

## 潜热通量缺失数据插补方法比较研究

杨强<sup>1</sup>, 李鑫豪<sup>2</sup>, 杜韬<sup>1</sup>

(1. 中国电力工程顾问集团华北电力设计院有限公司, 北京 100032; 2. 北京林业大学水土保持学院, 北京 100083)

**摘要:** 【目的】分析比较不同插补方法对生态系统潜热通量 ( $F_{LE}$ ) 缺失值的插补精度。【方法】利用涡度相关法于 2019 年对北京市松山国家级自然保护区典型天然落叶阔叶林生态系统  $F_{LE}$  与环境要素进行原位连续监测, 通过 3 种插补方法 (边缘分布抽样法、线性回归法、人工神经网络法) 对  $F_{LE}$  缺失数据 (0.5 h 数据中随机剔除) 进行插补, 分析实测  $F_{LE}$ 、插补  $F_{LE}$  与环境因子间的关系。【结果】3 种插补结果均低估了实测  $F_{LE}$ , 其中人工神经网络插补值最接近实测值 (决定系数  $R^2=0.40$ )。实测  $F_{LE}$  与空气温度 ( $T_a$ )、饱和水汽压差 ( $D_{VP}$ ) 间均呈指数关系。边缘分布抽样法插补  $F_{LE}$  与  $T_a$ 、 $D_{VP}$  间的关系最接近实测  $F_{LE}$ , 然而 3 种插补方法都不同程度改变了  $F_{LE}$  对  $T_a$  和  $D_{VP}$  的敏感性。【结论】人工神经网络法的插补结果与实测值最接近, 边缘分布抽样法的结果与环境因子间的关系最接近实测值与环境因子间的关系, 因此未来研究应依据研究目的选取合适的插补方法。图 5 表 1 参 41

**关键词:** 涡度相关; 潜热通量; 数据插补; 落叶阔叶林

中图分类号: Q945.11 文献标志码: A 文章编号: 2095-0756(2024)04-0810-10

## Comparative study of interpolation methods for missing latent heat flux data

YANG Qiang<sup>1</sup>, LI Xinhao<sup>2</sup>, DU Tao<sup>1</sup>

(1. North China Power Engineering Co., Ltd., China Power Engineering Consulting Group, Beijing 100032, China;

2. School of Soil and Water Conservation, Beijing Forestry University, Beijing 100083, China)

**Abstract:** [Objective] The aim of this study is to analyze and compare the interpolation accuracy of different interpolation methods for missing latent heat flux value ( $F_{LE}$ ) in ecosystem. [Method] Eddy-covariance (EC) method was used to continuously monitor  $F_{LE}$  and environmental factors of a typical natural deciduous broad-leaved forest ecosystem in Songshan National Nature Reserve, Beijing in 2019. Three interpolation methods, namely marginal distribution sampling method (MDS), linear regression method (REG), and artificial neural network method (ANN) were applied to interpolate the missing  $F_{LE}$  data (randomly removed from 0.5 h data), and the relationship between measured  $F_{LE}$ , interpolated  $F_{LE}$  and environmental factors was analyzed. [Result] All the three interpolation results underestimated the measured  $F_{LE}$ , among which ANN interpolation value was the closest to the measured one ( $R^2=0.40$ ). The measured  $F_{LE}$  showed an exponential relationship with air temperature ( $T_a$ ) and saturated vapor pressure deficit ( $D_{VP}$ ). MDS interpolated the relationship between  $F_{LE}$  and  $T_a$  and  $D_{VP}$ , and was the closest to the measured  $F_{LE}$ . All the three interpolation methods changed the sensitivity of  $F_{LE}$  to  $T_a$  and  $D_{VP}$  to varying degrees. [Conclusion] The interpolation results of ANN are the closest to the measured values. The relationship between the results of MDS and environmental factors is the closest to the relationship between measured  $F_{LE}$  and environmental factors. Therefore, appropriate interpolation methods

收稿日期: 2023-10-24; 修回日期: 2024-04-04

基金项目: 国家自然科学基金青年基金资助项目 (32101588)

作者简介: 杨强 (ORCID: 0009-0007-5317-402X), 工程师, 从事水土保持与林业生态工程研究。E-mail: yangqiang@ncpe.com.cn. 通信作者: 李鑫豪 (ORCID: 0000-0002-3507-7556), 工程师, 从事碳水循环及其调控机制研究。E-mail: 458819027@qq.com

should be selected in future research based on the research purpose. [Ch, 5 fig. 1 tab. 41 ref.]

**Key words:** eddy covariance; latent heat flux; data interpolation; deciduous broad-leaved forest

气候变化及其对生态系统造成的影响已经成为全球变化研究的重点问题<sup>[1]</sup>。随着气候变化强度的不断加剧,生态系统的结构和功能必定会受到影响<sup>[2]</sup>。森林生态系统约占陆地生态系统总面积的 30% 以上,是陆地生态系统的主体<sup>[3]</sup>,有降低大气二氧化碳(CO<sub>2</sub>)浓度、涵养和调节降水、减少地表侵蚀、调节区域小气候等重要作用<sup>[3-7]</sup>。因此,在气候变化背景下,研究森林生态系统过程,如生态系统潜热通量( $F_{LE}$ ),对深入理解环境因子对生态系统的调控机制具有重要意义<sup>[8-10]</sup>。

涡动相关法已成为从小时到年际时间尺度监测生态系统碳、水和能量通量的主要方法,并成为自下而上估算全球生态系统碳、水、平衡过程的支柱<sup>[11-13]</sup>。此外,涡度相关数据也越来越多地被用于生态系统模型的校准和验证<sup>[14-16]</sup>。为了定量估计生态系统过程,准确插补由于不利的气象状况或仪器故障等原因造成的缺失数据是非常重要的<sup>[17-18]</sup>。不同插补方法得到的结果可能有很大差别<sup>[19-21]</sup>。目前,大多数研究只关注插补结果,鲜有研究对插补方法进行比较分析,因此,对现有主要插补方法进行比较分析可以更好地理解生态系统对气候变化的响应。

如今在世界范围内利用涡动相关系统已经建立了许多通量数据集,尽管各大通量网系统(如 FLUXNET、加拿大通量网等)建立数据集的数据收集方法得到了统一,但组织间数据处理方法仍有所不同。例如 FLUXNET 通量网通常采用边缘分布抽样法对缺失数据进行插补<sup>[22]</sup>,而加拿大通量网通常利用线性与非线性回归法对缺失数据进行插补<sup>[23]</sup>。这导致相关数据可比性不强<sup>[24-25]</sup>。考虑到气候变化强度不断增强,增强通量数据的可比性是定量预测未来全球生态系统水文循环和改善森林管理决策的关键。

生态系统蒸散发( $E_T$ )作为全球水循环的第二大组成部分,在生态系统内部与外界进行能量和物质交换过程中发挥着重要作用,同时还影响着许多其他重要的生态过程<sup>[26-27]</sup>。本研究于北京松山自然保护区生态环境及生物多样性监测站开展,以天然落叶阔叶林生态系统为研究对象,通过涡度相关法连续监测数据,分析比较落叶阔叶林生态系统  $F_{LE}$  缺失值插补的不同方法。本研究假设:插补结果会低估  $F_{LE}$ ,同时,不同插补方法得到的  $F_{LE}$  模拟值有差异。本研究主要解决以下 2 个问题:①插补  $F_{LE}$  与实测  $F_{LE}$  是否存在偏差?②插补  $F_{LE}$  与环境因子间关系是否和实测  $F_{LE}$  与环境因子间关系相同?

## 1 研究地区与方法

### 1.1 研究区概况

松山自然保护区生态环境及生物多样性监测站位于北京市延庆区松山国家级自然保护区内(40°30'48"N, 115°47'11"E),离延庆区张山营镇中心约 10 km,离北京中心约 104 km,北侧与大海坨自然保护区相邻,东南侧为佛峪口水库。

研究区属于北温带,气候为大陆性季风气候。研究区受所处地形的影响,气温低,湿度高,是典型的山地气候。降水季节分布不均,保证率低,多年平均降水量为 450.0 mm,集中在 6—9 月;多年平均气温为 8.9 °C,极端最高气温为 38.1 °C,极端最低气温为 -28.6 °C,多年平均无霜期为 153 d。2018 年 12 月,以通量塔为中心,在冬奥延庆赛区外围松山国家级自然保护区设立 50 m×50 m 的样方(40°53'N, 116°18'E,海拔为 1161 m)。样地内主要乔木树种为胡桃楸 *Juglans mandshurica*,其他乔木树种为大果榆 *Ulmus macrocarpa*、白蜡 *Fraxinus chinensis* 等。灌木以绣线菊 *Spiraea salicifolia*、铁线莲 *Clematis florida* 为主。经样地调查,胡桃楸平均胸径为 10.96 cm,平均树高为 4.71 m;大果榆平均胸径为 7.20 cm,平均树高为 3.88 m;白蜡平均胸径为 7.64 cm,平均树高为 3.97 m。

### 1.2 涡动相关和气象测量

通量监测采用开路式涡度相关系统,仪器初始安装高度为 20 m。测量系统主要由三维超声风速仪(WindMaster, Campbell Scientific, Inc. Logan)、开路 CO<sub>2</sub>/H<sub>2</sub>O 红外气体分析仪(7500 DS, Campbell Scientific, Inc. Logan)和数据采集器(CR1000, Campbell Scientific, Inc. Logan)构成。同步测量的气象变量包括:净辐射[四分量辐射仪(CNR-4, Campbell Scientific, Inc. Logan)]、空气温度与湿度[空气温湿度传感器(ATMOS14, Campbell Scientific, Inc. Logan, )]、降雨量[雨量桶(ECRN-100, Campbell Scientific, Inc. Logan)],传感器

安装高度均为 20 m。采用 CR1000 数据采集器 (CR1000, Campbell Scientific, Inc. Logan) 采集气象数据, 采样频率与涡度相关监测频率同步。

通量塔 5 m 半径内随机布设的 5 套热通量板 (HFP01, Campbell Scientific, Inc. Logan) 监测土壤热通量, 传感器布设深度为 10 cm。土壤温度和土壤含水量 ( $C_{SW}$ ) 分别由通量塔 5 m 半径内随机布设的 5 套土壤温度传感器 (5TM, Campbell Scientific, Inc. Logan) 和土壤水分传感器 (5TM, Campbell Scientific, Inc. Logan) 测定, 布设深度为 10 cm。采用 CR1000 数据采集器 (CR1000, Campbell Scientific, Inc. Logan) 采集土壤监测数据。

### 1.3 数据处理

使用 EddyPro 软件 (version 4, LI-COR, Lincoln) 对原始 10 Hz 数据进行预处理, 处理步骤包括峰值去除、倾斜校正 (双轴旋转)、传感器滞后校正、光谱校正、去趋势化和 0.5 h 通量计算<sup>[20]</sup>。本研究采用 PAPALE 等<sup>[28]</sup>的方法剔除 0.5 h 通量数据异常值, 利用数据相对绝对中位偏差 ( $M_{AD}$ ) 的偏离程度识别异常数据。具体如下:

$$d_i = (F_{LEi} - F_{LEi-1}) - (F_{LEi+1} - F_{LEi})。$$

其中:  $F_{LEi}$ 、 $F_{LEi-1}$ 、 $F_{LEi+1}$  分别为第  $i$  时刻及其前后时刻的  $F_{LE}$  数值,  $d_i$  为  $F_{LEi}$  在数列中的相对位置,  $d_i$  如果满足以下条件则被定义为异常值:

$$d_i < M_d - \frac{z \times M_{AD}}{0.6745} \text{ 或 } d_i > M_d + \frac{z \times M_{AD}}{0.6745}。$$

其中:  $z$  为阈值, 不同的  $z$  可以用来评价该方法对数据的影响, 本研究中  $z$  取 4<sup>[28]</sup>。 $M_d$  为数列  $d_i$  的中位数。 $M_{AD}$  的计算公式如下:

$$M_{AD} = \text{median}(|d_i - M_d|)。$$

通过自举法确定摩擦风速阈值<sup>[28]</sup>, 夜间摩擦风速低于阈值的湍流均认为交换不均匀, 剔除对应  $F_{LE}$  值。本研究中摩擦风速阈值为  $0.204 \text{ m} \cdot \text{s}^{-1}$ 。另外, 为了避免由于红外气体分析仪光路受影响导致的测量误差, 光路信号值低于 70 的数据也被剔除。2019 年由于仪表故障、系统维护和质量控制等原因剔除掉的  $F_{LE}$  占全年数据的 37.96%。

### 1.4 缺失值插补方法

1.4.1 边缘分布抽样法 边缘分布抽样法 (MDS) 通过 RStudio 1.1.463 中的 REddyProc 程序包完成<sup>[29]</sup>。该算法主要分为 2 种不同的情况: ①只缺失主要数据, 其他气象数据可用; ②部分或所有气象数据缺失。在第 1 种情况下, 搜索 7 d 时间窗口内相似气象条件下是否存在数值, 如果有, 用满足条件数值的平均值插补缺失值。其中相似气象条件为总辐射 ( $R_G$ ,  $\text{W} \cdot \text{m}^{-2}$ )、空气温度 ( $T_a$ ,  $^{\circ}\text{C}$ ) 和饱和水汽压差 ( $D_{VP}$ , Pa) 相较于缺失值对应的  $R_G$ 、 $T_a$ 、 $D_{VP}$  变化幅度不超过  $50 \text{ W} \cdot \text{m}^{-2}$ 、 $2.5 \text{ }^{\circ}\text{C}$  和 500 Pa。如果在 7 d 时间窗口内没有相似气象条件, 则将时间窗口增加到 14 d。以此类推。

在第 1 种情况下, 14 d 时间窗口内相似条件下存在数值, 插补数据质量为高等; 21 d 时间窗口内相似条件下存在数值, 插补数据质量为中等; 28 d 时间窗口内相似条件下存在数值, 插补数据质量为低等。在第 2 种条件下,  $F_{LE}$  缺失数据的时间间隔小于 1 h, 插补数据质量为高等;  $F_{LE}$  缺失数据的时间间隔小于 1 d, 插补数据质量为中等;  $F_{LE}$  缺失数据的时间间隔大于 1 d, 插补数据质量为低等。

1.4.2 线性回归法 线性回归法主要采用 AMIRO 等<sup>[30]</sup>的方法。缺失值的时间间隔小于 2 h 采用线性内插法插补, 其中夜间数据 (即  $R_G < 20 \text{ W} \cdot \text{m}^{-2}$ ) 设为 0。缺失值的时间间隔超过 2 h 分为生长季和非生长季 2 种情况。生长季 ( $T_a > 0 \text{ }^{\circ}\text{C}$ ) 采用 240 个连续数据时间窗口内的  $F_{LE}$  与净辐射减去土壤热通量 (即可利用能量) 线性回归插补, 时间窗口每次移动 48 个时间点 (1 个时间点为 0.5 h)。非生长季 ( $T_a < 0 \text{ }^{\circ}\text{C}$ ) 内采用缺失值前 5 d 至后 5 d 对应时间的平均值插补。

1.4.3 人工神经网络法 人工神经网络 (ANN) 是一种可以映射输入与输出数据间非线性关系的模型。本研究使用 BP 神经网络, 模型为 4 层结构。输入层包含用于进行模型预测的自变量, 输出层包含需要预测的变量, 输入层和输出层由 2 个隐藏层连接。第 1 个隐藏层的激励函数为 tansig 函数, 第 2 个隐藏层的激励函数为线性函数, 训练方法为 Levenberg-Marquardt。1 层中的每个神经节点都与相邻层的所有神经节点相连。第 1 层隐藏层的神经节点数量范围由下式确定:

$$m = \sqrt{n+l} + \alpha。$$

其中： $m$  为第 1 层隐含层神经节点数， $n$  和  $l$  为输入层和输出层变量个数， $\alpha$  为 1~10 的正整数。第 2 层隐藏层神经元个数为 1。

本研究中输入变量为  $R_G$ 、 $T_a$ 、 $D_{VP}$  和  $C_{SW}$ 。为了强调昼夜造成的影响而加入  $R_G$ ，同时将  $R_G < 20 \text{ W} \cdot \text{m}^{-2}$  的值设置为 0。白天输入变量为  $R_G$ 、 $T_a$ 、 $D_{VP}$ 、 $C_{SW}$ ；夜晚  $R_G=0$ ，输入变量为  $T_a$ 、 $D_{VP}$ 、 $C_{SW}$ 。本研究采用均方误差 ( $E_{RMSE}$ ) 来衡量不同  $m$  构建的神经网络模型的优劣 (表 1)。 $m=9$  的神经网络模型的  $E_{RMSE}$  最低，故本研究中  $m$  取 9。

表 1 第 1 层隐藏层不同神经元个数拟合结果

Table 1 Result of the fitting of the number of different neurons in ANNs

| 隐藏层神经元个数 | 均方误差    |         |         |         |         | 平均值     |
|----------|---------|---------|---------|---------|---------|---------|
|          | 1       | 2       | 3       | 4       | 5       |         |
| 3        | 0.009 7 | 0.010 0 | 0.009 9 | 0.009 6 | 0.010 0 | 0.009 8 |
| 4        | 0.009 3 | 0.009 8 | 0.009 6 | 0.009 8 | 0.010 0 | 0.009 7 |
| 5        | 0.009 4 | 0.009 9 | 0.010 3 | 0.009 8 | 0.009 9 | 0.009 9 |
| 6        | 0.009 4 | 0.009 4 | 0.008 8 | 0.009 1 | 0.009 3 | 0.009 2 |
| 7        | 0.008 9 | 0.009 3 | 0.009 9 | 0.009 5 | 0.009 6 | 0.009 5 |
| 8        | 0.009 4 | 0.009 2 | 0.009 3 | 0.008 6 | 0.008 7 | 0.009 0 |
| 9        | 0.008 6 | 0.008 8 | 0.008 5 | 0.009 2 | 0.009 1 | 0.008 8 |
| 10       | 0.009 4 | 0.009 0 | 0.009 5 | 0.009 1 | 0.009 1 | 0.009 2 |
| 11       | 0.009 1 | 0.009 1 | 0.009 1 | 0.008 9 | 0.009 3 | 0.009 1 |
| 12       | 0.009 2 | 0.010 0 | 0.009 3 | 0.009 3 | 0.009 1 | 0.009 4 |

说明：均方误差下的 1、2、3、4、5 分别指神经网络模拟的重复数(即不同的重复)。

## 1.5 统计分析

本研究中涡度相关系统采用开路式红外气体分析仪。为了避免出现冬季由于仪器运作散发热量导致的错误通量<sup>[31]</sup>，数据分析时只使用 2019 年 4 月 11 日至 10 月 27 日的数据。

为了评估 3 种数据插补方法的优良程度，采用随机取点的方法去除 30% 的  $F_{LE}$  有效实测数据 (2019 年选定日期内数据共 11 040 个，其中有效实测数据 7 576 个，即随机去除其中 2 272 个数据)，分别采用上述 3 种方法插补。插补完成后与实测数据进行比较。

由于线性回归法将生长季夜晚  $F_{LE}$  的缺失值直接填补为 0，为了更好地表现线性回归法的结果，本研究将其分为 2 种情况：第 1 种为正常线性回归插补 (线性回归法 I)，第 2 种为去除夜晚  $F_{LE}$  数据插补 (线性回归法 II)。

通过分段平均法分析人为去除的有效实测  $F_{LE}$  数据及 3 种数据插补结果与环境因子间的关系，其中线性回归法采用去除夜间  $F_{LE}$  插补数据进行分析。 $T_a$  按 2.5 °C 分段，即每隔 2.5 °C 取 1 次  $F_{LE}$  均值； $D_{VP}$  按 0.2 kPa 分段，分段平均后再拟合。 $T_a$  与  $F_{LE}$  使用指数方程  $y=a\exp(bx)$  进行拟合，其中  $y$  为  $F_{LE}$ ， $x$  为  $T_a$ ， $a$ 、 $b$  为拟合参数； $D_{VP}$  与  $F_{LE}$  使用指数-二次方程  $y=\exp(a+bx+cx^2)$  进行拟合，其中  $y$  为  $F_{LE}$ ， $x$  为  $D_{VP}$ ， $a$ 、 $b$ 、 $c$  为拟合参数。

指数-二次方程可以很好地模拟  $F_{LE}$  随  $D_{VP}$  增大先升高后降低的趋势，可以根据方程计算出最适  $D_{VP}$  ( $D_{VPOpt}$ ) 为  $-b/2c$ 。

除 MDS 方法使用 RStudio 软件外，其余所有分析均在 MATLAB (Version 7.5.0, The MathWorks) 中进行。

## 2 结果

### 2.1 生态系统微气象季节变化

研究区 2019 年空气温度 ( $T_a$ )、10 cm 深度土壤温度 ( $T_s$ ) 及净辐射 ( $R_n$ ) 日均值的变化特征相似，春、冬季较低，夏、秋季较高。研究期内平均  $R_n$  为  $104.85 \text{ W} \cdot \text{m}^{-2}$ ， $R_n$  日均值为  $-13.93 \sim 211.82 \text{ W} \cdot \text{m}^{-2}$  (图 1A)。研究期内平均  $T_a$  为  $13.25 \text{ }^\circ\text{C}$ ， $T_a$  日均值为  $-10.63 \sim 25.00 \text{ }^\circ\text{C}$ 。平均  $T_s$  为  $12.08 \text{ }^\circ\text{C}$ ， $T_s$  的日均值为  $-0.34 \sim 19.28 \text{ }^\circ\text{C}$  (图 1B)。平均  $D_{VP}$  为  $0.71 \text{ kPa}$ ，日均值为  $0.02 \sim 1.98 \text{ kPa}$  (图 1C)。研究期内降雨总量为  $503.3 \text{ mm}$ ，

降雨季节分布不均, 降雨大多出现在6—9月, 占全年降雨总量的78.4%(图1D)。10 cm深度土壤体积含水量( $C_{sw}$ )季节变化范围为 $0.129 \sim 0.301 \text{ m}^3 \cdot \text{m}^{-3}$ 。土壤热通量( $G$ )在春冬季为负值, 温度上升后开始逐渐增加, 在4月7日达到最大值( $11.50 \text{ W} \cdot \text{m}^{-2}$ ), 研究期内均值为 $-12.26 \text{ W} \cdot \text{m}^{-2}$ 。

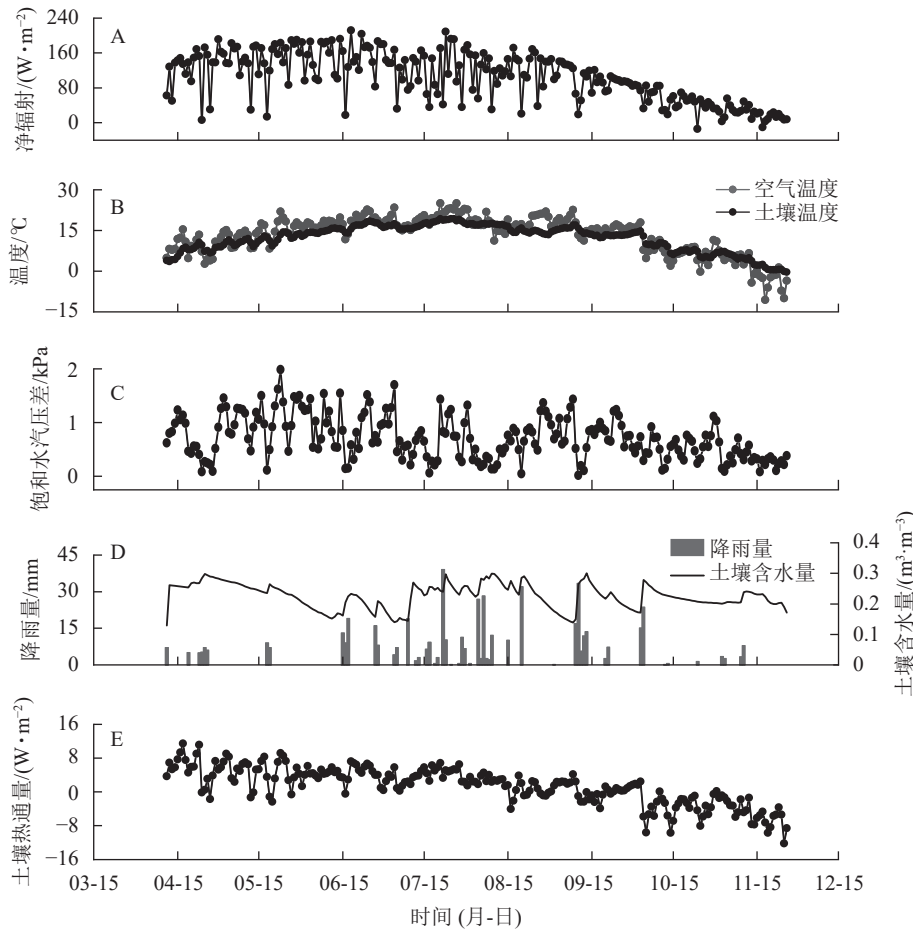


图1 2019年样地环境因子的季节变化

Figure 1 Dynamics in daily means of environmental variables in 2019 in the sample plots

## 2.2 3种插补方法结果及对环境因子的响应

实测 $F_{LE}$ 与3种插补方法模拟的 $F_{LE}$ 结果如图2所示, 其中插补 $F_{LE}$ 与实测 $F_{LE}$ 对比见图3。边缘分布抽样法、线性回归法及人工神经网络法得到的模拟值均低估了 $F_{LE}$ , 其中决定系数( $R^2$ )最高的为人工神经网络法, 回归斜率最高的为线性回归法I。插补 $F_{LE}$ 与实测 $F_{LE}$ 间回归直线截距均在0左右(图3)。

如图4所示: 3种插补结果中 $F_{LE}$ 与 $T_a$ 间拟合参数最接近原始数据的为边缘分布抽样法, 拟合效果最好( $R^2$ 最高)的为边缘分布抽样法。实测 $F_{LE}$ 及3种插补方法模拟的 $F_{LE}$ 均随 $T_a$ 增加而增加, 低温时 $F_{LE}$ 随 $T_a$ 上升而增加的幅度不大, 在 $15^\circ\text{C}$ 之后随 $T_a$ 上升有明显增加。线性回归法插补的 $F_{LE}$ 与 $T_a$ 间指数关系不明显。

3种插补结果中,  $F_{LE}$ 与 $D_{VP}$ 间分段平均后拟合曲线最接近原始数据的为边缘分布抽样法,  $R^2$ 最高的为人工神经网络法(图5)。实测 $F_{LE}$ 及3种插补 $F_{LE}$ 与 $D_{VP}$ 间关系均为指数-二次形式, 随 $D_{VP}$ 的增加先增加后减小。实测数据对应的最适 $D_{VP}(D_{VPopt})$ 为 $3.115 \text{ kPa}$ , 插补结果中 $D_{VP}$ 最小为边缘分布抽样法, 为 $3.538 \text{ kPa}$ , 最大 $D_{VPopt}$ 为线性回归法的 $4.178 \text{ kPa}$ ,  $D_{VPopt}$ 结果与原始数据最接近的为边缘分布抽样法。

## 3 讨论

### 3.1 插补数据与实测数据的比较

数据插补本质上是在类似气象条件下对有效数据的平均与模拟<sup>[8]</sup>。本研究中3种插补方法得到的 $0.5 \text{ h}$

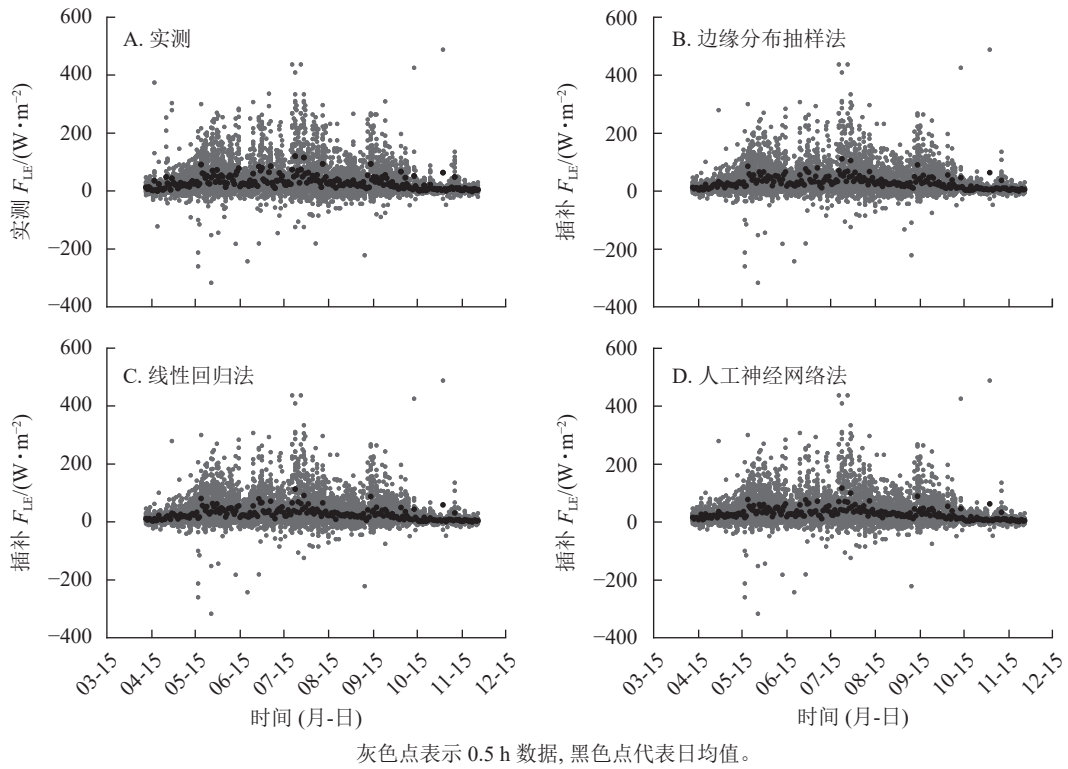


图 2 2019 年样地实测潜热通量 ( $F_{LE}$ ) 及 3 种插补方法结果

Figure 2 Results of measured latent heat flux ( $F_{LE}$ ) and the modeled  $F_{LE}$  with three gap-filling methods in 2019 in the sample plots

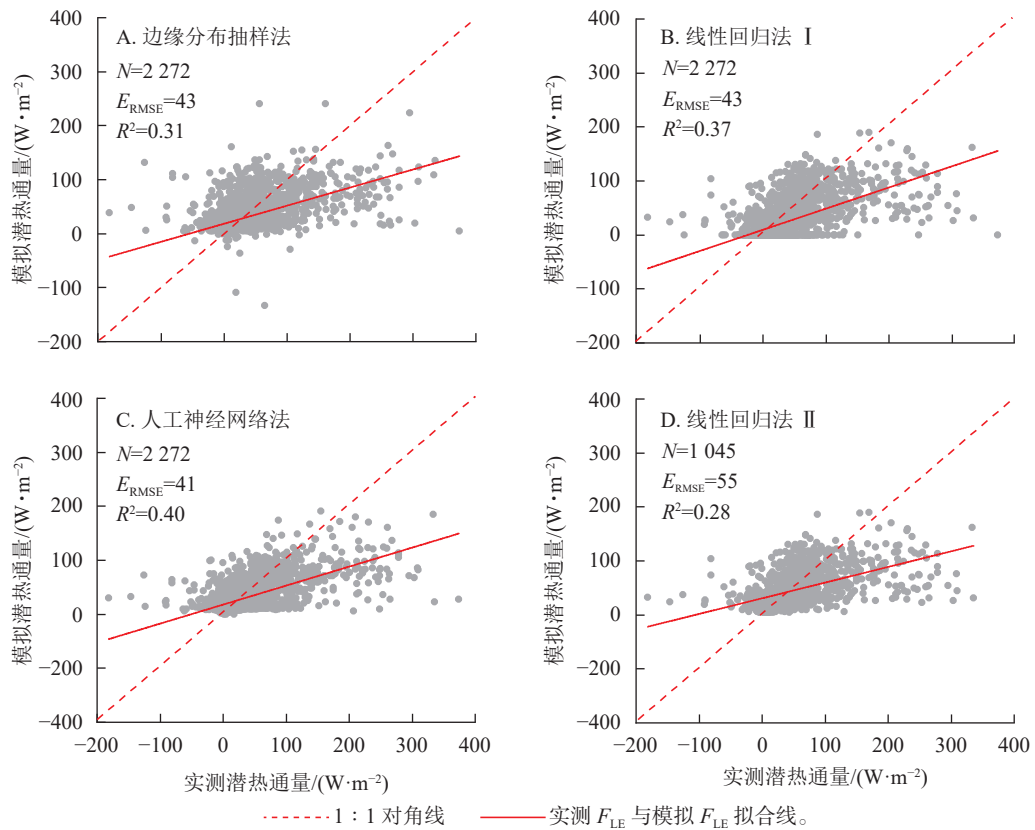
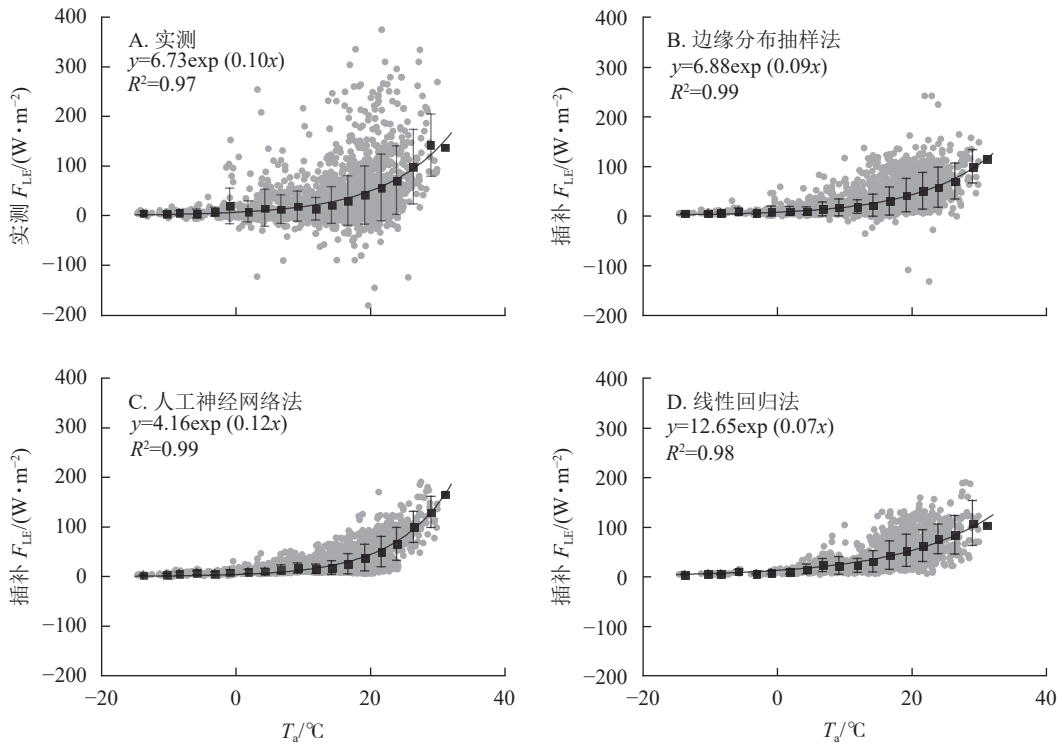


图 3 实测潜热通量 ( $F_{LE}$ ) 与模拟  $F_{LE}$  间关系

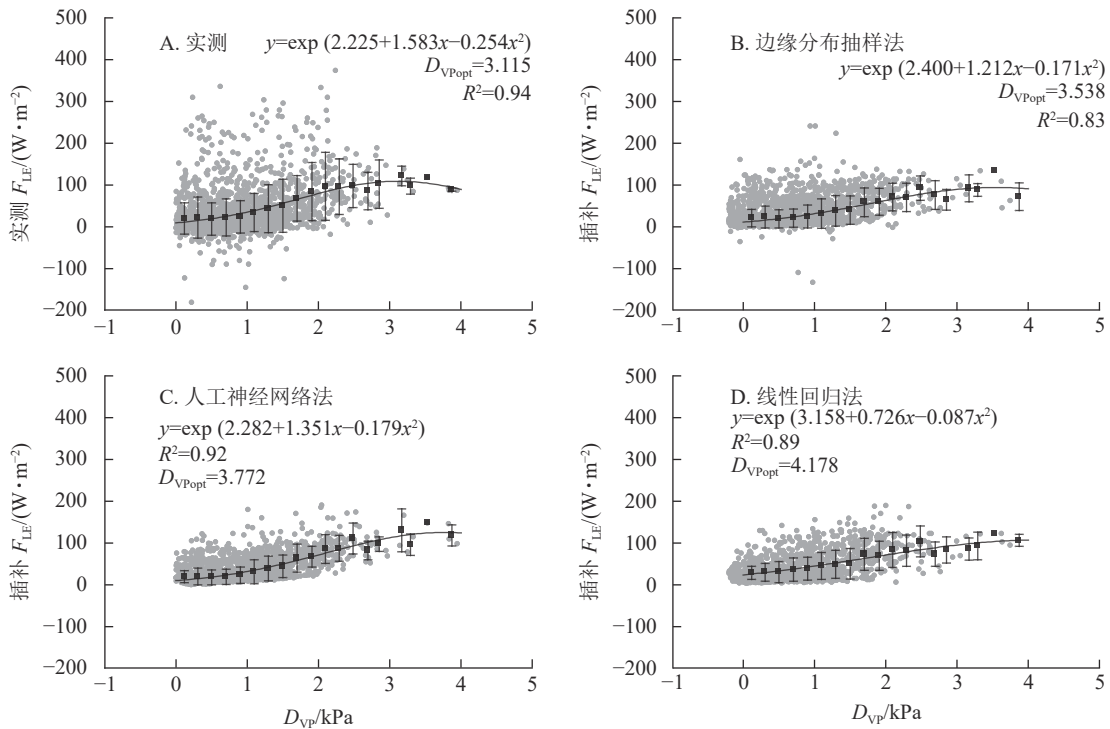
Figure 3 Relationship between the measured latent heat flux ( $F_{LE}$ ) and modeled  $F_{LE}$



灰点代表  $F_{LE}$ , 实线代表  $T_a$  与  $F_{LE}$  的拟合线。

图4 气温 ( $T_a$ ) 与实测潜热通量 ( $F_{LE}$ ) 及3种插补  $F_{LE}$  间的关系

Figure 4 Relationship between air temperature ( $T_a$ ) and measured latent heat flux ( $F_{LE}$ ) and the three modeled  $F_{LE}$



灰点代表  $F_{LE}$ , 实线代表  $D_{VP}$  与  $F_{LE}$  的拟合线。

图5 饱和水汽压差 ( $D_{VP}$ ) 与实测潜热通量 ( $F_{LE}$ ) 及3种插补  $F_{LE}$  间的关系

Figure 5 Relationship between vapor pressure deficit ( $D_{VP}$ ) and measured latent heat flux ( $F_{LE}$ ) and the three modeled  $F_{LE}$

数据与实测值间的回归斜率为 0.28~0.40, 即插补结果与实测  $F_{LE}$  相比均有一定程度的低估, 其中最接近实测数据的方法为人工神经网络法。涡度相关系统存在随机误差和系统误差(数据采集过程中的错误、仪器自身的偏差及空气的平流运动导致的误差等), 监测结果与真实值间会不可避免地出现偏差。该偏

差数学分布为标准正态分布，均值为 0<sup>[32]</sup>。

边缘分布抽样法的本质是取类似气象条件下实测有效的平均值进行插补<sup>[26]</sup>，插补值消除了随机偏差，但是这种方法并不能将缓慢变化的生物因素的影响考虑在内(如物候变化)。线性回归法同理。这可能是插补值与实测值间有偏差的原因之一。而人工神经网络法则是利用已有环境数据建模，再根据建立的模型进行反演，建模时已经将原始数据进行了最小二乘回归，即实测值的波动被消除一部分。本研究中插补值与实测值间的比较使用了 0.5 h 数据，插补值与实测值的日均值间差异可能很小。为了保证插补时有效数据的数量，本研究中无法连续剔除几天数据用于插补，否则会对插补结果造成影响，所以插补结果的日均值是否低估实测值，仍需进一步研究。

### 3.2 插补数据及实测数据与环境因子的关系

$F_{LE}$  与  $T_a$  呈指数形式。涡度相关中蒸散发由  $F_{LE}$  计算得到，蒸散发可以分为土壤蒸发和植物蒸腾 2 个部分，植物蒸腾主要受水势差控制，而土壤蒸发主要受  $T_a$  控制<sup>[33-34]</sup>。本研究中  $F_{LE}$  随  $T_a$  增加而增加，在气温较低时 ( $<15\text{ }^{\circ}\text{C}$ )  $F_{LE}$  随  $T_a$  增加幅度不大，而气温较高时增加幅度变大。这可能是由于气温较低时蒸散发组分增加都不明显，而气温较高时蒸散发开始以土壤蒸发为主，且研究地处于天然森林，土壤中水分充足，所以  $F_{LE}$  随  $T_a$  增加呈指数形式上升<sup>[35-36]</sup>。3 种插补结果中，边缘分布抽样法插补的  $F_{LE}$  和  $T_a$  间关系与实测值最接近，可能是因为边缘分布抽样法使用了实测数据进行平均插补，可以很好地重现  $F_{LE}$  与  $T_a$  间的关系。人工神经网络法与线性回归法分段平均后的曲线都与实测数据不符。这可能是因为人工神经网络法是先基于已有数据建模反演，建模时  $T_a$  权重较大，所以结果呈明显指数型。线性回归则是利用了  $R_n$  及  $G$  间的关系模拟  $F_{LE}$ ，所以  $F_{LE}$  与  $T_a$  间关系较差。

$F_{LE}$  与  $D_{VP}$  间关系呈指数-二次形式。 $D_{VP}$  较低时， $F_{LE}$  随  $D_{VP}$  增加趋势与  $T_a$  类似。 $D_{VP}$  由  $T_a$  与相对湿度决定，所以  $D_{VP}$  增加会促进生态系统蒸散发上升。随着  $D_{VP}$  的不断上升，植被蒸腾失水严重，植物为了保证自身正常生命活动，会主动关闭部分气孔降低自身水分损失，所以随着  $D_{VP}$  的不断上升， $F_{LE}$  会先升高后下降<sup>[37-39]</sup>。与  $T_a$  相同，边缘分布抽样法与  $D_{VP}$  间关系最接近实测值与  $D_{VP}$  的关系，人工神经网络法分段平均后  $F_{LE}$  随  $D_{VP}$  整体变化较快，线性回归法没有很好地模拟出  $F_{LE}$  随  $D_{VP}$  先平缓后增加的变化趋势。

### 3.3 3 种插补方法的比较

3 种插补方法各有优缺点，需视具体情况使用。研究表明：人工神经网络通常比其他方法更能准确插补通量缺失数据<sup>[40-41]</sup>。本研究中，将  $R_G$ 、 $T_a$ 、 $D_{VP}$ 、 $C_{SW}$  等影响  $F_{LE}$  的主要环境因素作为人工神经网络输入变量，模拟  $F_{LE}$  比线性回归与边缘分布抽样法更加贴近实测  $F_{LE}$ 。人工神经网络法是一种纯经验的“黑箱”模型，其预测能力取决于输入、输出数据集及模型训练过程，输入变量的选择对于其预测能力的影响很大<sup>[41]</sup>。因此，在预测变量(即气象数据)较全面、监测时序较长的情况下，优先选择人工神经网络对缺失数据进行插补，可以很好地重现  $F_{LE}$  动态变化。线性回归法从能量平衡的角度插补数据，这种方法显然没有考虑环境因子对  $F_{LE}$  的控制，无法重现与环境因子的关系<sup>[29]</sup>。边缘分布抽样法将气象条件与  $F_{LE}$  的变化同时考虑在内，可以较好地重现与环境因子的关系。因此，尽管线性回归法对  $F_{LE}$  动态变化的模拟结果略优于边缘分布抽样法，但是鉴于其对环境因子与  $F_{LE}$  关系的重现结果较差，在探究  $F_{LE}$  调控机制的研究工作中并不建议使用线性回归法插补  $F_{LE}$  缺失数据。当研究中驱动因子有效性不足时，可能需要使用边缘分布抽样法<sup>[28]</sup>；如果插补结果被用于模型验证，为了避免数据与模型间出现虚假或循环关系的现象，可能需要使用线性回归法等<sup>[27]</sup>。

## 4 结论

①3 种插补方法各有优缺点，得到的插补  $F_{LE}$  均对实测  $F_{LE}$  有一定低估，人工神经网络法插补结果与实测数据最为接近。②人工神经网络法会过分拟合  $F_{LE}$  与环境因子间的关系，线性回归插补结果与环境因子间的关系跟实测数据差别最大，边缘分布抽样法与环境因子间的关系最接近实测  $F_{LE}$ 。

## 5 参考文献

- [1] FRANK D, REICHSTEIN M, BAHN M, *et al.* Effects of climate extremes on the terrestrial carbon cycle: concepts,



- processes and potential future impacts [J]. *Global Change Biology*, 2015, **21**(8): 2861 – 2880.
- [2] JIANG Peng, LIU Hongyan, PIAO Shilong, *et al.* Enhanced growth after extreme wetness compensates for post-drought carbon loss in dry forests [J/OL]. *Nature Communications*, 2019, **10**: 195 [2023-09-24]. doi: 10.1038/s41467-018-08229-z.
- [3] MATSUMOTO K, OHTA T, NAKAI T, *et al.* Energy consumption and evapotranspiration at several boreal and temperate forests in the Far East [J]. *Agricultural and Forest Meteorology*, 2008, **148**(12): 1978 – 1989.
- [4] MCCAUGHEY J H, PEJAM M R, ARAIN M A, *et al.* Carbon dioxide and energy fluxes from a boreal mixedwood forest ecosystem in Ontario, Canada [J]. *Agricultural and Forest Meteorology*, 2006, **140**(1/4): 79–96.
- [5] KUME T, TANAKA N, KURAJI K, *et al.* Ten-year evapotranspiration estimates in a Bornean tropical rainforest [J]. *Agricultural and Forest Meteorology*, 2011, **151**(9): 1183 – 1192.
- [6] XIE Jing, ZHA Tianshan, ZHOU Caixian, *et al.* Seasonal variation in ecosystem water use efficiency in an urban-forest reserve affected by periodic drought [J]. *Agricultural and Forest Meteorology*, 2016, **221**: 142 – 151.
- [7] WU Jiabing, GUAN Dexin, HAN Shijie, *et al.* Energy budget above a temperate mixed forest in northeastern China [J]. *Hydrological Processes*, 2007, **21**(18): 2425 – 2434.
- [8] WILSON K B, GOLDSTEIN A, FALGE E, *et al.* Energy balance closure at flux net sites [J]. *Agricultural and Forest Meteorology*, 2002, **113**: 223 – 243.
- [9] HUANG Jianping, YU Haipeng, GUAN Xiaodan, *et al.* Accelerated dryland expansion under climate change [J]. *Nature Climate Change*, 2016, **6**: 166 – 171.
- [10] FORNER A, VALLADARES F, BONAL D, *et al.* Extreme droughts affecting mediterranean tree species' growth and water-use efficiency: the importance of timing [J]. *Tree Physiology*, 2018, **38**(8): 1127 – 1137.
- [11] BALDOCCHI D, FALGE E, GU Lianhong, *et al.* FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities [J]. *Bulletin of the American Meteorological Society*, 2001, **82**: 2415 – 2434.
- [12] PAPAIE D, VALENTINI R. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization [J]. *Global Change Biology*, 2003, **9**(4): 525 – 535.
- [13] JIA Xin, ZHA Tianshan, GONG Jinnan, *et al.* Energy partitioning over a semi-arid shrubland in northern China [J]. *Hydrological Process*, 2016, **30**(6): 972 – 985.
- [14] HANSON P J, AMTHOR J S, WULLSCHLEGER S D, *et al.* Carbon and water cycle simulations for an upland oak forest using 13 stand-level models: intermodel comparisons and evaluations against independent measurements [J]. *Ecological Monographs*, 2004, **74**(3): 443 – 489.
- [15] GONG Jinnan, JIA Xin, ZHA Tianshan, *et al.* Modeling the effects of plant-interspace heterogeneity on water-energy balances in a semiarid ecosystem [J]. *Agricultural and Forest Meteorology*, 2016, **221**: 189 – 206.
- [16] GONG Jinnan, WANG Ben, JIA Xin, *et al.* Modelling the diurnal and seasonal dynamics of soil CO<sub>2</sub> exchange in a semiarid ecosystem with high plant-interspace heterogeneity [J]. *Biogeosciences*, 2018, **15**(1): 115 – 136.
- [17] WUTZLER T, LUCAS-MOFFAT A, MIGLIAVACCA M, *et al.* Basic and extensible post-processing of eddy covariance flux data with REddyProc [J]. *Biogeosciences*, 2018, **15**(16): 5015 – 5030.
- [18] JIA Xin, MU Yu, ZHA Tianshan, *et al.* Seasonal and interannual variations in ecosystem respiration in relation to temperature, moisture, and productivity in a temperate semi-arid shrubland [J/OL]. *Science of the Total Environment*, 2020, **709**: 136210 [2023-09-24]. doi: 10.1016/j.scitotenv.2019.136210.
- [19] FALGE E, TENHUNEN J, BALDOCCHI D, *et al.* Phase and amplitude of ecosystem carbon release and uptake potentials as derived from FLUXNET measurements [J]. *Agricultural and Forest Meteorology*, 2002, **113**(1/4): 75 – 95.
- [20] MA Jingyong, ZHA Tianshan, JIA Xin, *et al.* Energy and water vapor exchange over a young plantation in northern China [J]. *Agricultural and Forest Meteorology*, 2019, **263**: 334 – 345.
- [21] LIU Peng, BLACK T A, JASSAL R S, *et al.* Divergent long-term trends and interannual variation in ecosystem resource use efficiencies of a southern boreal old black spruce forest 1999–2017 [J]. *Global Change Biology*, 2019, **25**(9): 3056 – 3069.
- [22] PASTORELLO G, TROTTA C, CANFORA E, *et al.* The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data [J/OL]. *Scientific Data*, **72** [2023-09-24]. doi: 10.1038/s41597-021-00851-9.
- [23] BARR A G, BLACK T A, HOGG E H, *et al.* Climatic controls on the carbon and water balances of a boreal aspen forest, 1994–2003 [J]. *Global Change Biology*, **13**(3): 561–576.
- [24] BALDOCCHI D, KNOX S, DRONOVA I, *et al.* The impact of expanding flooded land area on the annual evaporation of rice [J]. *Agricultural and Forest Meteorology*, 2016, **223**: 181 – 193.
- [25] TSURUTA K, KOSUGI Y, TAKANASHI S, *et al.* Inter-annual variations and factors controlling evapotranspiration in a temperate Japanese cypress forest [J]. *Hydrological Process*, 2016, **30**(26): 5012 – 5026.

- [26] MIAO Haixia, CHEN Shiping, CHEN Jiquan, *et al.* Cultivation and grazing altered evapotranspiration and dynamics in Inner Mongolia steppes [J]. *Agricultural and Forest Meteorology*, 2009, **149**(11): 1810 – 1819.
- [27] POWELL T L, STARR G, CLARK K L, *et al.* Ecosystem and understory water and energy exchange for a mature, naturally regenerated pine flatwoods forest in north Florida [J]. *Canadian Journal of Forest Research*, 2005, **35**: 1568 – 1580.
- [28] PAPALE D, REICHSTEIN M, AUBINET M, *et al.* Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: algorithms and uncertainty estimation [J]. *Biogeosciences*, 2006, **3**(4): 571 – 583.
- [29] REICHSTEIN M, FALGE E, BALDOCCHI D, *et al.* On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm [J]. *Global Change Biology*, 2005, **11**(9): 1424 – 1439.
- [30] AMIRO B D, BARR A G, BLACK T A, *et al.* Carbon, energy and water fluxes at mature and disturbed forest sites Saskatchewan, Canada [J]. *Agricultural and Forest Meteorology*, 2006, **136**(3/4): 237 – 251.
- [31] BURBA G G, MCDERMITT D K, GRELLA A, *et al.* Addressing the influence of instrument surface heat exchange on the measurements of CO<sub>2</sub> flux from open-path gas analyzers [J]. *Global Change Biology*, 2008, **14**(8): 1854 – 1876.
- [32] MAUDER M, CUNTZ M, DRÜE C, *et al.* A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements [J]. *Agricultural and Forest Meteorology*, 2013, **169**: 122 – 135.
- [33] 武夏宁, 胡铁松, 王修贵, 等. 区域蒸散发估算测定方法综述[J]. *农业工程学报*, 2006, **22**(10): 257 – 262.  
WU Xianing, HU Tiesong, WANG Xiugui, *et al.* Review of estimating and measuring regional evapotranspiration [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2006, **22**(10): 257 – 262.
- [34] 王怡宁, 张晓萌, 路璐, 等. 通径分析结合BP神经网络方法估算夏玉米作物系数及蒸散量[J]. *农业工程学报*, 2020, **36**(7): 109 – 116.  
WANG Yining, ZHANG Xiaomeng, LU Lu, *et al.* Estimation of crop coefficient and evapotranspiration of summer maize by path analysis combined with BP neural network [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2020, **36**(7): 109 – 116.
- [35] 张川, 闫浩芳, 大上博基, 等. 表层有效土壤水分参数化及冠层下土面蒸发模拟[J]. *农业工程学报*, 2015, **31**(2): 102 – 107.  
ZHANG Chuan, YAN Haofang, OUE Hiroki, *et al.* Parameterization of surface soil available moisture and simulation of soil evaporation beneath canopy [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015, **31**(2): 102 – 107.
- [36] 吴友杰, 杜太生. 基于氧同位素的玉米农田蒸散发估算和区分[J]. *农业工程学报*, 2020, **36**(4): 127 – 134.  
WU Youjie, DU Taisheng. Estimating and partitioning evapotranspiration of maize farmland based on stable oxygen isotope [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2020, **36**(4): 127 – 134.
- [37] 李鑫豪, 闫慧娟, 卫腾宙, 等. 油蒿资源利用效率在生长季的相对变化及对环境因子的响应[J]. *植物生态学报*, 2019, **43**(10): 889 – 898.  
LI Xinhao, YAN Huijuan, WEI Tengzhou, *et al.* Relative changes of resource use efficiencies and their responses to environmental factors in *Artemisia ordosica* during growing season [J]. *Chinese Journal of Plant Ecology*, 2019, **43**(10): 889 – 898.
- [38] 高冠龙, 冯起, 刘贤德, 等. 3种经验模型模拟荒漠河岸柽柳叶片气孔导度[J]. *生态学报*, 2020, **40**(10): 3486 – 3494.  
GAO Guanlong, FENG Qi, LIU Xiande, *et al.* Simulating the leaf stomatal conductance of the desert riparian *Tamarix ramosissima* Ledeb. based on three empirical models [J]. *Acta Ecologica Sinica*, 2020, **40**(10): 3486 – 3494.
- [39] 周文君, 查天山, 贾昕, 等. 宁夏盐池油蒿叶片水分利用效率的生长季动态变化及对环境因子的响应[J]. *北京林业大学学报*, 2020, **42**(7): 98 – 105.  
ZHOU Wenjun, ZHA Tianshan, JIA Xin, *et al.* Dynamics of water use efficiency of *Artemisia ordosica* leaf in growing season in response to environmental factors in Yanchi, Ningxia of northwestern China [J]. *Journal of Beijing Forestry University*, 2020, **42**(7): 98 – 105.
- [40] 冯新妍, 贾昕, 黄金泽, 等. ANN-BiLSTM 模型在温带荒漠灌丛碳通量数据缺失值插补中的应用[J]. *北京林业大学学报*, 2023, **45**(9): 62 – 72.  
FENG Xinyan, JIA Xin, HUANG Jinze, *et al.* Application of ANN-BiLSTM model to long-term gap-filling of carbon flux data in temperate desert shrub [J]. *Journal of Beijing Forestry University*, 2023, **45**(9): 62 – 72.
- [41] LIU Jianzhao, ZUO Yunjiang, WANG Nannan, *et al.* Comparative analysis of two machine learning algorithms in predicting site-level net ecosystem exchange in major biomes [J/OL]. *Remote Sensing*, 2021, **13**(12): 2242[2023-09-24]. doi: 10.3390/rs13122242.