

引用格式:应王敏,刘晓洁,房世峰,等.基于机器学习的日尺度短波净辐射气候资源遥感反演研究[J].资源科学,2020,42(10):1998-2009. [Ying W M, Liu X J, Fang S F, et al. Retrieval of daily net surface shortwave radiation climatic resources based on machine learning[J]. Resources Science, 2020, 42(10): 1998-2009.] DOI: 10.18402/resci.2020.10.16

基于机器学习的日尺度短波净辐射气候资源遥感反演研究

应王敏¹, 刘晓洁¹, 房世峰¹, 李秀娟¹, 赖明², 张旭振³, 吴骅^{1,4}

(1. 中国科学院地理科学与资源研究所, 北京 100101; 2. 中国地质大学(武汉)资源学院, 武汉 430000; 3. 中国地质调查局烟台海岸带地质调查中心, 烟台 264004; 4. 中国科学院大学资源与环境学院, 北京 100049)

摘要:日尺度地表短波净辐射(DNSSR)是大部分陆面过程模型、全球环流模型、陆-气交换过程模型和各种水文模型的重要输入参数,在自然资源调查、生态环境监测、能量平衡研究等领域具有重要的研究意义和实用价值。本文通过匹配MODIS双星遥感观测和FLUXNET日尺度地面观测数据,筛选出包含18个自变量总计15531对的有效样本,利用机器学习的随机森林方法构建了日尺度DNSSR遥感反演模型,并利用地面实测数据对模型结果进行了真实性检验。结果表明,构建的日尺度DNSSR遥感模型的偏差Bias为 -0.1 W/m^2 ,均方根误差RMSE为 27.8 W/m^2 ,决定系数 R^2 为0.90,表现出良好的精度。基于此过程,得到MODIS双星反演的DNSSR全球分布结果,并与不同季节下再分析ERA5数据扩展得到的DNSSR数据进行了对比,发现两者全球分布特征基本一致,且均与太阳能量随季节变化分布特点密切相关。为进一步证实实验验证的结果,将ERA5和地面站点实测数据作了进一步的对比,结果从侧面证实了本文构建的MODIS的DNSSR产品精度远高于ERA5的DNSSR,而且其产品空间分辨率也有了极大提升。研究结果证明,本文提出的基于MODIS双星观测与机器学习的日尺度DNSSR反演模型具有反演精度高、空间分辨率高、具备时间连续性等优点,能够有效移植至其他气候资源的遥感反演。

关键词:气候资源;日尺度地表短波净辐射(DNSSR);机器学习;遥感;反演

DOI:10.18402/resci.2020.10.16

1 引言

地表净辐射决定了陆地表面吸收的有效能量,其时空变化影响着水分和能量平衡,是全球气候和环境变化的重要驱动因素,也是自然资源综合观测的重要参数之一^[1-3]。地表短波净辐射(Net Surface Shortwave Radiation, NSSR),指到达地表的总太阳辐射与地表反射太阳辐射的差值,是地表净辐射中最为重要的主要组成部分^[4],在物质输送、地气能量交换、水汽流动、生物光合作用、气候变化中具有重

要作用^[3,5]。探究长时间序列生态过程的陆地表面模型、大气环流模型、地表辐射收支研究通常需要的是日尺度地表短波净辐射(Daily Net Surface Shortwave Radiation, DNSSR)^[6,7],定义为瞬时NSSR在一天范围内的积分值^[8]。因此,准确估算高精度高空间分辨率的DNSSR,对研究能量平衡^[9]、农业应用^[10]、生态环境、气候变化等具有重要意义^[11-14]。

卫星遥感观测因其大面积监测的优势已成为估算DNSSR关键来源,其核心思想是,先计算连续

收稿日期:2020-08-25;修订日期:2020-10-13

基金项目:全国自然资源要素综合观测体系规划与部署项目(DD20208063);黄河流域自然资源要素综合观测试点项目(DD20208066);自然资源要素综合观测数据集成与应用服务项目(DD20208067)。

作者简介:应王敏,男,浙江台州人,硕士,主要研究方向为地表短波净辐射遥感反演、机器学习在地理数据上的应用。E-mail: qsy@zju.edu.cn

通讯作者:吴骅,男,江苏无锡人,博士,研究员,主要研究方向为热红外遥感、尺度转换、真实性检验、无人机遥感、遥感大数据分析。E-mail: wuhua@igsnr.ac.cn

2020年10月

的瞬时 $NSSR$ 一天的积分值(即 $DNSSR$),再利用回归方程构建卫星过境时刻的通道星上反照率和 $DNSSR$ 关系。

表1展示了近几年 $DNSSR$ 遥感反演精度,其中 Kim 等^[8]将回归方程应用到 MODIS 单传感器上,算法涉及从 MODIS 观测时间到全天的推演,利用7个地面站点一年的数据验证其反演的 $DNSSR$,算法精度为 40.5 W/m^2 。该方法也分别被应用到陆地卫星 Landsat 传感器^[15]和风云卫星 FY3 传感器^[16]。Landsat 数据由于空间分辨率很高, $DNSSR$ 反演精度有了小幅提升,而 FY3 数据构建的 $DNSSR$ 精度略低于 MODIS 单传感器推演。这种利用单个卫星过境时刻 TOA 通道观测值回归 $DNSSR$ 的方法,需要假定一天内大气条件稳定,即没有考虑大气条件(大气可降水量、气溶胶、云属性等)的日内变化。 $DNSSR$ 的算法精度常常比较低^[8,15,16]。

为了提高 $DNSSR$ 的反演精度,学者们逐渐开始利用多次卫星过境观测进行 $DNSSR$ 的反演。Wang 等^[17]探究了卫星过境观测次数对推演 $DNSSR$ 精度的影响,发现利用同一天 MODIS 两次星上观测值(Terra 和 Aqua)构建回归方程推演 $DNSSR$,精度相比单次观测推演提升了很多,全天候反演精度可达 28.7 W/m^2 ,主要是因为双星观测考虑了大气条件的日变化情况,但其主要局限性为研究区域较小,方法移植至全球尺度的精度未知。同样基于 MODIS 双星数据,Tang 等^[18]采用了另外一种思路:首先利用混合型方法分别反演两个对应时刻的瞬时 $NSSR$,利用线性回归方程求得日尺度积分值,在有云条件下 $DNSSR$ 模型精度达 42.6 W/m^2 。然而,这种先反演两个时刻 $NSSR$ 再进行日尺度扩展的方法,会存在误差传递现象,并且其线性方程的应用受到区域、

地表覆盖类型和时间等限制,存在代表性不足、误差较大的弊端。

随着机器学习方法在 $NSSR$ 遥感反演领域的应用发展,一些研究也着手对 $DNSSR$ 构建机器学习遥感反演模型。例如 Wang 等^[19]使用 171 个站点日尺度实测数据和陆地卫星 Landsat 星上观测数据,利用多种机器学习方法来构建 $DNSSR$ 反演模型,全天候精度高达 18.8 W/m^2 。但由于其 Landsat 卫星的重返周期过长(16 天),无法构建时间连续的 $DNSSR$ 产品。MODIS 由于双星特性,在同一个区域一天至少会有两次过境,因此,利用机器学习方法直接构建 MODIS 双星遥感观测和 $DNSSR$ 之间的关系,是值得深入探究的工作。本文利用 MODIS 双星遥感观测产品和 FLUXNET 地面实测数据,构建了基于随机森林机器学习算法的 $DNSSR$ 遥感反演模型,并利用地面实测数据对其精度进行综合对比及评价,期望能够获取高精度、高空间分辨率、具备时间连续性的全球 $DNSSR$ 产品,为其他气候资源的遥感反演提供技术支持。

2 研究区与数据

2.1 FLUXNET 地面实测数据与研究区选择

FLUXNET 是一个以全世界广泛分布的通量塔为基础的全球通量观测网络^[20,21],遍布世界各大主要国家,在北美、欧洲、亚洲、非洲都有其子网络存在(如美洲通量网 Ameriflux,欧洲通量网 EuroFlux,亚洲通量网 Asiaflux 等)。FLUXNET 的最新数据集为 FLUXNET 2015,于 2016 年发布,其历史数据集为 FLUXNET Marconi (2000) 数据集和 FLUXNET LaThuile (2007) 数据集,下载网址为 <https://fluxnet.fluxdata.org/>。FLUXNET 2015 数据集中提供了 200

表 1 日尺度地表短波净辐射遥感反演精度统计表

Table 1 Statistics of daily net surface shortwave radiation retrieval accuracy

传感器	分辨率	重返周期/天	RMSE/(W/m^2)	站点数&时间跨度	方法类型	参考文献
MODIS/ Terra	1 km	1	40.5(全天候)	7个站点&1年	混合型	Kim et al ^[8]
Landsat 7	30 m	16	36.1(全天候)	6个站点&1年	混合型	Wang et al ^[15]
MODIS/ Combined	1 km	1	42.6(有云)	7个站点&1年	混合型	Tang et al ^[18]
FY3B/ MERSI	1 km	1	41.9(全天候)	7个站点&1年	混合型	Wang et al ^[16]
MODIS/ Combined	1 km	1	28.7(全天候)	7个站点&1年	混合型	Wang et al ^[17]
Landsat 7	30 m	16	18.8(全天候)	171个站点&数年	机器学习	Wang et al ^[19]

多个变量,包括地表短波净辐射变量(地表短波下行辐射、地表短波上行辐射、地表短波净辐射)、水文变量、热通量、质量控制等,其采样的时间间隔为半小时或一小时^[22,23]。本文选取了2014全年具有有效测量的95个FLUXNET站点为研究对象,图1展示了站点全球分布图及IGBP(International Geosphere-Biosphere Programme)地表分类,其中站点主要分布在美国、欧洲、澳洲区域,代表了不同的地表类型、生态系统条件、气候特征和地理环境。95个站点共覆盖了IGBP分类中的12种地表类型,包括常绿针叶林(Evergreen Needleleaf Forests, ENF)、常绿阔叶林(Evergreen Broadleaf Forests, EBF)、落叶针叶林(Deciduous Needleleaf Forests, DNF)、落叶阔叶林(Deciduous Broadleaf Forests, DBF)、混交林(Mixed Forests, MF)、郁闭灌丛(Closed Shrublands, CSH)、开放灌丛(Open Shrublands, OSH)、多树草原(Woody Savannas, WSA)、稀树草原(Savannas, SAV)、典型草原(Grasslands, GRA)、永久湿地(Permanent Wetlands, WET)、农田(Croplands, CRO)。表2为相应地表覆盖类型和数量统计。

2.2 MODIS遥感数据

搭载于美国 Terra 和 Aqua 极轨卫星平台上的中分辨率成像光谱仪(Moderate-resolution Imaging Spectroradiometer, MODIS),探测的光谱范围分布在0.405 um~14.385 um之间,总共分为36个谱段,覆盖了可见光-红外整个波谱^[24]。MODIS扫描宽度长达2330 km,平均1~2天就能够观测到整个地球表面。对于特定地面区域来说,搭载在双星上的MODIS每天最少可提供两次白天和两次黑夜的观测数据。以白天为例,MODIS/Terra在地方时上午10:30过境,提供着相应MOD系列产品;MODIS/Aqua在地方时下午13:30过境,提供着相应MYD系列产品。MODIS双星特性和丰富的地球产品为NSSR反演及日尺度扩展研究提供了可能性^[17]。本文同步下载了2014全年经过95个FLUXNET站点的MODIS遥感产品:MOD02和MYD02提供MODIS前7个通道的表现反射率(红通道R1、近红外通道R2、蓝通道R3、绿通道R4、3个短波红外通道R5-R7);MOD03和MYD03提供纬度、太阳天顶角、卫星观测天顶角等观测几何参数。

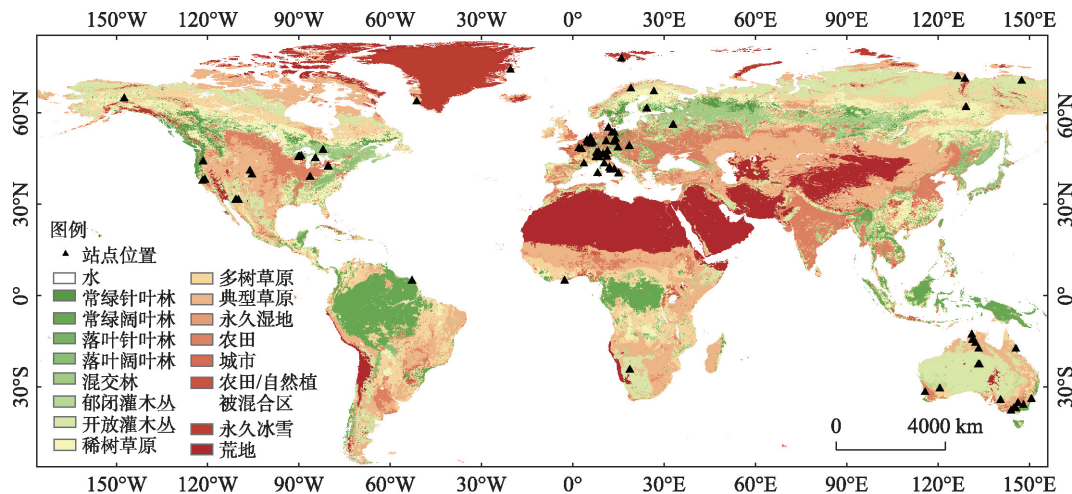


图1 选取的FLUXNET 95个站点分布及IGBP地表分类图

Figure 1 Distribution of the 95 selected sites from the FLUXNET 2015 dataset and the International Geosphere-Biosphere Programme (IGBP) surface types

表2 选取的FLUXNET 95个站点地表类型统计

Table 2 Statistics of surface type of selected FLUXNET 95 sites

地表类型	ENF	OSH	EBF	WSA	DNF	SAV	DBF	GRA	MF	WET	CSH	CRO
站点数量	19	4	9	4	1	4	11	14	6	12	1	10

2020年10月

2.3 ERA5再分析数据

ERA5数据是欧洲中期天气预报中心(EC-MWF)全球气候第五代大气再分析产品(<https://cds.climate.copernicus.eu/>)。再分析ERA5数据利用4D-Var数据同化技术,结合物理机制,将大量模型数据、全球站点实测数据融合在一起^[23]。ERA5数据集能够提供1979年以来全球大气演变的详细记录,变量包括温湿压、云风、气溶胶、能量、地表参数、土地类型、海洋因素等一系列环境因素^[25]。其空间分辨率为 0.25° ,时间分辨率为1h,根据需求提供多达137层的剖面数据或单层数据,其精细程度和覆盖程度在再分析数据中可算首屈一指。在本文中,使用了NetCDF文件格式的全球逐小时NSSR变量。

3 研究方法

3.1 日尺度短波净辐射定义

地表短波净辐射NSSR,指地表吸收的短波辐射通量($0.3\sim 5.0\ \mu\text{m}$)。根据地表辐射平衡方程^[25],地表短波净辐射由地表短波下行辐射减去地表短波上行辐射得到,或者由地表短波下行辐射结合地表反照率得到:

$$NSSR_{\text{sur}} = R_{\text{sur}}^{\downarrow} - R_{\text{sur}}^{\uparrow} = R_{\text{sur}}^{\downarrow} \cdot (1 - \text{albedo}) \quad (1)$$

式中: $NSSR_{\text{sur}}$ 为地表短波净辐射; $R_{\text{sur}}^{\downarrow}$ 为地表短波下行辐射; $R_{\text{sur}}^{\uparrow}$ 为地表短波上行辐射; albedo 为地表反照率。

日尺度地表短波净辐射DNSSR可定义为瞬时地表短波净辐射在一天内的积分值,并进行时间归一化:

$$DNSSR_{\text{sur}} = \frac{\int NSSR_{\text{inst}} dt}{\int dt} \quad (2)$$

式中: $DNSSR_{\text{sur}}$ 为日尺度地表短波净辐射; $NSSR_{\text{inst}}$ 为特定时刻瞬时地表短波净辐射; t 表示时间。

由于MODIS的双星特性,同一个区域一天至少会有两次过境时间(当地时间10:30和13:30),则理论上至少可以获取这两个时刻的瞬时NSSR,通过两个时刻估算值联合线性方程可以得到DNSSR^[18]。由于先反演两个时刻NSSR再进行日尺度扩展的方法,会存在误差传递现象,并且其线性方程存在代表性不足、误差较大的弊端,因此,本文利用机器学习方法直接构建MODIS双星产品瞬时

观测值和DNSSR之间的关系。

3.2 随机森林方法原理

机器学习方法中的随机森林(Random Forest, RF)方法可以从数据自身出发,自发学习数据间的本质联系,具有非常强大的非线性拟合能力,适用于DNSSR模型的反演工作。

RF是集成学习的一种典型算法,是由Breiman^[27]提出的一种基于分类与回归决策树CART集成得到的,是Bagging算法的一个扩展变体^[28]。CART是从一组无序的实例中推理出自上而下的树状分类规则,它代表着自变量属性与因变量之间的一种映射关系^[29]。RF模型中的每一棵CART使用的训练集都是从总的训练集中有放回随机抽取出来的;在训练每棵决策树的节点时,使用的特征是从所有特征中按照一定比例随机抽取的。因此,随机森林算法可以用以下数学表达式展示:

$$Y = \sum_{k=1}^N \{h(\mathbf{X}, \boldsymbol{\theta}_k)\}, k = 1, 2, \dots \quad (3)$$

式中: Y 是随机森林模型对应输入变量 X 的预测结果, N 表示决策树的总数, X 是输入变量组成的向量, $\{\boldsymbol{\theta}_k\}$ 是第 k 棵树的随机向量。

对于每一个输入向量 X , $\boldsymbol{\theta}_k$ 是第 k 棵树的参数向量,它会确定第 k 棵树对应输入变量 X 的预测结果,最后随机森林的预测结果 Y 是对多个CART预测结果取平均值得到。大量的理论和实证研究都证明了RF的主要特色在于对两种随机性的引入,使得随机森林对数据集具有较好的抗噪声能力,不容易陷入过拟合状态^[30]。该算法在运算量没有显著提高的前提下提高了预测精度,被誉为当前最好的算法之一。

3.3 DNSSR反演模型构建过程

特征变量的选择是随机森林算法的关键步骤之一。从地表净辐射的定义可以看出,通过估算地表短波下行辐射 R_s^{\downarrow} 和地表反照率 albedo 两个参数可以获取NSSR^[1,31,32]。但是,该方法的两种中间产品存在不确定性因素和时空匹配差异的问题,利用公式(1)估算NSSR会出现误差传递现象,且地面反照率 albedo 的估算方法基本只适用于晴天的环境^[33]。因此,越来越多的研究利用卫星窄波段星上观测数

据直接反演 $NSSR$, 其精度表现常常优于基于反照率两步法^[8,15,34]。

根据对前人选择变量的总结, 得出地理几何变量和 MODIS 星上光谱通道反射率是影响代表短波净辐射的关键核心变量, 因此本文也采取了经过时空匹配的 MODIS 双星产品 (MOD 产品+MYD 产品) 的地理几何变量和星上光谱通道表现反射率作为自变量, 并选来自 2014 全年 95 个 FLUXNET 瞬时观测值, 结合公式(2)得到的积分值 (即 $DNSSR$) 作为因变量。但由于受传感器自身硬件影响, 下午星 MODIS/Aqua 的第 6 个通道存在部分数据缺失的现象, 在匹配 MOD 和 MYD 产品变量时会造成大量样本无效, 故这里 $DNSSR$ 建模的时候舍弃了 MODIS 第 6 通道的星上反射率 (即短波红外 2 波段), 最终参与 $DNSSR$ 反演模型的变量有 MOD 光谱通道星上表现反射率 (6 个波段)、MOD 地理几何、MYD 光谱通道星上表现反射率 (6 个波段)、MYD 地理几何。图 2 展示了 $DNSSR$ 反演模型构建的总流程图。

在完成变量集的选择以后, 利用 Python 语言结合 Pandas (数据分析), Numpy (数值计算), Scipy (科学计算), Multiprocessing (多进程管理) 等常用包, 进行数据的预处理工作。首先, 采用多线程并行提取遥感影像中最邻近地面站点的像元值, 实现 MODIS 数据和地面站点数据空间位置上的对应。根据空间位置提取完遥感数据后, 需要与地面实测数据进行时间匹配。值得注意的是, FLUXNET 地面站点数据集提供的时间为地方时, 而 MODIS 数据采用协调世界时 UTC, 在匹配前需要做时区转换, 统

一转换成 UTC 时间。同时, 还需要经过异常值、空值剔除等一系列质量控制。其中, MODIS 产品的异常值常表现为 ‘-9999’ 或产品手册所指定填充值, 需要根据用户手册或质量控制数据集对数据做进一步的筛选。在经过几何校正、图像配准、数据筛选等数据预处理后, 得到高质量的自变量和因变量配套的数据集, 用于随机森林机器学习模型的训练。

4 结果与分析

4.1 $DNSSR$ 反演模型构建与验证

经过 FLUXNET 数据和 MODIS 双星产品的匹配、异常值剔除、标准化等预处理后, 有效数据集样本总数为 15531, 本文随机抽取了总样本的 70% 作为训练集 (样本量为 10871), 剩余的 30% 为测试集 (样本量为 4660), 在此基础上, 利用机器学习的随机森林方法, 进行 $DNSSR$ 遥感反演模型的构建。对于随机森林方法, 对模型的预测能力有较大影响的超参数主要有两个: $MAX-FEATURES$ 和 $N-ESTIMATORS$ 。首先, 在大范围内设置超参数的组合 ($MAX-FEATURES$ 6~17, 区间间隔为 1; $N-ESTIMATORS$ 100~2500, 区间间隔为 200)。然后, 根据 $RMSE$ 指标逐步缩小范围 ($RMSE$ 越低, 通常表示模型越好)。在逐步筛选之后, 确定 $MAX-FEATURES$ 的最优值为 10, $N-ESTIMATORS$ 的最优值为 2000。

将训练集代入参数最优的模型进行模型训练, 得到最终的 $DNSSR$ 反演模型。为了验证模型的精度, 本文将随机森林方法预估的 $DNSSR$ 与地面站点实测数据日扩展 $DNSSR$ 数据进行了对比, 并给出了

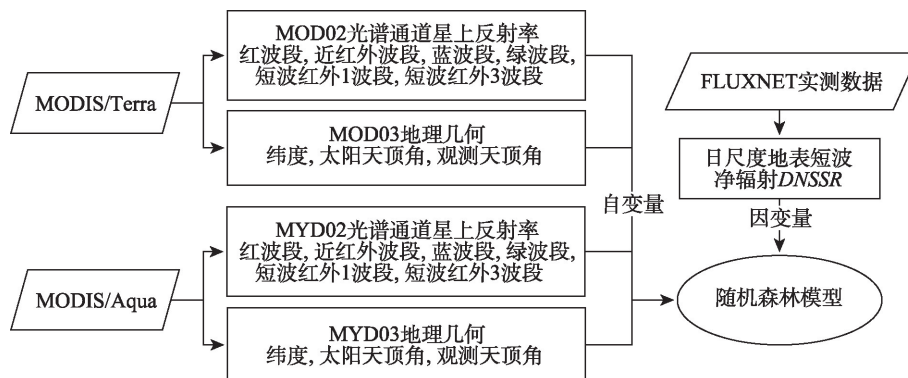


图2 $DNSSR$ 反演模型构建流程图

Figure 2 Flowchart for the development and validation of the daily net surface shortwave radiation ($DNSSR$) model

对比结果的散点分布图(图3)。其中,训练集的偏差 $Bias$ 为 0.0 W/m^2 , 均方根误差 $RMSE$ 为 26.3 W/m^2 , 决定系数 R^2 为 0.91 ; 测试集的偏差 $Bias$ 为 -0.1 W/m^2 , 均方根误差 $RMSE$ 为 27.8 W/m^2 , 决定系数 R^2 为 0.90 。从数据分布情况来看, $DNSSR$ 的取值范围基本在 400 W/m^2 以内, 并且大量样本点(红色散点)集中在低值区域。可以发现, 无论是训练集还是测试集, 其散点分布均匀分布在 1:1 参考线上下, 测试集表现出的良好精度表征了模型在建模过程中的高稳定性、高泛化能力和强鲁棒性。 $RMSE$ 达到 27 W/m^2 , 对比之前一些研究^[8,15,16,18], 精度大大提升。因此, 基于此过程, 本文给出了全球 $DNSSR$ 的反演结果。

4.2 $DNSSR$ 全球分布及季节特性

利用 FLUXNET 多站点观测值与 MODIS 双星数据构建完日尺度地表短波净辐射 $DNSSR$ 反演模型后, 可以应用得到全球范围的 $DNSSR$ 分布。这里选取了 4 个特定时刻 (20190101, 20190401, 20190701, 20191001), 分别代表了北半球的冬季、春季、夏季、秋季。由于采用的都是卫星 1 B 级别的产品, 表现为一景一景的 Swath 格式, 其条带之间会存在一定规律的空隙。Terra 卫星和 Aqua 卫星的扫描方向存在一定夹角, 同一个地区与 Terra 卫星轨道相交呈现为降交点, 而在 Aqua 卫星为升交点。因此, 其特殊的存储特性和卫星物理特性, 造成全球 $DNSSR$ 反演分布图中出现“叉”型空隙(图4)。此外, 由 MODIS 双星产品反演得到的 $DNSSR$ 全球分布, 可以达到非常高的 1 km 的空间分辨率。考虑到

实际应用需求, 这里对 $DNSSR$ 分布图中的海洋区域和南极洲进行了掩膜处理。

可以从 4 个特定时刻的 $DNSSR$ 分布图看出, 全球范围内的 $DNSSR$ 基本上在 300 W/m^2 以内。高纬度地区由于冰雪覆盖, 其地面反照率比较高, 导致 $DNSSR$ 较小。可以发现, 处于春季(20190401)和秋季(20191001)的时候, $DNSSR$ 的高值区域集中在赤道和低纬度区域(纬度 30° 以内)。当北半球为夏季的时候(20190701), 其 $DNSSR$ 高值区域则向北中纬度偏移 ($30^\circ\text{N}-60^\circ\text{N}$); 而当北半球为冬季的时候(20190101), 在南半球的中纬度区域呈现了 $DNSSR$ 高值。上述 $DNSSR$ 高值分布情况与太阳直射区域随着季节变化的特性相一致, 进一步证明了太阳能量是 $NSSR$ 的主要控制因素。

4.3 与再分析 ERA5- $DNSSR$ 对比

为了验证 MODIS 双星数据反演的全球 $DNSSR$ 数据, 本文选用了再分析 ERA5 数据与其进行对比。再分析 ERA5 数据仅提供了逐小时的瞬时 $NSSR$ 产品, 故首先利用公式(1)对其瞬时观测值进行了日尺度扩展, 选取的时刻为与反演的全球 $DNSSR$ 数据相同的 4 个季节特定时刻: 20190101, 20190401, 20190701, 20191001。

图5展示了全球范围内陆地区域 $DNSSR$ 分布情况, 与利用 MODIS 双星数据反演分布图相比, 其高值区域和低值区域分布情况基本相一致, 和太阳能量随季节变化相一致。ERA5 的全球 $DNSSR$ 最小值基本上接近 0 W/m^2 , 而 MODIS 双星数据反演的

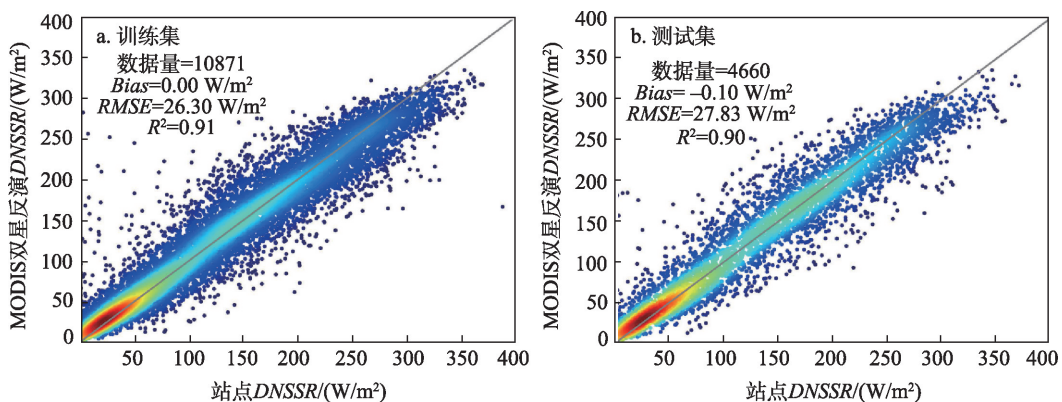


图3 $DNSSR$ 反演模型构建验证

Figure 3 Comparison of daily net surface shortwave radiation ($DNSSR$) values estimated using the random forest (RF) method with in-situ $NSSRs$

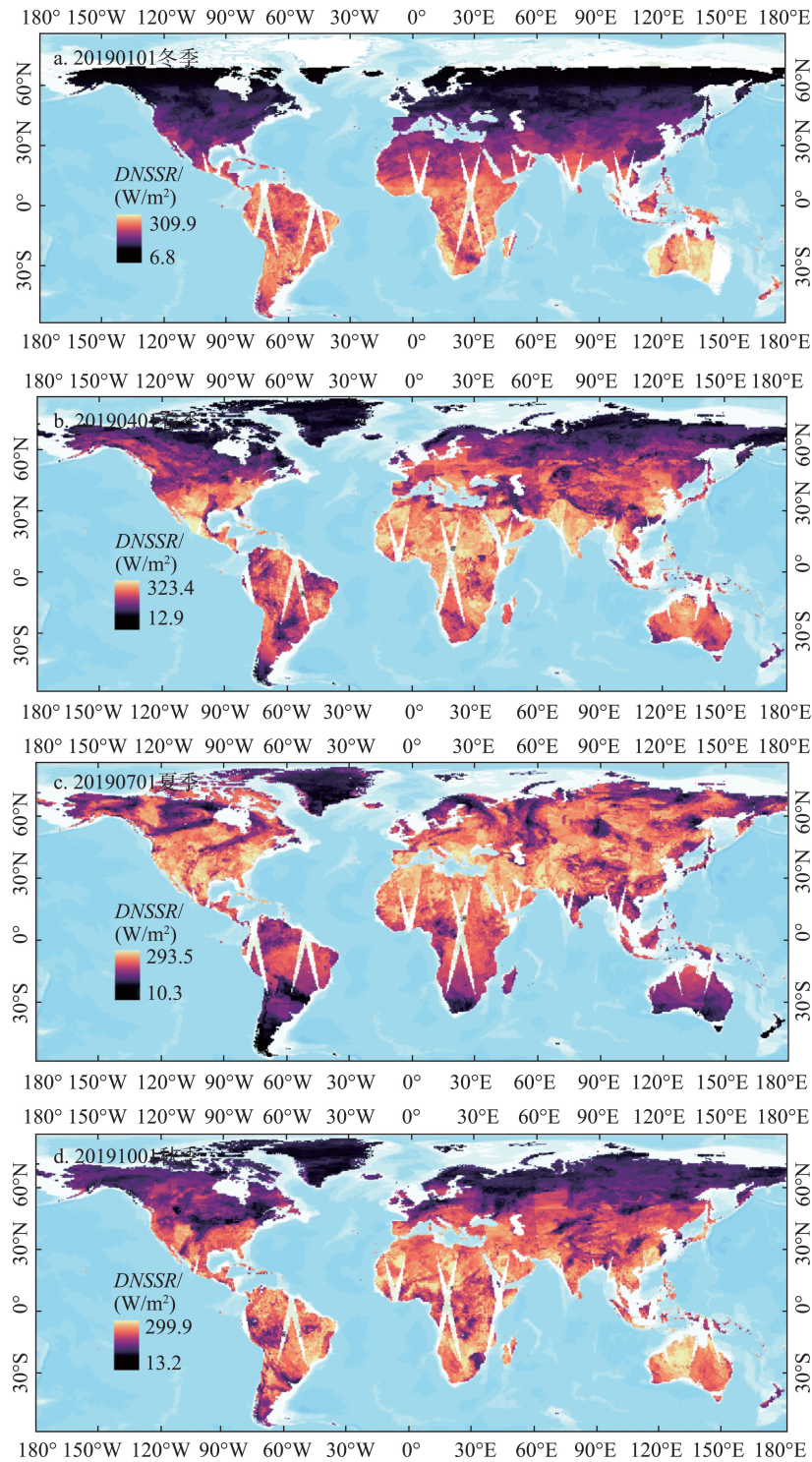


图4 全球DNSSR反演分布图

Figure 4 Distribution of global daily net surface shortwave radiation (DNSSR) retrieval

最小值差不多在 $10 W/m^2$, 存在一定不同。而从高值点出发, ERA5的DNSSR会略高于MODIS卫星数据反演的。由于ERA5数据为再分析产品, 主要由大气环流数值模型为主导, 其DNSSR分布变化会更

加光滑连续。但是, 由于相对粗糙的空间分辨率 0.25° , 使其难以融入到一些精细尺度的农业应用、气候模拟的领域中。

为了更加科学地对比ERA5再分析数据扩展得

2020年10月

到的 *DNSSR* 和 MODIS 双星数据反演的 *DNSSR*, 这里将 MODIS-*DNSSR* 数据进行空间插值至 ERA5-*DNSSR* 的空间分辨率, 并进行点对点的匹配。以 MODIS-*DNSSR* 数据为参照组, 利用相关系数 r 、偏

差 $Bias$ 、均方根误差 $RMSE$ 和决定系数 R^2 为统计指标, 对比其与 ERA5-*DNSSR* 之间的相关关系(表 3)。可以发现, 4 个时刻的相关系数 r 都相当高 (0.82~0.94), 其相应的置信度 p 都接近于 0, 符合显

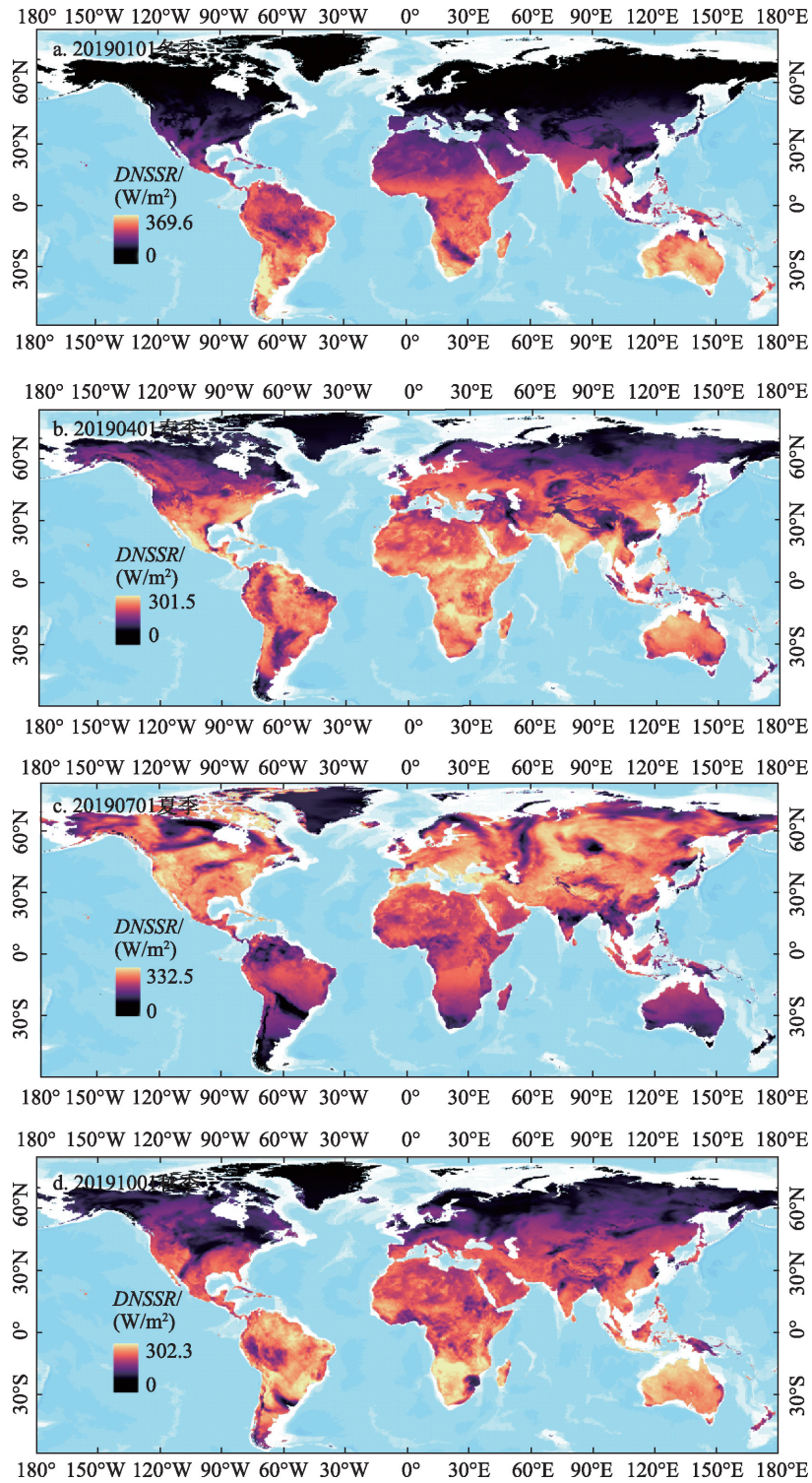


图5 ERA5再分析-全球 *DNSSR* 分布图

Figure 5 Distribution of global daily net surface shortwave radiation (*DNSSR*) from ERA5 data

著性检验需求。从误差统计的角度考虑,两者 *DNSSR* 的均方根误差大致在 35 W/m^2 左右,ERA5-*DNSSR* 在春季(20190401)和秋季(20191001)大于 MODIS-*DNSSR*,表现为正偏差,在冬季(20190101)和夏季(20190701)表现为负偏差,其中冬季两者偏差最小,仅有 1.2 W/m^2 ,秋季偏差最大,达到了 24.0 W/m^2 。

由于两者在秋季存在较大误差,所以需要验证结果作进一步的证明。考虑到 ERA5-*DNSSR* 和 MODIS-*DNSSR* 数据产生较大误差的原因可能来自

于 ERA5-*DNSSR* 数据自身的误差,因此,利用 2014 全年 95 个 FLUXNET 地面站点实际观测值对 ERA5-*DNSSR* 相应区域进行了验证(图 6)。通过验证分析,ERA5-*DNSSR* 在 4 个季节内均有较大的误差,两者的 *RMSE* 在冬季为 41.12 W/m^2 ,在春季为 57.98 W/m^2 ,在夏季为 49.01 W/m^2 ,在秋季为 36.43 W/m^2 ,均高于 MODIS 双星数据反演的 *DNSSR* 误差(27.8 W/m^2)。因此,ERA5-*DNSSR* 数据本身存在的较大误差,也从一定层面解释了表 3 的误差统计结果。

5 结论与展望

高精度高空间分辨率的日尺度地表短波净辐射对全球和区域气候变化、能量平衡、生态环境、陆面模型构建、大气循环探究等领域具有重要的实际意义。本文主要围绕使用随机森林算法,针对 *DNSSR* 遥感反演进行了相关研究,将日尺度扩展后的 FLUXNET 站点观测值和相对应 MODIS 双星产品进行匹配,构建了 15531 对的有效数据集,利用筛选出的数据集采用随机森林方法构建了 *DNSSR* 反

表 3 ERA5-*DNSSR* 与 MODIS-*DNSSR* 相关关系统计表

Table 3 Error statistics of ERA5-daily net surface shortwave radiation (*DNSSR*) correlation with MODIS-*DNSSR*

	20190101	20190401	20190701	20191001
相关系数 r	0.90	0.94	0.82	0.94
偏差 $Bias/(\text{W/m}^2)$	1.2	-13.3	10.1	-24.0
均方根误差 $RMSE/(\text{W/m}^2)$	32.2	31.9	40.6	36.9
决定系数 R^2	0.78	0.83	0.51	0.74

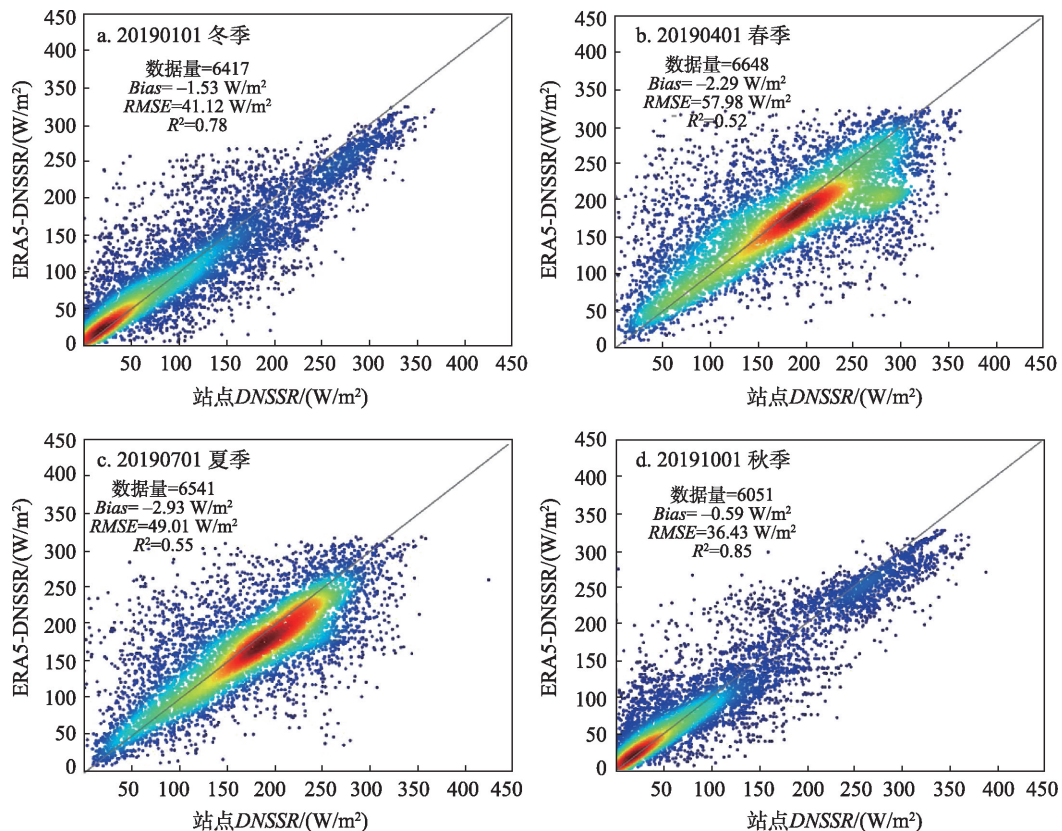


图 6 ERA5 再分析-*DNSSR* 地面验证散点图

Figure 6 Comparison of daily net surface shortwave radiation (*DNSSR*) from ERA5 data and FLUXNET observations

演模型,并利用地面站点 *DNSSR* 数据对反演结果进行了验证。得出以下结论:

(1)MODIS 双星产品结合机器学习方法适用于 *DNSSR* 的反演。反演 *DNSSR* 的偏差 *Bias* 为 -0.1 W/m^2 ,均方根误差 *RMSE* 为 27.8 W/m^2 ,决定系数 R^2 为 0.90,数据分布范围基本在 400 W/m^2 以内,具有良好的精度。

(2)结合 MODIS 卫星面状观测的特点,本文给出了不同季节(以北半球为标准)下的 *DNSSR* 的全球分布图,空间分辨率高达 1 km。4 个特定时刻的全球陆地 *DNSSR* 基本上在 300 W/m^2 以内,高纬度地区的 *DNSSR* 值则相对较小。春季和秋季条件下的 *DNSSR* 高值区域集中在赤道和低纬度区域(纬度 30° 以内),夏季时刻 *DNSSR* 高值区域则向北半球的中纬度区域偏移(30°N — 60°N),冬季时刻则在南半球的中纬度区域呈现了 *DNSSR* 高值。总之,全球 *DNSSR* 高值分布情况与太阳直射区域随着季节变化的特性相一致,进一步证明了太阳能量是地表短波净辐射的主要控制因素。

(3)本文构建的 *DNSSR* 反演结果和再分析 ERA5 数据扩展得到的 *DNSSR* 的对比结果显示,4 个季节特定时刻的 *DNSSR* 分布,其高值区域和低值区域分布情况基本相一致。为了进一步分析两者产生偏差原因,本文对比了 ERA5 和 95 个站点实测数据,结果表明两者存在的误差主要是由 ERA5-*DNSSR* 产品的精度较差造成。此外,再分析 ERA5 产品的空间分辨率十分粗糙(0.25°),难以融入到一些精细尺度的农业应用、气候模拟的领域中。总之,从产品精度、数据精细程度考虑,利用 MODIS 双星产品结合机器学习方法进行 *DNSSR* 反演具有空间分辨率高、精度高、操作简单等优点。

由于地球同步卫星可以提供高时间分辨率的连续观测值,未来可以将 *DNSSR* 算法推广到新一代的地球同步卫星传感器,期望利用多次观测数据,充分考虑短波净辐射日内变化情况,进一步提升 *DNSSR* 的反演精度。通过上述以 *DNSSR* 遥感反演为例的研究,有望将基于 MODIS 双星观测与机器学习的遥感反演方法移植至其他气候资源,为其他气候资源的高精度遥感反演提供技术支持。

参考文献(References):

- [1] Liang S L, Wang K C, Zhang X T, et al. Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2010, 3(3): 225-240.
- [2] 郭鹏, 武法东. 利用 Landsat8 数据估算干旱区晴天太阳瞬时和日间净辐射[J]. 干旱区地理, 2018, 41(1): 32-37. [Guo P, Wu F D. Estimating instantaneous solar radiation and daytime net radiation with Landsat8 data on clear-sky in arid areas[J]. Arid Land Geography, 2018, 41(1): 32-37.]
- [3] Jiang B, Zhang Y, Liang S L, et al. Empirical estimation of daytime net radiation from shortwave radiation and ancillary information [J]. Agricultural and Forest Meteorology, 2015, 211: 23-36.
- [4] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003. [Zhao Y S. Remote Sensing Theories and Applications[M]. Beijing: Science Press, 2003.]
- [5] Yang L, Zhang X T, Liang S L, et al. Estimating surface downward shortwave radiation over China based on the gradient boosting decision tree method[J]. Remote Sensing, 2018, DOI: 10.3390/rs10020185.
- [6] Hatzianastassiou N, Matsoukas C, Fotiadi A, et al. Global distribution of earth's surface shortwave radiation budget[J]. Atmospheric Chemistry and Physics, 2005, 10(5): 2847-2867.
- [7] Zhang X T, Liang S L, Zhou G Q, et al. Generating global land surface satellite incident shortwave radiation and photosynthetically active radiation products from multiple satellite data[J]. Remote Sensing of Environment, 2014, 152: 318-332.
- [8] Kim H Y, Liang S L. Development of a hybrid method for estimating land surface shortwave net radiation from MODIS data[J]. Remote Sensing of Environment, 2010, 114(11): 2393-2402.
- [9] Stephens G L, Li J L, Wild M, et al. An update on earth's energy balance in light of the latest global observations[J]. Nat Geoscience, 2012, 5(10): 691-696.
- [10] 汤绪, 杨续超, 田展, 等. 气候变化对中国农业气候资源的影响[J]. 资源科学, 2011, 33(10): 1962-1968. [Tang X, Yang X C, Tian Z, et al. Impacts of climate change on agro-climatic resources in China[J]. Resources Science, 2011, 33(10): 1962-1968.]
- [11] Zhang Y L, Li X, Bai Y L. An integrated approach to estimate shortwave solar radiation on clear-sky days in rugged terrain using MODIS atmospheric products[J]. Solar Energy, 2015, 113: 347-357.
- [12] Tang B H, Li Z L, Zhang R H. A direct method for estimating net surface shortwave radiation from MODIS data[J]. Remote Sensing Environment, 2006, 103(1): 115-126.

- [13] He T, Liang S L, Wang D D, et al. Estimation of high-resolution land surface net shortwave radiation from AVIRIS data: Algorithm development and preliminary results[J]. *Remote Sensing Environmental*, 2015, 167(15): 20-30.
- [14] Wang T X, Yan G J, Shi J C, et al. Topographic Correction of Retrieved Surface Shortwave Radiative Fluxes from Space under Clear-Sky Conditions[C]. Barcelona: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2014.
- [15] Wang D D, Liang S L, He T. Mapping high-resolution surface shortwave net radiation from Landsat data[J]. *IEEE Geoscience and Remote Sensing Letters*, 2014, 11(2): 459-463.
- [16] Wang D D, Liang S L, He T, et al. Surface shortwave net radiation estimation from FengYun-3 MERSI data[J]. *Remote Sensing*, 2015, 7(5): 6224-6239.
- [17] Wang D D, Liang S L, He T, et al. Estimation of daily surface shortwave net radiation from the combined MODIS data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(10): 1-11.
- [18] Tang B H, Li Z L, Wu H, et al. Estimation of Daily Net Surface Shortwave Radiation from MODIS Data[C]. Milan: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, 2015.
- [19] Wang Y Z, Jiang B, Liang S L, et al. Surface shortwave net radiation estimation from Landsat TM/ETM+ data using four machine learning algorithms[J]. *Remote Sensing*, 2019, 11(23): 2847-2847.
- [20] Balzarolo M, Vicca S, Nguy-Robertson A L, et al. Matching the phenology of net ecosystem exchange and vegetation indices estimated with MODIS and FLUXNET in-situ observations[J]. *Remote Sensing of Environment*, 2016, 174(1): 290-300.
- [21] Chaney N W, Herman J D, Ek M, et al. Deriving global parameter estimates for the NOAH land surface model using FLUXNET and machine learning[J]. *Journal of Geophysical Research Atmospheres*, 2016, DOI: 10.1002/2016JD024821.
- [22] Maes W H, Gentile P, Verhoest N, et al. Potential evaporation at eddy-covariance sites across the globe[J]. *Hydrology and Earth System Sciences Discussions*, 2018, DOI: 10.5194/hess-2017-682.
- [23] Alemohammad H, Fang B, Konings A G, et al. Water, energy, and carbon with artificial neural networks (WECANN): A statistically-based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence[J]. *Biogeosciences Discussions*, 2016, DOI: 10.5194/bg-2016-495.
- [24] Shrivastava R, Iyer I S, Hegde M N, et al. Application of remotely sensed data in the estimation of net radiation at the earth's surface in clear sky conditions[J]. *American Journal of Remote Sensing*, 2018, 6(1): 23-28.
- [25] Xue C D, Wu H, Jiang X G. Temporal and spatial change monitoring of drought grade based on ERA5 analysis data and BFAST method in the belt and road area during 1989-2017[J]. *Advances in Meteorology*, 2019, (428): 1-10.
- [26] 朱君, 唐伯惠. 利用MODIS数据计算中国地表短波净辐射通量的研究[J]. *遥感信息*, 2008, (3): 60-65. [Zhu J, Tang B H. Net surface shortwave radiation mapping over China using MODIS data[J]. *Remote Sensing Information*, 2008, (3): 60-65.]
- [27] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [28] Gislason P O, Benediktsson J A, Sveinsson J R. Random forests for land cover classification[J]. *Pattern Recognition Letter*, 2006, 27(4): 294-300.
- [29] Chi W R Y, Gombojav E, Boldbaatar B, et al. Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city[J]. *Environmental Pollution*, 2019, 245: 746-753.
- [30] Iverson L R, Prasad A M, Matthews S N, et al. Estimating potential habitat for 134 eastern US tree species under six climate scenarios[J]. *Forest Ecology Management*, 2008, 254(3): 390-406.
- [31] 梁师. 基于MODIS/Aqua数据反演地表短波净辐射[D]. 北京: 中国地质大学(北京), 2012. [Liang S. Estimate of Net Surface Shortwave Radiation from MODIS Data in Aqua Satellite[D]. Beijing: China University of Geosciences(Beijing), 2012.]
- [32] 王亚维, 宋小宁, 王睿馨. 晴天条件下基于FY-2E数据的地表短波净辐射的反演方法[J]. *中国科学院大学学报*, 2016, 33(6): 769-774. [Wang Y W, Song X N, Wang R X. Estimation of net surface shortwave radiation using Feng Yun-2E (FY-2E) data on cloud-free days[J]. *Journal of University of Chinese Academy of Sciences*, 2016, 33(6): 769-774.]
- [33] Liang S L. A direct algorithm for estimating land surface broadband albedos from MODIS imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2003, 41(1): 136-145.
- [34] Wu H, Ying W M. Benchmarking machine learning algorithms for instantaneous net surface shortwave radiation retrieval using remote sensing data[J]. *Remote Sensing*, 2019, 11(21): 2520.

Retrieval of daily net surface shortwave radiation climatic resources based on machine learning

YING Wangmin¹, LIU Xiaojie¹, FANG Shifeng¹, LI Xiujuan¹, LAI Ming²,
ZHANG Xuzhen³, WU Hua^{1,4}

(1. Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; 2. School of Earth Resources, China University of Geosciences (Wuhan), Wuhan 430000, China; 3. Yantai Coastal Zone Geological Survey Center, China Geological Survey, Yantai 264004, China; 4. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Daily net surface shortwave radiation (*DNSSR*) is one of the most important parameters in various global land process and hydrological models and is required in climate change, energy balance, ecological, and atmospheric circulation research. This study constructed a daily net surface shortwave radiation model using the random forest (RF) method and MODIS twin-satellite products. 15531 pairs of samples containing 18 independent variables were extracted by matching MODIS twin-satellite products and FLUXNET daily observations. The *Bias*, *RMSE* (root mean square error), and R^2 for the proposed *DNSSR* model using the RF method are $-0.1\text{W}/\text{m}^2$, $27.8\text{W}/\text{m}^2$, and 0.90, respectively. Based on the process, MODIS-*DNSSR* global distribution in different seasons were presented. Verification with field observations shows that the results are similar to the ERA5 reanalysis data, which are closely related to the seasonal distribution of solar energy. To further verify the results, ERA5-*DNSSR* were compared with the FLUXNET-*DNSSR*. The result shows that the proposed *DNSSR* model has also better accuracy and higher resolution than the ERA5 data. The RF-based *DNSSR* model has a good retrieval accuracy, high spatial resolution, and good temporal continuity. It can be effectively transplanted to the retrieval of other climatic resources.

Key words: climatic resources; daily net surface shortwave radiation (*DNSSR*); machine learning; remote sensing; retrieval