

# 基于 Prophet 模型与机器学习算法结合的全 国碳排放量预测研究

萧梓锋\*, 杨津锋, 赵馨悦, 刘颖恩  
广东白云学院 应用经济学院, 广东 广州 510450  
DOI:10.61369/ASDS.2025060017

**摘要 :** 为贯彻落实新发展理念, 准确预测全国碳排放对于实现碳中和的进程有着重要作用。本文提出 Prophet+ML 混合模型对碳排放的长期趋势进行预测。研究基于 2019–2024 年中国每日二氧化碳排放监测数据, 首先, 运用 Prophet 模型对碳排放数据进行趋势、季节性与节假日的预测并提取该残差信息, 然后, 分别用三种机器学习算法 RandomForest、XGBoost、LightGBM 对残差进行拟合, 用于提升该模型的预测效果。实验结果表明: 1. 相比于 ARIMA, Prophet 模型对具有季节性和长期趋势的数据有较好的拟合效果, MAE、MSE、RMSE、MAPE 四个偏差指标分别减少了 0.427107、1.054397、0.22992、1.472025。2. Prophet 与机器学习算法相结合的混合模型能明显的提高预测准确性, 其中 LightGBM 优化效果最好, 四个偏差指标分别减少了 0.759228、3.599075、1.107846、2.38763, 决定系数 R<sup>2</sup> 提高了 0.3918。

**关键词 :** 时间序列分析; 碳排放预测; Prophet; 机器学习; 混合模型

## Research on National Carbon Emission Prediction Based on the Combination of Prophet Model and machine Learning Algorithms

Xiao Zifeng\*, Yang Jinfeng, Zhao Xinyue, Liu Ying'en

School of Applied Economics, Guangdong Baiyun University, Guangzhou, Guangdong 510450

**Abstract :** To implement the new development concepts, accurately predicting the national carbon emissions plays a significant role in the process of achieving carbon neutrality. This paper proposes a Prophet+ML hybrid model to predict the long-term trend of carbon emissions. The research is based on the daily carbon dioxide emission monitoring data of China from 2019 to 2024. Firstly, the Prophet model is used to predict the trends, seasonality and holidays of the carbon emission data and extract the residual information. Then, Three machine learning algorithms, RandomForest, XGBoost, and LightGBM, were respectively used to fit the residuals to improve the prediction effect of the model. The experimental results show that :1. Compared with ARIMA, the Prophet model has a better fitting effect on data with seasonal and long-term trends. The four deviation indicators of MAE, MSE, RMSE, and MAPE have decreased by 0.427107, 1.054397, 0.22992, and 1.472025 respectively. The hybrid model combining Prophet with machine learning algorithms can significantly enhance prediction accuracy. Among them, LightGBM has the best optimization effect, with the four bias indicators reduced by 0.759228, 3.599075, 1.107846, and 2.38763 respectively, and the coefficient of determination R<sup>2</sup> increased by 0.3918.

**Keywords :** time series analysis; carbon emission prediction; Prophet; machine learning; hybrid model

## 引言

碳排放是工业文明的副产品也是人类可持续发展的关键约束条件, 为了维持经济发展与生态文明安全建设间的平衡关系。目前, 作为全球最大的碳排放国, 我国争取在 2030 年前二氧化碳排放达到峰值并争取在 2060 年前实现碳中和。

习近平总书记强调: “实现碳达峰, 碳中和, 是贯彻新发展理念、构建新发展格局、推动高质量发展的内在要求, 是党中央统筹国

### 作者简介:

杨津锋, 广东白云学院, 本科在读, 研究方向为机器学习;

赵馨悦, 广东白云学院, 本科在读, 研究方向为市场调查与预测;

刘颖恩, 广东白云学院, 本科在读, 研究方向为市场调查与预测。

通讯作者: 萧梓锋, 广东白云学院, 本科在读, 研究方向为应用统计、机器学习。

内国际两个大局作出的重大战略决策。”“双碳”政策是中国应对全球气候变化、推动绿色低碳转型的核心型国家级战略部署，对于构建清洁低碳安全高效的能源体系、推进产业结构调整和升级、提升生态环境质量等具有重要意义。

在传统的预测模型中，余航<sup>[1]</sup>（2018）、周树涛<sup>[2]</sup>等（2023）用 ARIMA 预测我国碳排放总量及减排潜力，钟进<sup>[3]</sup>等利用 SARI-MA 等模型综合考虑时间序列的趋势性，对比分析了不同模型对于碳排放数据的预测效果；而 Prophet 模型作为近几年流行的时间序列预测工具，也被广泛应用于各种时间序列分析。在环境方面，用 Prophet 对空气质量和 PM2.5 浓度进行预测。在能源方面，Prophet 进行电力负荷、天然气需求以及电网调度预测<sup>[4]</sup>等。在市场方面，利用 Prophet 进行产品销量的趋势预测和需求预测<sup>[5]</sup>等。SuS<sup>[6]</sup>等（2023）也把模型用到航空飞行风险预测和电离层扰动分析等非环境类时序中。

然而，传统的预测模型存在一定的局限性，无法考虑突变情况以及政策波动等非线性问题，因而学者们逐渐由传统模型转向用人工智能模型来预测的研究。从现有研究成果来看，张学清<sup>[7]</sup>等（2022）构建 CNN-LSTM 混合模型，通过发挥 LSTM 在时间序列依赖关系挖掘上的局部与长期特征优势，实现了对中国碳排放的实时动态预测；王琦<sup>[8]</sup>等（2021）则运用 LSTM 网络，成功解析了 SO<sub>2</sub> 浓度变化的复杂非线性污染过程，为空气质量预测提供了新路径；赵成柏<sup>[9]</sup>等（2012）创新性地将 ARIMA 模型与 BP 神经网络相结合，在碳排放强度预测中有效融合了传统统计方法与神经网络的技术优势，显著提升了预测效能。

在复杂数据预测任务中，由于单一模型的稳定性和精度等缺点，近年来学者们利用许多模型的融合、集成来弥补不足，提高预测能力。Prophet 具有良好的趋势分解能力，常常被学者用来和 XGBoost、LightGBM、随机森林等综合算法构建模型。常恬君<sup>[10]</sup>等（2019）利用 Prophet-随机森林提升空气质量指数的预测效果；高洁如<sup>[11]</sup>等（2024）构建的 Prophet-LightGBM 可以实现对 PM2.5 的精准预测；彭湃和刘敏<sup>[12]</sup>（2021）将 Prophet 和 LSTM 进行融合构建组合模型进行电力负荷短期预测；郑陈达和兰雁宁（2024）<sup>[13]</sup>提出的 Prophet-XGBoost 组合模型在天然气需求预测中取得很好的结果。

综上，通过对现有文献的系统梳理与分析可知，ARIMA、神经网络及集成模型等方法已在碳排放预测领域得到广泛应用，尽管 Prophet 在空气质量、电力负荷等领域取得良好效果，但目前专门针对碳排放长期趋势预测的相关研究仍较为匮乏。这一研究空白为本研究的开展提供了明确的切入点。

## 一、模型构建

### （一）Prophet 模型构建

Prophet 是由 Facebook 开发的一种基于加性模型的开源工具，用于时间序列预测的模型。这个基于加性模型工具特别适合于具有强季节性影响和多个历史数据季节的业务时间序列数据。Prophet 模型公式表示为：

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1)$$

$t$  为时间； $y(t)$  是时间序列，表示时间  $t$  时刻的观测值； $g(t)$  是趋势项，表示时间  $t$  时刻的趋势函数，描述数据的长期趋势； $s(t)$  是季节项，表示时间  $t$  时刻的季节性函数，捕捉周期性的波动，可以以周、月、季度、年为单位。 $h(t)$  是节假日效应，表示时间  $t$  时刻的假期效应，用于构建特殊事件或假期对数据的影响。

$\varepsilon_t$  是噪声项或误差项，通常假设误差服从正态分布。

#### 1. 趋势函数

趋势项是 Prophet 模型的核心，常见的趋势项有两种趋势：线性趋势和逻辑增长趋势。

（1）线性趋势用于随时间呈现线性变化的情况。线性趋势的公式为：

$$g(t) = (k + \delta(t))t + (m + \gamma(t)) \quad (2)$$

$k$  是线性增长速率， $\delta(t)$  是变化点在特定时间点发生速率变化的调整项， $m$  是截距， $\gamma(t)$  是截距变化的调整项。

（2）Logistic 增长趋势用于增长有上限的情况。logistic 增长趋势的公式为：

$$g(t) = \frac{C}{1 + \exp(-k(t - m))} \quad (3)$$

其中， $C$  表示函数的最大渐进值， $k$  表示增长率， $m$  表示曲线的中点。

#### 2. 季节性函数

季节性函数负责捕捉周期性波动。基于傅里叶级数来近似表达季节性变化。傅里叶级数的优势在于可以有效表达任何周期性波动，不管波动是简单的正弦曲线还是更复杂的周期性变化。其公式为：

$$s(t) = \sum_{n=1}^N \left( a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right) \quad (4)$$

$P$  是季节性的周期长度。 $a_n$  和  $b_n$  是傅里叶系数， $N$  是傅里叶级数的阶数。

#### 3. 节假日效应

$$h(t) = Z(t)\kappa = \sum_{i=1}^L \kappa_i \cdot I_{\{t \in D_i\}} \quad (5)$$

其中  $L$  表示节假日的个数， $\kappa_i$  表示对应节假日的影响范围， $I_{\{t \in D_i\}}$  这是一个示性函数，用于判断在时间  $t$  时，是否受到第  $i$  个节假日的影响。

#### 4. 残差项

$\varepsilon_t$  是误差项，表示模型和实际数据之间的差异。通常假设误差服从正态分布，即  $\varepsilon_t \sim N(0, \sigma^2)$  其中  $\sigma$  是标准差。Prophet 在拟合模型时，使用最大似然估计方法来找到最优参数，最小化模型的误差。

模型的具体工作原理图 1 如示。



图 1: Prophet 模型工作原理图

#### (二) Prophet-机器学习模型构建

Prophet 作为时间序列预测的常用工具，能快速把握数据的基本趋势和季节性变化，但是当数据呈现复杂的非线性特征或受到突发异常事件影响时预测结果往往捕捉不到这些信息，而机器学习模型有着强大的自适应和特征学习能力，能有效解决 Prophet 在复杂场景下的局限性问题，从而提高预测的准确性和稳定性。

本研究利用机器学习算法具有处理非线性等复杂关系的优势，采取了三种不同的机器学习算法对 Prophet 模型进行优化，弥补了 Prophet 模型的局限性，具体的步骤如图 2 所示：

第一步：将时间序列数据进行划分，将前 5 年的数据用 Prophet 模型进行拟合，得到未来 1 年的预测值，并计算预测数据的拟合偏差；

第二步：利用 Prophet 模型分解出来的趋势项、周期项、节假日项相加得到预测值，记为  $prediction_1$ ；

第三步：将 Prophet 模型的拟合偏差分别输入 3 个机器学习模型中，将残差转化成滞后特征进行预测，记为  $prediction_2$ ；

第四步：将预测值  $prediction_1$  和  $prediction_2$  利用网格搜索法，以 MAE 为评价指标进行网格搜索，确定最佳的权重系数进行组合相加，然后得到最终的预测值。

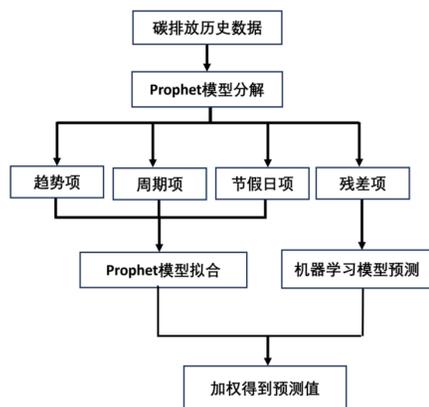


图 2: Prophet-机器学习模型构建流程

#### (三) 评价指标

本文使用平均绝对误差 MAE、均方误差 MSE、均方根误差 RMSE、平均绝对百分比误差 MAPE 四个指标作为模型与预测精度的评价指标，其计算原理如下

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

其中， $y_i$  和  $\hat{y}_i$  分别表示碳排放量的真实值与预测值。

## 二、数据获取与预处理

#### (一) 数据获取

本文数据来自于全球实时碳监测数据平台，选取中国 2019 年 1 月 1 日至 2024 年 12 月 31 日的每日二氧化碳排放量（单位：百万吨，Mt）。数据包含 Date（日期）和碳（Mt）（排放量），总共 2192 条记录。

#### (二) 数据预处理

##### 1. 数据检查与清洗

通过对数据的检查，发现时间序列连续完整，无日期断裂。原始数据中日期字段被标准化为 YYYY-MM-DD 格式，数据的时间间隔为连续的自然日，数据单位为“百万吨（Mt）”。

##### 2. 格式检查

Prophet 模型对数据变量名称有规定，时间的数据列的变量必须命名为 ds，其因变量碳排放量必须为 y。

##### 3. 数据划分

将 2019-1-1 到 2024-12-31 的数据分割成训练集和测试集，将前五年的数据作为训练集去训练模型，最后一年的数据作为验证集。

## 三、实证分析

#### (一) 基于 Prophet 模型的碳排放预测

本文将 2019-2023 年的数据作为训练集，使用 Prophet 模型对训练集的数据进行拟合，使其学习数据中的趋势、季节项等，并用 2024 年的数据进行测试，以此评估模型在新数据上的预测能力。模型的拟合结果（如图 3）所示，蓝色线代表的是真实值，黄

色线代表的是模型在训练集上的拟合值，通过拟合结果可以得出该模型在训练集上的拟合效果较好，基本能够掌握碳排放的波动趋势。

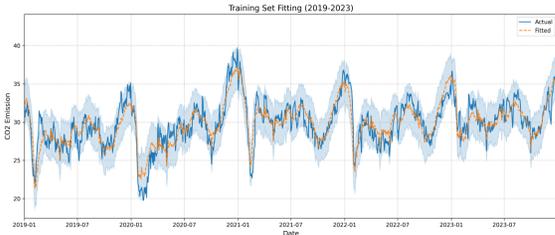


图3: Prophet 模型训练集拟合结果

接下来，将拟合好的Prophet 模型用于预测未来一年的数据，蓝色实线代表的是真实值，黄色虚线代表的是模型预测值，预测结果与实际值的对比如图4所示：

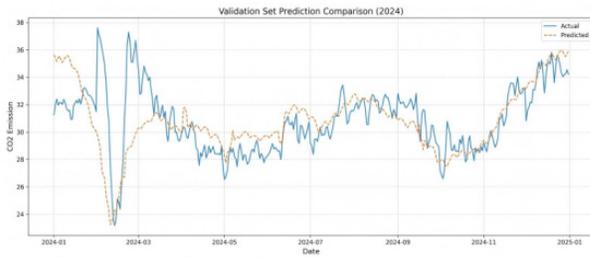


图4: Prophet 模型预测结果对比图

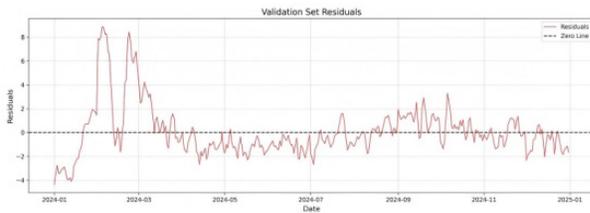


图5: Prophet 模型预测残差图

通过计算 Prophet 模型预测的残差，绘制了残差分析图（图5），并通过计算 MSE、RMSE、MAE、MAPE 四个预测偏差的指标来评估预测效果。如表1所示，从预测结果来看，模型在一定程度上能够把握大致的预测情况，但预测精度还存在偏差，比如2024年的2月到3月期间的预测偏差最大，通过调查发现这个期间中国开始全面推进疫情防控后的经济复苏，能源的需求不断增加，这导致了2月份初期排放量爆发式增长，而3月份太阳能与风能发电的不断扩大，以及建筑业等活动下降，中国在设法向清洁能源经济转型，因此碳的排放量又跌至最低点，由此可见该模型对于复杂的外部动态因素的处理能力不足。

表1: 预测偏差评价指标

评价指标	MSE	RMSE	MAE	MAPE	$R^2$
Prophet	1.461793	4.744903	2.17828	4.617975	0.3362

**(二) 不同模型的预测结果分析对比**

现将 Prophet、ARIMA 模型与基于贝叶斯时序模型与机器学习算法组合的预测结果进行对比分析，如表2、图6所示。

由2024年的预测结果可以看到，ARIMA 模型中各项偏差指标均为最高，这反映其预测值的误差比较大，这是因为该模型通

过对不平稳的原始数据进行差分处理来实现平稳这个假设，其在短期预测中的效果会比较好，而对长期趋势的预测来看，各种外界因素的动态变化会打破这个平稳性，因此效果并不好；相比于ARIMA，不难发现 Prophet 模型考虑趋势项、周期性、季节项的因素，可以得出 MAE、MSE、RMSE、MAPE 四个偏差指标分别减少了0.427107、1.054397、0.22992、1.472025，整体上来看该模型的预测效果优于 ARIMA 模型，但是 Prophet 模型同样无法处理复杂的非线性因素；鉴于机器学习算法能够处理复杂的非线性关系，本文通过引入机器学习算法拟合那些不可解释的残差，捕捉其 Prophet 模型未能预测的那部分误差，我们发现使用机器学习算法与 Prophet 模型相结合的组合优化算法能明显的提高预测准确性，其中 LightGBM 优化效果最好，MAE、MSE、RMSE、MAPE 四个偏差指标分别减少了0.759228、3.599075、1.107846、2.38763，除此之外，其决定系数  $R^2$  也提高了0.3918。

表2: 偏差评价指标

评价指标	MSE	RMSE	MAE	MAPE	$R^2$
ARIMA	1.8889	5.7993	2.4082	6.09	0.0372
Prophet	1.461793	4.744903	2.17828	4.617975	0.3362
Prophet+LGB	0.702565	1.145828	1.070434	2.230345	0.7280
Prophet+RF	0.957549	2.062008	1.435969	3.029652	0.7110
Prophet+XGB	0.752603	1.568132	1.252251	2.372883	0.7381

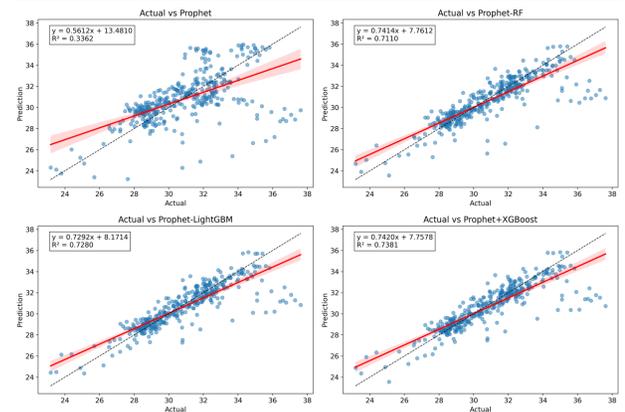


图6: 各模型散点图与回归直线图

**四、结论与展望**

**(一) 结论**

本研究基于2019-2024年的碳排放量数据建立了 Prophet 和机器学习混合模型，并加入 ARIMA 等模型进行对比，最终得出以下结论：

1. 相比于 ARIMA，Prophet 模型在预测二氧化碳排放时能有效捕捉变化趋势和季节性规律，对于预测长期的趋势具有一定的拟合效果，其 MAE、MSE、RMSE、MAPE 四个偏差指标分别减少了0.427107、1.054397、0.22992、1.472025。但是该模型对数据的假设非常理想化，因而面对实际的复杂的非线性特征或突发事件时存在局限性。

2. 将 RandomForest、XGBoost 和 LightGBM 等机器学习算

法与 Prophet 模型结合, 通过网格搜索法确定超参数和加权系数来优化 Prophet 模型, 通过对比 MAE、MSE 等偏差指标, 不难发现 Prophet+ 机器学习的组合预测模型明显比单一 Prophet 模型要好, 其中 Prophet+LightGBM 在优化残差、提高预测准确性方面表现突出, 四个偏差指标分别减少了 0.759228、3.599075、1.107846、2.38763, 决定系数  $R^2$  提高了 0.3918。

## (二) 展望

本文所构建的 Prophet 和机器学习混合模型在预测任务中取

得了一定的效果, 具有显著的潜力, 但需要注意的是, 该模型更擅长捕捉时间序列长期的趋势, 对于短期的动态波动以及外部复杂的考量存在不足, 为进一步提升模型的预测准确性与普遍适用性, 未来可将极端天气、疫情等突发状况作为外部特征加入到机器学习中作为特征项, 通过这些特征与 Prophet 模型分解的成分相结合进行混合建模, 继续优化该模型的预测能力。

## 参考文献

- [1] 余航. 基于 ARIMA 模型的我国碳排放量的预测 [J]. 经济师, 2018, (06): 59-60.
- [2] 周树涛, 邱畅, 姚均棠等. 基于 ARIMA 模型的碳排放预测及减排潜力 [J]. 中国新技术新产品, 2023, (02): 130-132.
- [3] 钟进, 李宗航. 基于趋势性时间序列的全国碳排放量预测研究 [J]. 运筹与模糊学, 2023, 13(4): 3870-3881.
- [4] Power Demand Forecasting Using ANN and Prophet Models for the Load Despatch Center in Andhra Pradesh, India [J]. International Journal of Soft Computing and Engineering (IJSCE), 2024, 14(1/).
- [5] Jia C. Prediction method of product market demand based on Prophet random forest [J]. International Journal of Product Development, 2024, 28(1-2): 60-72.
- [6] Su S, Sun Y, Zeng Y, et al. Aviation risk prediction based on Prophet - LSTM hybrid algorithm. Aircraft Engineering and Aerospace Technology, 2023, 95(7): 1054-1061.
- [7] 张学清, 李芳, 张绚, 等. 基于 CNN-LSTM 模型的中国碳排放量实时预测研究 [J]. 中阿科技论坛 (中英文), 2022, (10): 71-75.
- [8] 王琦, 赵静, 胡磊, 等. 基于机器学习 LSTM 网络的  $SO_2$  排放浓度预测 [J]. 自动化与仪表, 2021, 36(07): 77-80+85.
- [9] 赵成柏, 毛春梅. 基于 ARIMA 和 BP 神经网络组合模型的我国碳排放强度预测 [J]. 长江流域资源与环境, 2012, 21(06): 665-671.
- [10] 常恬君, 过仲阳, 徐丽丽. 基于 Prophet- 随机森林优化模型的空气质量指数规模预测 [J]. 环境污染与防治, 2019, 41(07): 758-761+766.
- [11] 高洁如, 魏霖静, 李玥, 等. 基于 Prophet-LightGBM 的  $PM_{2.5}$  浓度预测模型 [J]. 软件导刊, 2024, 23(07): 144-152.
- [12] 彭湃, 刘敏. 基于 Prophet-LSTM 组合模型的短期负荷预测方法 [J]. 电力系统及其自动化学报, 2021, 33(11): 15-20.
- [13] 郑陈达, 兰雁宁. 基于 XGboost 和 Prophet 组合模型的天然气需求预测 [J]. 城市燃气, 2024, (12): 32-36.