

陈琦, 苏国辉, 魏合龙, 等. 江苏盐城滨海湿地净生态系统碳交换量模拟参数选择[J]. 海洋地质前沿, 2023, 39(2): 56-65.

CHEN Qi, SU Guohui, WEI Helong, et al. Selection of parameters for simulation of net ecosystem carbon flux in Yancheng coastal wetland, Jiangsu[J]. Marine Geology Frontiers, 2023, 39(2): 56-65.

江苏盐城滨海湿地净生态系统碳交换量 模拟参数选择

陈琦^{1,2}, 苏国辉^{3*}, 魏合龙³, 叶思源³, 谢柳娟³, 丁喜桂³

(1 中国地质科学院 北京 100037; 2 中国地质大学(北京), 北京 100083; 3 中国地质调查局青岛海洋地质研究所, 青岛 266237)

摘要: 滨海湿地净生态系统碳交换量受到多种环境因素的影响, 在进行滨海湿地净碳交换量估算建模时, 参数的选择至关重要, 如何合理地选择输入参数不仅对于估算结果的精度有影响, 同时也会影响预测模型的适用性。本研究使用了 Pearson、Spearman、距离相关系数、最大互信息相关系数 4 种相关系数来计算各个环境因素与净碳交换量之间的相关性, 基于相关系数来选择最佳的输入参数组合。利用实际测得的江苏盐城盐沼湿地数据, 依次选择各个相关性中最高的 8 个参数组合, 基于卷积神经网络对江苏盐城滨海湿地 NEE 进行建模, 得到了 4 个预测模型, 并使用均方根误差和平均绝对值误差来进行模型精度的验证。研究表明, 使用基于最大互信息系数得到的参数组合进行滨海湿地 NEE 建模时模型的精度最好, 误差最小; 净光合有效辐射, 净辐射, 地表辐射与 NEE 在 4 个相关系数中都属于强相关, 表明这一类辐射类参数对滨海湿地 NEE 的影响要大于其他参数; 各参数与 NEE 之间的关系既包含线性关系也包含非线性关系, 传统的单一线性分析手段无法完整准确地反应各个环境参数与 NEE 之间的响应关系; 基于卷积神经网络的滨海湿地 NEE 预测模型在精度上要优于其它同类型模型, 这表明使用该模型在进行 NEE 预测建模时具有很好的适用性。

关键词: 卷积神经网络; 相关系数; 滨海盐沼湿地; 净生态系统碳交换量

中图分类号: P628.4

文献标识码: A

DOI: 10.16028/j.1009-2722.2022.247

0 引言

自人类进入工业世纪以来, 以二氧化碳为主的温室气体含量在大气中不断地增加, 由此导致的全球气候变暖已经对生态系统产生了重要影响。在此背景下, 世界各国正加紧出台一系列政策控制碳排放, 中国提出在 2030 年实现碳达峰, 2060 年实现

碳中和。为了实现这一目标, 既需要减少煤炭化石能源的消费, 加大清洁能源的投入, 也需要提升生态系统的碳汇能力, 发挥生态系统的固碳作用。滨海湿地作为海洋生态系统的重要组成部分, 具有巨大的碳吸收能力。据统计, 滨海湿地每平方千米的年碳埋藏量预计可达 0.22Gg C, 相当于 3.36×10^5 L 汽油燃烧所排放的二氧化碳量^[1], 尽管滨海湿地的面积仅占全球海洋面积的约 0.2%, 但湿地生态系统储存的碳却占海洋总碳量的 50%^[2], 因此, 量化滨海湿地的固碳能力对发挥滨海湿地的碳吸收作用至关重要。

1 净生态系统碳交换量

净生态系统碳交换量 (Net Ecosystem Carbon Exchange, NEE) 是目前量化生态系统固碳能力的一

收稿日期: 2022-09-05

资助项目: 国家重点研发计划“场景驱动的海洋科学大数据挖掘分析关键技术与应用”(2021YFF0704000); 青岛海洋科学与技术试点国家实验室山东省专项经费(2022QNLMO5032-4); 中国地质调查局项目(DD20221711)

作者简介: 陈琦(1998—), 男, 硕士, 主要从事地球信息科学技术方面的研究工作. E-mail: 760536077@qq.com

* 通讯作者: 苏国辉(1977—), 女, 硕士, 正高级工程师, 主要从事海洋地质信息化技术及应用研究. E-mail: sguohui@mail.cgs.gov.cn

个重要指标,反应生态系统中总呼吸作用和总生态系统生产力之间的动态平衡,当 NEE 为正值时反应该生态系统为净二氧化碳释放,为碳源,当 NEE 为负值时反应该生态系统为净二氧化碳吸收,为碳汇^[3]。

目前,对生态系统进行 NEE 建模的方法可以分为 2 类。①基于过程的生物模型,这类模型的代表是 DGVM 模型^[4]、IFUSE 模型^[5]、ECOSYS 模型^[6]等,着重于刻画碳通量复杂的时空变化机制,以生物圈的能量理论和假设为基础,使用数学关系来表达大气和地表之间的生物和物理过程^[7-8],这类模型的解释性很强,但往往过于复杂,难以应对大范围多尺度的情况;②基于数据驱动模型,代表是随机森林(RF)、支持向量机(SVM)、人工神经网络(ANN),可以从大量的数据中挖掘有效信息,不需要基于太多的假设,可以很好地刻画碳循环过程中的非线性关系,适用于多尺度大范围研究,近年来已经被成功用于解决生态系统和生理过程预测和建模问题^[9],在陆地上的如森林、稻田等生态系统的 NEE 预测建模中也得到了应用^[10-11],滨海湿地 NEE 建模方面相关研究仍较少。此外,由于 NEE 反应的是生态系统中碳总吸收和总排放之间的一种平衡,受到多种因素的影响,如光照辐射、气候条件、土壤环境等^[12],但是在实际建模中无法将所有的参数都纳入模型中,因此,如何合理地选择最佳的参数组合对于模型的建立至关重要。

在本次滨海湿地 NEE 建模研究中使用了 4 种相关系数方法来计算环境变量与 NEE 之间的相关性,选择每种相关系数中最高 8 个参数作为输入参数组合,使用卷积神经网络来进行预测模型的构建,将计算得到的 4 种参数组合输入到 NEE 预测卷积神经网络中进行训练,得到 4 种模型,通过比较预测值和实测值之间的误差,选择出最佳的参数组合,为未来的滨海湿地 NEE 建模研究提供基础。

2 研究地区与数据来源

本次研究区为江苏盐城滨海湿地(32°32′—34°25′N, 119°55′—121°50′E),位于江苏省沿海中部,苏北江淮平原东部,距盐城市中心约 50 km,湿地总面积近 $24.30 \times 10^4 \text{ hm}^2$ 。地处暖温带季风和亚热带季风的过渡地带,四季分明,雨热同期,年平均气温 13~16 ℃,受台风和梅雨季节的影响显著。湿地内

植被种类多样,主要的植被类型包括互花米草、芦苇、翅碱蓬和茅草,其中碱蓬盐沼湿地面积 23.8 km²、芦苇盐沼湿地面积 192.0 km²、米草盐沼湿地面积为 203.3 km²。

本研究数据通过涡度塔实地测得,数据观测时间为 2019 年 4 月 1 日至 2019 年 10 月 31 日,记录间隔为 0.5 h,共测得 5 397 条数据,观测指标共有 15 个(表 1)。

表 1 观测参数符号及名称
Table 1 Symbols and names of the variables

符号	参数名称	符号	参数名称
RG	地表辐射	WDir	风向
PPDF	光合有效辐射	APre	气压
RN	净辐射	Rain	降雨量
ATemp	气温	WSpee	风速
RH	相对湿度	Vapor	蒸汽压
TS	土壤温度	SWC	土壤含水量
Temp	地表温度	ALB	土壤反照率
NEE	净生态系统碳交换量		

3 基于数据驱动的 NEE 预测模型

在过去的几十年间,研究人员使用了多种方法来进行碳交换量的模拟估算,根据原理跟技术不同可以分为基于过程的模型和基于数据驱动的模型^[13]。基于过程的模型应用的时间最长,模型的理论基础也较完备,其特点在于使用数学关系来描述碳交换量时空变化的复杂物理过程,并且简化了选定过程和生态系统组成部分之间的关系^[14],解释性比较强。然而,近年也逐渐发现这种模型在具体应用上存在较大的不确定性^[15-16],且参数的确定过于复杂,模型之间结构差异明显,很难应用到不同尺度上。

基于数据驱动的模型近年得到了较多的关注,很大程度得益于该模型可以从大量数据中提取到有用信息,不需要太多先验知识与假设,同时随着以深度学习为首的数据挖掘技术的蓬勃发展,这类模型在精度和应用范围上都逐渐超过基于过程的模型,在生态系统研究和建模中也得到了大量的应用^[17-19]。

在本次研究中,从实际应用出发,选择基于过滤器的方法来进行参数的选取,主要是因为这类方法的计算方法成熟,独立于具体模型的特性也使得

这种方法可以在各个不同模型中通用,有利于后期进行比较分析。基于过滤器的方法包括部分互信息方法^[20]、条件互信息方法^[21]、相关系数方法^[22]、Gamma 测试集方法^[23]等。本次研究中使用的是基于相关系数的方法,在 NEE 与其它环境变量的响应机制研究中,相关系数分析是常用的方法之一,然而目前使用的相关系数方法大多数是线性的相关系数法,如 Pearson 相关系数、Spearman 相关系数等^[24-25],但是对于自然机制来说各个能量通量之间的响应机制不是简单的线性关系,这是由其自身的复杂性和不均匀性决定的,因此,使用这类系数可以对自然机制中的线性关系进行描述,却无法对于其中的非线性关系做出解释。为了解决这一问题,在本次研究中除了使用经典的线性相关系数外,还引入了距离相关系数和最大信息系数,来探究 NEE 与其他环境变量之间非线性关系。

3.1 数据的预处理与分析

本次数据的预处理主要包括缺省值的去除与标准化处理。由于观测或者其他客观条件的影响,实测数据中存在部分观测参数没有观测值或者观测值异常的情况,为了避免这些数据对后续处理的影响,先对存在缺省值的数据进行了剔除处理,得到了 1 443 条数据。此外由于各个参数之间的计量单位不同,为了将其纳入到同一个模型中,需要先对数据进行标准化处理。本次使用最小-最大标准化方法,即离散标准化,通过线性变化将数据值映射到 [0,1],同时不会对数据原始特征造成影响,转换方法如式(1)所示。

$$x^* = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

式中: x^* 为转换后数据;
 x 为原始数据。

3.2 相关系数的计算与输入参数的选取

相关系数是反映 2 个变量之间相关关系密切程度的统计指标,被广泛应用于数据驱动模型的参数选择中,该方法最早由统计学家卡尔皮尔逊提出,经过多年发展相关系数已经出现了多种类型,根据应用方向的不同,相关系数可以简单地分为线性相关系数和非线性相关系数,线性相关系数侧重于描述和计算参数之间的线性关系,主要代表有 Pearson 相关系数, Spearman 相关系数等;而非线性相关

系数着重于描述参数之间的非线性关系,包括距离相关系数和最大信息系数等,前者的出现时间较早,在数学建模分析中应用的也比较多,然而随着近些年人们研究的不断深入,这类相关系数的局限性正在逐渐显露,非线性相关系数得到了越来越多的关注。在本次研究中为了找出最佳的参数组合,同时更全面地分析各个参数与 NEE 之间的相关关系,选择了 4 种相关系数来进行计算,即 Pearson、Spearman、距离相关系数和最大信息系数,通过对计算结果进行排序的方法,来选择相关性最强的 8 个参数作为输入变量来进行模型的训练,

3.2.1 Pearson 相关系数的计算

Pearson 相关系数是经典的相关系数计算方法,该方法由英国统计学家皮尔逊提出,主要用于计算 2 组连续变量之间的线性关联程度,该系数定义为 2 个变量之间协方差和标准差之间的商,系数的正负号表示正相关还是负相关,结果趋于 [-1,1] 之间,系数的绝对值越接近 1,表明变量之间的线性关系越强。计算如式 2 所示:

$$R_p = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (2)$$

式中: R_p 为 Pearson 相关系数;

n 为样本数;

x_i 为预测值;

y_i 为实测值。

为了更直观的比较对于得到的值进行了求取绝对值处理,最终计算结果如表 2 所示。

表 2 Pearson 相关系数计算结果

参数名称	R_p	参数名称	R_p
RG	0.607 8	RN	0.599 3
PPDF	0.613 9	ALB	0.000 1
ATemp	0.446 1	RH	0.362 6
Temp	0.343 7	SWC	0.033 8
WDir	0.020 6	TS	0.120 6
Vapor	0.023 0	APre	0.056 1
WSpee	0.153 7	Rain	0.027 6

从图 1 可以看出该方法计算出来的相关性最强的 8 个参数为 PPDF、RG、RN、ATemp、RH、Temp、WSpee、TS。

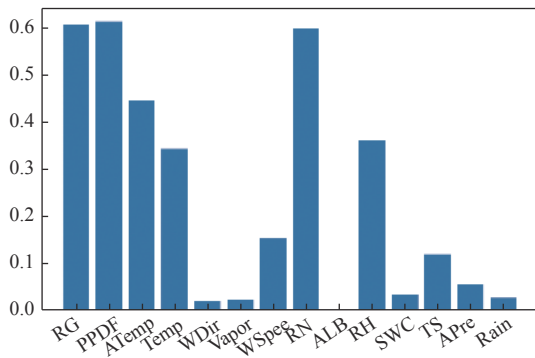


图 1 Pearson 相关性比较

Fig.1 Pearson correlation comparison chart

3.2.2 Spearman 相关系数的计算

斯皮尔曼相关系数,也被称为 Spearman 秩相关系数,与 Pearson 相关系数不同, Spearman 相关系数的计算基于每个变量的秩值来计算相关性,而非具体的数值,其值与 2 组相关变量的具体值无关,而仅仅与变量之间的大小关系有关,该方法基于 2 列成对登记的各队等级数之差来进行计算,与 Pearson 相关系数类似,系数的正负号表示正相关还是负相关,结果趋于 $[-1,1]$,系数的绝对值越接近 1,表明变量之间的线性关系越强。计算方法如下所示:

对变量 x 和变量 y 中的样本进行排序,得到每个样本的次序值 $Rg(x_i)$, $Rg(y_i)$

计算每个变量中对应数据次序值的差值 d_i (式 3)

$$d_i = Rg(x_i) - Rg(y_i) \quad (3)$$

计算得到 Spearman 相关系数 R_s (式 4)

$$R_s = 1 - \frac{\sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (4)$$

式中: n 为变量 x 和变量 y 中的样本个数。

与 Pearson 相关系数类似,在进行参数相关性排序时先对计算得到的 Spearman 相关系数进行了求取绝对值的处理,最终得到的计算结果如表 3 所示。

从图 2 可以看出使用 Spearman 相关系数计算得到的相关性最强的 8 个变量是: RG、PPDF、RN、ATemp、RH、Temp、ALB、WSpee。

3.2.3 距离相关系数的定义与计算

距离相关系数是由 SZEKELY 等^[26]在 2007 年提出的一种衡量变量之间相关性的一种新手段,距离协方差矩阵和距离相关类似于乘积矩阵协方差,但是与传统的相关系数不同,距离相关系数为 0 可

表 3 Spearman 相关系数计算结果

Table 3 Results of the Spearman coefficient calculation

参数名称	R_s	参数名称	R_s
RG	0.683 5	RN	0.638 8
PPDF	0.672 6	ALB	0.236 7
ATemp	0.484 5	RH	0.442 1
Temp	0.355 0	SWC	0.086 3
WDir	0.054 5	TS	0.134 5
Vapor	0.039 0	APre	0.035 0
WSpee	0.183 6	Rain	0.034 1

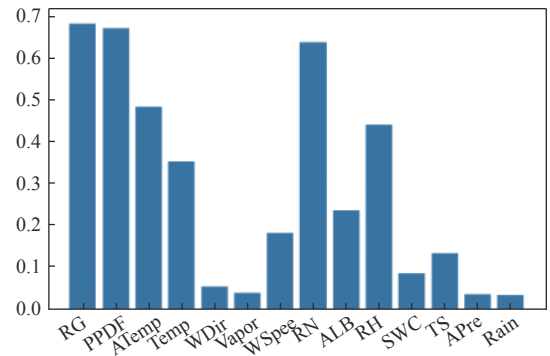


图 2 Spearman 相关性比较

Fig.2 Comparison in the Spearman correlation

以唯一确定变量之间的独立性,因此,该相关系数可以很好地对变量之间的非线性关系进行描述。距离相关系数取值在 $[0,1]$ 之间,值为 0 表示相互独立,完全无关,值越大,表明相关性就越强。

距离相关系数的计算基于距离协方差和距离方差,计算过程如下所示:

计算变量之间的范数距离矩阵(式 5、6)

$$a_{i,j} = \|x_i - x_j\|, i, j = 1, 2, \dots, n \quad (5)$$

$$b_{i,j} = \|y_i - y_j\|, i, j = 1, 2, \dots, n \quad (6)$$

式中: $a_{i,j}$, $b_{i,j}$ 为变量 x 和变量 y 各个样本之间的范数距离;

n 为变量 x, y 中样本的总个数。

对范数距离矩阵进行中心化处理(式 7、8)

$$A_{ij} = a_{i,j} - \bar{a}_j - \bar{a}_i + \bar{a} \quad (7)$$

$$B_{ij} = b_{i,j} - \bar{b}_j - \bar{b}_i + \bar{b} \quad (8)$$

式中: A_{ij} 为中心化处理后变量 x 的范数距离值;

$\bar{a}_j$ 为距离范数矩阵中第 j 列的平均值;

$\bar{a}_i$ 为第 i 行的平均值;

$\bar{a}$ 为整个矩阵中的平均值。

对于矩阵 B 来说计算同理。

计算变量之间以及变量自身的平方协方差(式 9~11)

$$dCov_n^2(x, y) := \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n A_{i,j} B_{i,j} \quad (9)$$

$$dCov_n^2(x) = \frac{1}{n^2} A_{i,j}^2 \quad (10)$$

$$dCov_n^2(y) = \frac{1}{n^2} B_{i,j}^2 \quad (11)$$

式中: A_{ij} 为中心化处理后变量 x 的范数距离值;
 B_{ij} 为中心化处理后变量 y 的范数距离值;
 $dCov_n^2(x, y)$ 为变量 x 与变量 y 之间的平方协方差;

$dCov_n^2(x)$ 为变量 x 自身的平方协方差;

$dCov_n^2(y)$ 为变量 y 自身的平方协方差;

n 为样本数。

计算得到距离相关矩阵(式 12)

$$dCor(x, y) = \frac{dCov(x, y)}{\sqrt{dCov(x) dCov(y)}} \quad (12)$$

式中: $dCov(x, y)$ 为变量 x, y 之间的协方差;

$dCov(x)$ 为变量 x 自身的协方差;

$dCov(y)$ 为变量 y 自身的协方差;

最终的计算结果如表 4 所示。

表 4 距离相关系数计算结果

Table 4 Results of distance coefficient calculation

参数名称	$dCor$	参数名称	$dCor$
RG	0.651 4	RN	0.638 2
PPDF	0.655 2	ALB	0.207 6
ATemp	0.491 0	RH	0.419 2
Temp	0.366 3	SWC	0.145 6
WDir	0.212 2	TS	0.179 9
Vapor	0.146 1	APre	0.117 2
WSpee	0.176 2	Rain	0.046 9

从图 3 可以看出,该方法计算得到的相关程度最高的 8 个变量分别是 PPDF、RG、RN、ATemp、RH、Temp、WDir、ALB。

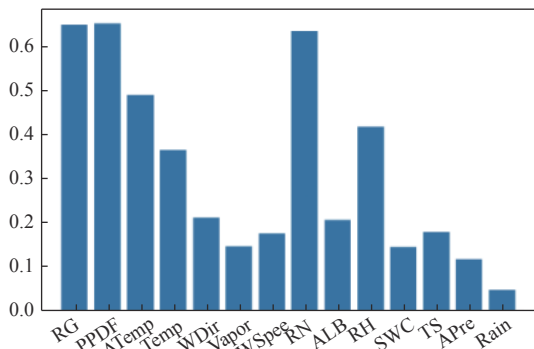


图 3 距离相关性比较

Fig.3 Comparison in distance correlation

3.2.4 最大信息系数定义与计算

最大信息系数(maximal information coefficients, MIC),该方法脱胎于互信息(mutual information, MI),用于衡量两个变量之间的关联程度,既包括线性关系也包括非线性关系,该方法通过将 2 个变量之间的关系离散在二维空间中,使用散点图来进行表示,将二维空间进行不断地划分,使用计算联合概率的方法来得到最终的结果,值越大表明该相关性就越强。

具体计算可以分为 3 步:

①在给定的 i, j 条件下,对变量 X , 和变量 Y 构成的散点图进行网格化,通过公式 2 计算得到互信息值,求取最大值(式 13):

$$I(x; y) = \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dx dy \quad (13)$$

其中 $p(x, y)$ 为变量 x, y 的联合概率。

②对得到的最大互信息值进行归一化处理(式 14)。

$$mic(x; y) = \frac{I(x; y)}{\log_2 \min(x, y)} \quad (14)$$

③选择不同 i, j 进行计算,从中选择出最大的值作为最大互信息系数。

最终计算结果如表 5 所示。

表 5 最大信息系数计算结果

Table 5 Results of the maximal information coefficients calculation

参数名称	mic	参数名称	mic
RG	0.539 6	RN	0.522 7
PPDF	0.494 7	ALB	0.403 6
ATemp	0.275 7	RH	0.247 4
Temp	0.213 4	SWC	0.195 0
WDir	0.170 8	TS	0.139 5
Vapor	0.121 6	APre	0.103 3
WSpee	0.099 0	Rain	0.078 1

从图 4 中可以看出,基于最大信息系数计算得到的相关性最强的 8 个参数为: RG、RN、PPDF、ALB、ATemp、RH、Temp、SWC。

最终,4 种相关系数计算得到的 4 种参数组合如表 6 所示。

3.3 基于卷积神经网络的 NEE 预测模型构建与训练

卷积神经网络(CNN)是人工神经网络(ANN)中的一种模型,由多个人工神经元组成并行处理系

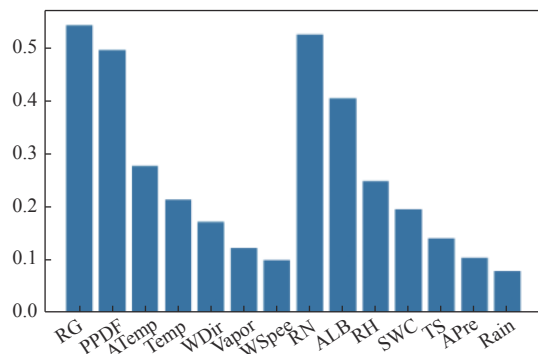


图 4 最大信息系数比较

Fig.4 Comparison in the maximal information coefficient

表 6 参数组合

Table 6 Combination of the parameters

相关系数计算方法	参数组合
Pearson相关系数	光合有效辐射, 地表辐射, 净辐射, 气温, 相对湿度, 地表温度, 风速, 土壤温度
Spearman相关系数	地表辐射, 光合有效辐射, 净辐射, 气温, 相对湿度, 地表温度, 土壤反照率, 风速
距离相关系数	光合有效辐射, 地表辐射, 净辐射, 气温, 相对湿度, 地表温度, 风向, 土壤反照率
最大信息系数	地表辐射, 净辐射, 光合有效辐射, 土壤反照率, 气温, 相对湿度, 地表温度, 土壤含水量

统,也是基于数据驱动模型中的代表模型,这类模型能很好地映射出输入量和输出变量之间的非线性关系,这一优势已经在多个领域得到了证明^[27-28]。在 NEE 估算领域,也引进了人工神经网络进行建模,如 MELESSE^[29]就使用了神经网络来对森林、草地、小麦田进行了 NEE 估测建模,选择了 6 个参数作为输入变量;WEN 等^[30]则使用了遗传神经网络(GNN)来对亚热带森林的碳交换量进行了估测预算,结果表明该模型在精度和相关性上都要优于基于过程的传统模型。

本研究构建了一个卷积神经网络来对江苏盐城滨海湿地的净碳交换量进行建模,该网络主要通过卷积运算进行处理,是一种具有深度结构的前馈神经网络^[31],该网络通过对输入数据进行卷积运算来提取关键特征,基于优化器来对模型的内部参数进行调整,通过多次迭代的方式来优化模型参数,最终达到提高模型精度的目的。已经在陆地生态系统碳交换量估算上得到了应用^[32-33]。

本次 NEE 估算构建的神经网络由 3 个卷积层,3 个激活层以及 1 个全连接层组成(图 5),激活函数选用 Relu 函数,使用 Adagrad 优化算法来进行参数的优化,Adagrad 优化是一种基于梯度的算法,该算法将学习速率与参数进行适应,在为不同的变量提供不同学习率的同时,还可以提高主要特征对模型的影响^[34],训练次数设置为 1 000 次,相关代码

由 Python3 语言编写,基于 Pytorch 架构来进行训练和构建。

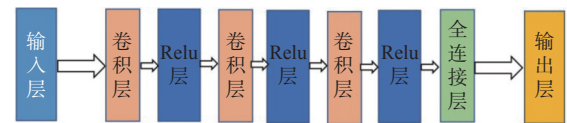


图 5 NEE 预测卷积神经网络结构

Fig.5 NEE prediction convolutional neural network structure

基于上文相关系数计算得到的结果,将 4 种参数组合依次输入到卷积神经网络中进行模型的训练,最终得到 4 个模型,依次命名为 CNN_Model1、CNN_Model2、CNN_Model3、CNN_Model4。

3.4 模型的验证与精度评价

为了验证不同参数组合训练得到的模型的具体精度差异,选取使用均方根误差和绝对值误差来对模型的预测误差进行量化,模型的均方根误差和绝对值误差越小,表明该模型的精度越高。

本次验证从数据中随机抽取 150 个样本数据,计算实测值和预测值之间的均方根误差(RMSE)和绝对值误差(ε),具体计算方法如式子 15、16 所示:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (NEE_i^* - NEE_i)^2}{n}} \quad (15)$$

$$\varepsilon = \frac{\sum_{i=1}^n (NEE_i^* - NEE_i)}{n} \quad (16)$$

式中: NEE^* 为 NEE 的预测值;

NEE 为实测值;

n 为预测样本的数量,计算结果如表 7 所示。

表 7 误差计算结果

Table 7 Results of calculation error

	均方根误差	绝对值误差
CNN_Model1	0.013 4	0.064 0
CNN_Model2	0.009 2	0.068 4
CNN_Model3	0.010 9	0.057 4
CNN_Model4	0.005 1	0.043 9

使用基于最大信息相关系数的参数组合得到的模型 CNN_Model4 无论是均方根误差还是绝对值误差在 4 个模型中都是最低的,表明该参数组合是最佳的。

同时,为了验证卷积神经模型在 NEE 预测精度上的优劣,将其与现有相关研究中的其他数据驱

动模型进行了比较,结果和相关参考文献如表 8 所示。可以看出,使用卷积神经网络结构的 4 个模型在精度上都要优于其他数据驱动模型,表明这一模型在滨海湿地 NEE 模拟预测上具有一定的适用性。

表 8 本次模型预测精度与其他模型精度的比较
Table 8 Comparison in prediction accuracy between CNN model and other data driven models

模型	均方根误差	参考文献
ELM模型	0.589 3	[35]
SVM	1.476 4	[35]
GNN	0.112 3	[36]
RF	0.612 0	[36]
CNN_Model1	0.005 1	
CNN_Model2	0.010 9	
CNN_Model3	0.009 2	
CNN_Model4	0.013 4	

4 讨论和分析

4.1 滨海湿地碳交换量影响因素

生态系统的碳交换量交换过程以植物光合作用主导的碳固定与以土壤有机质矿化和动植物呼吸为主导的碳释放为主,滨海湿地位于海洋生态系统和陆地生态系统之间的过渡地带,因此,滨海湿地生态系统碳交换过程更为复杂,除了陆地生态系统中的光合作用、矿化作用、动植物呼吸作用以外,湿地周期性的水位变化以及水文驱动的流入流出也会接受和释放碳交换量,同时滨海湿地自身又对气候变化以及相关因素又较为敏感,海平面的上升以及极端气候也会对滨海湿地的碳交换量交换产生影响^[37]。

现有研究关注的往往是单一或者少数变量与 NEE 之间的响应机制,如光照辐射、海平面上升^[38]、温度^[39] 等对 NEE 的影响,鲜有对多类型变量进行的研究,也少有对参数之间响应的差异性进行分析。本研究从相关系数出发,对 15 个参数都进行了相关性的计算,既包括辐射类参数,也包括了气候类参数,土壤类参数等,通过量化比较的方式可以很直观地对各个参数与 NEE 之间的响应机制做出解释。

4.2 各个生物环境参数与 NEE 之间的相关性差异比较

从相关系数的计算结果(图 6)可以发现各个参

数与 NEE 之间都存在一定的相关关系,表明 NEE 同时受到多种类型变量的影响,各个参数与 NEE 之间的相关性从计算结果上可以明显地分为 3 个梯次:第 1 梯次是以辐射参数为主的变量,具体包括净光合有效辐射、净辐射、地表辐射,这 3 个变量与 NEE 的相关性在 4 类相关系数中都是属于强相关,而且明显高于其他变量;第 2 梯次主要是包括气温、湿度、地表温度,这 3 个变量主要与气候有关,与 NEE 的相关性也比较强但是要低于辐射参数;第 3 梯次则主要包括风速、土壤温度、土壤含水率等参数,与 NEE 之间不论是线性关系还是非线性关系都显示低相关性。除此之外还发现,部分参数与 NEE 之间的非线性相关性在某种程度上要强于线性相关性,这表明不同的参数与 NEE 之间的相关性是存在差异的,有的可能倾向于显示线性相关关系,有的更多显示为非线性相关。这种差异性反应的可能是不同生理过程在滨海湿地碳循环中所扮演角色的差异,因为滨海湿地碳循环本身就涉及到多种类型的生理过程,如光合作用、呼吸作用、土壤矿化作用、水文作用等,在碳循环中的作用是不同的,在未来的研究中应当着重分析这种差异的来源以及这种差异对滨海湿地固碳作用的影响。

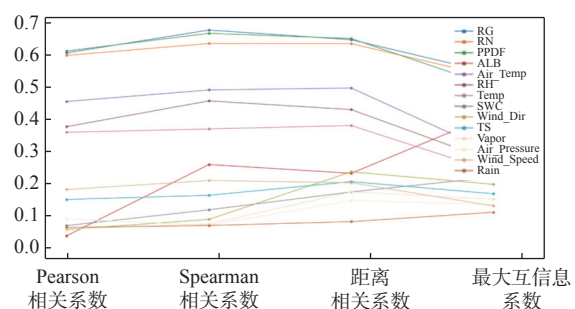


图 6 各个参数与 NEE 之间的相关系数

Fig.6 Correlation coefficient between each parameter and NEE

在进行滨海湿地 NEE 建模时,参数的选择本身是一个比较复杂的过程,既需要真实反映出各个参数与 NEE 之间的关系,同时也需要兼顾参数选取的代表性和合理性,精度上才能得到保证,在机制上也能很好地解释碳循环过程。除此之外,在分析各个环境变量与 NEE 的响应机制时,不能单纯地从线性或非线性的角度出发,单一研究方法很难发现差异,因为有的参数与 NEE 之间可能线性关系较强,而有的参数与 NEE 之间的非线性关系比较强。以反照率为例,该参数的 Pearson 相关系数计算结果为 0.000 14,显示弱相关,而使用 MIC 计

算结果为 0.403 6, 却显示强相关, 在进行参数选取时需要从多个角度进行分析。

5 结论

(1) 基于卷积神经网络的 NEE 预测模型最合适的输入参数组合为: 地表辐射、净辐射、光合有效辐射、土壤反照率、气温、相对湿度、地表温度、土壤含水率。

(2) 各个参数与 NEE 的相关性存在差异, 辐射类参数如净光合辐射、净辐射、地表辐射与 NEE 之间属于强相关, 在 4 类相关系数中都呈现出相似结果, 这可能在某一程度上表明辐射类参数在滨海湿地碳交换量交换过程中的影响大于其他参数的影响。

(3) 参数自身对于 NEE 之间的相关性也存在差异, 有的参数与 NEE 之间的线性相关性可能较强, 而有的参数与 NEE 之间可能表现得更多是非线性关系。

(4) 本次研究训练得到的基于卷积神经网络的模型在预测精度上都要优于其他同类型模型, 这类模型在滨海湿地 NEE 预测模拟上有很好的适用性。

参考文献:

- [1] 唐剑武, 叶属峰, 陈雪初, 等. 海岸带蓝碳的科学概念、研究方法以及在生态恢复中的应用[J]. 中国科学: 地球科学, 2018, 48(6): 661-670.
- [2] TANG J W, YE S F, CHEN X C, et al. Coastal blue carbon: concept, study method, and the application to ecological restoration[J]. Science China: Earth Sciences, 2018, 61(6): 637-646.
- [3] 吕亚香, 戚智彦, 刘伟, 等. 早春和夏季氮磷添加对内蒙古典型草原退化群落碳交换的影响[J]. 植物生态学报, 2021, 45(4): 11-13.
- [4] MÜLLER C, LUCHT W. Robustness of terrestrial carbon and water cycle simulations against variations in spatial resolution[J]. Journal of Geophysical Research: atmospheres, 2007, 112(D6): 19-26.
- [5] DESAI, ANKUR R. Climatic and phenological controls on coherent regional interannual variability of carbon dioxide flux in a heterogeneous landscape[J]. Journal of Geophysical Research. Biogeosciences, 2011, 116(G3): G00J02-1-G00J02-13.
- [6] GRANT R F, BALDOCCHI D D, MA S. Ecological controls on net ecosystem productivity of a seasonally dry annual grassland under current and future climates: modelling with Ecosys[J]. Agricultural and Forest Meteorology, 2012, 152: 189-200.
- [7] DOU X, YANG Y. Estimating forest carbon fluxes using four different data-driven techniques based on long-term eddy covariance measurements: model comparison and evaluation[J]. Science of the Total Environment, 2018, 627: 78-94.
- [8] CAI J, XU K, ZHU Y, et al. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest[J]. Applied Energy, 2020, 262: 1429-1442.
- [9] AGO E E, AGBOSSOU E K, GALLE S, et al. Long term observations of carbon dioxide exchange over cultivated savanna under a Sudanian climate in Benin (West Africa) [J]. Agricultural and Forest Meteorology, 2014, 197: 13-25.
- [10] HAO L, ZHANG Z. Mining the intrinsic trends of CO₂ solubility in blended solutions[J]. Journal of CO₂ Utilization, 2018, 26: 496-502.
- [11] ZHANG Z, LI H, CHANG H et al. Machine learning predictive framework for CO₂ thermodynamic properties in solution[J]. Journal of CO₂ Utilization, 2018, 26: 152-159.
- [12] WU J, ALBERT L P, LOPES A P et al. Leaf development and demography explain photosynthetic seasonality in Amazon evergreen forests[J]. Science, 2016, 351(6276): 972-976.
- [13] VELPURI M N, SENAY G B, SINGH R K, et al. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: using point and gridded FLUXNET and water balance ET[J]. Remote Sensing of Environment: An Interdisciplinary Journal, 2013, 139: 35-49.
- [14] CAI J, XU K, ZHU Y et al. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest[J]. Applied Energy, 2020, 262: 114566-114566.
- [15] WANG Y Y, XIE Z H, JIA B H, et al. Improving simulation of the terrestrial carbon cycle of China in Version 4.5 of the community land model using a revised V_c(cmax) Scheme[J]. Atmospheric and Oceanic Science Letters, 2015, 8(2): 88-94.
- [16] SCHAEFER K, SCHWALM C R, WILLIAMS C, et al. A model-data comparison of gross primary productivity: results from the north American carbon program site synthesis[J]. Journal of Geophysical Biogeosciences Research, 2012, 117(G3): G03010-1-G03010-15.
- [17] CHE S, TAN X, XIANG C, et al. Stand basal area modelling for Chinese fir plantations using an artificial neural network model[J]. Journal of Forestry Research, 2019, 30(5): 1641-1649.
- [18] JUNG M, REICHSTEIN M, MARGOLIS H, et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations[J]. Journal of Geophysical Research. Biogeosciences, 2011, 116(G3): G00J07-1-G00J07-16.
- [19] XIAO J F, ZHUANG Q L, BALDOCCHI D D, et al. Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data[J]. Agricultural and Forest Meteorology, 2008, 148(11): 1827-1847.
- [20] ROBERT J M, MAIER H R, DANDY J M, et al. Non-linear variable selection for artificial neural networks using partial mutual information[J]. Environmental Modelling & Software, 2008, 23(10/11): 1312-1326.

- [21] QUILTY J, ADAMOWSKI J, KHALIL B, et al. Bootstrap rank-ordered conditional mutual information (broCMI): a nonlinear input variable selection method for water resources modeling[J]. *Water Resources Research*, 2016, 52(3): 2299-2326.
- [22] GALELLI S, HUMPHREY G B, MAIER H R, et al. An evaluation framework for input variable selection algorithms for environmental data-driven models[J]. *Environmental Modelling & Software*, 2014, 62: 33-51.
- [23] NOORI R, KARBASSI A R, MOGHADDAMNIA A, et al. Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction[J]. *Journal of Hydrology*, 2011, 401(3/4): 177-189.
- [24] 李永强, 董智, 丁晨曦, 等. 山东省暖性草丛生态系统碳库现状和碳通量季节变化特征[J]. *植物生态学报*, 2018, 42(3): 11-13.
- [25] 王富强, 刘扬李, 杨欢, 等. 寒区滨河湿地生态系统碳通量特征及影响因素[J]. *水电能源科学*, 2018, 36(10): 68-71.
- [26] GJ SZÉKELY, RIZZO M L, BAKIROV N K. Measuring and testing dependence by correlation of distances [C]//ACM Symposium on Virtual Reality Software & Technology, 2007.
- [27] ABRAHART R J, ANCTIL F, Coulibaly P, et al. Two decades of anarchy Emerging themes and outstanding challenges for neural network river forecasting[J]. *Progress in Physical Geography*, 2012, 36(4): 480-513.
- [28] YADAV A K, CHANDEL S S. Solar radiation prediction using Artificial Neural Network techniques: a review[J]. *Renewable and Sustainable Energy Reviews*, 2014: 33-33.
- [29] MELESSE A M, HANLEY R S. Artificial neural network application for multi-ecosystem carbon flux simulation[J]. *Ecological Modelling*, 2005, 189(3/4): 305-314.
- [30] WEN X D, ZHAO Z H, DENG X W, et al. Applying an artificial neural network to simulate and predict Chinese Cross Mark fir (*Cunninghamia lanceolata*) plantation carbon flux in sub-tropical China[J]. *ECOL MODEL*, 2014, 294(5): 19-26.
- [31] GU, JIUXIANG, WANG, ZHENHUA, KUEN, JASON, et al. Recent advances in convolutional neural networks[J]. *Pattern Recognition: The Journal of the Pattern Recognition Society*, 2018: 77354-377.
- [32] HAMZEHIE M E, MAZINANI S, DAVARDOOST F, et al. Developing a feed forward multilayer neural network model for prediction of CO₂ solubility in blended aqueous amine solutions[J]. *Journal of Natural Gas Science and Engineering*, 2014, 21: 19-25.
- [33] BEIGZADEH R, RAHIMI M, SHABANIAN S R. Developing a feed forward neural network multilayer model for prediction of binary diffusion coefficient in liquids[J]. *Fluid Phase Equilibria*, 2012, 331: 48-57.
- [34] SLAVAKIS K, THEODORIDIS S, YAMADA I. Online Kernel-Based Classification Using Adaptive Projection Algorithms[J]. *IEEE Transactions on Signal Processing: a publication of the IEEE Signal Processing Society*, 2008, 56(7): 2781-2796.
- [35] ZHANG K, ZHU G F, MA J Z, et al. Parameter analysis and estimates for the MODIS evapotranspiration algorithm and multiscale verification[J]. *Water Resources Research*, 2019, 55(3): 2211-2231.
- [36] LIU Y L, ZHOU G M, DU H Q, et al. Response of carbon uptake to abiotic and biotic drivers in an intensively managed Lei bamboo forest[J]. *Journal of Environmental Management*, 2018, 223: 713-722.
- [37] LIU H X, YI Y J, YUE Y S, et al. Reducing the likelihood of carbon loss from wetlands by improving the spatial connections between high carbon patches[J]. *Journal of Cleaner Production*, 2020, 267: 121819.1-121819.12.
- [38] 韩智献, 全川, 刘白贵, 等. 模拟海平面上升和氮负荷增加对河口感潮沼泽湿地CO₂垂直交换的影响[J]. *环境科学学报*, 2021, 41(6): 2421-2429.
- [39] 陈小平, 刘廷玺, 王冠丽, 等. 温度和水分对科尔沁草甸湿地净生态系统碳交换量的影响[J]. *应用生态学报*, 2018, 29(5): 1523-1534.

Selection of parameters for simulation of net ecosystem carbon flux in Yancheng coastal wetland, Jiangsu

CHEN Qi^{1,2}, SU Guohui^{3*}, WEI Helong³, YE Siyuan³, XIE Liujuan³, DING Xigui³

(1 Chinese Academy of Geological Science, Beijing 100037, China;

2 Qingdao Institute of Marine Geology, China Geological Survey, Qingdao 266237, China;

3 China University of Geosciences, Beijing 100083, China)

Abstract: The net ecosystem carbon exchange (NEE) of coastal wetland is affected by various environmental factors. The selection of parameters is very important for estimating and modeling the NEE of coastal wetland. How to reasonably select the input parameters affects not only the accuracy of the estimation results, but also the applicability of the prediction model. Four correlation coefficients were used, including the Pearson correlation coefficient, the Spearman correlation coefficient, the distance correlation coefficient, and the correlation coefficient of maximum mutual information, to calculate the correlation between various environmental factors and NEE, according to which the best combination of input parameters was chosen. Using the measurement data of the Yancheng salt marsh wetland in Jiangsu Province, eight parameter combinations with the highest correlation were selected, then eight factors were input into the convolutional neural network for model training, and finally four prediction models obtained. The root mean square error and mean absolute error were used to verify the accuracy of the model. After calculation, the root mean square errors of the four models were 0.013 4, 0.009 2, 0.010 9, 0.005 1, and the absolute errors were 0.064, 0.068, 0.057 4, 0.043 9, respectively. This study shows that: 1) to model the NEE of coastal wetland with the parameter combination based on the maximum mutual information coefficient, the photosynthetic effective radiation, surface radiation, net radiation, photosynthetic effective radiation, soil albedo, air temperature, relative humidity, surface temperature, and soil moisture content, the accuracy of the modeling is the best and the error is the smallest. 2) Among 15 parameters used, the net photosynthetic effective radiation, net radiation, surface radiation, and NEE are strongly correlated in the four correlation coefficients, which showed that radiation parameter had a greater impact on wetland carbon cycle than other parameters. 3) Relationship between each parameter and NEE included both linear and nonlinear relationships. The conventional single linear analysis method cannot completely and accurately reflect the response relationship between each environmental parameter and NEE. In the future works, we shall not only study the linear relationship among variables but also pay more attention to the nonlinear mutual relationship. 4) The accuracy of the coastal wetland NEE prediction model based on convolutional neural network was better than other similar models', which shows that the model is applicable in NEE prediction modeling. This study provided a reference for NEE prediction modeling and analysis of coastal wetland in the future.

Key words: convolutional neural network; correlation coefficient; coastal wetland; NEE