

基于机器学习方法的农田下垫面近地层湍流通量订正研究

刘贞榕¹, 李煜斌^{1*}, 王泓¹, 刘昕晔², 高志球¹

1. 南京信息工程大学大气物理学院, 南京 210044

2. 浙江省杭州市桐庐县气象局, 桐庐 311500

摘要: 近地层湍流通量是表征地-气相互作用的重要变量, 在梯度观测或大气数值模式中通常基于 Monin-Obukhov 相似理论 (MOST) 进行计算。为进一步提升近地层湍流通量的计算精度, 本文以涡动相关法得到的湍流通量作为参考值, 使用随机森林和 XGBoost 对 MOST 得到的湍流通量进行了订正, 显著提高了特征温度和特征比湿的计算精度。为精简输入变量, 本文进一步对输入变量进行了前向、后向变量选择, 经变量选择后的最佳输入变量组合仍保持了较高的计算精度。此外, 将选择后的输入变量组合应用到另一个机器学习模型 (神经网络) 时, 其计算精度也在 MOST 结果的基础上取得了显著提高, 可见该变量选择方法可有效精简机器学习所需要的输入变量。

关键词 近地层 湍流通量 机器学习 Monin-Obukhov 相似理论 变量选择

On the Correction of Near-Surface Turbulent Fluxes over a Farmland Based on Machine Learning Methods

LIU Zhenrong¹, LI Yubin^{1*}, WANG Hong¹, LIU Xinye², Gao Zhiqiu¹

1. School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing 210044

2. Meteorological Bureau of Tonglu County, Hangzhou City, Zhejiang Province, Tonglu, 311500

Abstract: Turbulent fluxes in the surface layer are significant variables that characterize the interaction between the land and the atmosphere, which are typically calculated based on the Monin-Obukhov Similarity Theory (MOST) in gradient observations or atmospheric numerical models. To further enhance the calculational accuracy of turbulent fluxes in the surface layer, in this paper, the turbulent fluxes derived from the MOST are corrected by Random Forest and XGBoost, with fluxes obtained by the eddy covariance method as reference values, which significantly improves the calculational accuracy of the temperature scale and moisture scale. In order to reduce the number of input variables, this study further performs forward and backward variable selection on the input variables. The optimal combinations of input variables after selecting still maintain a high level of computational accuracy. Furthermore, when the

收稿日期: 2025-1-14, 网络预出版日期:

作者简介 刘贞榕, 女, 1999 年出生, 硕士研究生, 主要从事边界层气象研究. E-mail: 202212030011@nuist.edu.cn

通讯作者 李煜斌, E-mail: liyubin@nuist.edu.cn

资助项目 国家重点研发计划(2023YFC3706303)

Funded by The National Natural Science Foundation of China (No. 2023YFC3706303)

selected combinations of input variables are applied to the artificial neural network, their computational accuracy also show a significant improvement in comparison with the results of MOST. Therefore, the variable selection method of input variables is crucial for the machine learning method.

Keywords Surface layer, Turbulent fluxes, Machine learning, Monin-Obukhov Similarity Theory, Variable Selection.

文章编号: 2024077A

doi:10.3878/j.issn.1006-9895.2501.24077

1. 引言

近地层湍流动量、感热和潜热通量是描述地气系统交换的重要变量, 湍流通量的准确计算是地-气相互作用观测研究的基础, 也是大气数值模式获得合理天气预报和气候预测的基本前提 (Bourassa et al, 2013; Biermann et al, 2014; Helbig et al, 2021; Vilà et al, 2023; Jiang et al, 2024; 黄翊, 2017)。在湍流通量观测方法中, 涡动相关法因其直接测量湍流脉动而被认为是最可靠的方法 (Baldocchi, 2003; 严晓强, 2018; Liu, 2023), 当前世界知名通量观测网 (如 ChinaFlux, AmeriFlux 等) 皆将其用作于标准观测方法, 但其局限性在于仪器昂贵难以实现大面积覆盖。此外, 湍流通量还可以通过相对便宜的常规观测和莫宁奥布霍夫相似性理论 (MOST) 计算得到 (Dethloff et al, 2021; Gryanik, 2021; Schneider et al, 2022)。MOST 由 Monin 于 1954 年提出, 该方法在中性条件下的风廓线呈对数形式的基础之上构建了稳定度修正普适函数, 用于描述非中性条件下的通量-廓线关系, 即摩擦速度与风速廓线的关系、特征温度与温度廓线的关系以及特征比湿与湿度廓线的关系 (Monin, 1954)。稳定度修正普适函数最初以 BD 方程 (Businger, 1967; Dyer, 1970) 的形式为主, 之后学者们陆续提出了新的系数 (Paulson, 1970; Webb, 1970; Businger, 1971; Dyer 1974; Högström, 1988) 和新的形式 (Holtslag et al, 1988; Cheng and Brutsaert, 2005)。目前, 数值模式中广泛使用 MOST 方法计算湍流通量 (Sfyri, 2018; Finnigan, 2020)。然而 MOST 基于水平均匀的理想假设, 在非均一地形下的适用性存疑, 此外, 下垫面的非均一性常使得基于 MOST 计算的感热通量和潜热通量无法满足地表能量平衡方程 (刘鹏飞, 2010)。并且, 由于不符合 MOST 的均匀湍流假设, 稳定层结下基于 MOST 计算的摩擦速度与直接观测值之间也有较大差异 (Newman and Klein, 2014)。还有研究指出, MOST 未考虑非局地相互作用对湍流通量的影响 (Tong 和 Nguyen, 2015), 而源自近地层以外的涡流可能对近地层的热输送有重要贡献 (Li et al, 2018), 高层的动量同样对边界层动量有影响 (朱丽, 2020)。张蕴帅等人 (2021) 通过大涡模拟研究表明各种尺度的波对湍流通量具有累积贡献。已有研究表明, 辐射和土壤热通量信息的加入能够明显改善地表能量平衡情况 (刘鹏飞, 2010; 岳平, 2012)。

随着人工智能技术的发展, 机器学习方法开始在湍流通量的计算中得到应用。McCandless 等人 (2022) 基于直接观测数据构建了随机森林和神经网络机器学习模型, 用于计算湍流通量。相比于传统的 MOST 方法, 这些模型在感热和潜热通量的精度上取得了显著提升。Wulfmeyer 等人

(2022)研究了日间不稳定情况下的湍流通量,构建 XGBoost 和多层感知机模型。他们认为,通过增加湍流参数,可以改进湍流通量的计算精度。Yu 等人(2023)则使用随机森林模型计算不同稳定情况下的稳定度参数,建立非迭代参数化方案,显著降低了相对误差。

虽然机器学习在改进湍流通量计算精度方面表现出色,但不同机器学习模型的准确度特征尚不明确,并且机器学习模型对不同输入变量的敏感性也尚未明晰。由于气象条件、地表特性、地表加热及大气稳定性等影响,大气表层结构在时间和空间上呈显著的非线性变化,进而导致近地层湍流通量的非线性变化。简单的线性回归无法捕捉湍流通量与各种影响因子之间复杂的非线性关系,随机森林和 XGBoost 作为经典的机器学习集成模型,在数据科学中广泛使用并且在回归问题中表现出色,通过直接学习大量数据中的规律,其能够较好地捕捉和模拟输出目标与输入变量之间的非线性关系,在湍流通量的计算中呈现较好的效果(McCandless, 2022; Wulfmeyer, 2023)。因此,本文选取了随机森林(Random Forest: RF)和 XGBoost(XGB)这两种模型对机器学习湍流通量计算的准确度特征进行了分析,并基于 RF 和 XGB 模型对输入变量进行了前、后向变量选择,优化了机器学习模型的输入变量组合,实现了湍流通量计算所需输入变量的精简。所谓前向变量选择,即先把每个变量单独输入到机器学习模型中,取验证效果最优的变量作为第一个变量,再将剩下的每个变量都与第一个变量一起输入到模型里,取验证效果最优的第二个变量,以此类推,依次增加最优变量来达到重要性排序;而后向变量选择与之相反,通过依次剔除最不相关变量达到重要性排序。

本文主要针对摩擦速度($u_*(\text{m/s})$)、特征温度($T_*(\text{K})$)和特征比湿($q_*(\text{g/kg})$)进行研究,这三者分别是表征湍流动量通量、湍流感热通量和湍流潜热通量的特征量,表示了湍流垂直输送的强度。本文研究总共分为三部分,第一部分以涡动相关法计算的摩擦速度、特征温度和特征比湿为参考目标值,结合气象要素、辐射、土壤热通量等输入变量构建 RF 和 XGB 湍流通量计算模型,对 MOST 计算的摩擦速度、特征温度和特征比湿进行准确度订正。第二部分为基于机器学习模型的湍流通量输入变量敏感性研究,本研究针对第一部分的输入变量分别基于 RF 和 XGB 模型进行前、后向变量选择,从而对输入变量进行精简,获得优化的输入变量组合。第三部分为输入变量组合的适用性验证,将 RF 和 XGB 选取的变量组合用于第三个独立的机器学习模型——人工神经网络(Artificial Neural Network, NN),对第二部分优化后的输入变量组合进行验证。

2. 数据和方法

2.1 研究站点和数据

本研究的观测站点位于安徽省寿县站的国家气候中心观象台(32.42°N , 116.76°E),该站点海拔 27 米,下垫面为农田,由图 1(b)可见,为了方便农田管理,在农田之间存在着凸起的田埂,用于分界和蓄水,在观测塔的周边大约 10 至 40 米处设有田埂,90 米左右则靠近道路和房屋,但总体来看地势较为均匀。对于底层大气湍流的研究,下垫面的变化会起到决定性作用,定量地讨

论下垫面变化就显得尤其重要。本站点农田 11 月至 5 月下旬生长小麦，而从 6 月到 11 月则生长水稻。基于 2015 年至 2017 年的实地观测结果，小麦从播种至成熟期间，其最大生长高度约为 86 厘米；水稻在相同生长阶段最高可达到约 148 厘米。研究采用了该站点 2018 年 1 月 1 日至 2019 年 4 月 24 日的高频通量数据、气象数据、辐射数据和土壤数据。

高频通量数据是由涡动相关（EC）系统观测的高频三维风速、虚温、水汽密度，用于基于涡动相关法直接计算 30 分钟湍流通量。EC 系统包括一台三维超声风速仪（CSAT3A, Campbell Scientific Incorporation, USA）和一个 $\text{CO}_2/\text{H}_2\text{O}$ 开路红外气体分析仪（EC 150, Campbell Scientific Incorporation, USA），仪器架设高度为 2.5 米，采样频率为 10HZ。在涡动相关系统附近外接有一个温度传感器（109, Campbell Scientific Incorporation, USA）和地表气压传感器（PTB110, Vaisala, Inc, Helsinki, Finland），均使用 CR3000 数据采集器以 1HZ 的频率进行采样。此部分慢响应气温和气压用于结合高频水汽密度计算比湿。

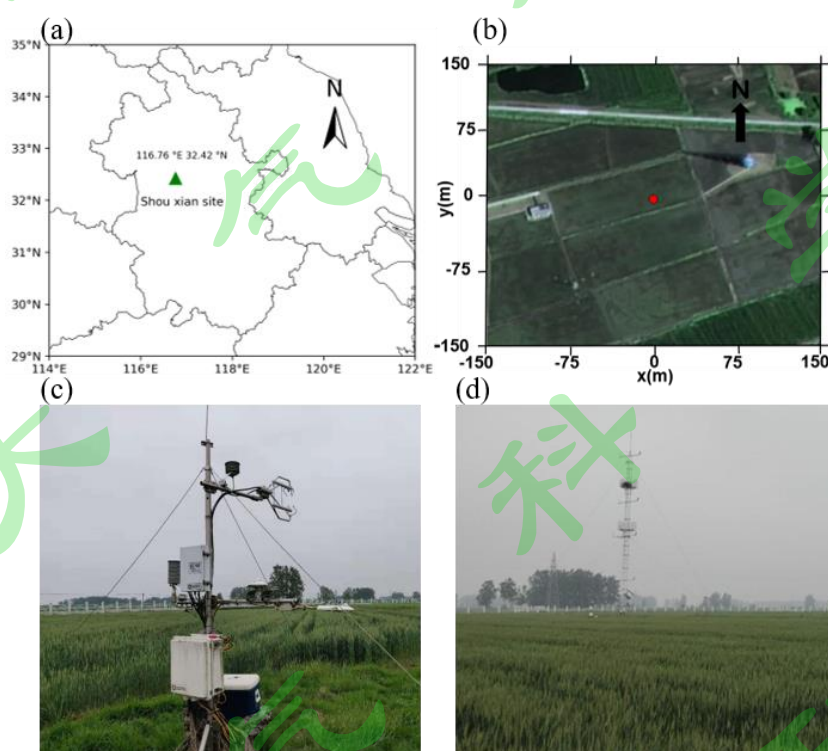


图 1 (a) 研究站点地理位置；(b) 站点卫星地图；(c) 通量塔；(d) 梯度塔

Figure 1 Geographic location of the research site (a); satellite map of the site (b); the flux tower (c) and the gradient tower (d)

气象数据包括风速、风向、温度、相对湿度、气压、降水等要素，梯度系统由温湿度传感器（HMP155A, Vaisala, Incorporation, Helsinki, Finland）、风速传感器（010C, Met One Incorporation., USA）、风向传感器（020C, Met One Inc., USA）和一座 20 米的梯度塔组成，20 米梯度塔位于 EC 通量塔以西 5 米处，详见图 1 (d)。在梯度塔 2 米、5 米、10 米、15 米、20 米处分别安装有温湿度传感器和风速传感器，而 10 米处安装有一个风向传感器。其中 5 米处的风速传感器损坏，在数据处理中将其排除。辐射数据来源于四分量净辐射计（CNR-4, Kipp & Zonen B.V., Delft, the Netherlands），该设备安装高度为 1.5 米。地表温度

由红外温度传感器（SI-111, Campbell Scientific Incorporation, USA）测量，同样安装在 1.5 米的高度。土壤数据包括土壤热通量和土壤温度，土壤热通量通过安装在土壤 5cm 深处的热通量盘（HFP01 heat flux plates, Hukseflux Thermal Sensors）进行观测，土壤温度由安装在土壤深 2cm 和 8cm 的土壤热电偶平均温度传感器（TCAV, Campbell Scientific Incorporation, USA）进行测量。以上各要素除了土壤热通量和土壤温度是使用 CR3000 数据采集器以 1HZ 的频率采集，其他要素以 0.1HZ 的频率采集，采集完后以 30 分钟平均值输出。降水量由翻斗式雨量计（52202, Young Traverse City MI, USA）测量，数据以 30 分钟累计量的形式记录。本文的观测数据可见于表 1，寿县站此次观测的全周期大气背景场见图 2，包括有六小时平均（a）温度，（b）相对湿度、（c）风速、（d）风向和（e）气压。

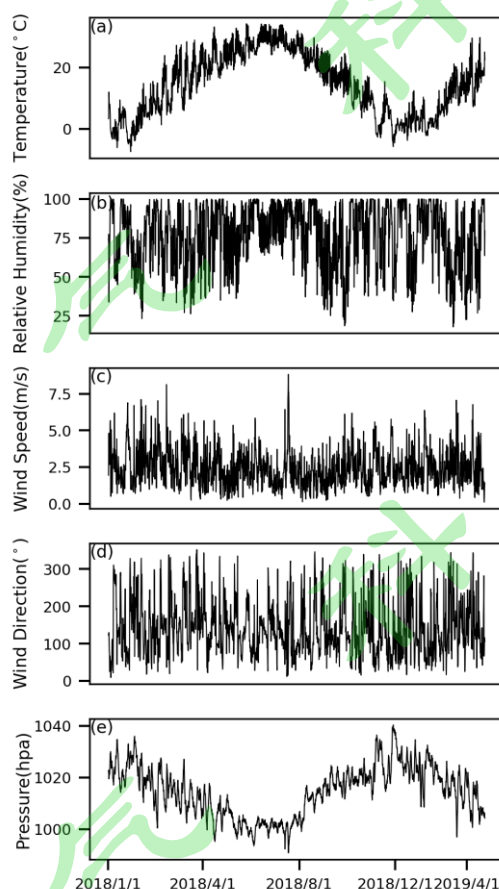


图 2 寿县站观测的全周期大气背景场，包括有六小时平均（a）温度，（b）相对湿度、（c）风速、（d）风向、和（e）气压

Figure 2: Atmospheric background field at Shouxian station during the observation period: 6-hour average (a) Temperature; (b) Relative humidity; (c) Wind speed; (d) Wind direction and (e) Air pressure

本研究对各项数据进行了数据预处理，包括去除野点、观测异常值（如水汽密度为负值时）并剔除降水前、后 1 小时的数据等（Liu, 2020; Li, 2023; Zhao, 2023），并且对湍流数据进行了坐标旋转和平稳性检验等质量控制（Foken, 2004; Li, 2023）。在未基于稳定度对数据分类的情况下，McCandless 等人(2022)使用机器学习方法计算了湍流通量，其结果与观测值之间的相关性较低（ R^2 小于 0.3）。因此，有必要按整体理查森数（Rib，定义见方程(2.1)，方程(2.1)中 z_1 是较低层， z_2 是较高层）将数据进行稳定度分类来提高机器学习的计算精度，其中， $Rib < 0$ 为不稳定情

况， $Rib > 0$ 为稳定情况（Launiainen, 1995; Li, 2010; Yu, 2023）。针对本次连续观测中湍流以外的慢响应观测原始数据进行了大于 4 倍标准差的异常值剔除，然后进行了半小时平均，并且慢响应观测的半小时平均数据分别与三种湍流量数据进行了时间上的匹配，不稳定和稳定情况下的摩擦速度及慢响应观测数据量分别为 13403 和 2351，特征温度及慢响应观测数据量分别为 11080 和 2047，特征比湿及慢响应观测数据量分别为 11151 和 1642。

$$Rib = \frac{2g[T(z_2) - T(z_1)](z_2 - z_1)}{[T(z_2) + T(z_1)][u(z_2) - u(z_1)]^2} \quad (2.1)$$

公式 2.1 中字母表示具体说明见表 2。

此外，本文还使用涡动相关法观测的摩擦速度和特征温度以及 2m 和 15m 的温度对 MOST 定义的稳定度参数 ζ （计算公式见方程 2.2）进行计算，将其与 Rib 进行对比统计，可见下图， Rib 和 ζ 划分稳定度和不稳定度时并不能完全一致，当 $Rib < 0$ 时， $\zeta > 0$ 的数据比例为 37%； $Rib > 0$ 时， $\zeta < 0$ 的数据比例为 7%，即不稳定时 Rib 与 ζ 同时小于 0 的情况占 63%，稳定时较为一致，同时大于 0 的情况占 93%。 Rib 与 ζ 符号不一致的原因包括：1）小扰动的影响（Mahrt, 2007; Durden, 2013）；2）两套观测系统的不一致产生的误差（Litt, 2015）。在涡动相关观测中， ζ 的符号由 T^* 所决定（见方程 2.2），因此当 ζ 和 Rib 的符号不一致时，为保证计算的合理性，本文将 T^* 调整为 0。

$$\zeta = \frac{z-d}{L} = \frac{2(z-d)\kappa g T^*}{[T(z_2) + T(z_1)]u_*^2} \quad (2.2)$$

公式(2.2)中字母表示具体说明见表 2。

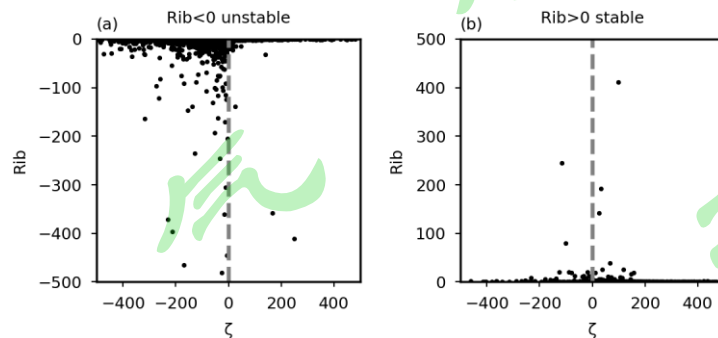


图 3（a）不稳定和（b）稳定情况下 Rib 与涡动观测计算的 ζ 对比

Figure 3 Comparison between Rib and vortex observation calculation of ζ under (a) unstable and (b) stable conditions

164 表 1 输入变量（观测频率小于等于 1HZ 的常规要素慢响应观测及基于 MOST 计算结果）与输出目标（涡动相关法

165 计算的湍流通量）说明

166 Table1 The description of input variables (slow response observations of conventional elements with frequencies

167 less than or equal to 1 Hz and MOST calculation results) and output targets (turbulent fluxes calculated by the eddy

168 covariance method)

变量说明	变量	观测高度 (m)	简写名称	均值	测量精度
输入变量	温度(°C) Temperature	2,5,10,15,20	$T_2, T_5, T_{10},$ T_{15}, T_{20}	14.40,14.80,14.60,14.90,14.97	$\pm (0.226-0.0028*T) ^\circ\text{C}$ in -80 ~ +20°C $\pm (0.055+0.0057*T) ^\circ\text{C}$ in +20 ~ +60°C
	相对湿度(%) Relative Humidity	2,5,10,15,20	$RH_2, RH_5,$ $RH_{10},$ RH_{15}, RH_{20}	81.46,78.46,76.69,71.82, 73.76	$\pm (1.4+0.032*RH) \%$ in -60 ~ -40% $\pm (1.2+0.012*RH) \%$ in -40 ~ -20% $\pm (1+0.008*RH) \%$ in -20 ~ +40% $\pm (1.2+0.012*RH) \%$ in +40 ~ 60%
	风速(m/s) Wind Speed	2,10,15,20	$W_{S2}, W_{S10},$ W_{S15}, W_{S20}	1.67,2.40,2.63,2.82	$\pm 1\%$
	风向(Deg) Wind Direction	10	W_{D10}	141.58	$\pm 3\%$
	地表温度(°C) Surface Temperature	0	T_{sf}	14.42	$\pm 0.2^\circ\text{C}$ in -10°C ~ 65°C $\pm 0.5^\circ\text{C}$ in -40°C~70°C
	土壤温度(°C) Soil Temperature	2-8(cm)	T_{soil}	14.38	$\pm 0.3^\circ\text{C}$
	土壤热通量(W/m²) Soil Heat Flux	5(cm)	H_{soil}	-0.54	$\pm 3\%$
	向上长波辐射(W/m²) Outgoing Longwave Radiation	1.5	R_{LWout}	875.91	$5 \sim 15\mu\text{V W}^{-1} \text{ m}^{-2}$
	向下长波辐射(W/m²) Ingoing Longwave Radiation	1.5	R_{LWin}	841.56	$5 \sim 15\mu\text{V W}^{-1} \text{ m}^{-2}$
	向上短波辐射(W/m²) Outgoing Shortwave Radiation	1.5	R_{SWout}	31.84	$5 \sim 20\mu\text{V W}^{-1} \text{ m}^{-2}$
	向下短波辐射(W/m²) Ingoing Shortwave Radiation	1.5	R_{SWin}	153.45	$5 \sim 20\mu\text{V W}^{-1} \text{ m}^{-2}$
	MOST 摩擦速度(m/s)	2-15	u^*Mo	0.15	-
	MOST 特征温度(°C)	2-15	T^*MO	-0.08	-
	MOST 特征比湿(g/kg)	2-15	q^*MO	-0.28	-
	MOST 动力和热力积分 稳定度修正函数	15	$\psi/m, \psi/h$	1.91,1.85	-
	MOST 稳定度参数	15	ζ	-0.005	-
输出目标	涡动摩擦速度(m/s)	2.5	u^*EC	0.21	u 和 v : $\pm 0.04 \text{ m s}^{-1}$
	涡动特征温度(°C)	2.5	T^*EC	-0.08	w : $\pm 0.02 \text{ mm s}^{-1}$
	涡动特征比湿(g/kg)	2.5	q^*EC	-0.14	T_s : $\pm 0.025^\circ\text{C}$ H_2O : <2%

169

170

2.2 MOST 方法

本文使用两层（2m 和 15m）风、温、湿半小时观测数据与不稳定情况下 Paulson 于 1970 年提出的稳定度修正函数（Paulson,1970, 本文称为 Pa70），以及稳定情况下 Cheng 和 Brutsaert 于 2005 提出的稳定度修正函数（Cheng and Brutsaert, 2005, 本文称为 Chen05），在近地面常通量假设的基础上，风、温、湿廓线方程（2.3）-（2.5）沿着有效观测高度积分即可得到摩擦速度、特征温度和特征比湿的方程。Pa70 和 Chen05 方案在数值模式中广泛使用，Pa70 与 BD 方程的形式一致，但系数不同，其在不稳定情况的准确度较高（Chen, 2023），Chen05 是针对稳定情况下改进后的方案，其在稳定情况下较为准确（Guo and Zhang, 2007; Jiménez, 2012）。由于潜热通量过程复杂且湿度梯度难以测量，目前没有潜热通量的稳定度修正普适函数，通常取 $\phi_q = \phi_h$ ， $\psi_q = \psi_h$ 。以下公式中各字母表示具体说明可见表 2。

$$\frac{\partial u}{\partial z} = \frac{u_*}{kz} \phi_m \left(\frac{z-d}{L} \right) \quad (2.3)$$

$$\frac{\partial T}{\partial z} = \frac{T_*}{kz} \phi_h \left(\frac{z-d}{L} \right) \quad (2.4)$$

$$\frac{\partial q}{\partial z} = \frac{q_*}{kz} \phi_q \left(\frac{z-d}{L} \right) \quad (2.5)$$

$$u_* = \frac{\kappa [u(z_2) - u(z_1)]}{\ln \left(\frac{z_2-d}{z_1-d} \right) - \psi_m \left(\frac{z_2-d}{L} \right) + \psi_m \left(\frac{z_1-d}{L} \right)} \quad (2.6)$$

$$T_* = \frac{\kappa [T(z_2) - T(z_1)]}{\ln \left(\frac{z_2-d}{z_1-d} \right) - \psi_h \left(\frac{z_2-d}{L} \right) + \psi_h \left(\frac{z_1-d}{L} \right)} \quad (2.7)$$

$$q_* = \frac{\kappa [q(z_2) - q(z_1)]}{\ln \left(\frac{z_2-d}{z_1-d} \right) - \psi_q \left(\frac{z_2-d}{L} \right) + \psi_q \left(\frac{z_1-d}{L} \right)} \quad (2.8)$$

$$L = \frac{[T(z_1) + T(z_2)] u_*^2}{2\kappa g T_*} \quad (2.9)$$

不稳定情况下使用的 Pa70 方案为：

$$\phi_m = (1 - A_m \zeta)^{\frac{1}{4}} \quad (2.10)$$

$$\phi_h = R(1 - A_h \zeta)^{\frac{1}{2}} \quad (2.11)$$

$$\psi_m(\zeta) = \ln \left[\left(\frac{1+x}{2} \right)^2 \left(\frac{1+x^2}{2} \right) \right] - 2 \arctan(x) + \frac{\pi}{2} \quad (2.12)$$

$$\psi_h(\zeta) = 2 \ln \left(\frac{1+y}{2} \right) \quad (2.13)$$

其中 $x = (1 - A_m \zeta)^{\frac{1}{4}}$, $y = R(1 - A_h \zeta)^{\frac{1}{2}}$, 各字母表示具体说明可见表 2。不稳定情况下即 Pa70 方案标定参数的观测数据来自澳大利亚平原地区 Kerang 的风速、温度、热通量和剪切应力的观测结果, 测量高度为 16 米 (Swinbank, 1964; Paulson, 1970)。Pa70 方案在数值模式中被广泛使用 (Chen, 2023)。

表 2 计算公式中各字母表示说明

Table2 Explanation of each letter in the calculation formulas

字母表示	含义	字母表示	含义	字母表示	含义
u_*	摩擦速度	A_m	16 (Pa70 方案动量稳定度修正函数参数)	z	观测高度 ($z_1=2m$, $z_2=15m$)
T_*	特征温度	A_h	16 (Pa70 方案热量稳定度修正函数参数)	L	奥布霍夫长度
q_*	特征比湿	R	1 (Pa70 方案热量稳定度修正函数参数)	κ	卡尔曼常数, 取 0.4
u	平均风速	a_0	6.1 (Chen05 方案动量稳定度修正函数参数)	ζ	稳定度参数
T	平均温度	b_0	2.5 (Chen05 方案动量稳定度修正函数参数)	q'	比湿脉动值
q	平均比湿	c_0	5.3 (Chen05 方案热量稳定度修正函数参数)	T'	温度脉动值
g	重力加速度, 9.8m/s	d_0	1.1 (Chen05 方案热量稳定度修正函数参数)	w'	笛卡尔坐标系 z 轴方向 风速脉动值
d	零平面位移	$\phi_{m,h,q}$	稳定度修正函数 (m, h, q 分别指动量通量、感热通量和潜热通量)	u'	笛卡尔坐标系 x 轴方向 风速脉动值
z_0	地表粗糙度	$\psi_{m,h,q}$	稳定度修正函数的 积分形式	v'	笛卡尔坐标系 y 轴方向 风速脉动值

稳定情况下使用的 Chen05 方案为:

$$\phi_m = 1 + a_0 \frac{\zeta + \zeta^{b_0} \left(1 + \zeta^{b_0} \right)^{\frac{1-b_0}{b_0}}}{\zeta + \left(1 + \zeta^{b_0} \right)^{\frac{1}{b_0}}} \quad (2.14)$$

$$\phi_h = 1 + c_0 \frac{\zeta + \zeta^{d_0} \left(1 + \zeta^{d_0} \right)^{\frac{1-d_0}{d_0}}}{\zeta + \left(1 + \zeta^{d_0} \right)^{\frac{1}{d_0}}} \quad (2.15)$$

$$\psi_m = -a_0 \ln \left(\zeta + \left(1 + \zeta^{b_0} \right)^{\frac{1}{b_0}} \right) \quad (2.16)$$

$$\psi_h = -c_0 \ln \left(\zeta + \left(1 + \zeta^{b_0} \right)^{\frac{1}{b_0}} \right) \quad (2.17)$$

以上各字母表示具体说明可见表 2。稳定情况下即 Chen05 方案标定参数的观测数据来自美国 Lenon, Kansas 附近一座 60 米塔上的观测，该地障碍物较少，地势较为平坦（Cheng and Brutsaert, 2005）。相比于 BD 方程，Chen05 方案提出了新形式和参数，经验证其在稳定情况下的计算结果更为准确（Guo and Zhang, 2007; Jiménez et al, 2012）。

有效高度由观测高度减去零平面位移得到，本文选取每个月前五天近中性条件下（ $-0.5 < Rib < 0.5$ ）的涡动摩擦速度，结合 2m 和 15m 的风速，使用风廓线（方程 2.18）列的二元一次方程来计算的零平面位移和地表粗糙度，由于零平面位移用于 MOST 计算湍流通量，因此，此部分数据仅用于计算零平面位移和地表粗糙度，不参与机器学习训练。由于下垫面植被高度会随季节生长变化，导致零平面位移高度也随时间变化，本文取每两个月的中位数作为该时段的零平面位移高度 d，这两个变量随时间变化的结果如表 3 所示，d 的变化范围在（0.24-1.44）， z_0 变化范围在（0.5-0.65）。

$$\frac{u}{u_*} = \frac{1}{k} \ln \left(\frac{z-d}{z_0} \right) \quad (2.18)$$

以上公式中字母表示可见表 2

表 3 零平面位移与地表粗糙度随时间变化
Table 3 Zero Plane Displacement and Surface Roughness over Time

	1-2 月	3-4 月	5-6 月	7-8 月	9-10 月	11-12 月
d(m)	0.6	0.24	1.25	1.44	1.35	0.52
zo(m)	0.65	0.58	0.57	0.5	0.6	0.54

本文通过对方程（2.6）-（2.8）进行迭代计算求得摩擦速度、特征温度和特征比湿（McCandless, 2022; 刘鹏飞, 2010），具体过程如下：

1) 首先将动量通量、感热通量和潜热通量的中性积分稳定度修正函数（ $\psi_m = \psi_h = \psi_q = 0$ ）分别代入方程（2.6）-（2.8）中，求得初始假设为中性情况下的摩擦速度、特征温度、特征比湿：

$$u_{*0} = \frac{\kappa \left[u(z_2 - d) - u(z_1 - d) \right]}{\ln \left(\frac{z_2 - d}{z_1 - d} \right)} \quad (2.19)$$

$$T_{*0} = \frac{\kappa \left[T(z_2 - d) - T(z_1 - d) \right]}{\ln \left(\frac{z_2 - d}{z_1 - d} \right)} \quad (2.20)$$

$$q_{*0} = \frac{\kappa [q(z_2 - d) - q(z_1 - d)]}{\ln \left(\frac{z_2 - d}{z_1 - d} \right)} \quad (2.21)$$

2) 将 u_{*0} 和 T_{*0} 用于计算 $\overline{w'T'} = -u_* T_*$ ，若 $\overline{w'T'} = 0$ ，则保留 1) 中的计算结果，否则再进行下一步，计算 z_1/L 和 z_2/L ，代入积分稳定度修正函数求得新的摩擦速度 u_* 、特征温度 T_* 、特征比湿 q_* 。当 $-2 \leq z_2/L < -10^{-6}$ 时代入 Pa70 方案的积分稳定度修正函数（方程 2.12 和 2.13），当 $10^{-6} < z_2/L \leq 1$ 时代入 Chen05 方案的积分稳定度修正函数（方程 2.16 和 2.17）。

3) 计算各通量总差值 $error = |u_* - u_{*0}| + |T_* - T_{*0}| + |q_* - q_{*0}|$ ，若 $error \geq 10^{-3}$ ，则 $u_{*0} = u_*, T_{*0} = T_*, q_{*0} = q_*$ 再回到 2) 重新计算摩擦速度 u_* 、特征温度 T_* 、特征比湿 q_* ，如此循环迭代直至 $error < 10^{-3}$ 时停止迭代。

2.3 涡动相关法

涡动相关法基于雷诺分解和湍流冻结假设（Foken et al., 2012; Stull, 1988），将变量分解为时间平均值和脉动值，实际观测中进行坐标旋转、平稳性选择（Foken, 2004）后通过计算两个变量的协方差（如 $\overline{w'u'}$ ）并进行必要的订正，即频率损失订正（Moore, 1986）、虚温订正（van Dijk et al., 2004）和密度订正（Webb et al., 1980）后得到各湍流通量：

$$u_* = \sqrt[4]{(\overline{w'u'})^2 + (\overline{w'v'})^2} \quad (2.22)$$

$$T_* = -\frac{\overline{w'T'}}{u_*} \quad (2.23)$$

$$q_* = -\frac{\overline{w'q'}}{u_*} \quad (2.24)$$

以上公式具体字母表示可见表 2。

2.4 机器学习回归分析方法

机器学习回归分析方法是一种基于经验的方法，通过构建输入变量和输出目标之间的关系以通过已知量估算未知量。机器学习模型可以捕捉输入变量与输出目标之间的非线性关系，从而提升回归分析的能力。本文使用经典的随机森林、XGBoost 和人工神经网络模型对涡动相关系统观测的摩擦速度、特征温度和特征比湿分别进行回归分析，候选输入变量包括慢响应观测变量（即不包括超声风温仪的观测部分，观测频率小于等于 1HZ）和基于 MOST 的计算量，具体见表 1。为了减轻模型的过拟合，本文使用贝叶斯优化方法（Frazier, 2018）在训练集上进行 5 折交叉验证（Cai, 2020）调整超参数。

2.4.1 训练和测试集构建

本文机器学习方法在训练集上构建模型，独立的测试集上验证模型效果。对于训练集和测试集的划分，将每个月划分为三部分，每部分取前八天数据作为训练集，剩下的作为测试集。训练集和测试集的所占比例分别约为 78%和 22%。

2.4.2 随机森林和 XGBoost 模型

随机森林和 XGBoost 是以决策树 (McGovern et al., 2017) 为基本学习单位的集成模型。决策树是一种二叉树结构的模型，包括一个根节点、多个内部节点和叶节点。构建决策树步骤 (任萍, 2020) 如下：

- 1) 根节点通过遍历所有变量，将所有可能切分的点切分为两部分，最小化估算与真实值之间平方误差找到最佳切分点 (某一个变量)，将输入空间 (训练集的输入变量，如风速、温度、辐射通量、土壤热通量等；输出目标如摩擦速度) 划分为两部分数据集。

- 2) 将 1) 中两部分数据集遍历各自变量，最小化估算与真实值之间的平方误差，找到最佳变量进行分枝，生成两个内部的子节点。

- 3) 对上述两个步骤进行递归，直至满足停止条件，生成叶节点即决策结果。本文的停止条件包括：树的深度，节点样本数量，由具体样本通过贝叶斯参数调整得到。

随机森林是通过 Bootstrap (即有放回的随机抽样：样本随机和变量随机) 获取变量信息，同时生长多棵决策树，平均多棵决策树结果的 Bagging 集成模型 (Breiman, 2001)。Bagging 算法原理如图 4 (a)，通过设置随机样本降低过拟合，各决策树之间互不影响分别得到决策结果。该模型在处理多维和多重共线性数据优势明显，具有良好的泛化能力，应用广泛 (Belgiu, 2016; 刘娣, 2024)。随机森林调整的超参数包括森林中树的数量 (`n_estimators`)、树的最大深度 (`max_depth`)、特征数量 (`max_features`)、节点分枝所需最小样本数量 (`min_samples_split`) 和分枝后节点最小样本数量 (`min_samples_leaf`)。

XGBoost (Friedman, 2001; Chen, 2016) 是通过逐一生长决策树，经过多次迭代后多棵决策树结果求和的 Boosting 集成模型。Boosting 算法原理如图 4 (b)，即通过将后一棵树的目标值设置为前一棵树的误差 (即目标值-估算值)，从而不断降低误差提升精度。XGBoost 因其优良效果和高效速度应用广泛 (韩念霏, 2022; 徐景峰, 2023)。XGBoost 调整的超参数包括森林中树的数量 (`n_estimators`)，树的最大深度 (`max_depth`)，学习率 (`learning_rate`)，构造每棵树的样本比例 (`subsample`)，L1 正则惩罚系数 (`reg_alpha`)，L2 正则惩罚系数 (`reg_lambda`)，节点分裂所需最小损失 (`gamma`)，每层分裂所用特征比例 (`colsample_bylevel`) 和构造每棵树的特征比例 (`colsample_bytree`)。

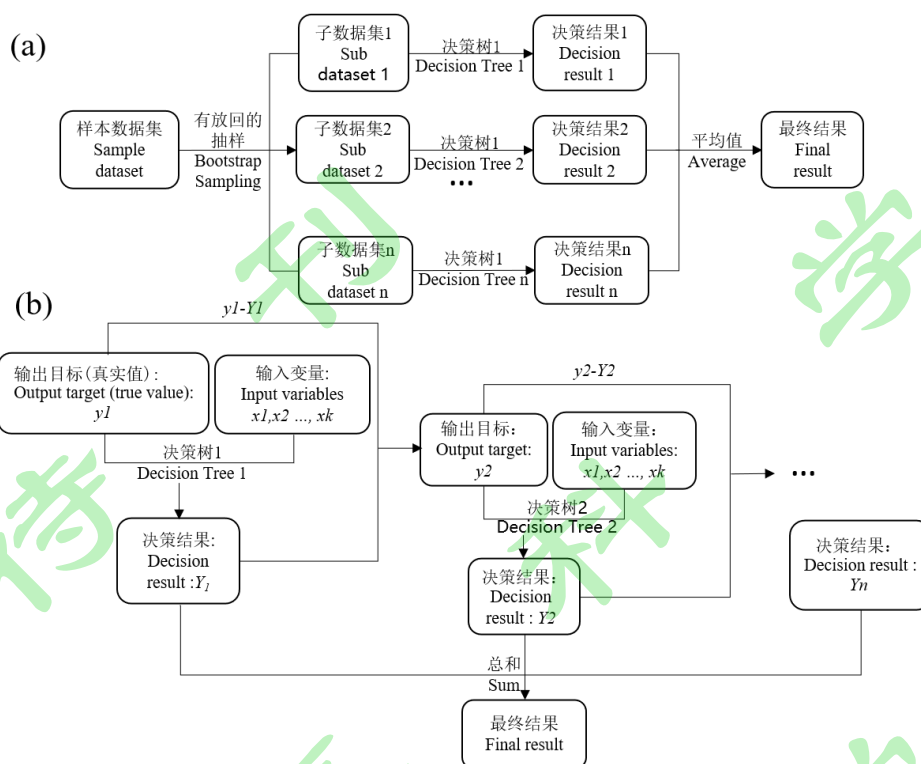


图 4 (a) Bagging 算法原理; (b) Boosting 算法原理示意图
Figure 4 (a) Schematic diagram of the Bagging algorithm principle; (b) Schematic diagram of the Boosting algorithm principle.

2.4.3 人工神经网络

人工神经网络模仿大脑神经系统处理信息的模式发展而来，由输入层、隐藏层和输出层多层神经元组成，神经元负责输入与输出（Uhlig, 1995）。本文使用的是全连接神经网络，输入层输入变量以后，每个输入值都与一个随机初始化的权重相乘后求和，再通过常用的非线性激活函数 ReLU 进行转换，引入非线性特性后再进入下一层神经元中。神经网络使用随机的小批量数据通过反向传播的随机梯度下降进行训练模型（即调整权重），不稳定和稳定情况批量大小分别为 1024 和 32，优化算法选择 Adam（Kingma and Ba, 2014），该方法易于实现，计算效率高，内存需求小，能够实现快速且稳定的收敛（McCandless, 2022）。在模型训练过程中使用了 Dropout（Srivastava, 2014）和权重衰减（Krogh and Hertz; 1991）的正则化方法以减轻过拟合：Dropout 通过随机失活神经元降低神经元之间的依赖，权重衰减通过在损失函数中增加惩罚项以限制权重大小。神经网络的超参数（失活率 dropout、权重衰减率 weight_decay、学习率 learning_rate、神经元的数目 num_layers 和隐藏层的数目 hidden_sizes）通过使用 Optuna 模型参数优化工具（Akiba, 2019）进行贝叶斯优化得到，训练过程中使用 5 折交叉验证。

本文对随机森林、XGBoost 和神经网络使用贝叶斯优化调整超参数，所使用的超参数空间设置情况见表 4 所示。

表 4 随机森林、XGBoost 和神经网络超参数空间设置

Table 4 Hyperparameter space settings for Random Forest, XGBoost, and Neural Networks

RF 超参数	空间设置	XGB 超参数	空间设置	NN 超参数	空间设置
n_estimators	(10,800)	n_estimators	(10,800)	dropout	(0,0.5)
max_depth	(1,100)	max_depth	(1,100)	weight_Decay	(e^{-5} , e^{-1})
max_features	(0.1,0.999)	learning_rate	(0.001,0.3)	learning_rate	(e^{-5} , e^{-1})
min_samples_split	(2,20)	subsample	(0.5,1)	num_layers	(2,4)
min_samples_leaf	(2,20)	reg_alpha	(0,10)	hidden_sizes	(32,64,128,256,512)
		reg_lambda	(0,30)		
		gamma	(0,10)		
		colsample_bylevel	(0.1,1)		
		colsample_bytree	(0.1,1)		

2.5 评估方法

模型调整超参数后，在最初划分出来的测试集上进行测试，本文使用均方根误差（RMSE）、相关性系数（R）和一致性指数（IA）（Willmott, 1981; Deo et al, 2015）对模型效果进行评估。均方根误差和相关系数广泛用于评估模型效果，均方根误差说明模型产生误差的实际大小，相关系数描述观测值和估算值之间的共线性。一致性指数是用于衡量模型误差程度的无量纲指标，对观测和估测之间共同变化幅度进行描述，对相关系数和均方根误差进行了补充（Willmott, 1981）。他们的计算方程如下：

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}} \quad (2.25)$$

$$R = \frac{\sum_{i=1}^N (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (P_i - \bar{P})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}} \quad (2.26)$$

$$IA = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N [|P_i'| - |O_i'|]^2} \quad (2.27)$$

其中 P_i 为估计值， O_i 为参考值，N 指总数据量。“ $\bar{P}$ ”指平均值，“ P_i' ”指减去平均值后的估计值，“ O_i' ”指减去平均值后的参考值。

3 结果与讨论

3.1 机器学习方法与 MOST 方法计算的湍流通量对比

本小节首先基于 MOST 方法对风、温、湿廓线方程积分迭代计算得到摩擦速度、特征温度和特征比湿。接着使用 RF 和 XGB 机器学习模型分别对涡动相关观测的摩擦速度、特征温度和特征比湿进行回归分析，机器学习模型的输入变量包括了慢响应观测部分（即不包括超声风温仪观测部分，观测频率小于等于 1HZ）和 MOST 的计算结果，具体见表 1。

图 5 为 MOST 方法计算的湍流通量与经机器学习方法订正后的湍流通量对比。由图 5 可知，与涡动相关法直接计算相比，基于 MOST 计算的摩擦速度和特征比湿呈明显的低估，摩擦速度相对较准确。MOST 计算的各湍流通量在稳定情况下比不稳定情况下更加分散，明显偏离 1:1 虚线。经机器学习方法订正后，不稳定和稳定情况下的摩擦速度、特征温度和特征比湿均明显向 1:1 虚线靠拢（见图 5）。

表 5 为机器学习订正前后评估指标对比，由表 5 可知，不稳定情况下，RF 和 XGB 明显提升 MOST 计算的湍流通量精度，摩擦速度、特征温度和特征比湿均方根误差分别由 0.066 降低至约 0.042、0.152 降低至约 0.105 以及 0.405 降低至约 0.12。RF 和 XGB 订正后的各湍流通量相关系数均超过了约 0.85，其中摩擦速度超过 0.9。订正的各湍流通量的一致性指数均超过 0.9，其中摩擦速度的一致性指数超过约 0.95。稳定情况下，RF 和 XGB 同样明显提升了 MOST 计算的湍流通量精度，摩擦速度、特征温度和特征比湿均方根误差分别由 0.061 降低至约 0.043、0.104 降低到约 0.036 以及 0.074 降低至约 0.02；摩擦速度、特征温度和特征比湿相关系数分别由 0.682 提升至超过 0.8，0.117 提升至超过 0.78 以及 0.095 提升至超过 0.63；一致性指数分别由 0.755 提升至超过 0.87、0.387 提升至超过 0.86 以及 0.294 提升至超过 0.76。不稳定情况下，RF 和 XGB 模型对各湍流通量的订正效果相近，稳定情况下 RF 模型计算精度相比 XGB 略高。机器学习模型对不稳定情况下的湍流通量计算精度高于稳定情况，这与 McCandless et al. (2022) 的研究结论相符。

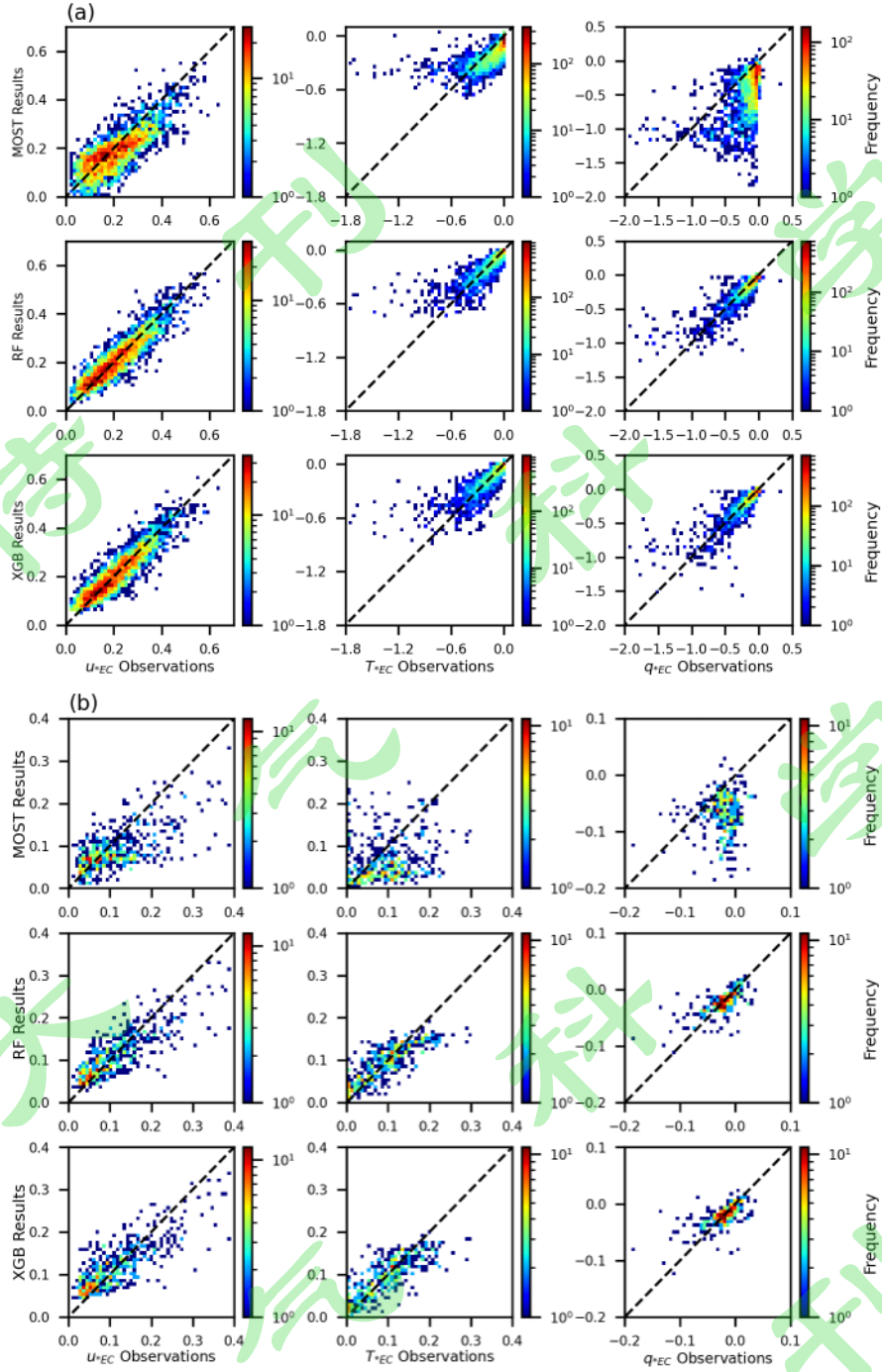


图 5 基于机器学习方法订正 MOST 方法计算 (a) 不稳定情况; (b) 稳定情况的湍流通量与 MOST 方法计算的湍流通量对比: x 轴为涡动相关法观测的湍流通量, y 轴为 MO 计算的湍流通量 (左)、RF 订正的湍流通量 (中) 和 XGB 订正的湍流通量 (右)。图中虚线为 1:1 线, 散点颜色代表出现频率。

Figure 5 Comparison of turbulent fluxes by the correction of machine learning methods on the MOST method and by the MOST method on (a) unstable conditions and (b) stable conditions: the x-axis label represents the turbulent flux observed by the eddy covariance method, and the y-axis labels represent the turbulent fluxes calculated by MO (left), corrected by RF (middle), and corrected by XGB (right). The dashed line in the figure represents the 1:1 line, and the colors of the scatter dots indicate the frequency of occurrence.

表 5 机器学习订正 MOST 前后在测试集上计算的评估结果对比，All 指表 1 中的全部变量组合，下划线部分为最佳方案

Table 5 Evaluation results comparison of calculations on the test set for MOST computational results and after machine learning correction, where 'All' refers to the complete variable combination in Table 1, and the underlined part indicates the best solution

		不稳定			稳定		
		MOST	RF 模型 +All	XGB 模 型+All	MOST	RF 模型 +All	XGB 模 型+ All
u^*	RMSE	0.066	<u>0.042</u>	<u>0.042</u>	0.061	<u>0.043</u>	0.045
	R	0.788	<u>0.91</u>	<u>0.91</u>	0.682	<u>0.833</u>	0.809
	IA	0.864	<u>0.951</u>	<u>0.951</u>	0.755	<u>0.886</u>	0.874
T^*	RMSE	0.152	<u>0.105</u>	0.107	0.104	<u>0.036</u>	0.038
	R	0.687	<u>0.855</u>	0.847	0.117	<u>0.811</u>	0.789
	IA	0.782	<u>0.908</u>	0.903	0.387	<u>0.867</u>	0.867
q^*	RMSE	0.405	<u>0.12</u>	0.123	0.074	<u>0.019</u>	0.02
	R	0.51	<u>0.86</u>	0.85	0.095	<u>0.676</u>	0.638
	IA	0.552	<u>0.914</u>	0.913	0.294	<u>0.793</u>	0.761

3.2 基于机器学习模型的输入变量敏感性研究

本文对机器学习所有输入变量进行了基于 RF 和 XGB 模型的前、后向变量选择研究各输入变量的敏感性。

前向变量选择的方法如下：使用同一组超参数的机器学习模型，先将变量单独输入模型，在训练集上进行 5 折交叉验证，使用相关系数 R 的平方评估，取所有变量中 R^2 最高的变量作为第一个变量，之后将剩下的每个变量与第一个变量进行组合，再次交叉验证，取所有组合中 R^2 最高的变量作为第二个变量。以此类推，直至所有变量都加入完毕。通过这种方法，可以将输入变量按重要性依次逐个提取得到，越早加入的输入变量相关系数越大，重要性越高。

后向变量选择的方法如下：使用同一组超参数的机器学习模型，将所有变量都输入模型后在训练集上进行 5 折交叉验证，同样使用 R^2 进行评估，每个变量都删除一次并进行交叉验证，删除后的 R^2 值最高的那个变量影响最小，将其剔除。将剩下的变量重复删除和交叉验证过程，得到第二个剔除的变量。以此类推，直至最后一个变量单独进行交叉验证。通过这种方法，可以将输入变量按次要性依次逐个剔除，最先剔除的输入变量相关系数最小，重要性越低。需要注意的是，由于输入变量的组合对机器学习方法具有一定影响，所以通过前向变量选择方法和后向变量选择方法得到的输入变量重要性排序并不会完全一致。

图 6 和图 7 分别展示了在前向和后向选择中，RF 和 XGB 模型随着输入变量的增加和减少从而产生的效果变化。由图 4-5 可知，在前向变量选择中，随着输入变量的增加， R^2 先快速上升，再缓慢上升，后几乎保持不变。在后向变量选择中，随着输入变量的减少， R^2 先近似保持不变，再缓慢上升，后几乎保持不变。

慢下降后快速下降。在这些输入变量中，使 R^2 快速上升或是下降的变量意味着它们对模型有较大影响，使 R^2 缓慢上升或下降的变量影响则较小，使 R^2 近似不变的变量影响最小。由图 4 和图 5 可以看出，在前向（后向）选择中不稳定情况下 RF 和 XGB 前（后）三个输入变量可使得摩擦速度、特征温度和特征比湿的 R^2 分别升高到约 0.83、0.65 和 0.68；稳定情况下分别升高到约 0.6、0.45 和 0.39。不稳定情况下 RF 前六个（后七个）输入变量和 XGB 前六个（后六个）输入变量可使得摩擦速度 R^2 升高到约 0.88，RF 前九个（后八个）输入变量和 XGB 前八个（后十一个）变量可使得特征温度 R^2 升高到约 0.75，RF 前六个（后七个）输入变量和 XGB 前七个（后九个）输入变量可使得特征比湿 R^2 升高到约 0.75。稳定情况下 RF 前十二个（后八个）输入变量和 XGB 前九个（后十二个）输入变量可使得摩擦速度 R^2 升高到约 0.68，RF 前八个（后十三个）输入变量和 XGB 前十个（后十个）变量可使特征温度 R^2 升高到约 0.55，RF 前九个（后九个）输入变量和 XGB 前十四个（后十四个）输入变量可使得特征比湿 R^2 升高到约 0.53。之后输入变量的增加或减少对 R^2 的影响极小（ R^2 变化小于 0.005）（见图 6 黑色虚线以右和图 7 黑色虚线以左的阴影部分）。

可知在各模型中变量信息有其上限，基于模型对变量进行选择，可以去除冗杂信息，避免过多的信息干扰模型计算。图 6 黑色虚线以右和图 7 黑色虚线以左的各输入变量对 R^2 影响小于 0.005（增幅或是降幅），影响极小，因此采用图 6 黑色虚线以左阴影部分和图 7 黑色虚线以右阴影部分的变量（即对模型影响较大的变量）来计算湍流通量，共四种变量组合：RF 前、后向变量组合（RF Forward; RF Backward）和 XGB 前、后向变量组合（XGB Forward; XGB Backward）。

3.2.2 机器学习前、后向特征选择的估算结果分析

将 RF 和 XGB 选择的变量组合分别输入到各自的模型中进行训练和测试。由图 5 和图 8 可知，RF 和 XGB 使用各变量组合在不稳定和稳定情况下都能较好地计算各湍流通量，其与使用全部变量的计算效果相近，相比基于 MOST 计算时效果显著提升，且使用变量组合计算甚至有优于使用全部变量计算的情况。由表 6 可知，计算特征温度时，RF 和 XGB 使用前、后向变量组合计算的一致性指数均高于使用全部变量时的一致性指数。稳定情况下，计算特征比湿时，使用前、后向变量组合 RF 和 XGB 的相关性系数（RF 模型+RF Forward: 0.757, XGB 模型+XGB Forward: 0.722, RF 模型+RF Backward: 0.77, XGB 模型+XGB Backward: 0.764）和一致性指数（RF 模型+RF Forward: 0.84, XGB 模型+XGB Forward: 0.831, RF 模型+RF Backward: 0.850, XGB 模型+XGB Backward: 0.857）均明显优于使用全部变量时的相关性系数（RF 模型+All: 0.676, XGB 模型+All: 0.638）和一致性指数（RF 模型+All: 0.793, XGB 模型+All: 0.761）。这主要是因为高维的数据中包括有较多不相关、冗余的变量，影响机器学习回归模型拟合，删除不相关变量可以提高准确性（Reunanen, 2004; Li, 2017; Venkatesh, 2019）。

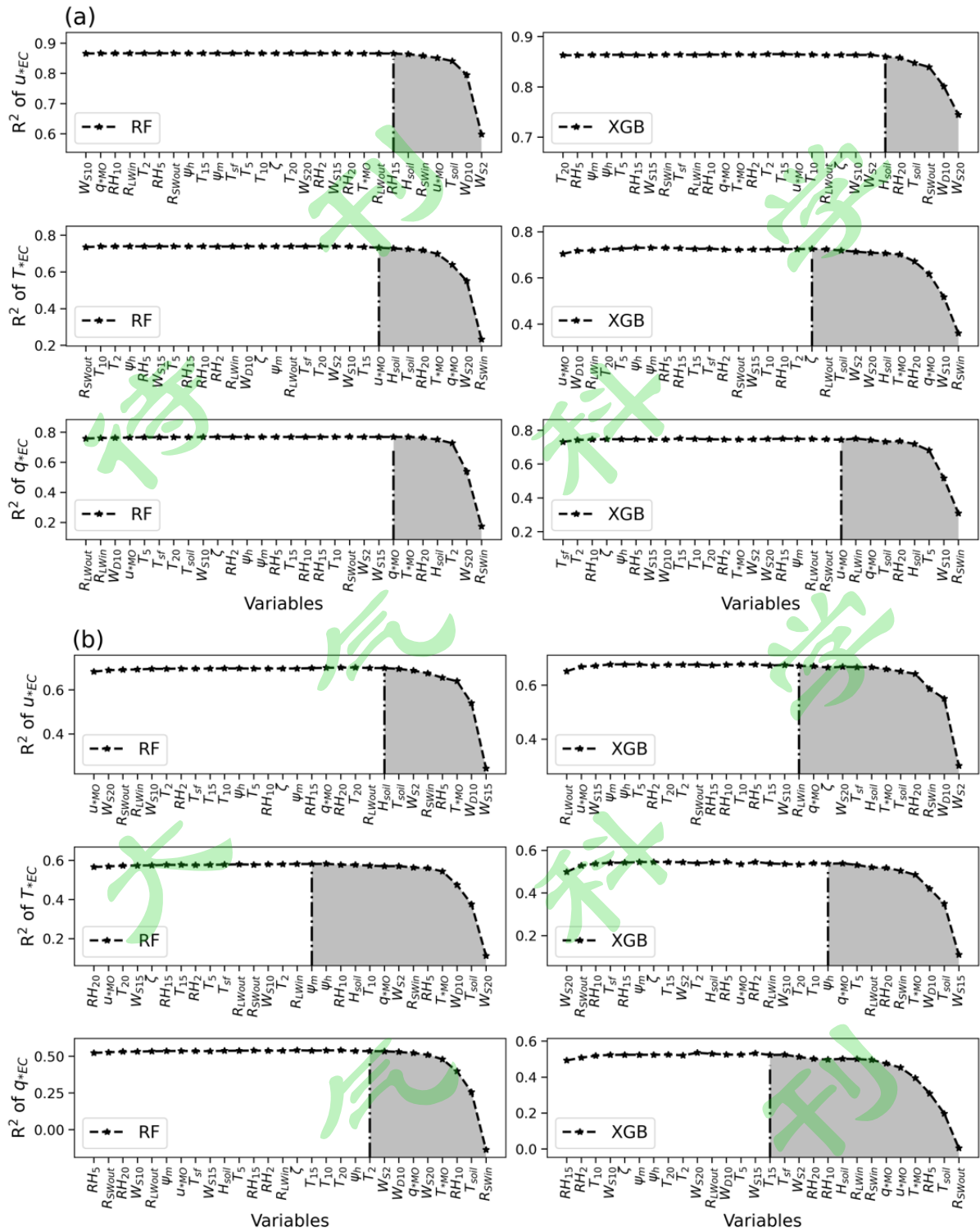


图 7 (a) 不稳定情况；(b) 稳定情况下端流通量 RF 和 XGB 机器学习模型后向变量选择。图中第一个点指使用横坐标所有变量的值，第二个点指删除横坐标第一个变量后剩下所有变量计算的结果，其他以此类推。横坐标变量具体含义见表 1，纵坐标为 R^2 。虚线以右的阴影部分为依据此图选择的变量组合。

Figure 7 Backward variable selection for turbulent fluxes in RF and XGB machine learning models on (a) unstable conditions and (b) stable conditions. In the chart, the first point indicates the value obtained using all variables on the horizontal axis, the second point indicates the result calculated after removing the first variable on the horizontal axis with all remaining variables, and so on for the rest. The specific meanings of the variables on the x-axis are found in Table 1, and the y-axis label represents R^2 . The variables groups selected based on this figure are the shaded areas, the left of the dashed line.

由表 6、图 6 和图 7 可知，各变量组合具有一定的差异，其在测试集上的应用之间也具有一定差异。不稳定情况下，摩擦速度计算最佳方案为 RF 模型及其后向变量组合；特征温度计算最佳方

案为 RF 模型及其后向变量组合；特征比湿计算最佳方案为 RF 模型及其前向变量组合。稳定情况下，摩擦速度为 XGB 模型及其后向变量组合；特征温度计算最佳方案为 RF 模型及其后向变量组合；特征比湿计算最佳方案为 XGB 模型及其后向变量组合。各个最佳方案的评估指标基本上超过了两种机器学习模型使用全部变量时的评估指标，可见合适变量选择的重要性。

3.3 机器学习前、后向变量组合适用性研究

由前、后向变量选择的步骤可知，前、后向变量选择在处理大量数据和变量时可能会非常耗时，因为每次增加或减少一个变量都需要遍历剩余的所有变量，现将选择的变量组合应用在人工神经网络（NN）中研究其适用性。

图 9 为 RF 和 XGB 选择的前、后向变量组合在 NN 模型上的应用，结合图 5 和图 8 可知，各变量组合在 NN 上的计算效果与在 RF 和 XGB 上的计算效果相近，均比基于 MOST 计算精度高。对比表 5、表 6 和表 7 可知，各变量组合在 NN 上的计算评估指标略低于在 RF 和 XGB 的计算评估指标，但相比 MOST，NN 评估指标仍有明显的提升。

图 10 显示了在使用各种变量组合进行计算时，相较于 MOST 方法，RF、XGB 和 NN 模型 RMSE 误差降低的百分比。不稳定情况下（黑色柱状图部分），各模型对特征比湿的误差降低效果明显（超过 63%，其中 RF 和 XGB 有超过 70% 的情况）。在稳定情况下（灰色柱状图部分），对特征温度和特征比湿都有明显的降低（特征温度 NN 超过 58%，XGB 和 RF 超过 60%；特征比湿均超过 70%）。不稳定和稳定情况下，RF 和 XGB 模型使用全部变量和使用各变量组合进行计算时，误差降低效果相近，各湍流通量的误差降低百分比差异都低于 5%。在不稳定情况下，相比于 RF 和 XGB，NN 模型使用大部分变量组合降低误差的效果略微下降，摩擦速度下降约 6%~9%，特征温度下降约 1%~6%，特征比湿下降约 0.4%~6%。在稳定情况下，与 RF 和 XGB 使用各变量组合相比，NN 使用变量组合计算摩擦速度降低误差的效果略微下降（其中 RF Backward 和 XGB Backward 降低误差的效果分别略微上升约 0.1% 和 1%），摩擦速度下降约 0.3%~13%，特征温度下降约 1%~6%，特征比湿下降约 0.4%~6%。

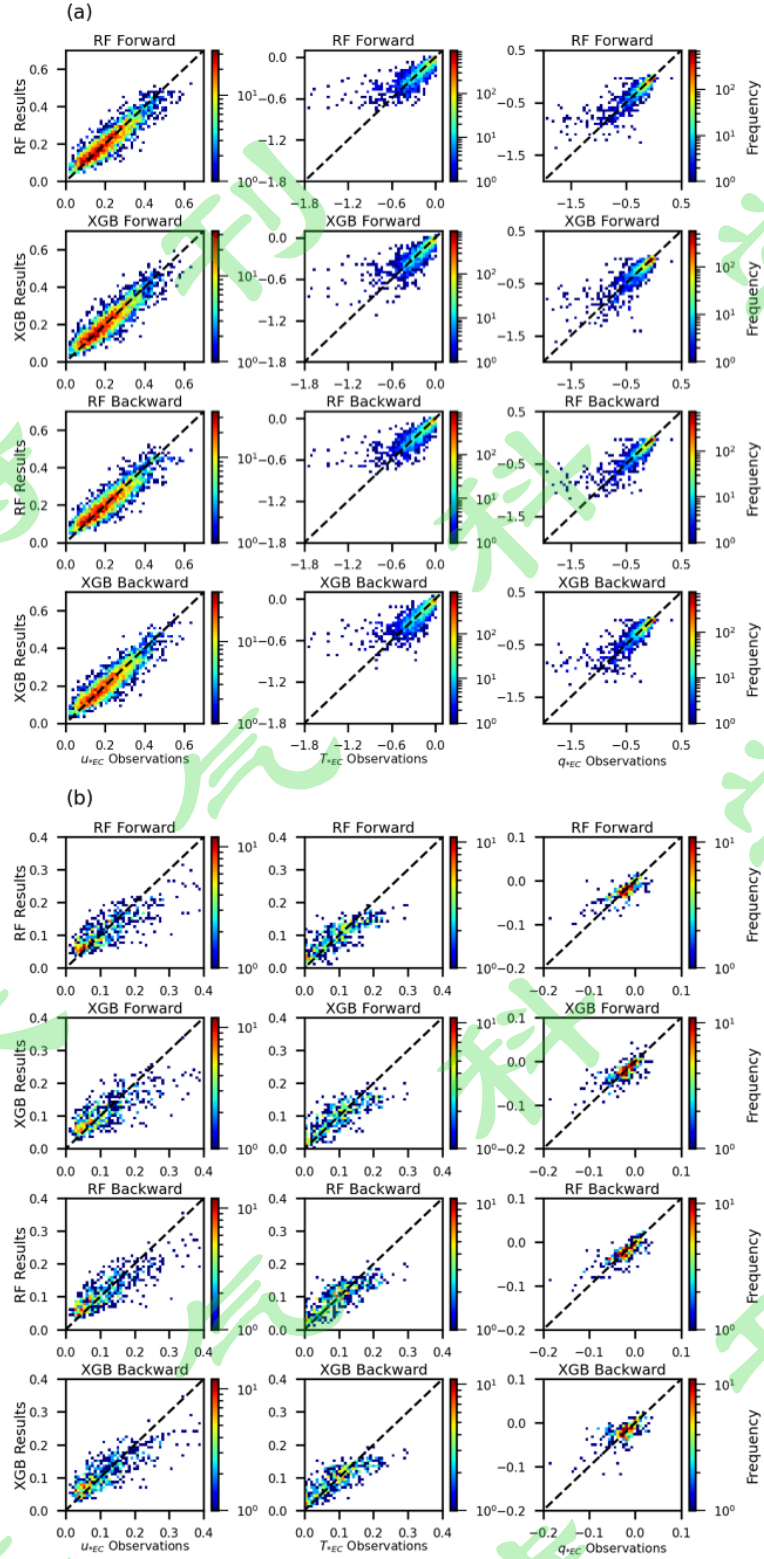


图 8 (a) 不稳定情况；(b) 稳定情况下 RF 和 XGB 选择的前、后向变量组合分别在 RF 和 XGB 模型上计算的湍流通量和观测对比。x 轴为观测的湍流通量，y 轴为机器学习方法计算的湍流通量。图中虚线为 1:1，散点颜色代表出现频率。

Figure 8 The comparison between observed turbulent fluxes and those calculated by the RF and XGB models using the forward and backward variables groups selected by RF and XGB on (a) Unstable conditions and (b) Stable conditions, respectively. The x-axis labels represent the observed turbulent fluxes, and the y-axis labels represent the turbulent fluxes calculated by the machine learning methods. The dashed line in the figure is the 1:1 line, and the colors of the scatter dots represent the frequency of occurrence.

表 6 不稳定和稳定情况下 RF 和 XGB 分别使用各自前、后向变量组合在测试集上计算结果评估对比，括号内的值为相对于使用全部变量得到的计算结果的变化量，下划线部分为最佳组合

Table 6 Evaluation comparison of the computational results on the test set using RF and XGB with their respective forward and backward variables groups under unstable and stable conditions, where the values in parentheses represent the change in computational results relative to using all variables, and the underlined parts indicate the best groups.

湍流通量		评估指标	RF 模型	XGB 模型	RF 模型	XGB 模型
			+RF Forward	+XGB Forward	+RF Backward	+XGB Backward
不稳定	u^*	RMSE	0.042	0.044	<u>0.042</u>	0.043
		(变化量)	(+0.001)	(+0.002)	<u>(0)</u>	(+0.001)
		R	0.908	0.899	<u>0.909</u>	0.904
		(变化量)	(-0.002)	(-0.011)	<u>(-0.001)</u>	(-0.007)
		IA	0.948	0.945	<u>0.950</u>	0.948
		(变化量)	(-0.003)	(-0.006)	<u>(-0.001)</u>	(-0.003)
		RMSE	0.104	0.106	<u>0.104</u>	0.107
		(变化量)	(0)	(-0.002)	<u>(-0.001)</u>	(0)
	T^*	R	0.855	0.849	<u>0.856</u>	0.846
		(变化量)	(0)	(+0.003)	<u>(+0.001)</u>	(-0.001)
		IA	0.909	0.911	<u>0.911</u>	0.906
		(变化量)	(+0.001)	(+0.008)	<u>(+0.004)</u>	(+0.002)
		RMSE	<u>0.120</u>	0.129	0.121	0.123
		(变化量)	<u>(0)</u>	(+0.006)	(+0.001)	(0)
		R	<u>0.860</u>	0.834	0.856	0.851
		(变化量)	<u>(0)</u>	(-0.016)	(-0.004)	(+0.001)
q^*	IA	<u>0.913</u>	0.903	<u>0.912</u>	0.909	
	(变化量)	<u>(-0.002)</u>	(-0.01)	<u>(-0.002)</u>	(-0.004)	
	u^*	RMSE	0.046	0.049	0.044	<u>0.043</u>
		(变化量)	(+0.003)	(-0.003)	(+0.001)	<u>(-0.002)</u>
R		0.822	0.785	0.835	<u>0.834</u>	
(变化量)		(-0.012)	(-0.024)	(+0.002)	<u>(+0.024)</u>	
IA		0.857	0.840	0.876	<u>0.888</u>	
(变化量)		(-0.029)	(-0.034)	(-0.010)	<u>(+0.013)</u>	
RMSE		0.037	0.036	<u>0.036</u>	0.039	
(变化量)		(0)	(-0.001)	<u>(-0.001)</u>	(+0.002)	
T^*	R	0.804	0.805	<u>0.820</u>	0.770	
	(变化量)	(-0.006)	(+0.016)	<u>(+0.010)</u>	(-0.019)	
	IA	0.867	0.874	<u>0.870</u>	0.847	
	(变化量)	(0)	(+0.008)	<u>(+0.003)</u>	(-0.020)	
	RMSE	0.017	0.018	0.017	<u>0.017</u>	
	(变化量)	(-0.002)	(-0.002)	(-0.003)	<u>(-0.003)</u>	
	R	0.757	0.722	0.770	<u>0.764</u>	
	(变化量)	(+0.080)	(+0.084)	(+0.093)	<u>(+0.126)</u>	
q^*	IA	0.840	0.831	0.850	<u>0.857</u>	
	(变化量)	(+0.047)	(+0.070)	(+0.056)	<u>(+0.096)</u>	

NN 计算各湍流通量的过程中对不同变量组合产生了一定的偏好，不稳定情况下，NN 模型计算摩擦速度时 RF 后向变量组合展现了最佳性能，而计算特征温度时 RF 前向变量组合成为了最优方案，计算特征比湿时则是 XGB 前向变量组合表现最佳；稳定情况下，NN 计算摩擦速度和特征温度时 XGB 前向变量组合展现了最佳性能，而计算特征比湿时在 RF 后向变量组合表现最佳。不稳定情况下的 NN 各个最佳方案的评估指标均比 3.2 节中的最佳方案评估指标略差，稳定情况下摩擦速度和特征比湿的 NN 最佳方案比 3.2 节中的最佳方案评估指标略差，但特征温度的 NN 最佳方案比 3.2 节中的最佳方案更加准确。可见变量选择有必要基于模型进行选择，但即便是使用其他模型选择的变量组合，独立的机器学习模型计算的湍流通量仍保持了较高的精度和较低的误差，均比传统的基于 MOST 方法计算的湍流通量精度更高、误差更小（图 9）。

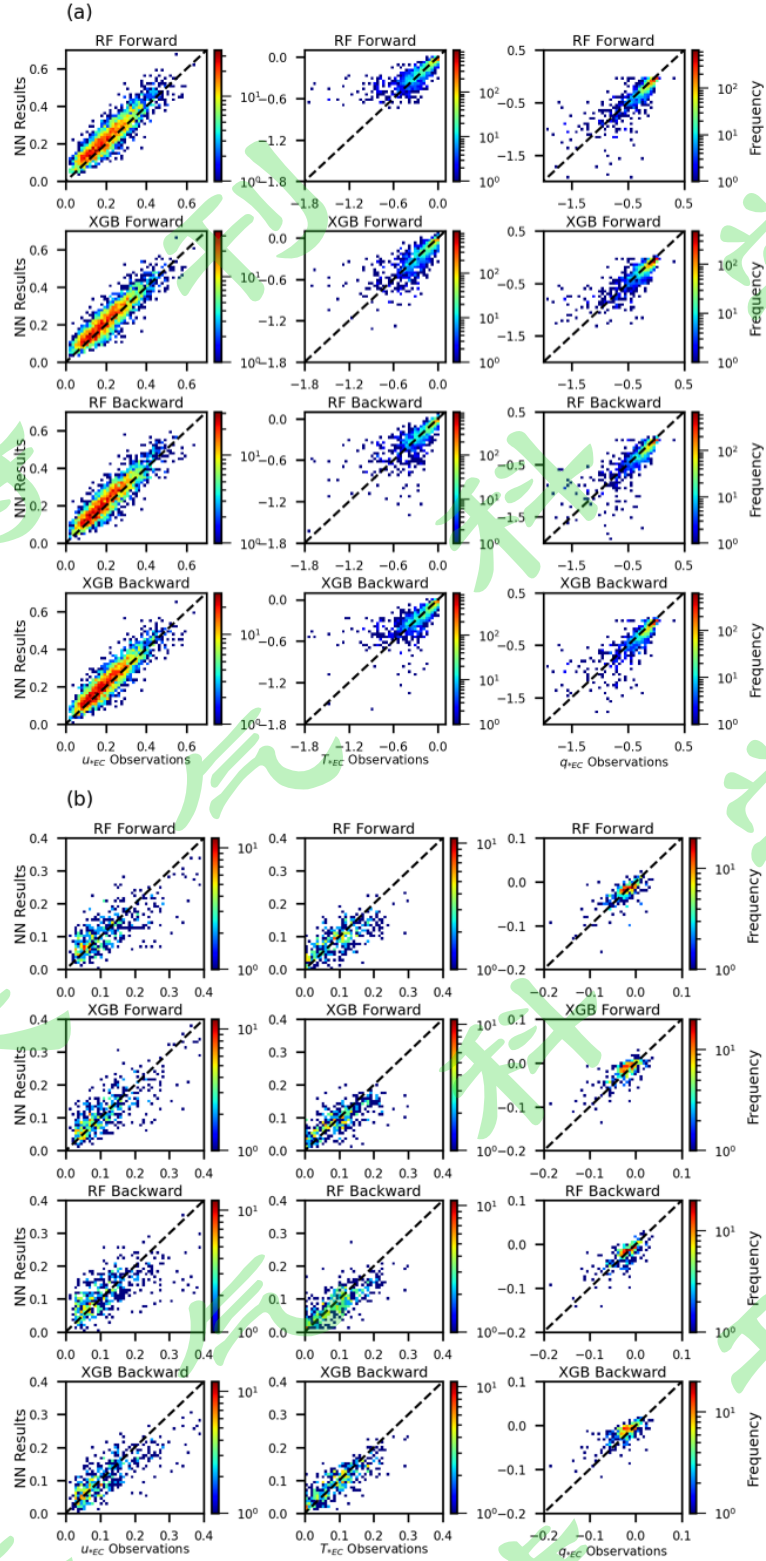


图9 (a) 不稳定情况；(b) 稳定情况下RF和XGB选择的前、后向变量组合在NN模型上计算的湍流量和观测对比。x轴为观测的湍流量，y轴为机器学习计算的湍流量。图中虚线为1:1，散点颜色代表出现频率。

Figure 9 The comparison between observed turbulent fluxes and those calculated by the NN model using the forward and backward variables groups selected by RF and XGB on (a) Unstable conditions and (b) Under stable conditions, respectively. The x-axis labels represent the observed turbulent fluxes, and the y-axis labels represent the turbulent fluxes calculated by the machine learning methods. The dashed line in the figure is the 1:1 line, and the colors of the scatter dots represent the frequency of occurrence.

表 7 RF 和 XGB 前、后向变量组合在 NN 上的应用结果评估对比, 括号中为与 MOST 结果的差值, 下划线部分为 NN 最佳方案

Table 7 Evaluation comparison of the application results of RF and XGB forward and backward variable groups on the NN, the values in parentheses being the differences from the results of MOST, and the underlined parts representing the best solutions for NN

湍流通量		NN 模型 + RF Forward	NN 模型 +XGB Forward	NN 模型 +RF Backward	NN 模型 +XGB Backward	
不稳定	u^*	RMSE	0.049	0.051	<u>0.046</u>	0.054
		(变化量)	(-0.017)	(-0.015)	<u>(-0.020)</u>	(-0.012)
		R	0.910	0.899	<u>0.907</u>	0.897
		(变化量)	(+0.122)	(+0.111)	<u>(+0.119)</u>	(+0.109)
	IA	0.940	0.936	<u>0.946</u>	0.929	
	(变化量)	(+0.076)	(+0.072)	<u>(+0.082)</u>	(+0.065)	
	RMSE	<u>0.109</u>	0.131	0.111	0.115	
	(变化量)	<u>(-0.043)</u>	(-0.021)	(-0.041)	(-0.037)	
	R	<u>0.845</u>	0.780	0.837	0.820	
	(变化量)	<u>(+0.158)</u>	(+0.093)	(+0.150)	(+0.133)	
	IA	<u>0.897</u>	0.877	0.893	0.891	
	(变化量)	<u>(+0.114)</u>	(+0.094)	(+0.110)	(+0.108)	
	RMSE	0.133	<u>0.130</u>	0.139	0.149	
	(变化量)	(-0.272)	<u>(-0.275)</u>	(-0.267)	(-0.256)	
	q*	R	0.831	<u>0.832</u>	0.811	0.800
	(变化量)	(+0.321)	<u>(+0.321)</u>	(+0.300)	(+0.290)	
IA	0.906	<u>0.893</u>	0.894	0.888		
(变化量)	(+0.299)	<u>(+0.286)</u>	(+0.287)	(+0.281)		
稳定	u^*	RMSE	0.054	<u>0.049</u>	0.050	0.050
		(变化量)	(-0.007)	<u>(-0.012)</u>	(-0.011)	(-0.010)
		R	0.716	<u>0.773</u>	0.768	0.755
		(变化量)	(+0.034)	<u>(+0.091)</u>	(+0.086)	(+0.073)
	IA	0.824	<u>0.871</u>	0.858	0.853	
	(变化量)	(+0.076)	<u>(+0.123)</u>	(+0.110)	(+0.105)	
	RMSE	0.043	<u>0.040</u>	0.041	0.040	
	(变化量)	(-0.061)	<u>(-0.063)</u>	(-0.063)	(-0.063)	
	R	0.740	<u>0.766</u>	0.748	0.752	
	(变化量)	(+0.623)	<u>(+0.649)</u>	(+0.635)	(+0.635)	
	IA	0.829	<u>0.860</u>	0.842	0.847	
	(变化量)	(+0.462)	<u>(+0.493)</u>	(+0.479)	(+0.479)	
	RMSE	0.019	0.019	<u>0.016</u>	0.017	
	(变化量)	(-0.055)	(-0.055)	<u>(-0.058)</u>	(-0.057)	
	q*	R	0.721	0.752	<u>0.798</u>	0.776
	(变化量)	(+0.626)	(+0.657)	<u>(+0.703)</u>	(+0.681)	
IA	0.837	0.795	<u>0.866</u>	0.847		
(变化量)	(+0.403)	(+0.403)	<u>(+0.474)</u>	(+0.456)		

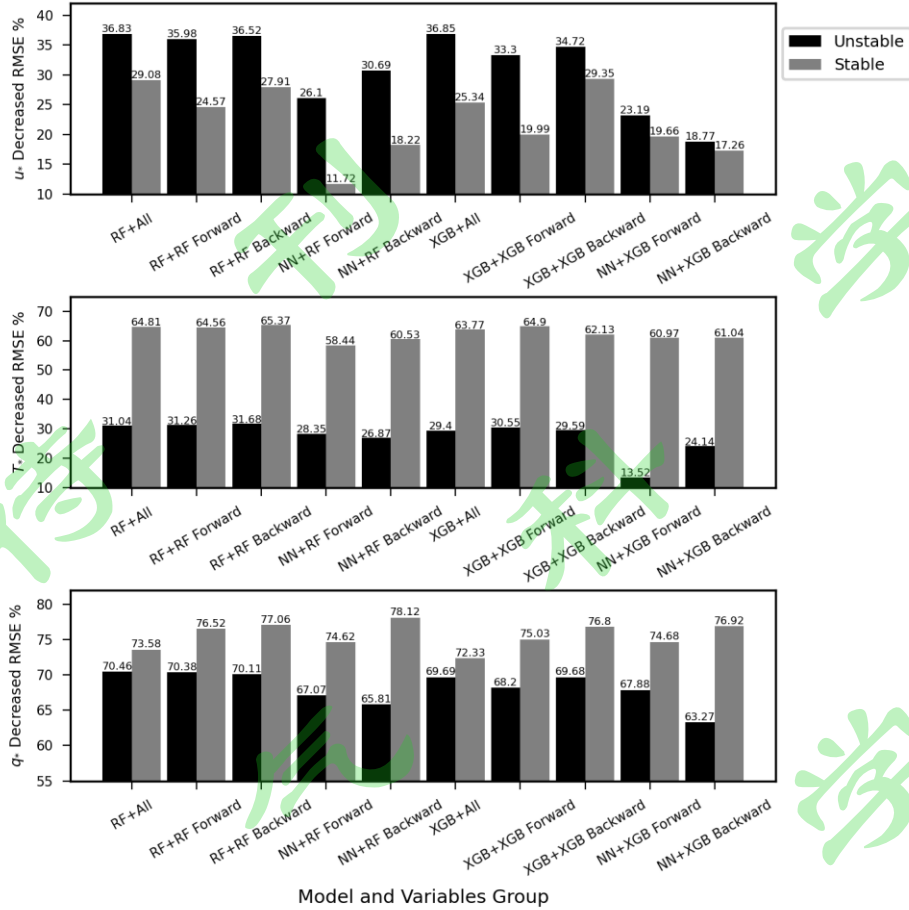


图 10 不稳定（黑色柱状图）和稳定（灰色柱状图）情况下机器学习计算的湍流量相比基于 MOST 降低的 RMSE 占比（例如， $\frac{(MOST - RF)}{MOST} \times 100\%$ ）。横坐标标签指模型+变量组合。

Figure 10 The proportion of RMSE reduction in the turbulent fluxes calculated by machine learning compared to those based on MOST under unstable (The black histograms) and stable (The gray histograms) conditions, for example, $\frac{(MOST - RF)}{MOST} \times 100\%$. The x-axis labels indicate the model + variables groups.

4. 总结和结论

近地层湍流量作为数值天气预报和气候预测的基本前提，通常基于莫宁奥布霍夫相似性理论（MOST）进行参数化。为改进 MOST 的计算结果，本文通过结合随机森林（RF）和 XGBoost（XGB）机器学习模型对基于 MOST 计算的湍流量进行了订正，提升了其准确度。同时，本文还使用前、后向变量选择方法分别对 RF 和 XGB 模型的输入变量进行选择，探究不同变量组合对计算结果的影响。在此基础之上进一步研究了 RF 和 XGB 选取的变量组合在另一机器学习模型——人工神经网络（NN）上的适用性。本文的主要结论如下：

1) 基于 MOST 计算的湍流量经 RF 和 XGB 机器学习模型订正后计算精度有显著提升，不稳定情况下各湍流量的一致性指数均超过 0.9，其中摩擦速度的一致性指数超过 0.95。各湍流量均方根误差均有明显降低，特征比湿甚至降低 70%。稳定情况下特征温度均方根误差降低超过 60%，特征比湿降低超过 70%。

2) RF 和 XGB 模型经前、后向变量选择得到的各变量组合之间有一定的差异, 其各自在测试数据集上的计算精度与使用全部变量时非常接近, 在模型与变量组合中, 最佳方案(最佳模型及其变量组合)基本优于使用全部变量时的最佳模型。其中以稳定情况下的特征比湿最为明显。RF 和 XGB 计算特征比湿时, 使用前、后向变量组合计算的相关性系数(RF 模型+RF Forward: 0.757, XGB 模型+XGB Forward: 0.722, RF 模型+RF Backward: 0.77, XGB 模型+XGB Backward: 0.764)和一致性指数(RF 模型+RF Forward: 0.840, XGB 模型+XGB Forward: 0.831, RF 模型+RF Backward: 0.850, XGB 模型+XGB Backward: 0.857)均明显优于使用全部变量时的相关性系数(RF 模型+All: 0.676, XGB 模型+All: 0.638)和一致性指数(RF 模型+All: 0.793, XGB 模型+All: 0.761)。这主要是因为在高维的数据中包括有较多不相关、冗余的变量, 影响机器学习回归模型拟合, 删除不相关变量可以提高准确性(Reunanen, 2004; Li, 2017; Venkatesh, 2019)。

3) 变量选择的适用性研究表明, RF 和 XGB 选择的大部分变量组合在 NN 上的计算精度仍可以保持, 其结果均明显比传统的基于 MOST 方法计算的通量精度更高、误差更小, 表明了本文变量选择方法的适用性。

致谢: 本论文的数值计算得到了南京信息工程大学高性能计算中心的计算支持和帮助, 特此致谢!

参考文献 (References)

- Akiba T, Sano S, Yanase T, et al. 2019. Optuna: A next-generation hyperparameter optimization framework[C]. //Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. New York, NY, USA: Association for Computing Machinery, 2623-2631. doi:10.1145/3292500.330701
- Baldocchi D. 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future[J]. Glob Chang Biol, 9(4): 479-492. doi:10.1046/j.1365-2486.2003.00629.
- Belgiu M, Drăguț L. 2016. Random forest in remote sensing: A review of applications and future directions[J]. ISPRS J. Photogramm. Remote Sens, 114: 24-31. doi:10.1016/j.isprsjprs.2016.01.011.
- Biermann, T., Babel, W., Ma, W., et al. 2014. Turbulent flux observations and modelling over a shallow lake and a wet grassland in the Nam Co basin, Tibetan Plateau[J]. Theor. Appl. Climatol, 116(1): 301-316. doi:10.1007/s00704-013-0953-6.
- Bourassa M A, Gille S T, Bitz C, et al. 2013. High-latitude ocean and sea ice surface fluxes: Challenges for climate research[J]. Bull. Am. Meteorol. Soc., 94(3): 403-423. doi:10.1175/BAMS-D-11-00244.1.
- Breiman L. 2001. Random forests [J]. Mach. Learn., 45(1): 5-32. doi:10.1023/A:1010933404324.
- Businger, J. A. 1966. Transfer of momentum and heat in the planetary boundary layer[C]// Proceedings of the Symposium on the Arctic Heat Budget and Atmospheric Circulation. California: University of California, 305-332.

- Businger, J. A., Wyngaard, J. C., Izumi, Y., et al. 1971. Flux-profile relationships in the atmospheric surface layer. [J] *J Atmos Sci*, 28(2): 181-189. [doi:10.1175/1520-0469\(1971\)028<0181:FPRITA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<0181:FPRITA>2.0.CO;2).
- Cai J, Xu K, Zhu Y, et al. 2020. Prediction and analysis of net ecosystem carbon exchange based on gradient boosting regression and random forest[J]. *Appl. Energy*, 262: 114566. [doi:10.1016/j.apenergy.2020.114566](https://doi.org/10.1016/j.apenergy.2020.114566).
- Cheng, Y. G., W. Brutsaert. 2005. Flux-profile relationships for wind speed and temperature in the stable atmospheric boundary layer [J]. *Bound. - Layer. Meteor.*, 114: 519-538. [doi:10.1007/s10546-004-1425-4](https://doi.org/10.1007/s10546-004-1425-4).
- Chen F, Yan Q, Liu C, et al. 2023. Evaluating Parameterizations for Turbulent Fluxes over the Landfast Sea-Ice Surface in Prydz Bay, Antarctica[J]. *Adv Atmos Sci*, 40(10):1816-1832. [doi:10.1007/s00376-023-2299-z](https://doi.org/10.1007/s00376-023-2299-z)
- Chen T, Guestrin C. 2016. Xgboost: A scalable tree boosting system[C]. //Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York, NY, USA: Association for Computing Machinery, 785-794. [doi:10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)
- Deo R C, Şahin M. 2015. Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia[J]. *Atmos Res*, 161: 65-81. [doi:10.1016/j.atmosres.2015.03.018](https://doi.org/10.1016/j.atmosres.2015.03.018)
- Dethloff K, Abegg C, Rinke A, et al. 2001. Sensitivity of Arctic climate simulations to different boundary-layer parameterizations in a regional climate model[J]. *Tellus A*, 53(1): 1-26. [doi:10.1034/j.1600-0870.2001.01073.x](https://doi.org/10.1034/j.1600-0870.2001.01073.x)
- Durden, D. J., Nappo, C. J., Leclerc, M. Y., et al. 2013. On the impact of wave-like disturbances on turbulent fluxes and turbulence statistics in nighttime conditions: a case study, *Biogeosciences*, 10: 8433-8443. [doi:10.5194/bg-10-8433-2013](https://doi.org/10.5194/bg-10-8433-2013).
- Dyer, A. J. 1967. The turbulent transport of heat and water vapour in an unstable atmosphere[J]. *Q. J. R. Meteorol. Soc.*, 93: 501-508. [doi:10.1002/qj.49709339809](https://doi.org/10.1002/qj.49709339809)
- Dyer, A. J. 1974. A review of flux-profile relationships[J]. *Bound. - Layer. Meteor.*, 7: 363-372. [doi:10.1007/BF00240838](https://doi.org/10.1007/BF00240838)
- Finnigan J, Ayotte K, Harman I, et al. 2020. Boundary-layer flow over complex topography[J]. *Bound. -Layer Meteor.*, 177: 247-313. [doi:10.1007/s10546-020-00564-3](https://doi.org/10.1007/s10546-020-00564-3)
- Foken, T., Göckede, M., Mauder, et al. 2004. Post-field data quality control[M]. //In: Lee, X., Massman, W., Law, B. (eds) *Handbook of Micrometeorology. Atmospheric and Oceanographic Sciences Library*, Berlin, German: Springer, Dordrecht, 29:181-208pp. [doi:10.1007/1-4020-2265-4_9](https://doi.org/10.1007/1-4020-2265-4_9)
- Foken, T., Aubinet, M., Leuning, R. 2012. The eddy covariance method[M]. //In: Aubinet, M., Vesala, T., Papale, D. (eds) *Eddy Covariance.*, Berlin, German: Springer, Dordrecht, 1-19 pp. [doi:10.1007/978-94-007-2351-1_1](https://doi.org/10.1007/978-94-007-2351-1_1)
- Frazier P I. 2018. Bayesian optimization[M]. //In *INFORMS TutORials in Operations Research*. Published online, 255-278. [doi:10.1287/educ.2018.0188](https://doi.org/10.1287/educ.2018.0188)
- Friedman J H. 2001. Greedy function approximation: a gradient boosting machine[J]. *Annals of statistics*, 29(5): 1189-1232. [doi:10.2307/2699986](https://doi.org/10.2307/2699986)

Gryanik V M, Lüpkes C, Sidorenko D, et al. 2021. A universal approach for the non-iterative parametrization of near-surface turbulent fluxes in climate and weather prediction models[J]. *J. Adv. Model. Earth Syst*, 13(8): e2021MS002590. doi:10.1029/2021MS002590

Guo, X. and Zhang, H. 2007. A performance comparison between nonlinear similarity functions in bulk parameterization for very stable conditions, *Environ. Fluid. Mech.*, 7: 239-257. doi:10.1007/s10652-006-9015-2.

韩念霏, 杨璐, 陈明轩等. 2022. 京津冀站点风温湿要素的机器学习订正方法. *应用气象学报*, 33(4):489-500. Han Nianfei, Yang Lu, Chen Mingxuan, et al. Machine learning correction of wind, temperature and humidity elements in Beijing-Tianjin-Hebei Region[J]. *J Appl Meteor Sci*, 2022, 33(4): 489-500. doi:10.11898/1001-7313.20220409.

黄翊, 彭新东. 2017. 边界层湍流参数化改进对雾的模拟影响[J]. *大气科学*, 41(3): 533-543. Huang Yi, Peng Xindong. 2017. The impact of an improved planetary boundary layer parameterization scheme on the simulation of fog [J]. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 41(3): 533-543. doi:10.3878/j.issn.1006-9895.1606.16152.

Helbig, M., Gerken, T., Beamesderfer, E. R., et al. 2021. Integrating continuous atmospheric boundary layer and tower-based flux measurements to advance understanding of land-atmosphere interactions[J]. *Agric For Meteorol*, 307: 108509. doi: 10.1016/j.agrformet.2021.108509.

Högström, U. 1988. Non-Dimensional Wind and Temperature Profiles in the Atmospheric Surface Layer: A Re-Evaluation[J]. *Bound. -Layer. Meteor*, 42: 55-78. doi:10.1007/BF00119875

Holtstag, A. A. M., H. A. R. de Bruin. 1988. Applied modeling of the nighttime surface energy balance over land[J]. *J. App. Meteorol*, 27: 689-704. doi:10.1175/1520-0450(1988)027<0689:AMOTNS>2.0.CO;2.

Jiang Y., Li Y., Lu Y., et al. 2024. Evaluating nine different air-sea flux algorithms coupled with CAM 6. *Atmos. Res*, 307: 107486. doi:10.1016/j.atmosres.2024.107486

Jiménez, P. A., Dudhia, J., González-Rouco, J. F., et al. 2012. A revised scheme for the WRF surface layer formulation, *Mon. Weather Rev.*, 140: 898-918. doi: 10.1175/MWR-D-11-00056.1

Kingma D, Ba J. 2014. Adam: A Method for Stochastic Optimization[C]. *International Conference on Learning Representation s. CORR*. doi:10.48550/arXiv.1412.6980.

Krogh A, Hertz J. 1991. A simple weight decay can improve generalization[C]. // *Advances in neural information processing systems 4*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, 950-957. doi:10.5555/2986916.2987033

刘鹏飞, 刘树华, 胡非等. 2010. 湍流通量计算方法和误差的比较研究[J]. *气象学报*, 68(4):487-500. Liu Pengfei, Liu Shuhua, HuFei, et al. 2010. A comparison of the different methods for estimating turbulent fluxes and their errors[J]. *Acta. Meteorol. Sin*, 68(4): 487-500. doi:10.11676/qxxb2010.047.

刘娣, 孙佳倩, 余钟波. 2024. 基于机器学习模型的多层土壤湿度反演[J]. *河海大学学报 (自然科学版)*, 52(3):7-14. Liu Di, Sun Jiaqian, Yu Zhongbo. Multi-layer soil moisture inversion based on machine learning models[J]. *Journal of Hohai University (Natural Sciences)*, 2024, 52(3):7-14. doi:10.3876/j.issn.1000-1980.2024.03.002.

- Launiainen J, 1995. Derivation of the relationship between the Obukhov stability parameter and the bulk Richardson number for flux-profile studies. *Bound. - Layer. Meteor.* 76:165-179. [doi:10.1007/BF00710895](https://doi.org/10.1007/BF00710895).
- Litt, M., Sicart, J.-E., and Helgason, W., 2015. A study of turbulent fluxes and their measurement errors for different wind regimes over the tropical Zongo Glacier (16°S) during the dry season. *Atmos. Meas. Tech.*, 8, 3229-3250, [doi:10.5194/amt-8-3229-2015](https://doi.org/10.5194/amt-8-3229-2015).
- Liu, C., Li, Y., Gao, Z., et al, 2020. Improvement of drag coefficient calculation under near-neutral conditions in light winds over land[J]. *J. Geophys.*: 125(24): e2020JD033472. [doi:10.1029/2020JD033472](https://doi.org/10.1029/2020JD033472)
- Liu C, Yang Q, Shupe M D, et al.2023.Atmospheric Turbulent Intermittency Over the Arctic Sea-Ice Surface During the MOSAiC Expedition[J]. *J. Geophys.*: 128(15): e2023JD038639. [doi:10.1029/2023JD038639](https://doi.org/10.1029/2023JD038639)
- Li J, Cheng K, Wang S, et al, 2017. Feature selection: A data perspective[J]. *ACM computing surveys (CSUR)*, 50(6): 1-45. [doi:10.1145/3136625](https://doi.org/10.1145/3136625)
- Li, Y., Gao, Z., Lenschow, D. H., et al. 2010. An Improved Approach for Parameterizing Surface-Layer Turbulent Transfer Coefficients in Numerical Models. *Bound. - Layer. Meteor.* 137(1):153-165. [doi:10.1007/s10546-010-9523-y](https://doi.org/10.1007/s10546-010-9523-y).
- Li, Y., Wu, Y., Tang, J, et al,2023. Quantitative evaluation of wavelet analysis method for turbulent flux calculation of non-stationary series[J]. *Geophys.*, 50(5): e2022GL101591. [doi:10.1029/2022GL101591](https://doi.org/10.1029/2022GL101591)
- Li Q, Gentile P, Mellado J P, et al, 2018. Implications of nonlocal transport and conditionally averaged statistics on Monin–Obukhov similarity theory and Townsend’s attached eddy hypothesis[J]. *J. Atmos. Sci.*, 75(10): 3403-3431. [doi:10.1175/JAS-D-17-0301.1](https://doi.org/10.1175/JAS-D-17-0301.1).
- Mahrt, L., 2007. The influence of nonstationarity on the turbulent flux–gradient relationship for stable stratification. *Bound. - Layer. Meteor.* 125(2):245-264. [doi:10.1007/s10546-007-9154-0](https://doi.org/10.1007/s10546-007-9154-0)
- Monin A S, Obukhov A M. 1954. Basic laws of turbulent mixing in the surface layer of the atmosphere[J]. *Contrib. Geophys. Inst. Acad. Sci. USSR*, 24(151):163-187.
- McCandless T, Gagne D J, Kosović B, et al, 2022. Machine learning for improving surface-layer-flux estimates[J]. *Bound. - Layer Meteor.* 185(2): 199-228. [doi:10.1007/s10546-022-00727-4](https://doi.org/10.1007/s10546-022-00727-4)
- Moore CJ.1986. Frequency response corrections for eddy correlation systems[J]. *Bound. - Layer Meteor.* 37(1), 17-35. [doi:10.1007/BF00122754](https://doi.org/10.1007/BF00122754)
- McGovern A, Elmore K L, Gagne II D J, et al. 2017. Using artificial intelligence to improve real-time decision–making for high-impact weather [J]. *Bull. Amer. Meteor. Soc.*, 98(10): 2073-2090. [doi:10.1175/BAMS-D-16-0123.1](https://doi.org/10.1175/BAMS-D-16-0123.1)
- Newman JF, Klein PM. 2014. The impacts of atmospheric stability on the accuracy of wind speed extrapolation methods[J]. *Resources*, 3(1):81-105. [doi:10.3390/resources3010081](https://doi.org/10.3390/resources3010081)
- Obukhov AM. 1946. Turbulentnost v temperaturnoj – neodnorodnoj atmosfere (Turbulence in an Atmosphere with a non-uniform Temperature) [J]. *Trudy. Inst. Theor. Geofiz. AN. SSSR*, 1:95-115 [doi:10.1007/BF00718085](https://doi.org/10.1007/BF00718085)
- Paulson, C. A. 1970.The mathematical representation of wind speed and temperature in the unstable a

654 atmospheric surface layer [J]. J. Appl. Meteor, 9: 857-861. [doi:10.1175/1520-0450\(1970\)009<0857:TMRO](https://doi.org/10.1175/1520-0450(1970)009<0857:TMRO)
655 [WS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1970)009<0857:TMRO)

656 任萍, 陈明轩, 曹伟华等. 2020. 基于机器学习的复杂地形下短期数值天气预报误差分析与
657 订正[J]. 气象学报, 78(06):1002-1020. Ren Ping, Chen Mingxuan, Cao Weihua, et al. 2020. Error anal
658 ysis and correction of short-term numerical weather prediction under complex terrain based on machine lea
659 rning. Acta. Meteorol. Sin, 78(6):1002-1020. [doi:10.11676/qxxb2020.060](https://doi.org/10.11676/qxxb2020.060).

660 Reunanen J, 2003. Overfitting in making comparisons between variable selection methods[J]. J. Che
661 m. Inf. Comput., 3: 1371-1382. [doi:10.1021/ci0342472](https://doi.org/10.1021/ci0342472)

662 Schneider T, Lüpkes C, Dorn W, et al. 2022. Sensiti
663 vity to changes in the surface-layer turbulence parameterization for stable conditions in winter: A case stud
664 y with a regional climate model over the Arctic[J]. Atmos. Sci. Lett, 23(1): e1066. [doi:10.1002/asl.1066](https://doi.org/10.1002/asl.1066)

665 Sfyrí E, Rotach M W, Stiperski I, et al, 2018. Scalar-flux similarity in the layer near the surface over m
666 ountainous terrain[J]. Bound. - layer Meteor, 169: 11-46. [doi:10.1007/s10546-018-0365-3](https://doi.org/10.1007/s10546-018-0365-3)

667 Srivastava N, Hinton G, Krizhevsky A, et al, 2014. Dropout: a simple way to prevent neural networks
668 from overfitting[J]. J. Mach. Learn. Res, 15(1): 1929-1958. [doi:abs/10.5555/2627435.2670313](https://doi.org/10.5555/2627435.2670313).

669 Swinbank, W. C., 1964: The exponential wind profile. Quarl. J. Roy. Meteor. Soc., 90(384):119-135.
670 [doi:10.1002/qj.49709038402](https://doi.org/10.1002/qj.49709038402)

671 Stull, R. B. (1988). An introduction to boundary layer meteorology [M]. Springer Science & Business
672 Media. [doi:10.1007/978-94-009-3027-8](https://doi.org/10.1007/978-94-009-3027-8)

673 Tong C, Nguyen K X. 2015. Multipoint Monin–Obukhov similarity and its application to turbulence s
674 pectra in the convective atmospheric surface layer[J]. J. Atmos. Sci, 72(11): 4337-4348. [doi:10.1175/JAS-](https://doi.org/10.1175/JAS-D-15-0134.1)
675 [D-15-0134.1](https://doi.org/10.1175/JAS-D-15-0134.1).

676 Uhrig R. E. 1995. Introduction to artificial neural networks[C]. //Proceedings of IECON '95-21st Ann
677 ual Conference on IEEE Industrial Electronics. Orlando, FL, USA: IEEE, 1: 33-37, [doi:10.1109/IECON.19](https://doi.org/10.1109/IECON.1995.483329)
678 [95.483329](https://doi.org/10.1109/IECON.1995.483329).

679 van Dijk A, Moene AF, De Bruin HAR. 2004. The principles of surface flux physics: Theory, practice
680 and description of the EC Pack library[M]. Wageningen, The Netherlands: Meteorology and Air Quality
681 Group, Wageningen University, 99pp.

682 Venkatesh B, Anuradha J, 2019. A review of feature selection and its methods[J]. Cybernetics and info
683 rmation technologies, 19(1): 3-26. [doi:10.2478/cait-2019-0001](https://doi.org/10.2478/cait-2019-0001)

684 Vilà-Guerau de Arellano J, Hartogensis O, Benedict I, et al. 2023. Advancing understanding of land–a
685 tmosphere interactions by breaking discipline and scale barriers[J]. Ann. N. Y. Acad. Sci, 1522(1): 74-97. [doi:10.1111/nyas.14956](https://doi.org/10.1111/nyas.14956)

686 Webb, E. K. 1970. Profile Relationships: the Log-Linear Range, and Extension to Strong Stability[J].
687 Quart. J. Roy. Meteorol. Soc, 96(407): 67-90. [doi:10.1002/qj.49709640708](https://doi.org/10.1002/qj.49709640708)

688 Webb EK, Pearman GI, Leuning R. 1980. Correction of flux measurements for density effects due to
689 heat and water transfer[J]. Q. J. R. Meteorol. Soc, 106(447): 85-100. [doi:10.1002/qj.49710644707](https://doi.org/10.1002/qj.49710644707)

690 Willmott C J. 1981. On the validation of models[J]. Phys. Geogr., 2(2): 184-194. [doi:10.1080/027236](https://doi.org/10.1080/02723646.1981.10642213)
691 [46.1981.10642213](https://doi.org/10.1080/02723646.1981.10642213)

Wulfmeyer V, Pineda J M V, Otte S, et al, 2023. Estimation of the surface fluxes for heat and momentum in unstable conditions with machine learning and similarity approaches for the LAFE data set[J]. *Boundary-Layer Meteorol*, 186(2): 337-371. [doi:10.1007/s10546-022-00761-2](https://doi.org/10.1007/s10546-022-00761-2)

徐景峰, 宋林烨, 陈明轩等. 2023. 冬奥会复杂山地百米尺度 10 m 风速预报的机器学习订正对比试验 [J]. *大气科学*, 47(3): 805–824. Xu Jingfeng, Song Linye, Chen Mingxuan, et al. 2023. Comparative Machine Learning-Based Correction Experiment for a 10 m Wind Speed Forecast at a 100 m Resolution in Complex Mountainous Areas of the Winter Olympic Games [J]. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 47(3):805–824. [doi:10.3878/j.issn.1006-9895.2209.22117](https://doi.org/10.3878/j.issn.1006-9895.2209.22117)

严晓强, 胡泽勇, 孙根厚等. 2018. 那曲高寒草地上四种地表通量计算方法的对比[J]. *高原气象*, 37(2): 358-370. Yan Xiaoqiang, Hu Zeyong, Sun Genhou, et al, 2018. Comparison of four methods for calculating surface fluxes on alpine grassland at Naqu [J]. *Plateau Meteorol*, 37(2): 358-370. [doi:10.7522/j.issn.1000-0534.2017.00067](https://doi.org/10.7522/j.issn.1000-0534.2017.00067)

岳平, 张强, 牛生杰等. 2012. 半干旱草原下垫面能量平衡特征及土壤热通量对能量闭合率的影响[J]. *气象学报*, 70(01): 136-143. Yue Ping, Zhang Qiang, Niu Shengjie, et al. 2012. Characters of surface energy balance over a semi-arid grassland and effects of soil heat flux on energy balance closure[J]. *Acta Meteorol. Sin*, 70(1):136-143 [doi:10.11676/qxxb2012.013](https://doi.org/10.11676/qxxb2012.013)

Yu Y, Gao C Y, Li Y, et al, 2023. Two Random Forest Models for the Non-Iterative Parametrization of Surface-Layer Turbulent Fluxes[J]. *Geophys.*, 50(21): e2023GL105923. [doi:10.1029/2023GL105923](https://doi.org/10.1029/2023GL105923)

张蕴帅, 黄倩, 马耀明, 等. 2021. 黄河源区鄂陵湖湖面和湖边草地对流边界层湍流结构特征的大涡模拟研究 [J]. *大气科学*, 45(2): 435–455. Zhang Yunshuai, Huang Qian, Ma Yaoming, et al. 2021. Large Eddy Simulation Study of the Turbulent Structure Characteristics of the Convective Boundary Layer over Ngoring Lake and Surrounding Grassland in the Source Region of the Yellow River [J]. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 45(2): 435–455. [doi:10.3878/j.issn.1006-9895.2009.20111](https://doi.org/10.3878/j.issn.1006-9895.2009.20111)

朱丽, 苗峻峰, 高阳华. 2020. 重庆城市热岛环流结构和湍流特征的数值模拟 [J]. *大气科学*, 44(3): 657–678. Zhu Li, Miao Junfeng, Gao Yanghua. 2020. A Numerical Simulation of Urban Breeze Circulation Structure and Turbulence in Chongqing [J]. *Chinese Journal of Atmospheric Sciences (in Chinese)*, 44(3): 657–678. [doi:10.3878/j.issn.1006-9895.2003.19239](https://doi.org/10.3878/j.issn.1006-9895.2003.19239)

Zhao Y, Wu J, Guo C, et al, 2023. Comparing the eddy covariance and gradient methods for measuring water and heat fluxes in paddy fields[J]. *Agric. Water. Manag*, 284: 108340. [doi:10.1016/j.agwat.2023.108340](https://doi.org/10.1016/j.agwat.2023.108340)