

# 山东禹城站叶面积指数无线传感器网络数据集(2020) 研发方法与验证结果

李若溪<sup>1,2</sup>, 周翔<sup>1\*</sup>, 吕婷婷<sup>1</sup>, 陶醉<sup>1</sup>, 王锦<sup>1</sup>, 谢富泰<sup>1,2</sup>

1. 中国科学院空天信息创新研究院遥感卫星应用国家工程实验室, 北京 100101;

2. 中国科学院大学电子电气与通信工程学院, 北京 100049

**摘要:** 随着通讯技术的发展, 基于无线传感器网络技术的地面测量成为获取地表参数的重要手段。无线传感器网络以其长时间、多点同步观测的优势, 为遥感卫星叶面积指数产品的真实性检验提供了可靠的数据支撑。然而, 意外的电池、通信故障或天气影响会导致测量数据的不稳定, 这就需要从大量原始数据中处理得到能够代表地面相对真值的数据。本文针对2020年5–11月在山东禹城站布置3个节点(0803、0804、0805)运用叶面积指数传感器网络系统SBLX-034进行观测。首先选择观测时间在上午10时至下午15时的数据, 筛选出有效数据的时刻。然后依据站点间的时空相关性利用时序神经网络NARX建模, 剔除模型预测误差大于1的异常时刻数据, 并利用LSTM神经网络对处理后的数据规律进行检验。最后对每天的数据进行平均处理得到长时间序列的LAI实测相对真值, 用于为相关研究提供数据支撑。本数据集为处理后的山东禹城站0803、0804和0805站点的无线传感器网络数据, 具体包括: (1)禹城站3个无线传感器网络节点的地理位置数据; (2)2020年5–11月3个节点、每日叶面积指数数据。该数据集存储为.xlsx、.shp和.kml数据格式, 由10个文件组成, 数据量为49.1 KB(压缩为1个文件, 42.5 KB)。

**关键词:** 山东禹城; 叶面积指数; 观测节点; 日平均; 地面观测

**DOI:** <https://doi.org/10.3974/geodp.2021.02.04>

**CSTR:** <https://cstr.escience.org.cn/CSTR:20146.14.2021.02.04>

## 数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志(中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2021.03.01.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2021.03.01.V1>.

## 1 前言

叶面积指数(Leaf Area Index, LAI)通常定义为单位地表面积上绿叶表面积总和的一半<sup>[1]</sup>, 是描述植被冠层结构及功能的重要参数<sup>[2]</sup>。由于叶片是地表层与大气边界层中能量与

收稿日期: 2021-01-14; 修订日期: 2021-05-10; 出版日期: 2021-06-25

基金项目: 中华人民共和国科学技术部(2018YFE0124200); 中国科学院(2020)

\*通讯作者: 周翔 L-7359-2016, 中国科学院空天信息创新研究院, [zhouxiang@radi.ac.cn](mailto:zhouxiang@radi.ac.cn)

数据引用方式: [1] 李若溪, 周翔, 吕婷婷等. 山东禹城叶面积指数无线传感器网络数据集(2020)研发方法与验证结果[J]. 全球变化数据学报, 2021, 5(2): 135–142. <https://doi.org/10.3974/geodp.2021.02.04>.  
<https://cstr.escience.org.cn/CSTR:20146.14.2021.02.04>.

[2] 李若溪, 周翔, 吕婷婷等. 山东禹城站观测节点日平均叶面积指数数据集(2020)[J/DB/OL]. 全球变化数据仓储电子杂志, 2021. <https://doi.org/10.3974/geodb.2021.03.01.V1>.  
<https://cstr.escience.org.cn/CSTR:20146.11.2021.03.01.V1>.

物质（如水分，二氧化碳等）交换的主要通道，因此 LAI 也是大多陆表生态模型的关键输入因子<sup>[3-6]</sup>。随着卫星技术以及传感器性能的快速发展，卫星遥感数据利用多种反演模型已经生成了多种全球 LAI 产品。然而，受限于传感器获取遥感数据自身的精度限制以及反演模型的假设和不稳定性，LAI 产品本身蕴含着一定程度的误差<sup>[7,8]</sup>。为了更好地评价数据质量，扩展应用范围，需要对 LAI 产品进行真实性检验。遥感产品真实性检验需要获取能够代表地面目标的相对真值，受到地面观测成本的限制，传统 LAI 实地测量无法满足连续长时间的测量要求。随着通讯技术的发展，无线传感器网络被用于试验场观测过程中。基于无线传感器网络技术的地面测量方法能保证对参数实现长时间、稳定的多点同步观测，且便于进行多站点数据之间的对比分析，为真实性检验提供比较可靠的地面观测手段。

植被生长是一个复杂的生物过程，受多种环境因素的影响。在田间试验中，野外综合场获取的叶面积指数无线传感网络数据每 5 分钟测量一次。不同自然条件下植物叶片的表现是不同的，即使在上午 10 时至下午 15 时之间，风速、风向和太阳光照条件等因素都会影响叶片的张开程度、倾斜程度以及张角，这就会导致叶面积指数无线传感器网络的测量结果不稳定。目前国内外利用植被无线传感器网络数据进行的研究中，对时间分辨率要求不高的情况下，通常采取设置聚集滑动窗口的方法<sup>[9]</sup>，对多天数据进行处理；其他以天或小时为周期的研究中大部分采用固定时间内插值平均的方法<sup>[10-13]</sup>，但过程中对异常值的剔除存在一定的主观性。针对这一问题，本文阐明了 2020 年 5 月至 11 月山东禹城综合试验场的无线传感器网络数据的研发方法和验证结果，其中包括筛选有效数据，剔除异常值，保证原有变化趋势的前提下，更好地保留原始数据等流程。

## 2 数据集元数据简介

《山东禹城站观测节点日平均叶面积指数数据集(2020)》<sup>[14]</sup>的名称、作者、地理区域、数据年代、时间分辨率、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表 1。

## 3 数据研发方法

### 3.1 叶面积指数传感网络系统工作原理

叶面积指数传感器网络系统 SBLX-034 依托先进的鱼眼摄影技术，利用图像处理技术对植被冠层进行快速分析，实时获取结构冠层信息，利用自行研发的图像分割技术对图像进行准确分割，极大的消除了强光照条件下耀斑的影响，提高到了分析精度，获取到包括叶面积指数在内的多种植被参数。传感器网络系统的体系由处理中心、节点管理器和部署在监测区域内的传感器节点组成。监测区域内高密度分布大量传感器节点，自动组网。传感器节点接收节点管理器的命令，采集数据并进行初步处理，然后通过其他节点逐级传送至节点管理器。节点管理器收集网络中所有传感器节点的监测数据并整理，最后通过外部网络到达数据中心，由管理员进行批量处理。所有的仪器节点都采用自动观测模式，每隔 5 分钟获取一次观测数据。

本系统主要采用单角度法估算植物叶面积指数。冠层孔隙度与植物叶面积指数和植物叶倾角有关，可以间接推导出植物叶面积指数的表达式，常见植物叶倾角分布一般可以分为如下 5 种情况：①叶片呈水平分布，即叶片倾斜角度都为 0°；②叶片垂直分布，叶片倾斜的角度全部为 90°；③叶片呈锥面分布，叶片倾斜角度都在 0°–90°之间；④叶片呈球面分布，这是一种非常均匀的分布方式，在这种分布方式中所有的叶片没有特定的角度，就

表 1 《山东禹城站观测节点日平均叶面积指数数据集(2020)》元数据简表

条 目	描 述
数据集名称	山东禹城站观测节点日平均叶面积指数数据集(2020)
数据集短名	LAI_YuCheng_2020_0501-1108
作者信息	李若溪 ABH-7136-2020, 中国科学院空天信息创新研究院, liruoxi19@mailsucas.ac.cn 周翔 L-7359-2016, 中国科学院空天信息创新研究院, zhouxiang@radi.ac.cn 吕婷婷 R-8978-2016, 中国科学院空天信息创新研究院, lvt@radi.ac.cn 陶醉 L-4530-2016, 中国科学院空天信息创新研究院, taozui@radi.ac.cn 王锦 ABH-9051-2020, 中国科学院空天信息创新研究院, wangjin01@radi.ac.cn 谢富泰 ABH-7123-2020, 中国科学院空天信息创新研究院, xieft@radi.ac.cn
地理区域	中国科学院叶面积指数地面观测网络山东禹城综合试验站
数据年代	2020 年 5–11 月
时间分辨率	1 天
数据格式	.xlsx、.shp、.kml
数据量	42.5 KB
数据集组成	(1) 3 个无线传感器网络节点的地理位置数据 (2) 2020 年 5–11 月 3 个节点的每日叶面积指数地面相对真值数据
基金项目	中华人民共和国科学技术部（2018YFE0124200）；中国科学院（2020）
数据计算环境	MATLAB，中国科学院空天信息创新研究院内平台
出版与共享服务平台	全球变化科学研究数据出版系统 <a href="http://www.geodoi.ac.cn">http://www.geodoi.ac.cn</a>
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）、通过《全球变化数据仓储电子杂志（中英文）》发表的实体数据集和通过《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 <sup>[15]</sup>
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS/ISC, GEOSS

是呈连续随机分布；⑤叶片呈椭球面分布，叶片倾斜的角度全是呈连续的椭球面分布。在这五种情况中，方位角都是随机，而且第五种情况是一般情况，其他四种情况都视为特殊的情况。Campbell 用椭球分布函数模拟叶倾角分布<sup>[16]</sup>，将投影函数展开如式(1)所示：

$$\begin{cases} G(\theta) = \frac{(x^2 + \tan^2 \theta)^{\frac{1}{2}} \cos \theta}{x + \frac{\sin^{-1} \varepsilon_1}{\varepsilon_1}}, x \leq 1 \\ G(\theta) = \frac{(x^2 + \tan^2 \theta)^{\frac{1}{2}} \cos \theta}{x + \frac{1}{2\varepsilon_2} \ln \left( \frac{1 + \varepsilon_2}{1 - \varepsilon_1} \right)}, x > 1 \end{cases} \quad (1)$$

式中,  $G(\theta)$  表示投影面积, 投影面积的计算方式是以  $\theta$  方向为准, 投影到该方向上;  $x$  是椭圆水平半轴和垂直半轴的比值,  $x$  越大表示冠层叶倾角越倾向于水平分布,  $x$  越小表示冠层叶倾角越倾向于垂直分布。Wang 根据叶倾角密度函数等关系推出  $x$  与平均叶倾角  $\theta_L$  的关系<sup>[17]</sup>, 如式(2)所示:

$$x = 3 + \left( \frac{\theta_L}{9.65} \right)^{-0.6061} \quad (2)$$

当视角为  $57.5^\circ$  角的时候, 总有投影函数  $G(\theta)$  的值约等于 0.5, 也就是说当入射视角达到此  $57.5^\circ$  角时, 投影函数的值可以近似看成是与植物叶倾角无关的。单角度法是根据投影函数这个特性来推导出植物叶面积的计算公式, 如式(3)所示:

$$LAI = \frac{-\ln T(57.5^\circ) \cos(57.5^\circ)}{0.5} \quad (3)$$

式中,  $T$  表示透射率, 根据式(3)即可计算出植物叶面积指数。

### 3.2 原始数据采集

禹城综合试验场位于山东省禹城市, 属暖温带半湿润季风气候带, 农业系统以种植冬小麦-夏玉米为主, 在黄淮海平原的农业类型具有典型性和代表性。禹城站无线传感器网络布设如图 1。本数据集获取了 2020 年 5 月 1 日至 11 月 8 日的叶面积指数无线传感器网络的全部有效原始数据, 共 30 万余条, 测量数据包括: 节点经纬度、数据采集时间、空气温度和叶面积指数。由于意外的电池、通信故障或天气影响, 导致部分数据缺失或无效, 因此选择数据质量相对更好的 3 个节点 (0803、0804、0805) 进行处理, 有效原始数据共计 9 万余条。

### 3.2 数据处理方法

《山东禹城站叶面积指数无线传感器网络数据集(2020)》处理流程如图 2 所示。

首先, 选择 3 个站点观测时间在上午 10 时至下午 15 时的数据, 共计 20,610 条原始数据。由于各个节点受到环境不同程度的影响, 因此存在不同情况的数据缺失, 需要筛选出 3 个站点均存在有效数据的时刻, 共计 4,576 个时刻, 13,728 条原始数据。

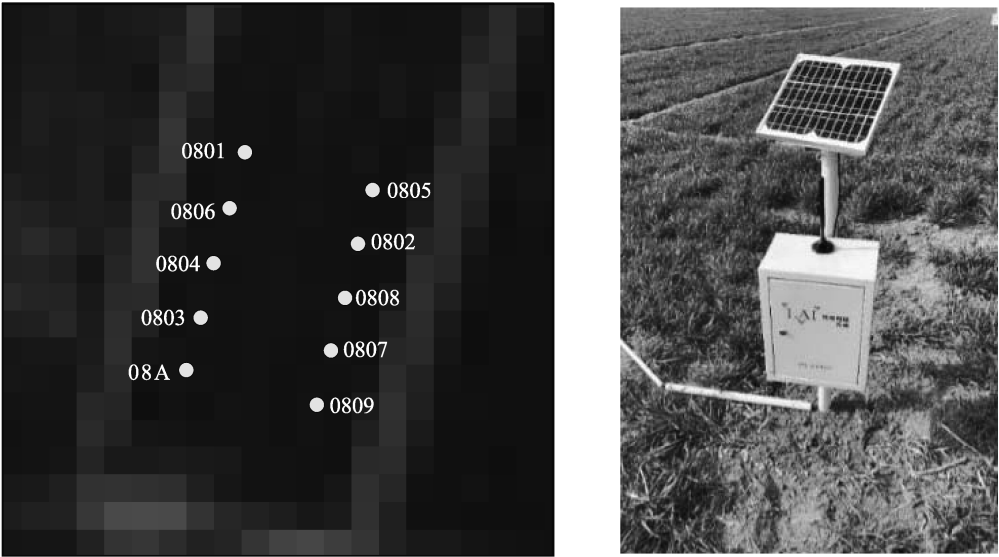


图 1 禹城站无线传感器网络节点分布

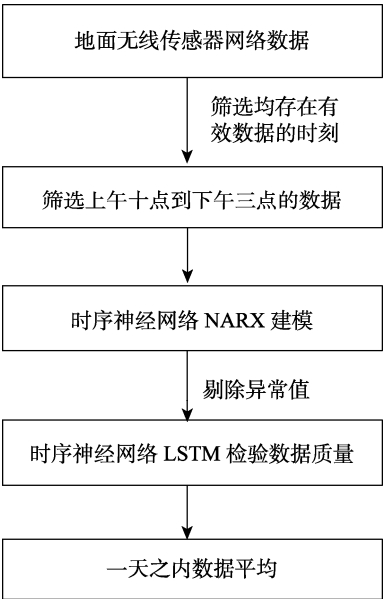


图 2 数据处理流程

然后，依据站点间的时空相关性利用时序神经网络 NARX 建模，经过不断调整模型使之误差最小，最终确定的参数设置如下：80%的数据被用于神经网络的训练，10%的数据用来验证网络是否正在泛化，并在此之前停止训练，10%的数据将用作外部验证；隐含层 10 个节点；延迟数设为 2 个时刻；模型的建立核函数为贝叶斯正则化。神经网络模型建模精度如图 3 所示。训练集模型精度为 0.935,39，检验集模型精度为 0.920,53，整体建模精度为 0.933,91，说明 WSN 实测 LAI 数据具有较强的时间规律性，能够通过 NARX 模型表征

这一规律并且模拟效果较好。剔除神经网络模型预测误差大于 1 的时刻数据，0803 节点筛选出 65 个异常时刻，0804 节点筛选出 57 个异常时刻，0805 节点筛选出 28 个异常时刻，3 个节点合并后，共剔除 130 个时刻。剔除后共计 4,446 个时刻，13,338 条有效原始数据。

最后，将上述处理后每天的有效原始测量数据进行平均处理作为每天的叶面积指数地面测量相对真值，共计 170 天，510 条站点数据，从而更好地利用每天的数据进行时间序列上产品的真实性检验。

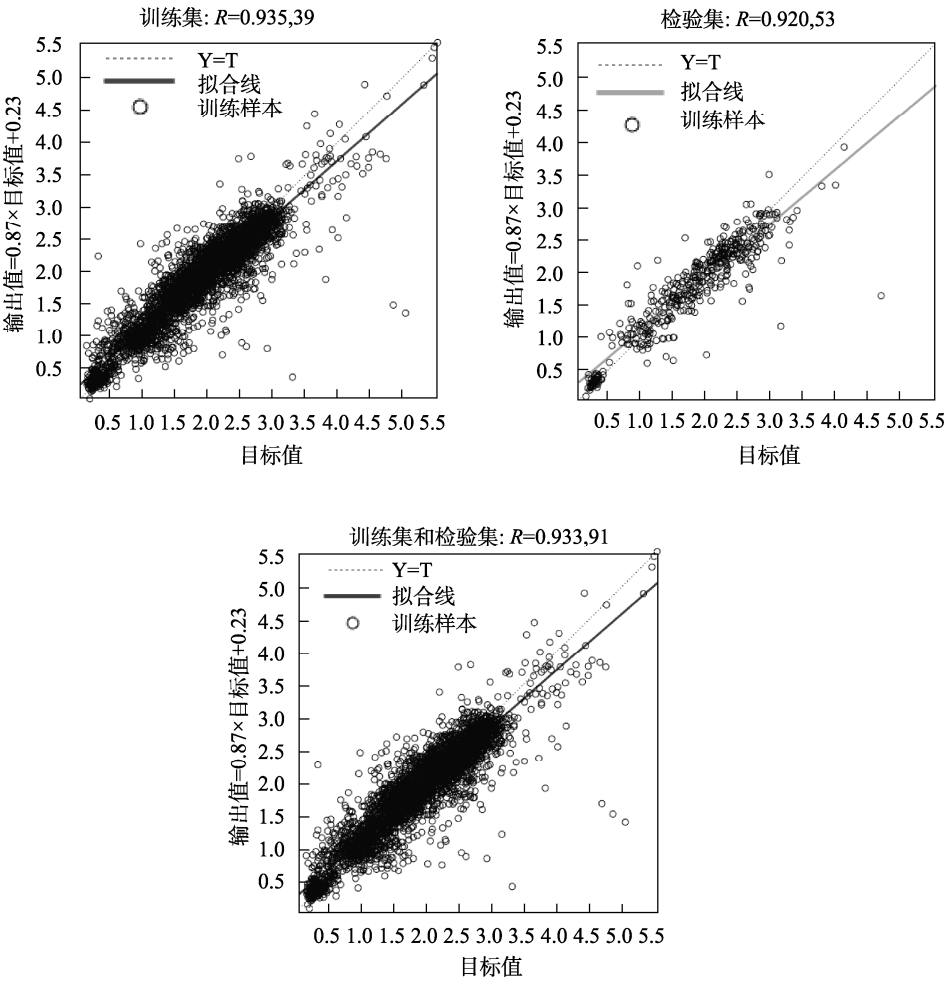


图 3 神经网络建模精度

4 数据结果与验证

4.1 数据集组成

本数据集包括：

- (1) 禹城站 3 个节点的地理位置；
- (2) 2020 年 5-11 月 3 个节点的每日叶面积指数数据。

## 4.2 数据结果

2020年5月至11月山东禹城综合试验场0803(36°50′2.4174″N, 116°34′56.1″E)、0804(36°50′3.048″N, 116°34′56.316″E)和0805(36°50′4.7682″N, 116°34′54.84″E)3个节点叶面积指数数据,时间分辨率为1天。

## 4.3 数据结果验证

为了保证处理后的数据仍具有较强的时空相关性及数据规律,采用LSTM时序神经网络对剔除异常值后的13,338条有效原始数据进行建模。如果模型精度好,预测误差小,预测相关性高的情况下,那么可以认为剔除后的数据质量较高。参数设置如下:以连续时间序列的前80%用于训练,后20%用于测试;求解器并进行250轮训练;为了防止梯度爆炸,将梯度阈值设置为1;指定初始学习率0.005,在125轮训练后通过乘以因子0.2来降低学习率;指定LSTM层有200个隐含单元。训练模型的预测精度为0.83。可以认为处理后的无线传感器网络数据保留了原始实测数据的变化情况与相关性。

## 5 讨论和总结

本文收集了2020年山东禹城站叶面积指数无线传感器网络数据,通过时序神经网络NARX设定误差阈值,考虑到尽量多的保留地面实测数据,本文处理数据时设定的误差阈值剔除了3%的原始数据,即极端异常点,通过LSTM神经网络建模检验剔除后具有较好的数据规律,最后在有效时间范围内进行平均,从而得到长时间序列的连续实测地面相对真值作为相关研究的数据支撑。该数据集可以应用于不同尺度叶面积指数产品的点对点真实性检验,并且在长时间序列的检验中具有很明显的优势,如研究不同植被生长阶段卫星产品精度的变化情况。另外,由于3个节点所处的环境不同,缺失数据也是不同时刻的,因此处理后的有效数据只占原始数据的64%。其中,外界因素导致的数据缺失是不可抗力的,但随着通信技术与深度学习原理的不断发展,如何利用已有数据更好地插补缺失值从而提高数据的利用率是个值得研究的问题。

**作者分工:**周翔、吕婷婷对数据集的开发做了总体设计;李若溪处理了叶面积指数无线传感器网络数据并对数据进行验证;所有作者共同撰写了数据论文。

**利益冲突声明:**本研究不存在研究者以及与公开研究成果有关的利益冲突。

## 参考文献

- [1] Chen, J. M., Black, T. A. Defining leaf area index for non-flat leaves [J]. *Plant, Cell & Environment*, 1992, 15(4): 421–429.
- [2] Garrigues, S., Lacaze, R., Baret, F., *et al.* Validation and intercomparison of global leaf area index products derived from remote sensing data [J]. *Journal of Geophysical Research*, 2008, 113: G02028.
- [3] Liu, J., Chen, J. M., Cihlar, J., *et al.* A process-based boreal ecosystem productivity simulator using remote

- sensing inputs [J]. *Remote Sensing of Environment*, 1997, 62(2): 158–175.
- [4] Andrew, D., Richardson, R. S., Anderson, M., *et al.* Terrestrial biosphere models need better representation of vegetation phenology: results from the North American carbon program site synthesis [J]. *Global Change Biology*, 2012, 18(2): 566–584.
- [5] Sellers, P. J., Dickinson, R. E., Randall, D. A., *et al.* Modeling the exchanges of energy, water, and carbon between continents and the atmosphere [J]. *Science*, 1997, 275(5299): 502–509.
- [6] Bonan, G. B. Land-Atmosphere interactions for climate system models: coupling biophysical, biogeochemical, and ecosystem dynamical processes [J]. *Remote Sensing of Environment*, 1995, 51(1): 57–73.
- [7] Friedl, M. A., Davis, F. W., Michaelsen, J., *et al.* Scaling and uncertainty in the relationship between the NDVI and land surface biophysical variables: an analysis using a scene simulation model and data from FIFE [J]. *Remote Sensing of Environment*, 1995, 54(3): 233–246.
- [8] 丁艳玲. 植被覆盖度遥感估算及其真实性检验研究[D]. 长春: 中国科学院研究生院(东北地理与农业生态研究所), 2015.
- [9] Shi, Y. C., Wang, J. D., Qin, J., *et al.* An upscaling algorithm to obtain the representative ground truth of LAI time series in heterogeneous land surface [J]. *Remote Sensing*, 2015, 7(10): 12887–12908.
- [10] Zhang, J. L., Liu, Q., Li, X. H., *et al.* Calibration and data validation of wireless sensor network [P]. *Intelligent Earth Observing Systems*, 2015.
- [11] Dou, B. C., Wen, J. G., Li, X. H., *et al.* Wireless sensor network of typical land surface parameters and its preliminary applications for coarse-resolution remote sensing pixel [J]. *International Journal of Distributed Sensor Networks*, 2016, 12(4): 55–60.
- [12] 周雨石. 基于改进热惯量模型土壤含水量时空变化研究[D]. 开封: 河南大学, 2018.
- [13] Qu, Y., Zhu, Y., Han, W., *et al.* Crop leaf area index observations with a wireless sensor network and its potential for validating remote sensing products [J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2014, 7(2): 431–444.
- [14] 李若溪, 周翔, 吕婷婷等. 山东禹城站观测节点日平均叶面积指数数据集(2020) [J/DB/OL]. 全球变化数据仓储电子杂志, 2021. <https://doi.org/10.3974/geodb.2021.03.01.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2021.03.01.V1>.
- [15] 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]. <https://doi.org/10.3974/dp.policy.2014.05> (2017年更新).
- [16] Campbell, G. S. Extinction coefficients for radiation in plant canopies calculated using an ellipsoidal inclination angle distribution [J]. *Agricultural and Forest Meteorology*, 1986, 36(4): 317–321.
- [17] Wang, Y. P., Jarvis, P. G. Mean leaf angles for the ellipsoidal inclination angle distribution [J]. *Agricultural and Forest Meteorology*, 1988, 43(3): 319–321.