 HydroLAKES 数据的湖泊和水库图层进行相交，从而筛选出湖泊和水库的稳定水体边界。由于湖泊图层包含了自然湖泊和小面积的水库/坑塘，本文将该图层面积小于 500 ha 的斑块删除，仅保留大面积湖泊。使用该数据，能够生成湖泊样本点和水库样本点，共生产湖泊样本点 59090 个、水库样本点 14854 个，如图 3 (b) 所示。然后，

使用GRWL河网数据生产河流样本点。将GRWL的河流中线数据与80%积水频率边界进行相交,筛选河流边界,同时参照水库数据,剔除河流中的水库边界。使用该数据筛选水体样本,共生成河流样本25503个,如图3(c)所示。再次,使用GSHHG-WVS海岸线数据建立15 km缓冲区,使用该矢量数据对剩余样本点进行分类,生成内陆水体样本点16230个,滨海水体样本点9192个,如图3(d)所示。最后,使用Google Earth和Collect

Earth软件,对滨海水体进行目视解译,生产河口水域、潟湖、盐田/养殖池的样本;对内陆水体进行目视解译,生产养殖池、运河/水渠样本。河口水域是位于滨海区的河流水体,与浅海水域相连;潟湖是位于滨海区的湖泊,呈面状分布,至少有一个出水口与滨海水体相连;盐田/养殖池形状规则,面积较小,大都位于河流和滨海区附近;运河/水渠呈线性分布,宽度较窄,大都有修筑的人工护岸设施。

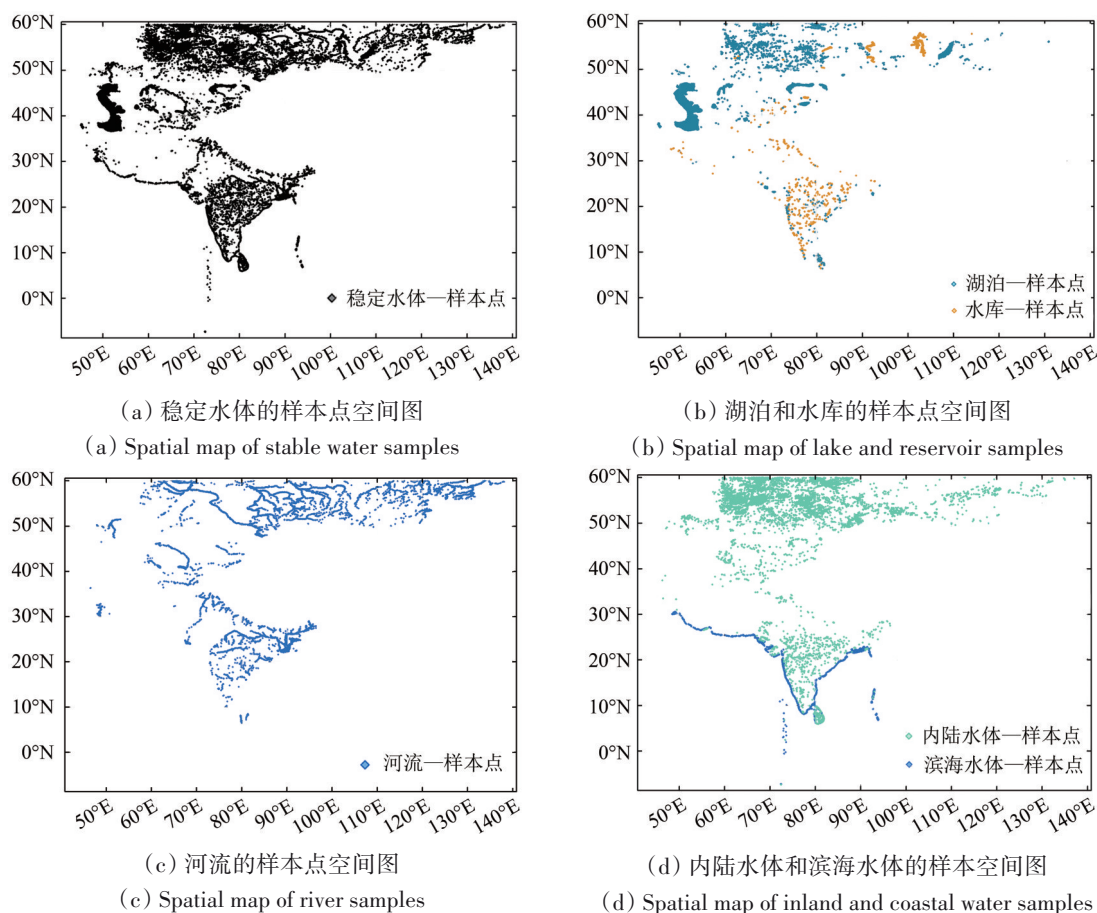


图3 水体类型的湿地样本点空间图

Fig. 3 Spatial maps of wetland samples of water types

4.2 其他内陆湿地的样本数据生产

其他内陆湿地是指植被覆盖度低于10%的区域,通常为河流和湖泊周围漫滩、河流冲积扇、藓类湿地等。考虑其他内陆湿地的植被特征和地理特征,本小节使用JRC水体数据、3.1小节的河流与湖泊边界数据、GlobeLand30数据以及GEE平台统计的NDVI时间序列,并结合目视解译的方式,生产其他内陆湿地样本点。首先,使用JRC数据生产种子样本点:将积水频率为5%—40%之

间的区域作为潜在非水体湿地区,以最小间隔距离为300 m生产种子样本点。其次,基于NDVI时间序列筛选潜在在其他内陆湿地样本点:考虑其他内陆湿地的低植被特征以及其NDVI数值的异常波动,将满足“ $NDVI > 0.12$ 的次数大于200或 $NDVI > 0.2$ 的次数大于100”样本点剔除,剩余点为潜在在其他内陆湿地样本点,共计13747个。再次,使用GlobeLand30数据筛选其他潜在在内陆湿地样本点:该数据的湿地图层包含了洪泛湿地、草本沼泽、

本沼泽等类型,使用该图层与潜在在其他内陆湿地样本点相交,进一步筛选潜在在其他内陆湿地样本点。然后,使用河流和湖泊最大水体边界对潜在在其他潜在在内陆湿地样本点进行筛选,从而得到河流与湖泊周边的潜在在其他湿地样本点。经过以上过程,并删除滨海区的潜在在其他内陆湿地点,生

产了9884个潜在在其他内陆湿地样本点,如图4(a)所示。最后,通过目视解译的方式,对筛选的潜在在其他内陆湿地进行解译,确认为其他内陆湿地样本点共8410个。目视解译过程中,对稀疏样点区或典型区也进行样本增选,共876个。最终的样本点如图4(b)所示。

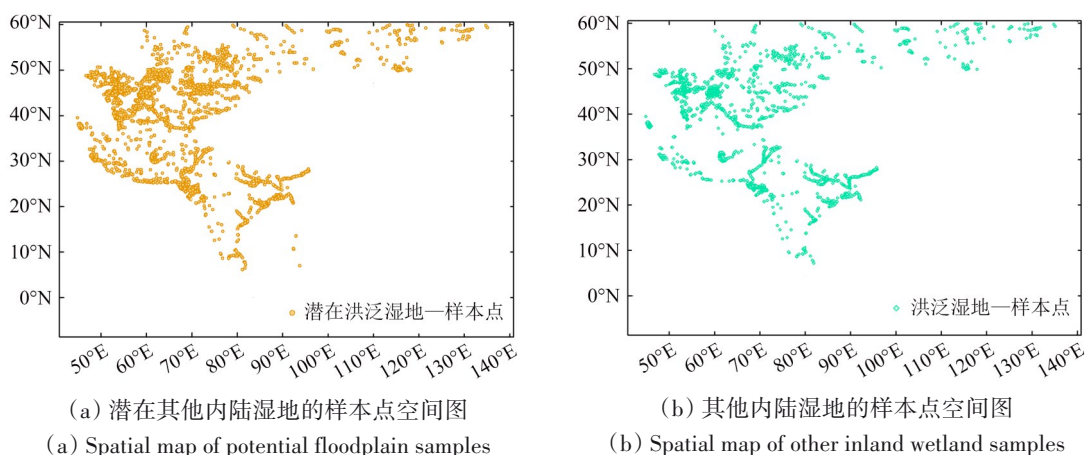


图4 潜在在其他内陆湿地和其他湿地的样本点空间图

Fig. 4 Spatial maps of potential other inland wetland samples and potential other inland wetland samples

4.3 有植被湿地的样本数据生产

分析湿地分类体系可知,有植被湿地包括内陆/滨海草本沼泽、内陆/滨海木本沼泽,这些湿地类型大多是由季节性水体淹没的植被。本小结探究结合湿地专题数据和人工目视解译的方式生产有植被湿地样本。首先,基于JRC数据生产潜在种子样本点:在积水频率为5%—40%的区域生成样本点,样本点的最小间隔为300 m。其次,基于

NDVI时间序列筛选潜在有植被湿地样本点:将样本点的NDVI>0.2次数超过200次的样本点,作为潜在湿地样本点,共计34145个,如图5(a)所示。再次,使用GlobeLand30数据进一步筛选有植被湿地样本点。根据GlobeLand30的湿地图层定义,该图层包含沼泽地、红树林、森林/灌丛湿地等。本研究使用该数据对潜在有植被湿地进行筛选,生成有植被湿地样本点,共计8164个,如图5(b)。

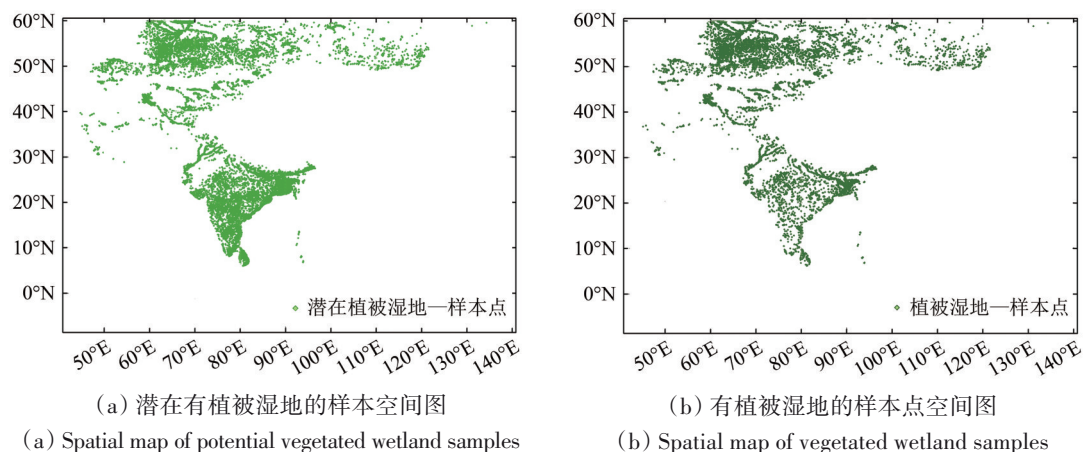


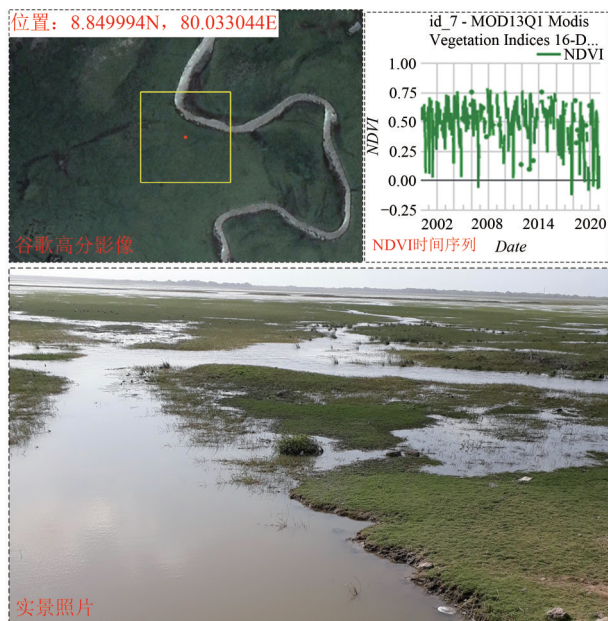
图5 潜在有植被湿地和有植被湿地的样本空间图

Fig. 5 Spatial maps of potential vegetated wetland samples and vegetated wetland samples

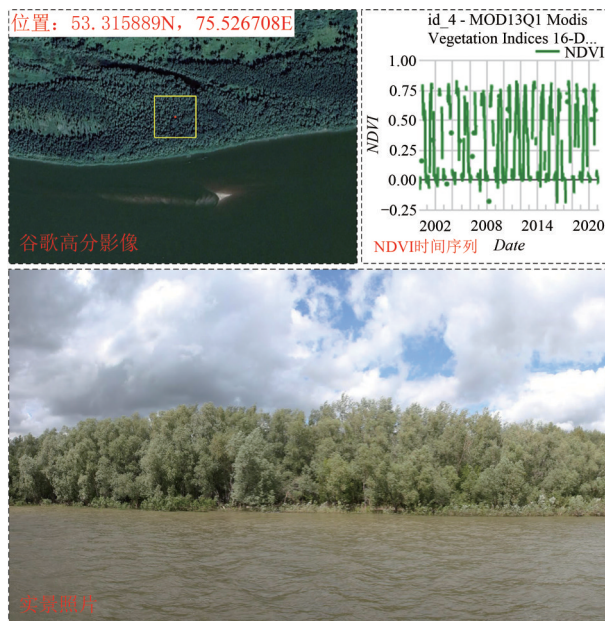
将以上有植被湿地样本点导入Collect Earth平台中,进行目视解译判读木本沼泽和草本沼泽。由于部分草本沼泽和木本沼泽具有相似光谱特征和NDVI时间序列特征,本研究将有植被湿地样点分两部分进行解译。首先,使用GLWD的“淡水草本沼泽/洪泛湿地”图层对有植被湿地进行筛选,选择的样本点大都为草本沼泽类型,导入到Collect Earth平台进行目视解译。这种方式解译,能够对草本沼泽和木本沼泽提供先期认知。其次,将剩余有植被湿地样点导入到Collect Earth平台,进行目视解译。

通过有植被湿地的样本解译,总结了草本沼

泽和木本沼泽的颜色、纹理和NDVI时间序列特征。对于草本沼泽:(1)颜色:生长季为青绿色、浅青色;非生长季为灰色、红褐色;(2)纹理:纹理较为匀质;(3)NDVI时间序列特征:NDVI数值在0—0.5间周期变化,但部分草本沼泽样本点的NDVI在0—0.8间周期变化。草本沼泽的高分影像、NDVI变化图与实景照片如图6(a)所示。对于木本沼泽:(1)颜色在生长季为绿色、青绿色;非生长季为褐色、红褐色;(2)纹理:纹理粗糙,有颗粒状;(3)NDVI时间序列特征:NDVI数值在0—0.8之间震荡。木本沼泽的高分影像、NDVI变化图与实景照片如图6(b)所示。



(a) 草本沼泽的样本点示意图
(a) The diagram of marsh sample



(b) 木本沼泽的样本点示意图
(b) The diagram of swamp sample

图6 草本沼泽和木本沼泽的样本示意图

Fig. 6 The diagram of marsh sample and swamp sample

基于以上过程,完成对草本沼泽和木本沼泽的目视解译。同时对稀疏样点区或典型湿地地区,在Google Earth平台经进行样本增选,并使用海岸线缓冲区进行区分内陆和滨海的湿地类型。共生产内陆草本沼泽16747个,内陆木本沼泽10786个,滨海草本沼泽701个,滨海木本沼泽378个,如图7(a)所示。此外,本研究使用GMW红树林数据生产红树林样本点。为了保证样本的准确性和稳定性,本研究使用2007年、2008年、2010年、2015年、2016年的5个时相数据生产样本点。将这些数据进行空间相交,在公共相交区域内生成红树林样本

点,共生成红树林样本点5000个,如图7(b)所示。将生产的红树林样本点归为滨海木本沼泽。

4.4 滩涂的样本数据生产

滩涂是指潮间带地区的植被覆盖度低于10%的区域,包括沙滩、岸岩等。本研究使用Murray等(2019)生产的全球滩涂数据集生产本研究区的滩涂样本点。该数据是基于长序密集的Landsat影像生产的,空间分辨率为30 m,数据的总体精度为82.3%。本研究使用2014年—2016年时期的数据产品生产样本点,样本点的最小间隔为300 m,共生产了2500个样本点,如图8所示。

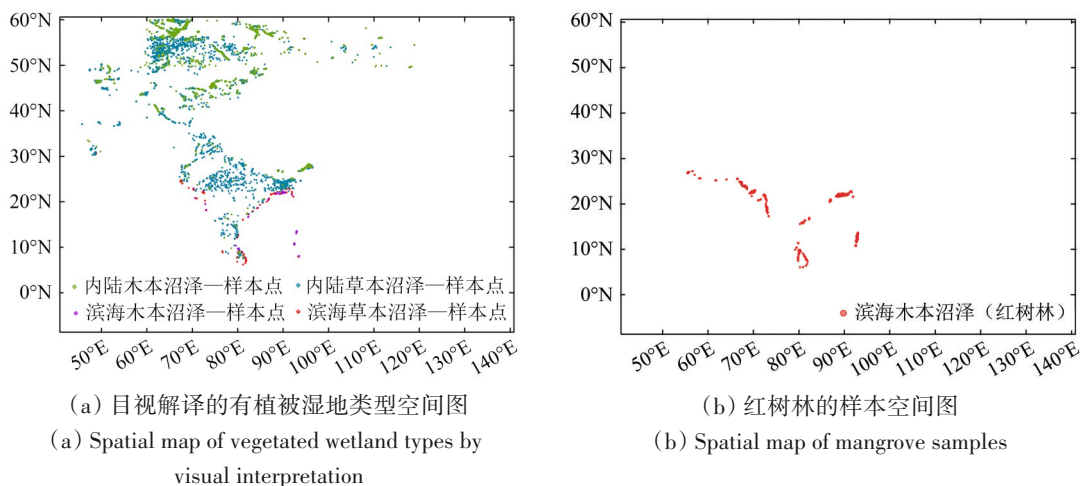


图7 有植被湿地类型的样本空间图
Fig. 7 Spatial map of vegetated wetland samples

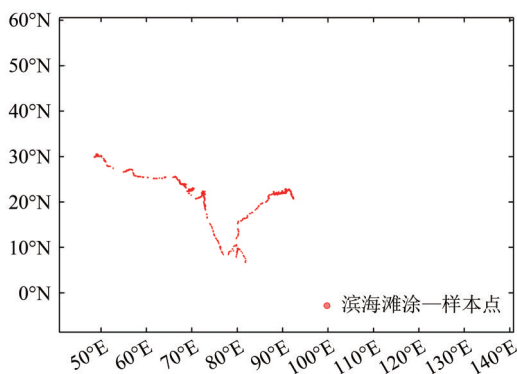


图8 滨海滩涂样本点的空间图
Fig. 8 Spatial map of coastal tidal flat samples

5 讨论

现阶段, 湿地制图研究大多侧重于湿地遥感分类研究, 而将湿地样本点生产作为研究基础之一。该研究范式适用于小尺度湿地遥感研究, 但对于大尺度的湿地制图研究, 如何准确、高效地生产高质量的样本点, 是保证其研究结果准确性的核心工作之一。本研究系统梳理了国内外的湿地分类体系, 选择了适用于大区域尺度湿地遥感制图的分类系统; 然后基于国内外已有的湿地数据、土地利用数据以及其他数据, 针对性的提出适用于大尺度的密集湿地样本生产方法。该方法充分利用了已有的数据基础, 生产了河流、湖泊、水库、滨海木本沼泽(红树林)、滩涂等多个湿地类型样本, 同时筛选了有植被湿地、其他水体类型和其他内陆湿地等样本点, 能够在保证样本精度和质量的前提下, 极大减少人工目视解译

的工作量。本研究可为开展湿地遥感制图的样本生产提供一定的技术参考, 对于开展大区域的湿地资源调查具有重要意义。

本研究生产了大洲尺度的全湿地类型样本点, 但仍存在有待改进的地方。首先, 本研究将基于规则筛选和目视解译的样本点全部作为湿地类型样本点, 未考虑湿地类型样本均衡性的问题。未来的湿地分类研究中, 可根据算法要求和各湿地类型占比, 选择合适比例的湿地子类型样本点。其次, 缺少对生产湿地样本的精度验证。尽管本文生产了可靠、稳定的湿地类型样本, 但样本生产过程中难免存在误差。在未来的研究中, 可选择典型区的样本点进行二次人工目视解译, 以验证样本的可靠性和精度。

此外, 本研究的湿地样本生产方法也存在一定的不确定性。首先, 人工目视解译难以判别某些草本沼泽和木本沼泽, 某些草本沼泽表现为高植被特征, 纹理和颜色特征与木本沼泽相似, 从而导致样本解译中存在一些人工误差。其次, 本文仅使用JRC水体数据来框定大多数湿地类型样本点, 未能在水体淹没最大范围以外采集样本点。最后, 由于湿地专题数据本身存在一定的数据误差, 基于专题数据生产的湿地样本同样存在一定的不确定性。然而, 对于大尺度遥感分类研究, 其样本点不一定要全部覆盖到所有湿地类别区, 只要样本满足均匀分类、数量合理即可。同时, Gong等(2019)的研究表明, 样本数量减少至原样本量的60%, 其分类精度仅下降1%左右, 一定程度的样本生产误差对分类研究的影响是可接受

的。综上所述,本文湿地样本生产的技术框架,能够满足大尺度湿地遥感制图的要求,是一种准确、高效的湿地样本生产方法。

6 结 论

本研究以多源湿地专题数据为基础,以 Google Earth Engine 大数据云平台、Google Earth 和 Collect Earth 软件等为工具平台,提出了一种准确、高效的大尺度密集湿地样本生产方法。相比于传统的目视解译方法和自动样本生产方法,本研究的样本方案能够充分利用已有的数据基础,有效避免盲目目视解译导致的工作量大、效率低的问题,能够为大尺度精细化湿地制图提供可靠密集的全湿地类型样本数据。

本文开展了大洲尺度的全湿地类型样本数据生产。研究结果表明:本研究共生产了 150668 个湿地样本点,其中内陆湿地样本有 121412 个,滨海湿地样本有 11563 个,人工湿地样本有 17693 个。13 个湿地子类型中,湖泊样本数量占比最大 (39.22%),潟湖样本数量占比最小 (0.19%)。同时,河流、水库、内陆草本沼泽和内陆木本沼泽的样本占比也较大。本研究生产了精度可靠、数量充足的全湿地类型样本点,能够为大洲尺度的湿地制图和精度验证提供可靠的数据基础。

此外,本文的样本生产方法及其样本生产结果仍存在有待改进的地方。首先,本文将规则筛选和目视解译的湿地样本全部作为样本集,未充分考虑湿地类型的样本均衡问题。未来的湿地分类研究中,应根据算法要求和湿地类型占比,选择合适比例和数量的样本点。其次,对于某些区域的草本沼泽和木本沼泽,其相似的颜色和纹理特征导致样本解译存在一些人工粗差。未来的湿地样本解译,可加强对解译人员的训练,以增强其对湿地类型的认知。最后,多源辅助数据本身存在一定的误差,基于其生产的样本点存在一定的不确定性。未来的样本生产,可以通过多期数据叠加生产稳定样本、二次人工目视解译等方式,提高样本点质量与精度。

志 谢 感谢清华大学宫鹏教授领导全球湿地制图研究,提供全球湿地制图的科学思想。感谢中国科学院东北地理与农业生态研究所王宗明研究员,对本文的湿地分类体系的确定和湿地专

题数据的收集提供的帮助和指导。感谢中国地图出版集团左伟编审对地图制图提供的帮助和指导。

参考文献(References)

- Allen G H and Pavelsky T M. 2018. Global extent of rivers and streams. *Science*, 361(6402): 585-588 [DOI: 10.1126/science.aat0636]
- Bunting P, Rosenqvist A, Lucas R, Rebelo L M, Hilarides L, Thomas N, Hardy A, Itoh T, Shimada M and Finlayson C M. 2018. The global mangrove watch – a new 2010 global baseline of mangrove extent. *Remote Sensing*, 10(10): 1669 [DOI: 10.3390/rs10101669]
- Calderón-Loor M, Hadjikakou M and Bryan B A. 2021. High-resolution wall-to-wall land-cover mapping and land change assessment for Australia from 1985 to 2015. *Remote Sensing of Environment*, 252: 112148 [DOI: 10.1016/j.rse.2020.112148]
- Chen J S, Han Y, Chen G and Zhang J. 2014. Land utilization mapping in Guangdong Province based on integration of optical and SAR remote sensing data. *Acta Ecologica Sinica*, 34(24): 7233-7242 (陈劲松, 韩宇, 陈工, 张瑾. 2014. 基于多源遥感信息融合的广东省土地利用分类方法——以雷州半岛为例. *生态学报*, 34(24): 7233-7242) [DOI: 10.5846/stxb201310162499]
- Deng Y, Jiang W G, Tang Z H, Ling Z Y and Wu Z F. 2019. Long-term changes of open-surface water bodies in the Yangtze River Basin based on the Google Earth Engine cloud platform. *Remote Sensing*, 11(19): 2213 [DOI: 10.3390/rs11192213]
- Finlayson C M. 2018. Ramsar convention typology of wetlands//Finlayson C M, Everard M, Irvine K, McInnes R J, Middleton B A, van Dam A S and Davidson N C, eds. *The Wetland Book: I: Structure and Function, Management, and Methods*. Dordrecht: Springer: 1529-1532 [DOI: 10.1007/978-90-481-9659-3_339]
- Giri C, Ochieng E, Tieszen L L, Zhu Z, Singh A, Loveland T, Masek J and Duke N. 2011. Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography*, 20(1): 154-159 [DOI: 10.1111/j.1466-8238.2010.00584.x]
- Gong P, Liu H, Zhang M N, Li C C, Wang J, Huang H B, Clinton N, Ji L Y, Li W Y, Bai Y Q, Chen B, Xu B, Zhu Z L, Yuan C, Suen H P, Guo J, Xu N, Li W J, Zhao Y Y, Yang J, Yu C Q, Wang X, Fu H H, Yu L, Dronova I, Hui F M, Cheng X, Shi X L, Xiao F J, Liu Q F and Song L C. 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Science Bulletin*, 64(6): 370-373 [DOI: 10.1016/j.scib.2019.03.002]
- Han Q Q and Niu Z G. 2020. Construction of the long-term global surface water extent dataset based on water-NDVI spatio-temporal parameter set. *Remote Sensing*, 12(17): 2675 [DOI: 10.3390/rs12172675]
- Hansen M C, Potapov P V, Moore R, Hancher M, Turubanova S A, Tyukavina A, Thau D, Stehman S V, Goetz S J, Loveland T R, Kommareddy A, Egorov A, Chini L, Justice C O and Townshend J R G. 2013. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160): 850-853 [DOI: 10.1126/science.1244693]

- He Z X, Zhang M, Wu B F and Xing Q. 2019. Extraction of summer crop in Jiangsu based on Google Earth Engine. *Journal of Geo-Information Science*, 21(5): 752-766 (何昭欣, 张森, 吴炳方, 邢强. 2019. Google Earth Engine支持下的江苏省夏收作物遥感提取. *地球信息科学学报*, 21(5): 752-766)
- Hu S J, Niu Z G and Chen Y F. 2017. Global wetland datasets: a review. *Wetlands*, 37(5): 807-817 [DOI: 10.1007/s13157-017-0927-z]
- Huang H B, Wang J, Liu C X, Liang L, Li C C and Gong P. 2020. The migration of training samples towards dynamic global land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 161: 27-36 [DOI: 10.1016/j.isprsjprs.2020.01.010]
- Jia M M, Mao D H, Wang Z M, Ren C Y, Zhu Q D, Li X C and Zhang Y Z. 2020. Tracking long-term floodplain wetland changes: a case study in the China side of the Amur River Basin. *International Journal of Applied Earth Observation and Geoinformation*, 92: 102185 [DOI: 10.1016/j.jag.2020.102185]
- Jun C, Ban Y F and Li S N. 2014. Open access to earth land-cover map. *Nature*, 514(7523): 434-434 [DOI: 10.1038/514434c]
- Lehner B and Döll P. 2004. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, 296(1/4): 1-22 [DOI: 10.1016/j.jhydrol.2004.03.028]
- Li C C, Gong P, Wang J, Zhu Z L, Biging G S, Yuan C, Hu T Y, Zhang H Y, Wang Q, Li X C, Liu X X, Xu Y D, Guo J, Liu C X, Hackman K O, Zhang M N, Cheng Y Q, Yu L, Yang J, Huang H B and Clinton N. 2017. The first all-season sample set for mapping global land cover with Landsat-8 data. *Science Bulletin*, 62(7): 508-515 [DOI: 10.1016/j.scib.2017.03.011]
- Liu H and Gong P. 2021. 21st century daily seamless data cube reconstruction and seasonal to annual land cover and land use dynamics mapping-iMap (China) 1.0. *National Remote Sensing Bulletin*, 25(1): 126-147 (刘涵, 宫鹏. 2021. 21世纪逐日无缝数据立方体构建方法及逐年逐季土地覆盖和土地利用动态制图—中国智慧遥感制图iMap (China) 1.0. *遥感学报*, 25(1): 126-147)
- Mahdavi S, Salehi B, Granger J, Amani M, Brisco B and Huang W M. 2018. Remote sensing for wetland classification: a comprehensive review. *GIScience and Remote Sensing*, 55(5): 623-658 [DOI: 10.1080/15481603.2017.1419602]
- Mao D H, Wang Z M, Du B J, Li L, Tian Y L, Jia M M, Zeng Y, Song K S, Jiang M and Wang Y Q. 2020. National wetland mapping in China: a new product resulting from object-based and hierarchical classification of Landsat 8 OLI images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 164: 11-25 [DOI: 10.1016/j.isprsjprs.2020.03.020]
- Messenger M L, Lehner B, Grill G, Nedeva I and Schmitt O. 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7(1): 13603 [DOI: 10.1038/ncomms13603]
- Murray N J, Phinn S R, DeWitt M, Ferrari R, Johnston R, Lyons M B, Clinton N, Thau D and Fuller R A. 2019. The global distribution and trajectory of tidal flats. *Nature*, 565(7738): 222-225 [DOI: 10.1038/s41586-018-0805-8]
- Pekel J F, Cottam A, Gorelick N and Belward A S. 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633): 418-422 [DOI: 10.1038/nature20584]
- Peng K F, Jiang W G and Deng Y. 2019. Identification of wetland damage degree and analysis of its driving forces in Wuhan urban agglomeration. *Journal of Natural Resources*, 34(8): 1694-1707 (彭凯锋, 蒋卫国, 邓越. 2019. 武汉城市圈湿地受损程度识别及驱动因素分析. *自然资源学报*, 34(8): 1694-1707) [DOI: 10.31497/zrzyxb.20190810]
- Pickens A H, Hansen M C, Hancher M, Stehman S V, Tyukavina A, Potapov P, Marroquin B and Sherani Z. 2020. Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series. *Remote Sensing of Environment*, 243: 111792 [DOI: 10.1016/j.rse.2020.111792]
- Sun C, Li J L, Liu Y X, Liu Y C and Liu R Q. 2021. Plant species classification in salt marshes using phenological parameters derived from Sentinel-2 pixel-differential time-series. *Remote Sensing of Environment*, 256: 112320 [DOI: 10.1016/j.rse.2021.112320]
- Tian F Y, Wu B F, Zeng H W, He Z X, Zhang M and Bofana J. 2019. Identifying soybean cropped area with Sentinel-2 data and multi-layer neural network. *Journal of Geo-Information Science*, 21(6): 918-927 (田富有, 吴炳方, 曾红伟, 何昭欣, 张森, Bofana J. 2019. 基于多层神经网络与 Sentinel-2 数据的大豆种植区识别方法. *地球信息科学学报*, 21(6): 918-927)
- Wang P, Wan R R and Yang G S. 2017. Advance in classification and biomass estimation of plants in wetlands based on multi-source remote sensing data. *Wetland Science*, 15(1): 114-124 (王鹏, 万荣荣, 杨桂山. 2017. 基于多源遥感数据的湿地植物分类和生物量反演研究进展. *湿地科学*, 15(1): 114-124) [DOI: 10.13248/j.cnki.wetlandsci.2017.01.017]
- Weise K, Höfer R, Franke J, Guelmami A, Simonson W, Muro J, O'Connor B, Strauch A, Flink S, Eberle J, Mino E, Thulin S, Philipson P, Van Valkengoed E, Truckenbrodt J, Zander F, Sánchez A, Schröder C, Thonfeld F, Fitoka E, Scott E, Ling M, Schwarz M, Kunz I, Thürmer G, Plasmeijer A and Hilarides L. 2020. Wetland extent tools for SDG 6.6.1 reporting from the Satellite-based wetland observation service (SWOS). *Remote Sensing of Environment*, 247: 111892 [DOI: 10.1016/j.rse.2020.111892]
- Wessel P and Smith W H F. 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. *Journal of Geophysical Research: Solid Earth*, 101(B4): 8741-8743 [DOI: 10.1029/96JB00104]
- Wu N, Shi R H, Zhuo W, Zhang C, Zhou B C, Xia Z L, Tao Z, Gao W and Tian B. 2021. A classification of tidal flat wetland vegetation combining phenological features with Google Earth Engine. *Remote Sensing*, 13(3): 443 [DOI: 10.3390/rs13030443]
- Xie S, Liu L Y, Zhang X, Yang J N, Chen X D and Gao Y. 2019. Automatic land-cover mapping using Landsat time-series data based on Google Earth Engine. *Remote Sensing*, 11(24): 3023 [DOI: 10.3390/rs11243023]
- Xu H Z Y, Wei Y C, Liu C, Li X and Fang H. 2019. A scheme for the long-term monitoring of impervious-relevant land disturbances using high frequency Landsat archives and the Google Earth Engine. *Remote Sensing*, 11(16): 1891 [DOI: 10.3390/rs11161891]
- Xu P P, Herold M, Tsensbazar N E and Clevers J G P W. 2020. Towards a comprehensive and consistent global aquatic land cover characterization framework addressing multiple user needs. *Remote Sensing of Environment*, 250: 112034 [DOI: 10.1016/j.rse.2020.112034]
- Xue B, Xiao X, Li J Z, Xie X and Lu C P. 2019. Spatio-temporal evolution of water areas and croplands in the three provinces of

- Northeast China based on remote sensing data. *Chinese Journal of Ecology*, 38(5): 1444-1452 (薛冰, 肖骁, 李京忠, 谢潇, 逯承鹏. 2019. 基于遥感数据的东北三省水域与农田用地时空演变. *生态学杂志*, 38(5): 1444-1452) [DOI: 10.13292/j.1000-4890.201905.002]
- Yan W, Zhou W, Yi L L and Tian X. 2019. Research progress of remote sensing classification and change monitoring on forest types. *Remote Sensing Technology and Application*, 34(3): 445-454 (颜伟, 周雯, 易利龙, 田昕. 2019. 森林类型遥感分类及变化监测研究进展. *遥感技术与应用*, 34(3): 445-454) [DOI: 10.11873/j.issn.1004-0323.2019.3.0445]
- Zhang D J, Pan Y Z, Zhang J S, Hu T G, Zhao J H, Li N and Chen Q. 2020. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. *Remote Sensing of Environment*, 247: 111912 [DOI: 10.1016/j.rse.2020.111912]
- Zhang L, Gong Z N, Wang Q W, Jin D D and Wang X. 2019. Wetland mapping of yellow river delta wetlands based on multi-feature optimization of Sentinel-2 images. *Journal of Remote Sensing*, 23(2): 313-326 (张磊, 宫兆宁, 王启为, 金点点, 汪星. 2019. Sentinel-2 影像多特征优选的黄河三角洲湿地信息提取. *遥感学报*, 23(2): 313-326) [DOI: 10.11834/jrs.20198083]

Dense wetland sample production at large scale by combining multi-source thematic datasets and visual interpretation

PENG Kaifeng^{1,2}, JIANG Weiguo¹, HOU Peng³, LING Ziyan^{1,4}, NIU Zhenguo⁵,
MAO Dehua⁶, HUANG Zhuo¹

1. State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China;
2. School of Geographical and Environmental Science, Tianjin Normal University, Tianjin 300387, China;
3. Satellite Environment Center, Ministry of Ecology and Environment, Beijing 100094, China;
4. School of Geography and Planning, Nanning Normal University, Nanning 530001, China;
5. Aerospace Information Research Institute, China Academy of Sciences, Beijing 100094, China;
6. Key Laboratory of Wetland Ecology and Environment, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun 130102, China

Abstract: Sample collection is one of key research foundations for wetland mapping. It plays an important role in classifier training and accuracy validation. Generally, wetland samples are produced by visual interpretation based on high-spatial-resolution images or automatic generation based on multi-source existing dataset. The visual interpretation is time and labor consuming and cannot meet the demand for large-scale wetland classification. The automatic sample-generation method is unsuitable to detailed-type wetland mapping due to the diversity of wetlands and classification-scheme inconsistency of existing wetland datasets. Thus, an efficient and accurate sampling method is in demand for large-scale and detailed-type wetland mapping.

In our study, we collected a series of auxiliary datasets and developed an efficient solution for continental-scale wetland sample generation by combining automatic sampling method and visual interpretation. In the first part, the samples of five wetland types can be automatically generated by rule filtering based on multi-source existing datasets. River, lake, and reservoir samples were created using the JRC Global Surface Water, Global River Widths from Landsat and HydroLAKES datasets. Coastal swamp (mangrove) samples were produced by using Global Mangrove Watch dataset. Tidal flat samples were generated using the Global Intertidal Change dataset. In the second part, by combining time series of MODIS NDVI images and existing auxiliary datasets, we first produced potential wetland samples for coarse wetland types (i.e., vegetated wetland samples and inundated wetland samples). Then, we identified them by visual interpretation based on the Google Earth Engine platform, Google Earth software, and Collect Earth software. We applied our sample method in our study area, and produced continental-scale and detailed-type wetland samples.

Results indicated that the total wetland samples in our study area was 150688, among which 141412 points were inland wetland samples, 11563 were coastal wetland samples, and 17693 were human-made wetland samples. Among the 13 wetland sub-categories, lake accounted for the largest proportion (39.22%) and primarily distributed the northern and central of study area, whereas lagoon accounted for the smallest proportion (0.19%), mostly scattered in coastal region of the study area. Samples of river, reservoir, inland swamp, and inland marsh also shared a considerable amount, accounting for 16.93%, 9.86%, 7.16%, and 11.12% of total wetland samples, respectively. River samples were primarily distributed north and south of the study area, and reservoir samples were primarily scattered south of the study area. Meanwhile, inland swamp and inland marsh samples were mostly distributed northwest and south of the study area.

This study successfully produced stable and high-quality wetland samples at continental scale. The generated samples shared sufficient quantities and reasonable spatial distribution, which can lay a good foundation for classifier training and accuracy validation. Meanwhile, by combining the multi-source thematic datasets and multiple platform, our designed sample solution can make full use of the existing database and greatly reduce manual workload. It can also create high-quality samples for complex wetland types, such marsh, swamp and floodplain. Overall, the designed sample method in our study was efficient and reliable, which has significance for large-scale wetland mapping.

Keywords: remote sensing, Wetland, sample production, multi-source thematic data, visual interpretation, Google Earth Engine, Google Earth, Collect Earth

Supported by National Natural Science Foundation of China (No. 42071393, 42301413, 41571077)