DOI:10.12171/j.1000−1522.20240187

TSIT-PatchTST 模型:一种净生态系统交换量（NEE）缺失值插补方法

齐建东[1,2]　石成城[1]　吴　鹏[1]
（1. 北京林业大学信息学院，北京 100083；2. 国家林业和草原局林业智能信息处理工程技术研究中心，北京 100083）

摘要:【目的】净生态系统交换量（NEE）是评估陆地生态系统在全球碳循环中作用的重要指标，NEE 原始观测数据缺失值的插补精度会直接影响生态系统关键参数的可靠性和精确性。为提高不同植被 NEE 在长时间连续性数据缺失情景下的插补精度，提出一种融合时间序列表征向量的 TSIT-PatchTST 深度学习模型。**【方法】**以全球长期通量观测网络站点的碳通量因子数据为研究对象，通过构造短缺失（1 d）、中缺失（7 d）、长缺失（30 d）3 种随机连续数据缺失场景，评估边际分布采样法（MDS）、PatchTST 模型、TS2Vec-PatchTST 模型和 TSIT-PatchTST 模型在 8 种不同植被类型下 NEE 的插补结果。**【结果】**在短缺失场景下，4 种插补方法都表现出最优的性能。随着连续缺失天数的增多，MDS 的插补精度逐渐下降，该方法在长缺失场景下已不能对 NEE 进行有效插补，而其他 3 种深度学习模型能够有效地插补 NEE 缺失数据。综合 3 种缺失场景，TSIT-PatchTST 模型表现出最优的插补性能，尤其在长缺失场景下该模型具有较高的插补精度。长缺失场景下，TSIT-PatchTST 模型对 31 个站点插补结果的平均均方误差（MSE）为 0.942 μmol/(m^2·s)，平均绝对误差（MAE）为 0.628 μmol/(m^2·s)，平均 R^2 为 0.457。与 PatchTST 模型的插补结果相比，TSIT-PatchTST 模型平均 MSE 下降了 53.3%，平均 MAE 下降了 39.7%，平均 R^2 相持平。**【结论】**综合 8 种植被类型和 3 种缺失场景的应用结果，得出 TSIT-PatchTST 模型的插补效果最佳，并具有适应性。TSIT-PatchTST 模型可应用于时间序列数据缺失场景以提高数据插补精度。

关键词：深度学习；模型开发；数据插补；TSIT-PatchTST 模型；碳循环；植被类型；净生态系统交换量（NEE）
中图分类号：S718.5；TP181　　**文献标志码：**A　　**文章编号：**1000−1522(2025)02−0105−14
引文格式：齐建东，石成城，吴鹏. TSIT-PatchTST 模型：一种净生态系统交换量（NEE）缺失值插补方法［J］. 北京林业大学学报，2025，47(2)：105−118. Qi Jiandong, Shi Chengcheng, Wu Peng. TSIT-PatchTST model: a missing value interpolation method for net ecosystem exchange (NEE)［J］. Journal of Beijing Forestry University, 2025, 47(2): 105−118.

TSIT-PatchTST model: a missing value interpolation method for net ecosystem exchange (NEE)

Qi Jiandong[1,2]　Shi Chengcheng[1]　Wu Peng[1]
(1. School of Information, Beijing Forestry University, Beijing 100083, China;
2. Engineering Research Center for Forestry-Oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China)

Abstract:［**Objective**］Net ecosystem exchange (NEE) is an important indicator for evaluating the role of terrestrial ecosystems in the global carbon cycle. The accuracy of imputation of missing values in NEE raw observation data directly affects the reliability and precision of key ecosystem parameters. To enhance the imputation accuracy of NEE in scenarios of continuous long-term data gaps across different vegetation types, a TSIT-PatchTST model was proposed based on deep learning.［**Method**］Using carbon flux factor data from sites within the global long-term flux observation network as the research object, three types of

收稿日期: 2024−06−11　**修回日期:** 2024−12−29
基金项目: 国家自然科学基金项目（32071842），国家重点研发计划项目（2022YFD2200304）。
第一作者: 齐建东，博士，教授。主要研究方向：生态信息学、智能信息处理。Email：qijd@bjfu.edu.cn　地址：100083 北京市海淀区清华东路 35 号北京林业大学信息学院。
本刊网址: http://j.bjfu.edu.cn；http://journal.bjfu.edu.cn

random continuous data gap scenarios were constructed, including short missing (1 d), medium missing (7 d), and long missing (30 d). The imputation results of marginal distribution sampling (MDS) method, PatchTST model, TS2Vec-PatchTST model, and TSIT-PatchTST model under eight different vegetation types were evaluated. [**Result**] In the scenario of short missing, all imputation methods demonstrated optimal performance. As the number of consecutive missing days increased, the imputation accuracy of MDS method gradually declined, and it was no longer effective for imputing NEE in the long missing scenario. In contrast, the three deep learning models were capable of effectively imputing missing NEE data. Considering all three missing scenarios, the TSIT-PatchTST model exhibited the best imputation performance, particularly with a high accuracy in long missing scenarios. In the long missing scenario, the TSIT-PatchTST model achieved an average mean squared error (MSE) of 0.942 μmol/(m^2·s), an average mean absolute error (MAE) of 0.628 μmol/(m^2·s), and an average R^2 of 0.457 across 31 sites. Compared with PatchTST model, the TSIT-PatchTST model reduced the average MSE by 53.3%, average MAE by 39.7% and the average R^2 remained unchanged. [**Conclusion**] Integrating the performance across eight vegetation types and three missing scenarios, the TSIT-PatchTST model demonstrates the best imputation effect and adaptability. It can be applied to the problem of missing data in time series to improve the accuracy of data imputation.

Key words: deep learning; model development; data imputation; TSIT-PatchTST model; carbon cycle; vegetation types; net ecosystem exchange (NEE)

净生态系统交换量（net ecosystem exchange，NEE）是指陆地生态系统与大气之间的 CO_2 通量差值，反映了陆地生态系统对大气 CO_2 的吸收和释放能力，是评估陆地生态系统在全球碳循环中作用的重要指标。涡度相关法是目前最常用的，能直接、长期观测 NEE 的技术[1]。然而，原理、仪器构造和观测系统的故障会导致原始观测数据出现一定程度的缺失。缺失比例大致在总数据量的 17% ~ 50% 之间，其中甚至会出现较长时间段的连续数据缺失情况[2]。长期连续观测数据是生态学研究的基础，因此需要对缺失的观测数据进行插补。不同插补方法的计算结果会直接影响到诸如 NEE 年积累量等生态系统关键参数的可靠性和精确性[3−4]，因此选择合适的方法至关重要。当前，国际上主要的生态观测网络中，全球通量网 FLUXNET、CarboEurope 和 AmeriFlux 采用平均值插补方法或边际分布采样法（marginal distribution sampling，MDS）[5]；ChinaFLUX、日本通量网采用 MDS、非线性回归方法或人工神经网络进行插补[6]。上述大型观测网络从一般性、普遍性出发，主要采用 MDS 数据插补方法，并未充分考虑观测数据的缺失长度、植被类型，以及观测数据的内在数据分布特征等重要因素。

目前，针对 NEE 通量数据的插补方法可分为两大类：一是基于生态机理模型[7]，需要明确的生物物理、生物化学等理论支撑[8]，而这有赖于学者对自然规律的逐步发现；二是以统计学习、机器学习为主的数据驱动模型，数据驱动模型又可进一步分为以 MDS 为代表的统计模型[3]，以人工神经网络为代表的机器学习模型[9]、深度学习模型[10]。随着长期连续观测工作的推进，不断累积的大数据使得数据驱动模型逐渐获得越来越多的关注[11]。从数据科学角度，数据量越多，数据质量越好，数据插补或预测的效果就越好。鉴于此，以随机森林为代表的传统机器学习模型已经在通量数据插补中得到应用，但仍存在诸如插补效果在不同场景下存在显著差异等问题[12−13]。

近年来，深度学习模型在时间序列数据分类、预测等研究领域取得长足进步，在捕获和利用时间序列数据中的数据长期依赖关系、多因子间非线性关系建模等方面表现优异。因此，基于深度学习的生态监测数据插补或预测等研究工作得到关注。Guo 等[14]基于涡度相关观测数据，使用长短期记忆网络（long short term memory，LSTM）估算全球 30 m 分辨率的陆地潜热通量，为全球陆地潜热通量估算提供了更准确的方法。冯新妍等[15]使用 Bi-LSTM 模型开展了针对温带荒漠灌丛碳通量数据缺失值插补的研究工作，模型计算结果表明，在数据缺失长度 > 30 d 时，均方根误差和平均绝对误差结果较好，模型提高了数据插补结果的精度。Qian 等[16]使用卷积神经网络（convolutional neural network，CNN）、LSTM 等模型，在 FLUXNET 中选取 241 个站点开展了针对蒸散发数据的缺失数据插补工作，从不同的纬度、土地覆盖类型和气候类型等方面对研究站点进行划分，评估了 LSTM 模型的插补精度。Jerse 等[17]使用

LSTM 模型针对太阳射电流量进行分析预测，并评估了在不同时间滞后和太阳活动水平下的模型性能。Chen 等[18]使用 LSTM 模型并引入迁移学习框架 TrAdaBoost，实现了从完整序列（源域）到不完整序列（目标域）的知识迁移，从而提高了水质缺失数据的插补精度。目前，大多数工作主要集中在使用 ANN、CNN、LSTM 等模型上，在面向时间序列数据分析的深度学习模型探索和应用上存在不足[19]。本研究拟融合 PatchTST 模型、TS2Vec 方法和 ITransformer 架构，保留各响应因子与 NEE 之间的非线性关系，并考虑 NEE 在时间上的自相关性，设计并实现一种深度学习模型 TSIT-PatchTST，旨在更准确地模拟 NEE 变化的内在机制，提高不同数据缺失场景下的 NEE 数据插补精度和适用性。

1　数据源和研究区域

1.1　数据源

数据资料来自 FLUXNET2015 数据集的 FULLSET 数据产品[20]。FLUXNET2015 数据集包含来自 206 个全球分布站点的公开数据，数据以每半小时为尺度，包括经过质量控制的 NEE 通量，以及相关的气象和生态学变量。使用 Python 程序对 FLUXNET2015 上的各站点原始数据集进行过滤，截取各站点连续时间最长的一段数据用于本研究。通过检验，其中 60 个站点缺失需要的特征指标，115 个站点连续时间段内数据不足 8 000 条，即站点可用数据量少于半年数据量，不满足本研究 NEE 观测数据的季节性变化特征需要。最终选择 31 个 FLUXNET 站点的数据集。

1.2　研究区域

各站点所在区域共计 8 种植被类型 31 个站点，分别为 1 个农田站点、3 个落叶阔叶林站点、10 个常绿针叶林站点、9 个草地站点、1 个混合林站点、2 个开放灌木地站点、3 个稀树草原站点、2 个木本灌草丛站点，分别分布在欧洲、大洋洲、北美洲（表 1）。

2　研究方法

2.1　NEE 响应因子选择

NEE 是植物光合作用和呼吸作用的产物，它的产生过程主要受植物的生长环境影响。在文献[12,21]基础上，本研究在 FLUXNET 2015 数据集中选取了包括短波辐射（SW_IN_F）、饱和水汽压差（VPD_F_MDS）、空气温度（TA_F_MDS）、净辐射能量（NETRAD）、风速（WS）、风向（WD）、土壤热通量（G_F_MDS）、土壤温度（TS_F_MDS）、相对湿度（RH）、土壤含水量（SWC_F_MDS）共 10 个 NEE 响应因子的站点及数据，各指标详情如表 2 所示。31 个站点数据集中包含了 MDS 所需的 3 种响应因子，因此使用 MDS 进行数据插补时，无需再对数据集进行区分。使用 FLUXNET 数据处理协议中的默认方法对响应因子中的缺失值进行插补[20]，上述因子中包含“*_MDS”的即为插补后的因子。本研究选择数据产品 NEE_VUT_REF 中的数值作为实验 NEE 目标值，NEE_VUT_REF 是使用跨年度的恒定 U*阈值（constant ustar threshold，CUT）和基于模型效率（model efficiency，MEF）选择的基准 NEE 数值。

2.2　不同的插补方法

本研究采用 MDS、深度学习模型 PatchTST 及其改进模型 TS2Vec-PatchTST、TSIT-PatchTST 共 4 种方法对缺失的 NEE 数据进行插补。在使用上述 4 种方法对缺失的通量数据进行插补时，设计 3 种不同的缺失情况进行实验，分别为短缺失（连续缺失 1 d）、中缺失（连续缺失 7 d）、长缺失（连续缺失 30 d）[2,12,22−23]。以短缺失为例，将原始数据集随机制造连续 48 行（1 d）的缺失，用 4 种方法分别对剩余的数据进行训练，实验中随机连续缺失概率为 50%。

2.2.1　边际分布采样法（MDS）

MDS 是平均昼夜变化法和查表法的一种集合应用[4]。在空气温度、总辐射量和饱和水汽压差都可用时，该方法利用一定时间窗口（缺失数据前后 14～28 d）内相似环境条件下的通量数据均值对缺失数据进行插补；仅净辐射能量数据可用时，将插补的时间窗口缩短至前后 14 d；在观测数据全部缺失时，使用平均昼夜变化法对缺失数据进行插补。MDS 是 FLUXNET 观测网络所使用的插补方法。

2.2.2　PatchTST 模型

PatchTST 是一种基于 Transformer 的多变量时间序列预测和自监督表示学习模型[24]。它通过两个关键函数实现高效设计：补丁分段和通道独立性。这种设计减少了模型参数量，并促使不同通道之间的信息交流和共享，让模型能够更好地捕获数据的局部模式和特征，可以对缺失值进行更准确的插补。补丁设计还显著减少了计算负载和内存使用，有利于处理大规模的时间序列数据，尤其是长时间连续缺失的数据。

2.2.3　TS2Vec 方法

TS2Vec 是一种时间序列通用表示学习框架[25]，它采用了层次化对比学习的方法，利用上下文一致性原则，将相同时间戳的表示作为正样本，不同时间戳的表示作为负样本。通过引入实例对比损失，模型学习到具有区分性的实例级表示，因此 TS2Vec

表 1 站点数据集植被分类

Tab. 1 Vegetation classification of site dataset

站点	纬度	经度	时间区间(年月日时刻)	数据量/条	植被类型
罗林斯 US-GLE	41°21′59″N	106°14′24″W	201205241430 − 201406221530	36 435	常绿针叶林
梅托利乌斯 US-Me2	44°27′08″N	121°33′27″W	200401181300 − 200601041530	34 422	常绿针叶林
尼沃特岭 US-NR1	40°01′58″N	105°32′47″W	201009281530 − 201208101900	32 744	常绿针叶林
内华达 US-Var	38°24′48″N	120°57′03″W	201303280000 − 201407070030	22 370	草地
埃尔福特 DE-Geb	51°05′59″N	10°54′53″E	201202171030 − 201303141130	18 771	农田
波特科尔伯恩 CA-TP3	42°42′25″N	80°20′54″W	200512160530 − 200612260930	18 009	常绿针叶林
科奇斯 US-SRG	31°47′22″N	110°49′40″W	201101270030 − 201201160300	16 998	草地
坦普斯通 US-Whs	31°44′38″N	110°03′08″W	201007281300 − 201106081130	15 118	开放灌木地
埃森纳赫 DE-Hai	51°04′45″N	10°27′08″E	200602280300 − 200612240230	14 352	落叶阔叶林
默里桥 AU-Lox	34°28′14″S	140°39′18″E	200808191630 − 200906091000	14 100	落叶阔叶林
诺福克 CA-TP1	42°39′39″N	80°33′34″W	200512311330 − 200610131230	13 727	常绿针叶林
库特韦克 NL-Loo	52°09′59″N	5°44′37″E	200205220830 − 200303031230	13 689	常绿针叶林
内珀维尔 US-IB2	41°50′26″N	88°14′28″W	200508221300 − 200605310200	13 515	草地
韦兰德 CA-TP4	42°42′37″N	80°21′27″W	200207201400 − 200304271900	13 499	常绿针叶林
斯普林斯 AU-TTE	22°17′13″S	133°38′24″E	201305220130 − 201401180230	11 571	草地
先锋市 US-Ton	38°25′51″N	120°57′58″W	201305081200 − 201312312200	11 397	木本灌草丛
沙欣尼 US-LWW	34°57′38″N	97°58′44″W	199802181030 − 199810040000	10 924	草地
罗尔斯顿 AU-Emr	23°51′31″S	148°28′29″E	201202062100 − 201209191500	10 837	草地
巴梅拉 AU-Cpr	34°00′08″S	140°35′21″E	201206020930 − 201301041100	10 372	稀树草原
鲍威 US-Wkg	31°44′12″N	109°56′31″W	200506171000 − 200601170230	10 258	草地
科菲维尔 US-Goo	34°15′17″N	89°52′25″W	200510170000 − 200605161030	10 150	草地
埃尔奥拜德 SD-Dem	13°16′58″N	30°28′42″E	200810141400 − 200905031400	9 649	稀树草原
拉龙日 CA-SF1	54°29′06″N	105°49′04″W	200603130130 − 200609202100	9 208	常绿针叶林
诺德豪森 DE-Lnf	51°19′42″N	10°22′04″E	200302221030 − 200308311030	9 121	落叶阔叶林
凯瑟琳 AU-DaS	14°09′34″S	131°23′17″E	201104081600 − 201110142230	9 086	稀树草原
西姆科 CA-TP2	42°46′28″N	80°27′32″W	200505171330 − 200511190730	8 917	常绿针叶林
西斯特斯 US-Me4	44°29′57″N	121°37′21″W	200006081300 − 200012070630	8 724	常绿针叶林
伍德拉夫 US-Syv	46°14′31″N	89°20′52″W	200305310000 − 200311241930	8 536	混合林
塞拉维斯塔 US-SRM	31°49′17″N	110°51′58″W	200910211530 − 201004131300	8 348	木本灌草丛
纽伦比 AU-DaP	14°03′48″S	131°19′05″E	200905081630 − 200910272100	8 266	草地
比格里弗 CA-SF3	54°05′30″N	106°00′20″W	200405051330 − 200410230830	8 199	开放灌木地

注：混合林由树木主导，覆盖百分比超过60%，高度超过2 m。包括交错混合或其他4种森林类型的树木，且任何一种植被类型都不超过60%。

表 2 NEE 响应因子字段

Tab. 2 NEE response factor field

字段名称	中文名称	单位
SW_IN_F	短波辐射	W/m^2
VPD_F_MDS	饱和水汽压差	hPa
TA_F_MDS	空气温度	℃
NETRAD	净辐射能量	W/m^2
WS	风速	m/s
WD	风向	
G_F_MDS	土壤热通量	
TS_F_MDS	土壤温度	℃
RH	相对湿度	%
SWC_F_MDS	土壤含水量	%

能捕捉 NEE 数据的周期性、趋势性和季节性特征，有利于提高模型的插补精度。

2.2.4 ITransformer 方法

ITransformer 通过维度倒置的方式改进了 Transformer 的结构[26]。它将注意力和前馈网络应用于整个序列的变量标记，实现对多变量关系更好地捕捉，并利用前馈网络来学习时间序列的非线性特征。该模型的回溯窗口不是固定的大小，它可以从任意长度的回溯窗口中提取有用信息，从而增强模型的适用性。为了处理不同特征分布的变量之间的差异，该模型在每个变量标记的表示上应用了层归一化技术。

2.2.5　TSIT-PatchTST 模型

为进一步提高插补精度，本研究基于 PatchTST 网络结构设计了 TSIT-PatchTST 模型。一方面，通过将 TS2Vec 与 PatchTST 模型相结合，对数据预处理进行改进，TS2Vec 方法取消了原始 PatchTST 模型批量数据的归一化操作，并对原始 NEE 数据进行了时序特征编码；另一方面，将 PatchTST 模型的编码器与解码器替换为 ITransformer，在解码器阶段使用线性层进行替换。改进后的模型可以通过 ITransformer 的自注意力机制捕获多变量相关性，并可以利用层归一化和前馈网络模块学习时间序列的全局表示。TSIT-PatchTST 模型整体结构主要分为 3 部分（图 1）。第一部分为 TS2Vec 模型处理时间序列数据，采用层次化对比学习方法，以捕获时间序列的多尺度上下文信息；第二部分为 ITransformer 模型，使用 ITransfoemr 架构替代 PatchTST 模型原有的 Transformer 编码器，通过自注意力捕获多变量相关性，并利用层归一化和前馈网络模块学习更好的时间序列全局表示；第三部分为 PatchTST 模型，经过前两种模型处理的拟合结果将进入 PatchTST 模型的展平层和线性头进行运算，最终得到 NEE 的插补值。

TSIT-PatchTST 模型在第一部分处理时序数据的过程如下：对于输入的时间序列数据 $x_i \in R^{T\times F}$，其中 T 是序列长度，F 是特征维度，首先通过一个全连接层（线性投影）将输入映射到高维潜在向量 $\boldsymbol{z}_{i,t}$。向量 $\boldsymbol{z}_{i,t}$ 计算公式为

$$\boldsymbol{z}_{i,t} = \boldsymbol{W}x_{i,t} + \boldsymbol{b} \tag{1}$$

式中：$\boldsymbol{W} \in R^{F'\times F}$ 是权重矩阵，$\boldsymbol{b} \in R^{F'}$ 是偏置向量，F' 是潜在向量的维度，且 $F' > F$。

对高维潜在向量 $\boldsymbol{z}_{i,t}$ 进行掩码处理，用于模拟数据缺失的情况，从而增强模型对有用信息和噪音的区分。此步骤通过在时间轴上应用二进制掩码 $m \in \{0,1\}^T$ 来实现，该掩码是从伯努利分布 Bernoulli($p = 0.5$) 中独立采样。

为了降低模型对时间序列中数据位置的依赖性，TS2Vec 对输入数据进行了随机截取，生成了新的数据视图。这有助于模型学习到与数据位置无关的特征表示。此步骤通过随机选择两个重叠的时间片段 $[a_1,b_1]$ 和 $[a_2,b_2]$ 来实现，其中 $0 < a_1 \leqslant a_2 \leqslant b_1 \leqslant b_2 \leqslant T$。重叠时间段 $[a_2,b_1]$ 上的数据应该在两个上下文中保持一致。

使用具有扩张卷积的网络块捕捉时间序列中的长距离依赖关系。每个网络块包含两个一维卷积层，卷积层的扩张参数是 2 的幂次方，幂次方的大小取决于网络块的索引。

模型通过层次化对比学习，即在实例级和时间维度上进行对比学习，调整模型的参数，使得相似的样本在特征空间中更接近，不相似的样本更远离。通过比较正样本对（相似的时间点）和负样本对（不相似的时间点）之间的相似度，学习区分不同时间点的表示。学习到的时间序列表示将作为新的特征向量应用于 TSIT-PatchTST 改进模型的插补任务中。时间对比损失的计算公式为

$$l_{(i,t)}^{\text{temp}} = -\log \frac{\exp\left(r_{i,t}\cdot r'_{i,t}\right)}{\sum_{t'\in\theta}\exp\left(r_{i,t}\cdot r'_{i,t'}\right) + 1_{[t=t']}} \tag{2}$$

式中：i 为输入时间序列样本的索引，t 为特定的时间戳，$r_{i,t}$ 为时间序列样本 i 在时间戳 t 的表示，t' 为从同一时间序列的不同时间戳抽取的表示，$r_{j,t}$ 为批次中其他时间序列样本 j 在时间戳 t 的表示，θ 是两个子序列重叠的时间段集合。

实例级对比损失的计算公式为

$$l_{(i,t)}^{\text{inst}} = -\log \frac{\exp\left(r_{i,t}\cdot r'_{i,t}\right)}{\sum_{j=1}^{B}\left(\exp\left(r_{i,t}\cdot r'_{i,t}\right) + 1_{[i=j]}\exp\left(r_{i,t}\cdot r_{j,t}\right)\right)} \tag{3}$$

式中：B 是批次大小。

2.3　评价指标

使用平均绝对误差（mean absolute error，MAE）、均方误差（mean squared error，MSE）和决定系数（R^2）作为评价指标[27]，以评估插补结果的准确性和性能。最终得到 31 个站点在不同数据缺失场景下的 MAE、MSE、R^2 值。为便于最终结果分析和观察整体表现，对 31 个站点中同一植被类型的 MAE、MSE、R^2 取平均值，对同一植被类型多站点的插补评价指标取平均值也可以降低偶然性。较小的 MAE 和 MSE 数值表示插补结果与真实值更接近，表明插补方法具有较高的准确性和插补性能。

MAE 用于量化插补结果的平均偏差，衡量插补值与实际观测值之间的平均误差大小。其计算公式为

$$\text{MAE} = \frac{1}{N}\sum|p_i - u_i| \tag{4}$$

式中：p_i 为各插补方法预测的 NEE，μmol/(m^2·s)；u_i 为实际观测的 NEE，μmol/(m^2·s)；N 为样本数量。

MSE 用于评估插补结果的平均方差，即插补值与实际观测值之间的整体差异程度。其计算公式为

$$\text{MSE} = \frac{1}{N}\sum(p_i - u_i)^2 \tag{5}$$

R^2 用于判断插补值与实际观测值之间的匹配程

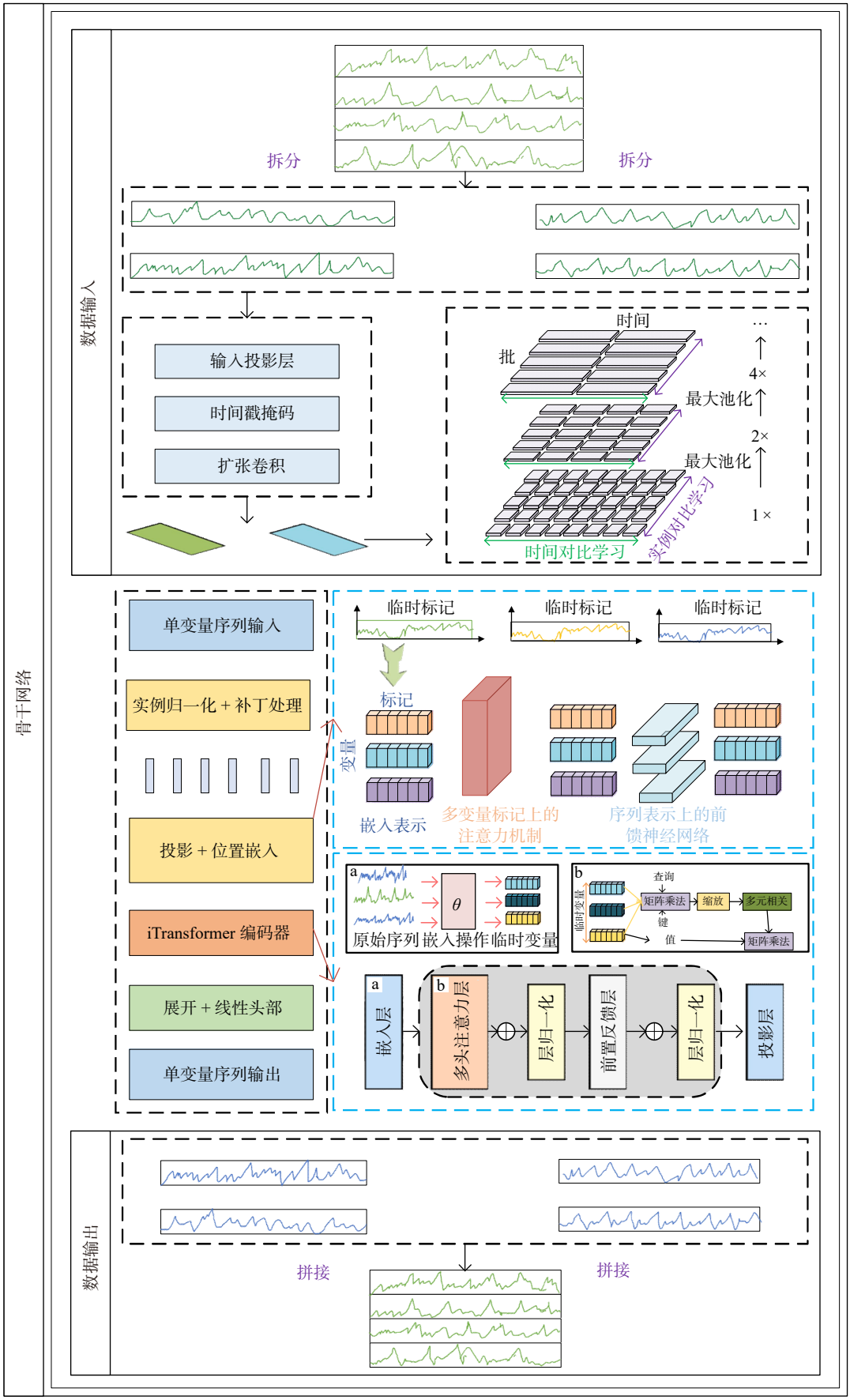


图 1 TSIT-PatchTST 模型结构

Fig. 1 TSIT-PatchTST model structure

度，越接近 1 表示插补结果与原始数据的拟合度越高。其计算公式为

$$R^2 = \frac{\left[\sum (p_i - \overline{p})(u_i - \overline{u})\right]^2}{\sum (p_i - \overline{p})^2 \sum (u_i - \overline{u})^2} \tag{6}$$

式中：$\overline{p}$为各插补方法预测 NEE 的均值，μmol/（m^2·s）；$\overline{u}$为实际观测 NEE 的均值，μmol/（m^2·s）。

2.4　实验环境和参数配置

使用 python3.8 作为开发语言，集成开发环境为 Anaconda3。PatchTST 和 TimesNet 模型的编写和调参由 torch1.7.1、numpy1.23.5、pandas1.5.3、matplotlib3.7.0、scikit-learn1.2.2、einops0.4.0、patool1.12、scipy1.10.1、sktime0.16.1、tqdm4.64.1、reformer-pytorch1.4.4 完成。为保证不同模型对插补效果的公平性，实验中各深度学习模型（PatchTST、TS2Vec-PatchTST、TSIT-PatchTST）采用相同的超参数，模型主要超参数如表 3 所示。所有的深度学习实验均使用自适应器 2015 优化器，将（β_1，β_2）的默认超参数配置为（0.900，0.999）。

表 3　模型主要超参数设置

Tab. 3　Primary hyper-parameter setting of models

超参数	含义	设定值
top_k	准确率	5
layers	层	2
d_{min}, d_{max}	深度范围	（64, 128）
batch_size	批量大小	1 500
learning_rate	学习率	0.001
epoch	迭代次数	10

注：因长缺失场景需要制造连续1 440条数据缺失，故模型批量大小设置为1 500；上述超参数设置可以满足模型对NEE缺失数据插补的需要。

3　结果与分析

3.1　NEE 插补方法的整体性能评估

分别对 31 个站点构造 NEE 数据短缺失、中缺失、长缺失的 3 种不同场景，对每个站点的 NEE 缺失数据分别使用 4 种方法进行插补，进而获取插补结果的 3 种评价指标 MSE、MAE 和 R^2 的平均值（表 4）。

在短缺失场景下，MDS 在开放灌木地和稀树草原的 NEE 插补效果较好，但是其他植被类型插补效果不如 3 种深度学习模型。中缺失和长缺失场景下，MDS 方法的插补性能显著下降，在 8 种植被类型下的插补精度不如 3 种深度学习模型。TS2Vec-PatchTST 模型通过引入 TS2Vec 方法作为嵌入层，与 PatchTST 模型的插补结果相比，MSE 和 MAE 显著降低，但是 R^2 指标也有所降低。

TS2Vec-PatchTST 模型相对于 PatchTST 模型在 3 种缺失场景下，8 种植被类型插补结果的 MSE 和 MAE 均得到了提高。在短缺失场景下，该模型对 31 个站点的平均 MSE 和 MAE 为 1.361 和 0.820 μmol/（m^2·s），相较于 PatchTST 模型的 1.805 和 0.995 μmol/（m^2·s），分别降低了 24.6% 和 17.6%。在中缺失场景下，该模型的平均 MSE 和 MAE 为 1.397 和 0.824 μmol/（m^2·s），相较于 PatchTST 模型的 1.798 和 0.992 μmol/（m^2·s），分别降低了 22.3% 和 16.9%。在长缺失场景下，该模型的平均 MSE 和 MAE 为 1.387 和 0.826 μmol/（m^2·s），相较于 PatchTST 模型的 1.804 和 0.995，分别降低了 23.1% 和 17.0%。

相较于 PatchTST 模型，TS2Vec-PatchTST 模型在评价指标 MSE 和 MAE 上有显著的提升，但 R^2 下降。R^2 的降低表明模型对数据拟合程度的不足。为解决这一问题，本研究在 TSIT-PatchTST 模型中，使用 ITransformer 替换 PatchTST 模型原有的 Transformer 架构，以增强模型对时间序列长期依赖性的捕捉能力，期望在保持 MSE 和 MAE 优势的同时，提升 R^2，提高模型的插补精度。

TSIT-PatchTST 和 TS2Vec-PatchTST 模型在 3 种缺失场景及 8 种植被类型下的 NEE 数据插补结果的对比结果表明，TSIT-PatchTST 模型在 3 种评价指标上均有较大优势。在短缺失场景下，TSIT-PatchTST 模型的平均 MSE 和 MAE 分别降至 0.947 和 0.636 μmol/（m^2·s），与 TS2Vec-PatchTST 模型相比，分别降低了 30.4% 和 22.4%。在中缺失场景下，TSIT-PatchTST 模型的平均 MSE 和 MAE 进一步降至 0.938 和 0.620 μmol/（m^2·s），相较于 TS2Vec-PatchTST 模型，分别降低了 32.9% 和 24.8%。而在长缺失场景下，TSIT-PatchTST 模型的平均 MSE 和 MAE 为 0.942 和 0.628 μmol/（m^2·s），相较于 TS2Vec-PatchTST 模型，分别降低了 32.1% 和 24.0%。

在评价指标 R^2 方面，短缺失场景下，TSIT-PatchTST 模型的平均 R^2 为 0.447，相较于 TS2Vec-PatchTST 模型的 0.331，提升了 35.1%。在中缺失场景下，TSIT-PatchTST 模型的平均 R^2 为 0.475，相较于 TS2Vec-PatchTST 模型的 0.341，提升了 39.3%。而在长缺失场景下，TSIT-PatchTST 模型的平均 R^2 为 0.457，相较于 TS2Vec-PatchTST 模型的 0.327，提升了 39.8%。这些结果表明，TSIT-PatchTST 模型在各种缺失场景下均能显著提高 NEE 缺失数据插补的精度。

短缺失场景下，PatchTST、TS2Vec-PatchTST、TSIT-PatchTST 模型的插补结果泰勒图如图 2 所

表 4 数据插补结果

Tab. 4 Data imputation results

缺失情况	植被类型	PatchTST			TS2Vec-PatchTST			TSIT-PatchTST			边际分布采样法 MDS		
		MSE/(μmol·m^{-2}·s^{-1})	MAE/(μmol·m^{-2}·s^{-1})	R^2	MSE/(μmol·m^{-2}·s^{-1})	MAE/(μmol·m^{-2}·s^{-1})	R^2	MSE/(μmol·m^{-2}·s^{-1})	MAE/(μmol·m^{-2}·s^{-1})	R^2	MSE/(μmol·m^{-2}·s^{-1})	MAE/(μmol·m^{-2}·s^{-1})	R^2
短缺失	农田	1.547	0.950	**0.886**	0.736	0.570	0.641	**0.191**	**0.230**	0.879	0.233	0.293	0.002
	落叶阔叶林	1.550	0.936	**0.611**	0.824	0.657	0.450	**0.457**	**0.465**	0.596	2.201	0.813	0.014
	常绿针叶林	1.208	0.865	**0.448**	0.913	0.690	0.323	**0.617**	**0.518**	0.439	4.792	1.014	0.045
	草地	2.057	1.046	**0.329**	1.804	0.935	0.247	**1.428**	**0.798**	0.309	17.340	1.542	0.039
	混合林	2.596	1.223	0.834	1.176	0.771	0.591	**0.377**	**0.392**	**0.847**	2.307	0.677	0.064
	开放灌木地	1.116	0.856	0.423	1.007	0.745	0.304	**0.750**	0.592	**0.439**	1.114	**0.521**	0.025
	稀树草原	2.232	1.175	0.388	2.043	1.085	0.284	**1.534**	0.895	**0.409**	3.551	**0.864**	0.032
	木本灌草丛	3.853	1.289	0.531	2.141	1.026	0.391	**1.149**	**0.726**	**0.536**	18.586	1.816	0.023
中缺失	农田	1.501	0.939	**0.875**	0.903	0.607	0.633	**0.237**	**0.224**	0.874	0.810	0.529	0.026
	落叶阔叶林	1.577	0.942	0.647	0.738	0.621	0.476	**0.403**	**0.426**	**0.652**	5.777	1.308	0.034
	常绿针叶林	1.198	0.862	0.475	0.910	0.683	0.351	**0.588**	**0.502**	**0.475**	14.790	1.574	0.119
	草地	2.046	1.044	0.313	1.888	0.954	0.218	**1.452**	**0.796**	**0.316**	16.691	1.547	0.102
	混合林	2.610	1.223	**0.850**	1.178	0.757	0.604	**0.371**	**0.355**	0.844	2.370	0.684	0.087
	开放灌木地	1.084	0.844	**0.612**	0.967	0.718	0.399	**0.590**	**0.486**	0.601	1.027	0.516	0.045
	稀树草原	2.159	1.160	**0.403**	2.277	1.136	0.293	**1.615**	**0.914**	0.394	2.804	0.755	0.091
	木本灌草丛	3.927	1.290	**0.529**	2.078	1.026	0.372	**1.141**	**0.732**	0.528	14.882	2.472	0.055
长缺失	农田	1.538	0.948	0.901	0.835	0.598	0.648	**0.177**	**0.209**	**0.903**	1.248	0.733	0.008
	落叶阔叶林	1.559	0.939	0.603	0.695	0.607	0.447	**0.398**	**0.436**	**0.607**	6.709	1.539	0.013
	常绿针叶林	1.206	0.864	**0.442**	0.912	0.690	0.323	**0.610**	**0.518**	0.439	20.316	2.492	0.047
	草地	2.052	1.046	0.292	1.877	0.954	0.204	**1.458**	**0.801**	**0.297**	38.533	2.238	0.041
	混合林	2.603	1.223	**0.851**	1.136	0.778	0.615	**0.354**	**0.401**	0.850	4.565	1.006	0.065
	开放灌木地	1.106	0.852	**0.589**	1.028	0.745	0.389	**0.640**	**0.506**	0.579	2.098	0.699	0.021
	稀树草原	2.205	1.170	**0.423**	2.290	1.138	0.310	**1.593**	**0.907**	0.422	8.449	1.237	0.031
	木本灌草丛	3.878	1.290	**0.548**	2.006	1.014	0.385	**1.092**	**0.714**	0.544	37.945	2.534	0.013

注:加粗字体表示在同一植被类型下插补效果最佳方法。下同。

示。泰勒图中的散点代表插补模型,图上散点颜色越浅,则 MAE 越小;越靠近原点,则 MSE 越小;与 X 轴夹角越小,则 R^2 越大。TSIT-PatchTST 模型在多个生态系统中的插补性能较好。该模型的散点更接近原点,表明其 MSE 较低,具有更高的插补精度;散点与 X 轴的夹角最小,说明 R^2 最大,对数据的拟合度更佳;散点颜色较浅,说明 MAE 较小。

中缺失场景下,PatchTST、TS2Vec-PatchTST、TSIT-PatchTST 模型的插补结果泰勒图如图 3 所示。TSIT-PatchTST 模型的散点普遍更接近原点,说明其 MSE 较低,有着更高的插补精度;散点与 X 轴的夹角较小,反映出模型的 R^2 较大;散点颜色较浅,说明 MAE 较小。综合考虑这些评价指标,TSIT-PatchTST 模型在插补性能上优于 PatchTST 和 TS2Vec-PatchTST 模型。

长缺失场景下,PatchTST、TS2Vec-PatchTST、TSIT-PatchTST 模型的插补结果泰勒图如图 4 所示。TSIT-PatchTST 模型在泰勒图中的散点更接近原点,MSE 更小;散点与 X 轴的夹角较小,该模型的 R^2 较高,对数据的拟合程度更好;散点颜色较浅,即 MAE 较小。TSIT-PatchTST 模型在处理长时间连续缺失的 NEE 数据时的插补精度更高。

TSIT-PatchTST 模型插补效果折线图(图 5)显示:该模型在稀树草原和落叶阔叶林等植被类型的插补值与真实值之间具有高度一致性,在捕捉 NEE 的变化趋势上表现最佳。在常绿针叶林和草地等植被类型中,因 NEE 数据波动较大,模型的预测精度有所下降。在农田和开放灌木地这两种植被类型中,NEE 数据出现较大波动,不利于模型插补。在 NEE 数据较为稳定的植被类型中,模型插补效果较

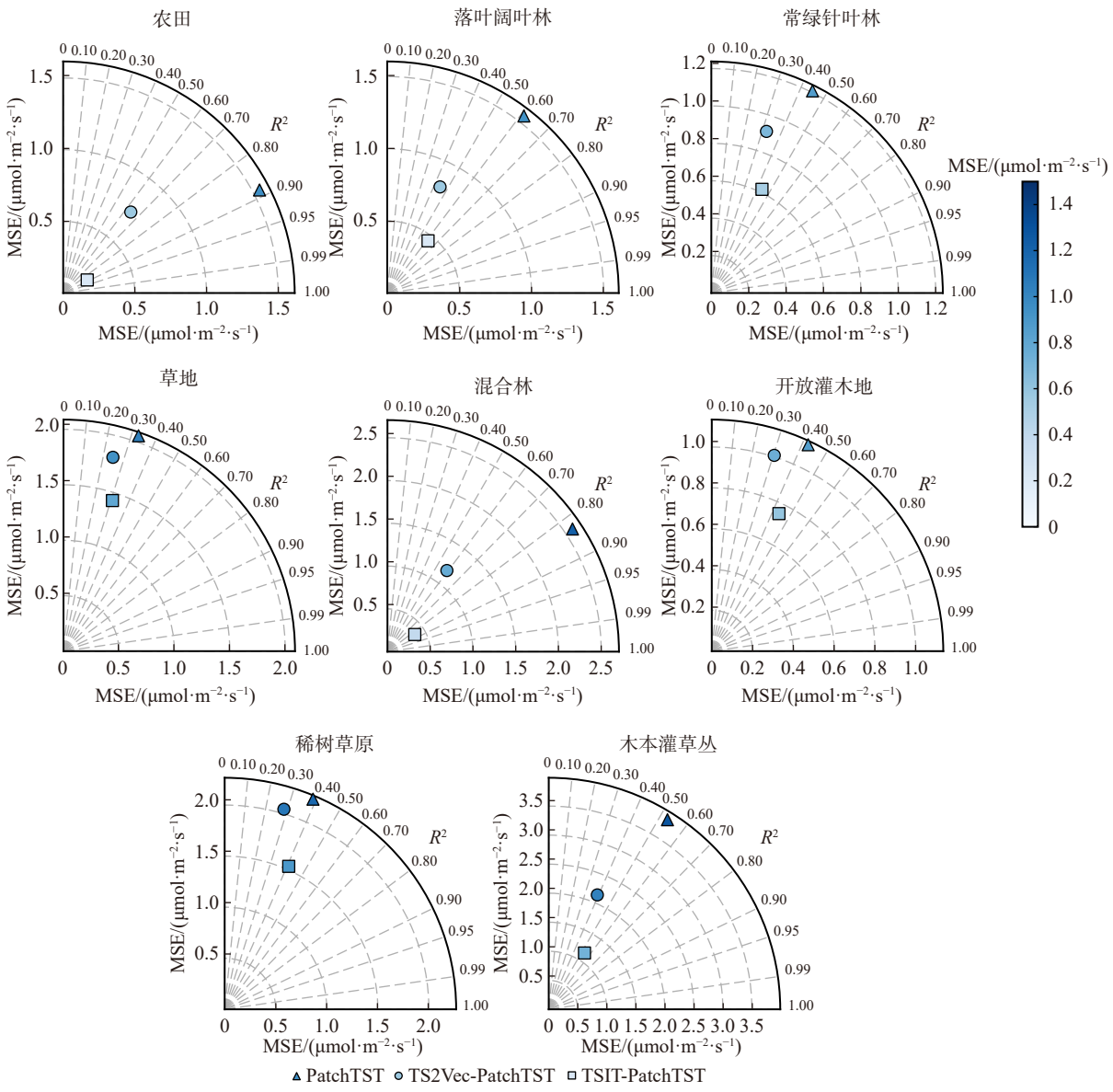


图 2　短缺失泰勒图

Fig. 2　Short missing Taylor diagram

好，但在 NEE 数据波动较大时，插补值与真实值之间存在一定偏差。

Zhu 等[12]提出的 RFR_{10} 机器学习模型，在多种植被类型的站点进行 NEE 缺失数据插补实验，提升了 MDS 在长缺失场景下插补精度，但对于部分植被类型 RFR_{10} 模型的插补效果较差。TSIT-PatchTST 模型在一定程度上解决了这一问题。TSIT-PatchTST 模型和 RFR_{10} 模型在长缺失场景下插补结果的平均 R^2 对比如表 5 所示。与 RFR_{10} 模型相比，TSIT-PatchTST 模型在农田、混合林、开放灌木地和木本灌草丛场景下提高了插补精度。

3.2　日间与夜间的插补效果

在 NEE 数据插补的过程中，日间与夜间的差异性是一个不可忽视的因素。为避免不同站点数据量的差异对计算结果产生影响，以数据量最大的 US-GLE 站点为研究对象，计算日间与夜间对 NEE 数据插补效果。日间与夜间 NEE 插补结果如表 6 所示。

3.3　生长季与非生长季的插补效果

在生态系统碳循环的研究中，应使用准确的 NEE 数据来评估陆地生态系统在全球碳平衡中的作用。以 US-GLE 站点为例，生长季与非生长季的 NEE 数据在短、中、长 3 种不同缺失场景下的插补结果如表 7 所示。

4　讨　　论

4.1　植被类型对 NEE 插补结果的影响

通过对比不同植被类型下 NEE 的插补结果（表 4），与 MDS 相比，深度学习模型在不同植被类型下的插补性能较好，其适用范围更广。不同植被类型的 NEE 观测数据之间存在一定差异，数据的差

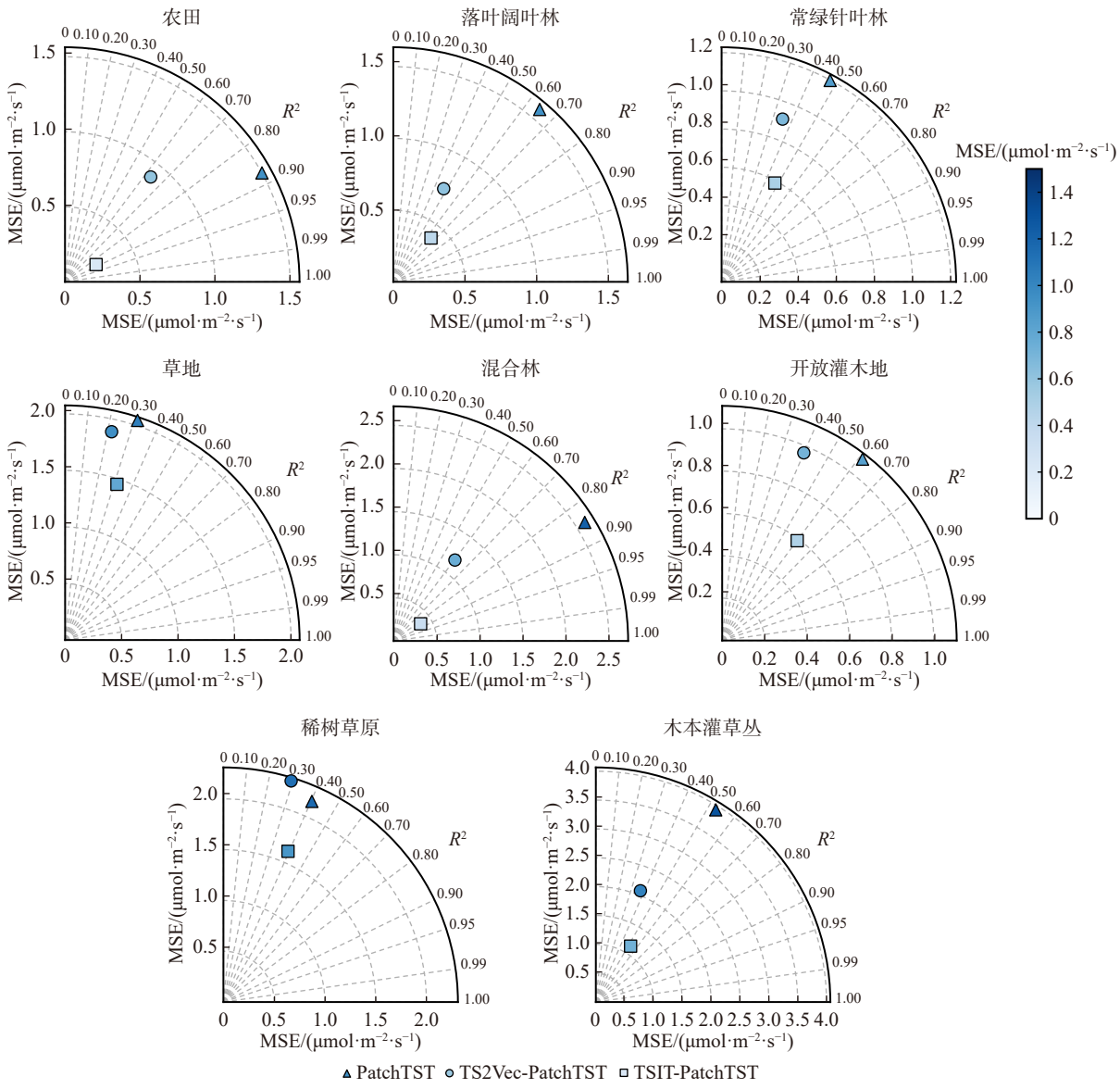


图 3 中缺失泰勒图

Fig. 3 Middle missing Taylor diagram

异会造成数据分布特征不同，从而会对插补方法的性能产生较大影响，这与 Rivera-Muñoz 等[28]研究结果类似。在短缺失场景下，深度学习模型在混合林、农田、落叶阔叶林和常绿针叶林的插补效果较好，在开放灌木地、稀树草原、草地和木本灌草丛这 4 种植被类型的插补效果较差。同样在短缺失场景下，MDS 更适用于开放灌木地；深度学习模型的适用范围较广，不同植被类型对其插补性能的影响较弱，对木本灌草丛的插补效果更好，对开放灌木地的插补效果一般。

4.2 昼夜变化对 NEE 插补结果的影响

在对 NEE 数据进行插补时，研究发现日间与夜间的插补效果存在显著差异（表 6），这与周宇等[4]研究结果相似。具体来说，日间插补的 NEE 数据与实测数据的 R^2 低于夜间，同时 MSE 明显高于夜间，表明日间插补的效果不如夜间。这一现象可能与日间 NEE 的复杂性有关。

日间 NEE 受到多种环境因素的共同作用，包括温度、湿度、风速[29]和辐射[30]等，这些环境因素每时每刻都会产生变化，各种环境因素的变化会堆叠在一起作用于 NEE，使得 NEE 的日间变化波动强烈，导致数据稳定性不如夜间。例如：植物因太阳辐射会在正午达到光合作用的峰值，此时 NEE 会快速上升；午后快速衰减的辐射量则会引起 NEE 向下快速波动。NEE 的这种剧烈变化使得模型难以精确插补缺失值，当缺失时间段较短时，可以利用缺失值前后的波动范围进行插补，但当连续缺失天数较长时，会导致拟合结果不准确。这可以解释在短缺失时日间 R^2 较高，而随着缺失天数的增加，R^2 会迅速下降，直至不如夜间的现象。日间生物活动频繁，一定程度

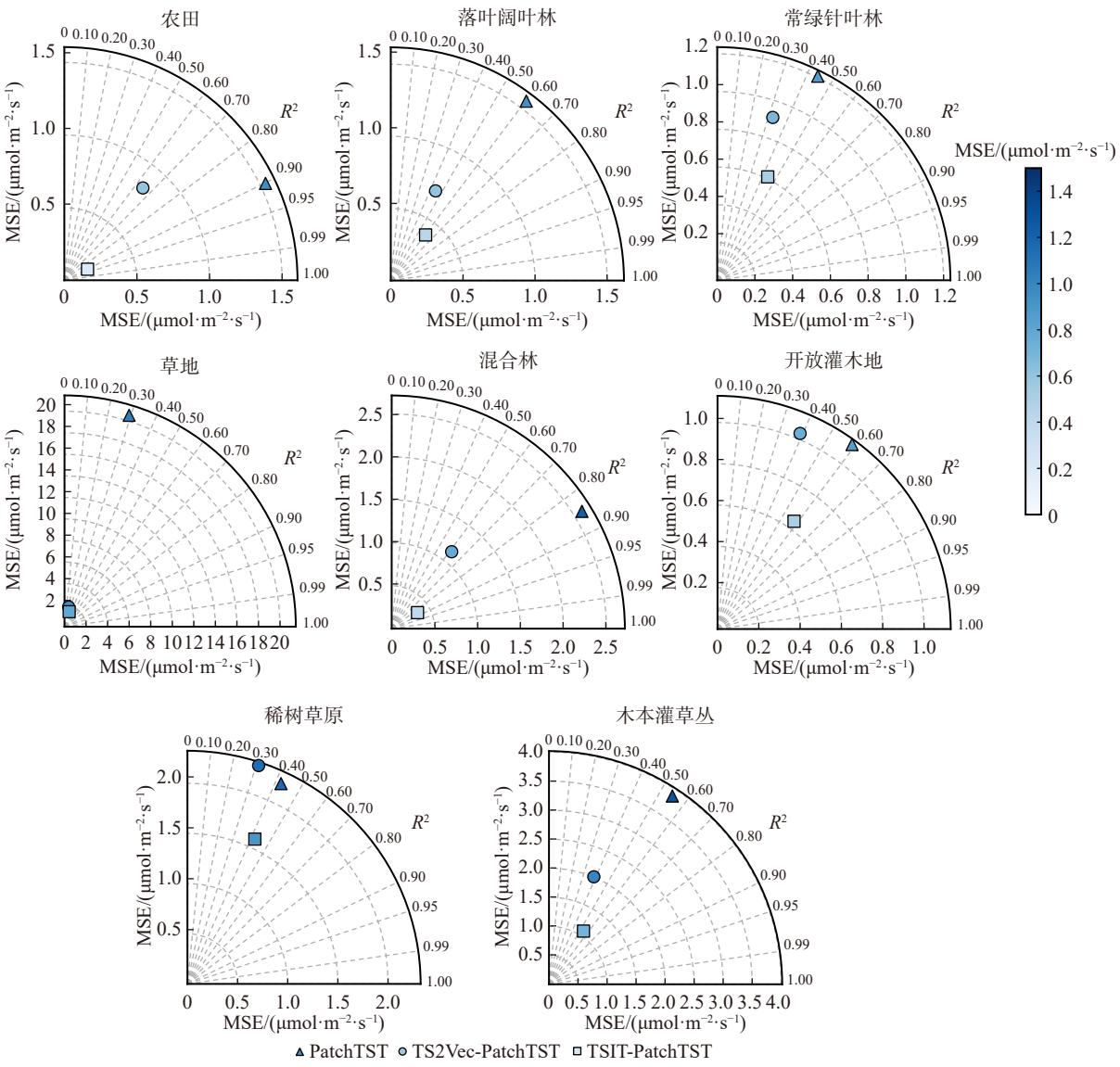


图 4　长缺失泰勒图

Fig. 4　Long missing Taylor diagram

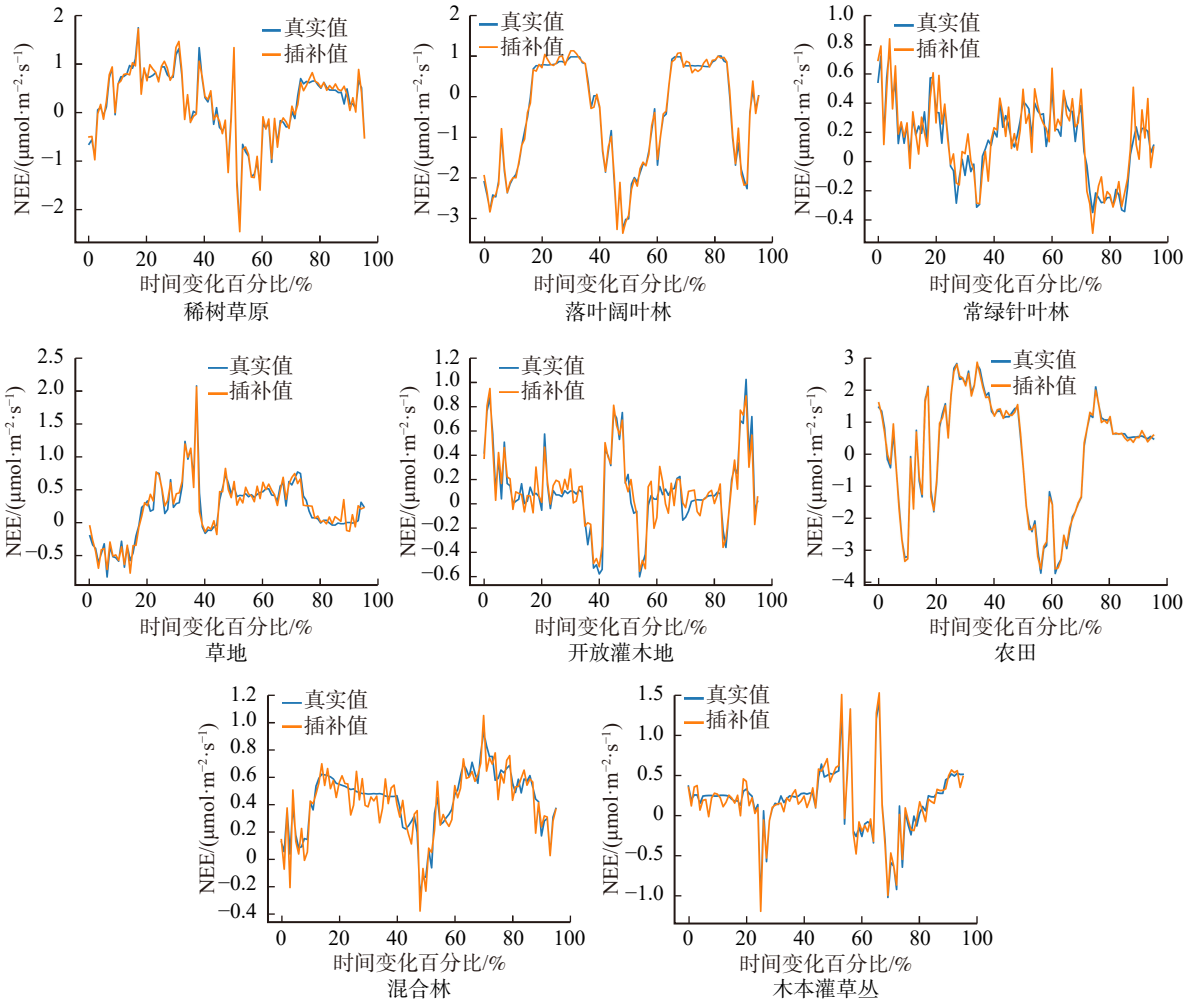
上会影响生态监测设备捕获 NEE 的准确度，导致原始数据质量不如夜间，增加了插补模型填补缺失值的不确定性，这会导致模型日间的 MSE、MAE 数值略大于夜间。

综上，日间 NEE 数据插补效果不如夜间是由多方面原因导致。为提高模型对 NEE 缺失数据的插补精度，应充分考虑昼夜因素影响，可以利用不同时间段（日间和夜间）分别进行插补；或者是分别针对日间、夜间数据建模，以提高模型对昼夜变化的捕捉能力，从而提高模型的插补精度。

4.3　生长季变化对 NEE 插补结果的影响

根据本研究结果，生长季插补所得的 NEE 数据与实测数据的 R^2 高于非生长季，同时 MSE 明显低于非生长季（表 7），这表明生长季的 NEE 插补效果显著优于非生长季。在生长季，植被的生理活动较为活跃，光合作用强烈，生态系统的碳吸收能力得到增强，表现为 NEE 的负值，即生态系统作为有效的碳汇。这一时期的植被生理活动对 NEE 的影响显著，且与环境响应因子如光合有效辐射和土壤温度的动态变化密切相关。这种植被生长状态与环境因素相协同的变化方式为插补模型提供了更多的特征变量和有效响应因子，有助于提高模型插补 NEE 缺失值的精度。这与游桂莹等[31]研究结果相似：在非生长季植物主要依靠呼吸作用，NEE 受光合作用影响较弱，而受土壤温度、土壤湿度影响较强。在非生长季，插补模型更多地依赖土壤温度和土壤湿度等响应因子来插补缺失值，可用特征因子数目较少，从而导致模型的插补效果不如生长季。

除了植被生长状态和环境因素的影响外，NEE 数据还可能会受到凋落物分解等其他生物、微生物与植被相互作用的影响，导致 NEE 数据内在分布特征不规律变化，增加了模型对 NEE 缺失数据的插补难度。



稀树草原以 AU-Cpr 站点为例，落叶阔叶林以 DE-Lnf 站点为例，常绿针叶林以 US-GLE 站点为例，草地以 US-SRG 站点为例，开放灌木地以 US-Whs 站点为例，农田以 DE-Geb 站点为例，混合林以 US-Syv 站点为例，木本灌草丛以 US-Ton 站点为例。

图 5 TSIT-PatchTST 模型插补效果折线图

Fig. 5 Effect line chart of TSIT-PatchTST model interpolation

表 5 长缺失下数据插补结果的平均 R^2 对比

Tab. 5 Comparison of average R^2 of data imputation results under long missing

插补方法	农田	落叶阔叶林	常绿针叶林	草地	混合林	开放灌木地	稀树草原	木本灌草丛
TSIT-PatchTST	**0.903**	0.607	0.439	0.297	**0.850**	**0.579**	0.422	**0.544**
RFR_{10}[20]	0.700	**0.623**	**0.547**	**0.638**	0.689	0.397	**0.784**	0.437

表 6 US-GLE 站点日间与夜间 NEE 插补结果

Tab. 6 Daytime and nighttime NEE imputation results of US-GLE station

时间段	评价指标	短缺失	中缺失	长缺失
日间	MSE	0.509	0.635	0.709
	MAE	0.429	0.514	0.502
	R^2	0.479	0.285	0.321
夜间	MSE	0.448	0.537	0.498
	MAE	0.387	0.418	0.404
	R^2	0.439	0.386	0.429

注：日间时间段为上午10点至下午16点，夜间时间段为下午16点至次日上午10点。

表 7 US-GLE 站点生长季与非生长季 NEE 插补结果

Tab. 7 NEE imputation results during growing and non-growing seasons of US-GLE station

时期	评价指标	短缺失	中缺失	长缺失
生长季	MSE	0.424	0.429	0.451
	MAE	0.402	0.411	0.416
	R^2	0.486	0.398	0.392
非生长季	MSE	1.005	0.913	0.647
	MAE	0.552	0.578	0.480
	R^2	0.191	0.105	0.204

注：生长季为4月到10月，非生长季为10月到次年4月。

生长季植被周围环境因素比非生长季更稳定，模型可利用更多的响应因子来插补 NEE 缺失值，因此生长季 NEE 的插补精度高于非生长季。在进行 NEE 数据插补时，应考虑植被生长状态变化对生态系统碳交换过程的影响，以及对模型插补准确性的影响。

4.4　TSIT-PatchTST 模型在 NEE 插补中的优势

在 NEE 缺失数据插补中，与其他插补方法相比，TSIT-PatchTST 模型表现出较强的插补能力和适用能力。3 种评价指标显示该模型在 3 种缺失场景和 8 种植被类型下，均能高精度地插补 NEE 缺失数据，其插补精度显著优于 PatchTST 模型。这说明通过引入 TS2Vec 和 ITransformer 方法，TSIT-PatchTST 模型能够更有效地处理时间序列数据，其优势在于以下 3 方面。

（1）TSIT-PatchTST 模型具备识别和理解 NEE 数据内在分布特征的能力。通过层次化对比学习和上下文一致性原则，捕捉时间序列数据在不同时间尺度上的上下文信息。这种特征提取能力使得模型能够更准确地理解 NEE 数据的内在分布特征，包括趋势性、周期性和季节性，能够较好地适应不同植被类型的 NEE 数据。

（2）TSIT-PatchTST 模型通过对比学习与数据增强技术，区分信息与噪声。它将相似的站点 NEE 数据归为正样本，不相似的归为负样本，利用不同站点间的相似变化规律来优化模型训练。该模型利用数据增强技术中的时间戳遮蔽和随机裁剪，增强了对噪声和数据缺失的适应性。即使在日间或非生长季，模型也能准确识别可用信息，提升 NEE 数据插补的准确性。

（3）TSIT-PatchTST 模型加强了处理长时间连续性缺失数据的能力。将时间序列中的每个时间点作为独立的变量标记进行处理，学习并重建了 NEE 时间序列的全局模式，从而加强模型在长时间连续性缺失时的插补能力和插补精度。

5　结　　论

本研究通过引入 TS2Vec 和 ITransformer 方法，对 PatchTST 模型进行改进，设计了一种深度学习模型 TSIT-PatchTST，以提高插补精度。8 种植被类型数据下的 3 种评价指标显示：TSIT-PatchTST 模型可显著提高 NEE 缺失数据插补的精度，尤其是在数据长缺失场景下。TSIT-PatchTST 模型具有学习 NEE 数据内在分布特征，区分可用信息和噪声，处理长时间连续性缺失数据的能力，与其他插补方法对比具有优势。下一步研究将扩大数据站点覆盖范围，提升模型的适用性和鲁棒性。

参　考　文　献

[1] 田容才, 文双雅, 阳会兵. 基于涡度相关法的农田生态系统碳通量研究进展 [J]. 激光生物学报, 2019, 28(5): 415−420.
Tian R C, Wen S Y, Yang H B. Research progress on carbon flux in agro-ecosystem based on eddy covariance system[J]. Acta Laser Biology Sinica, 2019, 28(5): 415−420.

[2] 王秀英, 周秉荣, 苏淑兰, 等. 青藏高原高寒草甸和荒漠碳交换特征及其气象影响机制 [J]. 生态学报, 2023, 43(3): 1194−1208.
Wang X Y, Zhou B R, Su S L, et al. Carbon exchange characteristics and meteorological influence mechanism of alpine meadow and desert in Qinghai-Tibet Plateau[J]. Acta Ecologica Sinica, 2023, 43(3): 1194−1208.

[3] Bopucki R, Kiersztyn A, Pitucha G, et al. Handling missing data in ecological studies: ignoring gaps in the dataset can distort the inference[J]. Ecological Modelling, 2022, 468: 109964.

[4] 周宇, 黄辉, 张劲松, 等. 森林生态系统涡度相关法碳通量长时间连续性缺失数据插补方法的比较 [J]. 中国农业气象, 2021, 42(4): 330−343.
Zhou Y, Huang H, Zhang J S, et al. Comparison of gap-filling methods for long-term continuous missing data in carbon flux observation by eddy covariance method of forest ecosystem[J]. Chinese Journal of Agrometeorology, 2021, 42(4): 330−343.

[5] 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.

[6] 王少影, 张宇, 孟宪红, 等. 机器学习算法对涡动相关缺失通量数据的插补研究 [J]. 高原气象, 2020, 39(6): 1348−1360.
Wang S Y, Zhang Y, Meng X H, et al. Fill the gaps of eddy covariance fluxes using machine learning algorithms[J]. Plateau Meteorology, 2020, 39(6): 1348−1360.

[7] Xiao Q, Geng G, Cheng J, et al. Evaluation of gap-filling approaches in satellite-based daily PM2.5 prediction models[J]. Atmospheric Environment, 2021, 244: 117921.

[8] Mahabbati A, Beringer J, Leopold M, et al. A comparison of gap-filling algorithms for eddy covariance fluxes and their drivers[J]. Geoscientific Instrumentation, Methods and Data Systems, 2021, 10(1): 123−140.

[9] Soloway A D, Amiro B D, Dunn A L, et al. Carbon neutral or a sink? uncertainty caused by gap-filling long-term flux measurements for an old-growth boreal black spruce forest[J]. Agricultural and Forest Meteorology, 2017, 233: 110−121.

[10] Yang L, Fan X, Zhang Z. Your time series is worth a binary image: machine vision assisted deep framework for time series forecasting[J/OL]. arXiv: 2302.14390v1. [2024−07−14]. https://arxiv.org/pdf/2302.14390.

[11] Irvin J, Zhou S, Mcnicol G, et al. Gap-filling eddy covariance methane fluxes: comparison of machine learning model predictions and uncertainties at FLUXNET-CH4 wetlands[J]. Agricultural and Forest Meteorology, 2021, 308−309: 108528.

[12] Zhu S, Clement R, Mccalmont J, et al. Stable gap-filling for longer eddy covariance data gaps: a globally validated machine-learning approach for carbon dioxide, water, and energy fluxes[J]. Agricultural and Forest Meteorology, 2022, 314: 108777.

[13] Vu M T, Jardani A, Massei N, et al. Reconstruction of missing groundwater level data by using long short-term memory (LSTM) deep neural network[J]. Journal of Hydrology, 2021, 597: 125776.

[14] Guo X, Yao Y, Tang Q, et al. Multimodel ensemble estimation of Landsat-like global terrestrial latent heat flux using a generalized deep CNN-LSTM integration algorithm[J]. Agricultural and Forest Meteorology, 2024, 349: 109962.

[15] 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 X Y, Jia X, Huang J Z, 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.

[16] Qian L, Wu L, Zhang Z, et al. A gap filling method for daily evapotranspiration of global flux data sets based on deep learning[J]. Journal of Hydrology, 2024, 641: 131787.

[17] Jerse G, Marcucci A. Deep learning LSTM-based approaches for 10.7 cm solar radio flux forecasting up to 45-days[J]. Astronomy and Computing, 2024, 46: 100786.

[18] Chen Z, Xu H, Jiang P, et al. A transfer learning-based LSTM strategy for imputing large-scale consecutive missing data and its application in a water quality prediction system[J]. Journal of Hydrology, 2021, 602: 126573.

[19] Lucarini A, Cascio M L, Marras S, et al. Artificial intelligence and eddy covariance: a review[J]. Science of the Total Environment, 2024, 950: 175406.

[20] Trotta C, Canfora E, Cheah Y, et al. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data[J]. Scientific Data, 2020, 7(1): 225−227.

[21] Vekuri H, Tuovinen J, Kulmala L, et al. A widely-used eddy covariance gap-filling method creates systematic bias in carbon balance estimates[J]. Scientific Reports, 2023, 13(1): 1720.

[22] 宋亮, 万建洲. 缺失数据插补方法的比较研究 [J]. 统计与决策, 2020, 36(18): 10−14.
Song L, Wan J Z. Comparative research on interpolation method of missing data[J]. Statistics and Decision, 2020, 36(18): 10−14.

[23] 郑小波. 基于机器学习的 MIMIC 数据库缺失数据插补研究 [D]. 合肥: 安徽大学, 2021.
Zheng X B. Study on missing data imputation in MIMIC database based on machine learning [D]. Hefei: Anhui University, 2021.

[24] Nie Y, Nguyen N H, Sinthong P, et al. A time series is worth 64 words: long-term forecasting with transformers[J/OL]. arXiv: 2211.14730v2. [2024−07−14]. https://arxiv.org/pdf/2211.14730.

[25] Yue Z, Wang Y, Duan J, et al. TS2Vec: towards universal representation of time series[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(8): 8980−8987.

[26] Liu Y, Hu T, Zhang H, et al. iTransformer: inverted transformers are effective for time series forecasting[J/OL]. arXiv: 2310.06625v3. [2024−08−14]. https://arxiv.org/pdf/2310.06625v3.

[27] Wu H, Hu T, Liu Y, et al. TimesNet: temporal 2D-variation modeling for general time series analysis[J/OL]. arXiv: 2210.02186v3. [2024−08−14]. https://arxiv.org/pdf/2210.02186.

[28] Rivera-Muñoz L M, Giraldo-Forero A F, Martinez-Vargas J D. Deep matrix factorization models for estimation of missing data in a low-cost sensor network to measure air quality[J]. Ecological informatics, 2022, 71: 101775.

[29] 李润东, 范雅倩, 冯沛, 等. 北京松山天然落叶阔叶林生态系统净碳交换特征及其影响因子 [J]. 应用生态学报, 2020, 31(11): 3621−3630.
Li R D, Fan Y Q, Feng P, et al. Net ecosystem carbon exchange and its affecting factors in a deciduous broad-leaved forest in Songshan, Beijing, China[J]. Chinese Journal of Applied Ecology, 2020, 31(11): 3621−3630.

[30] 张弥, 于贵瑞, 张雷明, 等. 太阳辐射对长白山阔叶红松林净生态系统碳交换的影响 [J]. 植物生态学报, 2009, 33(2): 270−282.
Zhang M, Yu G R, Zhang L M, et al. Effects of solar radiation on net ecosystem exchange of broadleaved-Korean pine mixed forest in changbai mountain, china[J]. Chinese Journal of Plant Ecology, 2009, 33(2): 270−282.

[31] 游桂莹, 张志渊, 张仁铎. 全球陆地生态系统光合作用与呼吸作用的温度敏感性 [J]. 生态学报, 2018, 38(23): 8392−8399.
You G Y, Zhang Z Y, Zhang R D. Temperature sensitivity of photosynthesis and respiration in terrestrial ecosystems globally[J]. Acta Ecologica Sinica, 2018, 38(23): 8392−8399.

（责任编辑 吴 娟
责任编委 陈志泊）