时空三极环境大数据平台

**机器学习方法融合北半球长时间序列逐日雪深数据集（1980-2019）**

英文标题：Long-term series of daily snow depth dataset over the Northern Hemisphere based on machine learning （1980-2019）

1、摘要

在地球大数据科学工程专项时空三极环境项目第一课题“三极大数据共享与集成” (XDA19070100)资助下，中国科学院西北生态环境资源研究院车涛课题组利用机器学习方法结合多源雪深产品数据、环境因子变量及地面观测雪深数据等制备了北半球长时间序列逐日雪深数据集。  
首先将人工神经网络、支持向量机和随机森林方法在积雪深度融合的适用性进行对比研究，发现随机森林方法在雪深数据融合上表现出较强优势。其次，利用随机森林方法，结合AMSR-E，AMSR2，NHSD和GlobSnow等遥感雪深产品及ERA-Interim和MERRA2等再分析资料格网雪深产品和环境因子变量等作为模型的输入自变量，用中国气象台站数据（945）、俄罗斯气象台站（620）、俄罗斯积雪调查数据（514）和全球历史气象网络逐日数据（41261）等43340个地面观测站点的雪深数据作为参考真值对模型训练与验证，在专项“地球大数据科学工程”提供的云平台上制备1980~2019年积雪水文年（上一年9月1日至本年度5月31日）的逐日格网雪深数据集。由于1980~1987年微波亮温数据为隔日数据，所以这段时间的数据会出现少量条带缺失现象。利用全球积雪模型对比计划及独立的地面观测数据进行验证，融合数据集的质量在整体上有所提升。利用地面观测数据及融合前的雪深产品对比来看，融合数据的决定系数（R2）从6种融合前产品中最高的0.23（GlobSnow雪深产品）提升至0.81，而相应的均方根误差（RMSE）和平均绝对误差（MAE）也减小至7.7 cm 和2.7 cm。

2、关键词

主题关键词：积雪深度,积雪  
学科关键词：冰冻圈  
地点关键词：北半球  
时间关键词：1980-2019

3、数据细节

1.比例尺：None

2.投影：

3.文件大小：2834.81MB

4.数据格式：None

4、空间范围

|  |  |  |
| --- | --- | --- |
| - | 北：0.0 | - |
| 西：-180.0 | - | 东：180.0 |
| - | 南：90.0 | - |

5、时间范围1980-08-31 16:00:00+00:00--2019-05-30 16:00:00+00:00

6、引用方式

数据的引用:

车涛, 胡艳兴, 戴礼云, 肖林. 机器学习方法融合北半球长时间序列逐日雪深数据集（1980-2019）. 时空三极环境大数据平台, DOI:10.11888/Snow.tpdc.271701, CSTR:18406.11.Snow.tpdc.271701, 2021.[HU Yanxing, XIAO Lin, DAI Liyun, CHE Tao. Long-term series of daily snow depth dataset over the Northern Hemisphere based on machine learning （1980-2019）. A Big Earth Data Platform for Three Poles, DOI:10.11888/Snow.tpdc.271701, CSTR:18406.11.Snow.tpdc.271701, 2021]

文章的引用:

Hu,Y.X., Che, T., Dai, L.Y., & Xiao, L. (2021). Snow depth fusion based on machine learning methods for the Northern Hemisphere. Remote Sensing, 13,1250.

7、资助项目信息

地球大数据科学工程专项时空三极环境项目(XDA19000000)

8、数据资源提供者

姓名: 车涛  
单位: 中国科学院西北生态环境资源研究院  
电子邮件: chetao@lzb.ac.cn  
  
姓名: 胡艳兴  
单位: 中国科学院西北生态环境资源研究院  
电子邮件: huyanxing@lzb.ac.cn  
  
姓名: 戴礼云  
单位: 中国科学院西北生态环境资源研究院  
电子邮件: dailiyun@lzb.ac.cn  
  
姓名: 肖林  
单位: 四川农业大学  
电子邮件: xiaolin@lzb.ac.cn