作品编号:TJJM20250228008100

2025 年（第十一届）全国大学生统计建模大赛

参 赛 作 品

参赛学校：	哈尔滨工业大学
论文题目：	黑龙江省综合碳排放评估与预测——基于熵权-TOPSIS-耦合协调度评估与 BO 优化的随机森林算法研究
参赛队员：	肖家伟 朱俊彰 胡伟豪
指导老师：	周永春

黑龙江省综合碳排放评估与预测——基于熵权-TOPSIS-耦合协调度评估与 BO 优化的随机森林算法研究

摘　要

在全球气候治理加速推进与中国“2030 碳达峰、2060 碳中和”战略背景下，区域碳排放的精准评估与路径优化成为实现低碳转型的核心命题。黑龙江省作为我国重要的老工业基地与农业大省，其碳排放呈现“总量趋稳但结构矛盾突出”的特征：重工业占比高、冬季严寒导致供暖能耗显著、可再生能源（如风能、光伏）受季节性波动制约，现有宏观模型难以直接适配。然而，针对东北地区碳排放系统性评估与预测的研究仍存在显著空白，亟需结合区域禀赋构建科学工具。

本文首先以黑龙江省为对象，整合国家统计局、碳核算数据库及黑龙江省统计年鉴的 2014—2024 年的 11*23 维度面板数据，提出“评估-预测-优化”三位一体的研究框架。接着引入三次样条插值法补全缺失数据，增强模型鲁棒性。然后创新设计了“熵权-TOPSIS-耦合协调度”多层次综合碳排放评估模型，从碳排放、经济、工业、农业、人居与居民生活六大维度构建包含 23 项指标的指标体系，量化区域发展与碳排放的动态协同效应。最后针对小样本、非线性数据特性，提出贝叶斯超参数优化的随机森林算法（PURE-RF）解决了小样本数据场景的预测难题，通过高斯过程代理模型（GP）与期望提升函数协同寻优。

研究结果显示，黑龙江省综合碳排放评估值呈现“先降后升再波动”的非线性趋势，揭示了传统高耗能路径存在系统脆弱性。模型优化方面，BO 优化方法效果显著，PURE-RF 预测精度较 RAW-RF 方法有较大提升（R^2 均值从 0.8104 增值 0.9982，增长 81.7%），验证了其在复杂非线性数据中的强泛化能力。基于以上结果，本文建议黑龙江省加速工业低碳转型，同时增强系统韧性，进一步完善动态监测与政策适配。此外，后续研究可扩展数据维度，并探索多模型融合，加强韧性路径设计，为黑龙江省绿色转型提供更多科学手段。

关键词 碳排放评估；熵权-TOPSIS；贝叶斯优化；随机森林；耦合协调度；

目录

表格与插图清单

一、绪论

（一）研究背景

气候变化已成为全球可持续发展的重要挑战。2016 年，全球 178 个缔约方共同签署的《巴黎协定》提出的温控目标要求各国加速低碳转型。中国在 2020 年明确提出了“2030 年前碳达峰、2060 年前碳中和”的战略目标。尤其在 2021 年两会，“碳达峰”“碳中和”被首次写进政府工作报告；会议上也将“加快推动绿色低碳发展”列入“十四五”规划中。碳中和是贯彻新发展理念，构建新发展格局，进一步推进产业结构转型，使我国走上以创新为驱动的绿色、低碳、循环的发展路径，实现高质量发展的必经之路。

近年来（2021-2024 年），黑龙江省碳排放呈现总量趋稳定减少，但结构性矛盾突出的特征。通过高耗能行业技术改造和清洁能源推广等措施，2016-2019 年全省碳排放累计下降 18.7%；2020 年黑龙江省二氧化碳排放总量为 2.7873 亿吨，人均排放量 8.79 吨，均处于全国中等水平。但我国能源的需求会持续增长以满足经济的合理增长，因此我国实现碳达峰碳中和时间紧迫并且任务繁重。黑龙江省自 2021 年起强化政策引导，根据《黑龙江省“十四五”节能减排综合工作实施方案》，通过十大重点工程推动产业绿色升级。根据黑龙江统计年鉴数据，黑龙江省能源消费 2021 年来呈下降趋势，但仍然保持在较高水平。

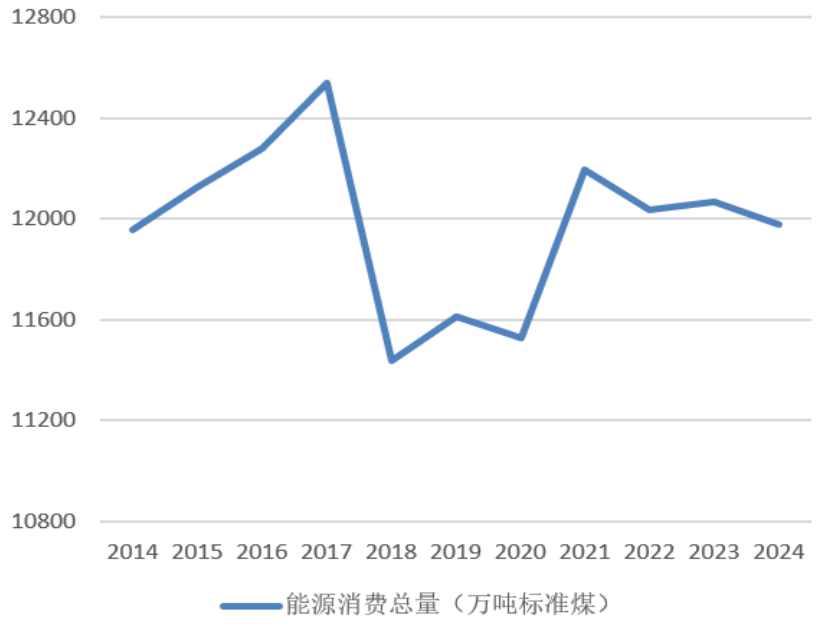


图 1 黑龙江省 2014-2024 年能源消费总量变化曲线

现有研究多聚焦国家或省级尺度，专门针对东北地区的综合碳排放评估和预

测研究存在显著空白。黑龙江省碳排放结构性矛盾依然突出。黑龙江省冬季严寒导致供暖能耗高，可再生能源（如风能、光伏）的季节性波动显著，现有碳排放评估、路径优化模型难以直接套用。

因此本文立足黑龙江省相关碳排放及关联指标统计数据，基于 Topsis-熵权法和耦合度评估方法构建多层次碳排放评估模型，并搭建基于贝叶斯超参数优化的随机森林预测算法，实现对黑龙江省综合碳排放预测。

（二）研究意义

在全球气候治理加速推进的背景下，科学评估与调控碳排放已成为构建可持续发展模式的核心命题。同时，我国区域经济版图中能源禀赋与产业基础的差异化客观特征，要求北、东、中、西部在产业梯度转移、能源结构优化及生态补偿机制等方面形成合理的动态平衡。在上述背景下，寻找因地制宜且多维度的碳排放评估办法及相应的综合碳排放预测方法对于实现 “双碳目标”尤为重要。

本文基于区域具体禀赋设计碳排放评估方法，有助于补充现有评估碳排放方法框架，为政府与企业提供科学决策框架；同时对于碳排放预测的相关研究也对于推动产业链低碳化具有借鉴价值；对于人口分布、经济情况评估、污染物估计等社会领域指标的评估、预测也有一定借鉴意义。

（三）研究思路和技术路线

本文整体的研究思路和技术路线如图 2 所示，在此结构支撑下，逐步展开针对黑龙江省多层次碳排放评估模型及碳排放预测模型的相关研究。

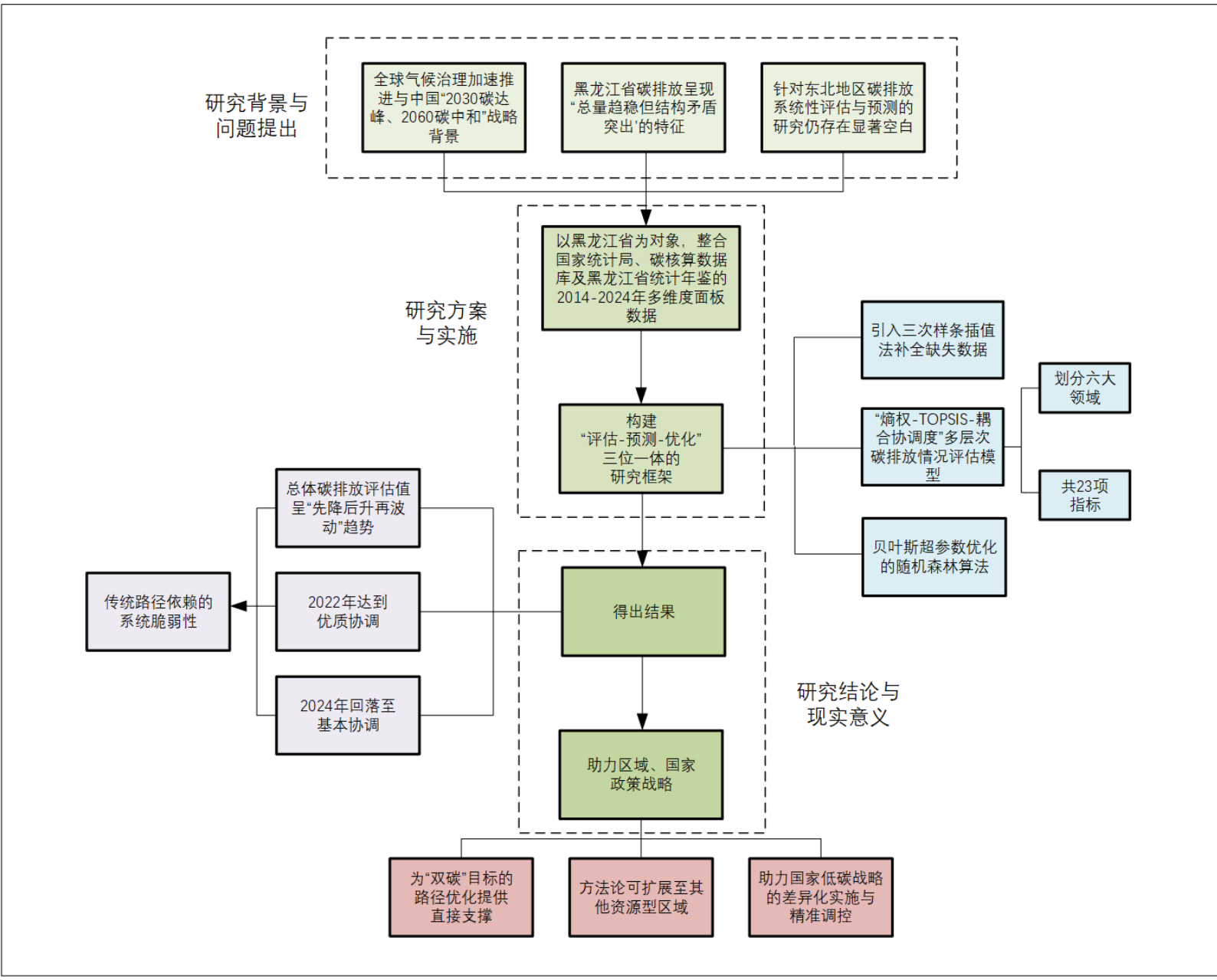


图 2 本文研究思路和技术路线

二、文献综述

（一）综合碳排放评估方法的相关研究文献

学界目前综合碳排放评估主要采用的是经典方式，即利用数据集进行分析得出碳排放量化值。针对碳排放评估过程中的碳排放影响因素，学界常见因素分解模型有 LMDI 分解法(Logarithmic Mean Disivia Index)、Kaya 公式、IPAT 模型、STIRPAT 模型、拉斯尔斯指数法等。Lin 等(2015)选取经济增长、能源消耗、研发、金融发展、外商直接投资、贸易开放、工业化和城市化作为输入变量，利用神经网络研究这 9 个对碳排放强度的影响。Xiong 等(2016) 采用对数均值除

指数分解方法，将农业碳排放分解为效率因子、结构因子、经济因子和劳动因子，具体研究了新疆不同阶段农业碳排放变化及其影响因素。Dong 等(2016)选择城镇化(城市人口/总人口，URB)和能源结构(煤炭消费/能源消费，EM)作为关键自变量进行了相关研究。

（二）碳排放预测的相关研究文献

目前，学者主要采用情景分析方法、智能算法等对碳排放情况进行预测。情景分析法是指在合理假设的基础上，对可能的未来情景加以描述，同时将一些有关联的单独预测集合形成总体的预测方法。赵亚涛等(2018)运用情景分析法分析煤电行业 CO,排放量、燃煤发电比重以及 CO_2,排放强度之间的关系。CB.Wu 等(2018)发现 STIRPAT 模型可以用于青岛市未来 CO_2排放量的预测，并采用情景分析法获取了青岛市 2015-2030 年期间的 CO_2排放量 。韩楠等(2022)构建碳排放系统动力学模型，并设置六种情景方案，模拟其对京津碳达峰时间、峰值以及减排潜力的影响。

影响综合碳排放的元素体系是一个复杂、多层次的非线性结构，随着计算水平的不断提高，包括神经网络、粒子群优化等在内的智能算法越来越多地被应用于碳排放预测。Junhong Hao 等(2022)基于过去 40 年中国的碳排放数据建立了碳排放动力学模型在多情景下预测了未来 40 年中国的碳排放量。Feng Ren 等(2021)构建了一种经过鸡群优化改进的快速学习网络预测方法，预测 2020-2060 年的碳排放量，并在模拟的 9 种情景中探索广东省碳达峰碳中和的路径。张国兴等(2020)运用对数平均权重分解法对各项驱动因素进行分解以构建拓展STIRPAT模型预测不同情境下的碳排放趋势 。

（三）综述总结

虽然目前已经存在通过结合碳排放因素以评估、预测碳排放的相关研究，但现有研究多聚焦宏观层面或特定地区，缺乏对区域间异质性（如资源禀赋、产业结构）的系统考量和对于地区本身的发展特点与客观环境关系的研究，尤其是针对黑龙江省的碳排放评估与预测的相关研究仍有不足。

因此，本文设计了“熵权-TOPSIS-耦合协调度”多层次综合碳排放评估模型针对关于黑龙江省综合碳排放进行评估，并采用基于贝叶斯超参数优化的随机森林算法，对黑龙江省综合碳排放进行预测。

三、指标体系搭建

本文借鉴现有研究成果，遵循指标选取的基本原则，以黑龙江省区域发展特点及整体发展现状为基础，构建包括碳排放、经济发展、工业发展、农业发展、人居环境及居民生活六类指标的评价指标体系，全面反映黑龙江省碳排放水平与区域发展情况。

最终统计指标体系如表 1 所示。

表 1 黑龙江省统计指标体系

一级指标	二级指标	单位	指标属性
碳排放	能源消费总量	万吨标准煤	+
	表观总碳排放	万吨	+
	单位 GDP 碳排放强度	吨/万元	-
经济发展	GDP 总量	万亿元	+
	人均 GDP	万元	+
	第三产业占比	百分比%	+
工业发展	规模以上工业总资产增加值	亿元	+
	工业产值增加量	亿元	+
	工业能源消耗	万吨标准煤	-
	单位工业产值能耗	吨标准煤/万元	-
	焦炭产量	万吨	+
农业发展	农业总产值	亿元	+
	耕地面积	千公顷	+
	单位耕地化肥施用量	公斤/公顷	-
	水利种植畜牧碳排放	万吨	-
	农业机械总动力	万千瓦	+
人居环境	造林总面积	千公顷	+
	人均公园绿地面积	平方米/人	+
	建成区绿地覆盖率	百分比%	+
居民生活	居民人均可支配收入	元	+
	电力蒸汽热水供应碳排放	万吨	-
	居民生活用电量	亿千瓦时	+
	民用汽车保有量	万辆	+

四、黑龙江省各面板数据的预处理

（一）三次样条插值法数据补全

搜集所得数据中由于统计年限等原因，存在某些指标的数据缺失（约占总数据 5%）。本文结合实际数据特点，采用三次样条插值（Cubic Spline Interpolation，Spline）方法进行缺值补全。具体通过scipy.interpolate库中的CubicSpline函数实现计算a_i，b_i，c_i，d_i的值。处理流程图如图 3 所示。

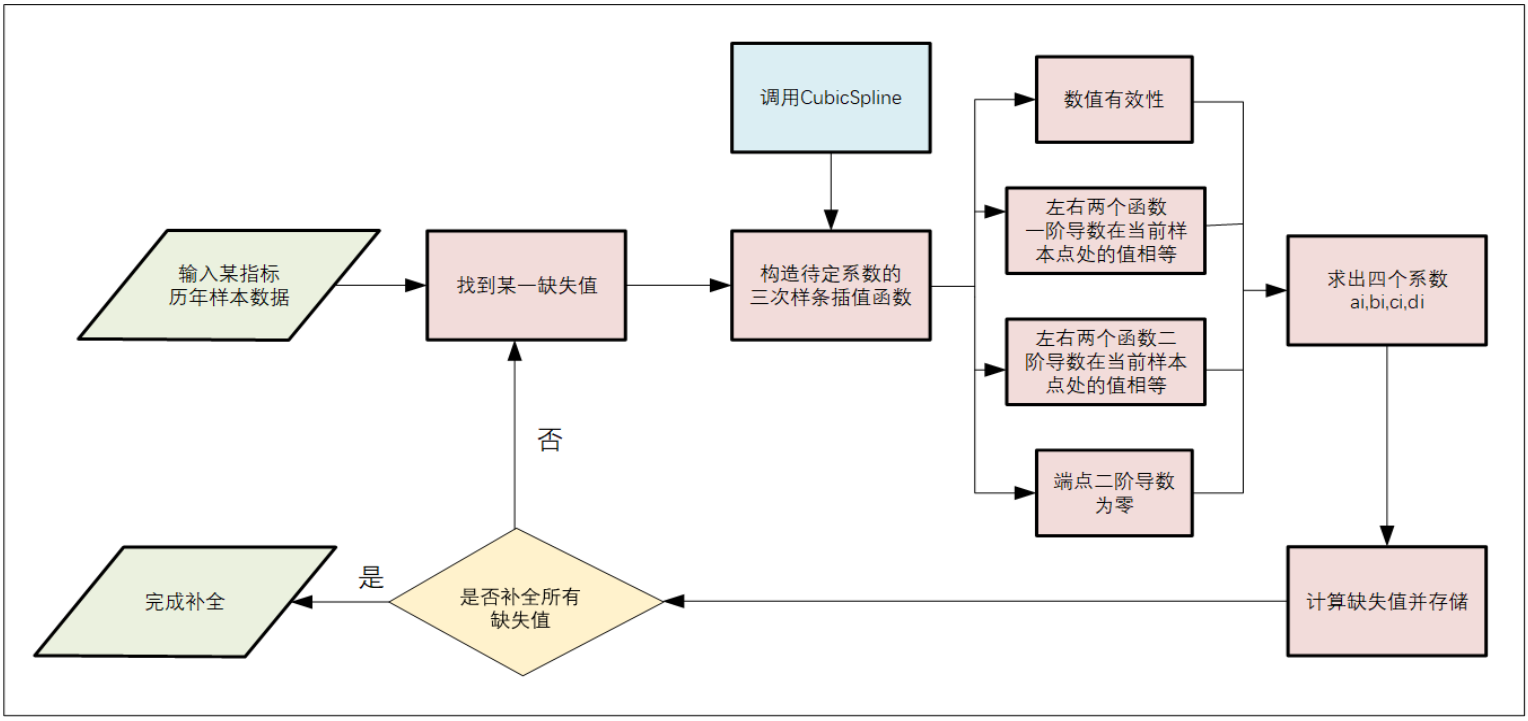


图 3 三次样条插值补全数据流程图

以样本年份t_i和数据x_i构建i个二元实数对(t_i, x_i)。

定义在$[t_i, t_{i+1}]$上的三次样条插值函数为：

$$S_i(t) = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3 \qquad (1.1)$$

其中t为自变量，n为该列有效时间点总数。

可见相比于传统 Lagrange 插值方法，Spline 插值方法通过限制多项式次数的方法，克服了增加多项式次数过程中舍入误差易扩散的缺点，避免了龙格现象(Runge phenomenon)。

区间$[t_i, t_{i+1}]$内有 4 个未知数a_i，b_i，c_i，d_i，至少需要四个约束条件进行求解，具体如下：

在每个有效时间点t_i处，需满足数值有效性：

$$S_i(t_i) = x_{ij} \quad (1.2)$$

在相邻区间$[t_i, t_{i+1}]$和$[t_{i+1}, t_{i+2}]$的交界点t_{i+1}处，要求左右两个三次函数的一、二阶导数在当前样本点处的值相等：

$$S_i'(t_{i+1}) = S_{i+1}'(t_{i+1}) \quad (1.3)$$

$$S_i''(t_{i+1}) = S_{i+1}''(t_{i+1}) \quad (1.4)$$

在整个区间的端点t_1和t_n处，要求满足自然边界条件（Natural boundary conditions），即端点二阶导数为0：

$$S_1''(t_1) = S_{n-1}''(t_n) = 0 \quad (1.5)$$

其中$i \in [1, n-1]$，n为有效时间点总数。

经约束，$S(t)$在整个区间$[t_0, t_n]$上为平滑三次函数连接而成，故可以取缺失数据对应时间点t_{void}的函数值$S(t_{void})$补充缺失值。

（二）数据标准化

原始数据矩阵$X = (x_{ij})_{n\times m}$（n为样本数，m为指标数），采用极差法消除量纲差异得到标准化后的数据矩阵$X' = (x'_{ij})_{n\times m}$，满足$x'_{ij} \in [0,1]$，为后续熵权-TOPSIS 算法提供同向数据基础。

定义极差 (Range)：

$$max(x_j) - min(x_j) \quad (1.6)$$

极差以R表示，用来表示统计数据中的变异量数(Measures of variation)。

标准化处理，对于正向指标（效益型）：

$$x'_{ij} = \frac{x_{ij} - min(x_j)}{R} \quad (1.7)$$

对于负向指标（成本型）：

$$x'_{ij} = \frac{max(x_j) - x_{ij}}{R} \quad (1.8)$$

五、“熵权-TOPSIS-耦合协调度”多层次综合碳排放评估模型

（一）模型原理概述

本文采用熵权-TOPSIS 算法，对于黑龙江省各指标数据通过客观赋权与空间距离测度的方法实现多维指标的量化计算，进行两级指标处理和耦合度、协调度分析。整体原理如图 4 所示。

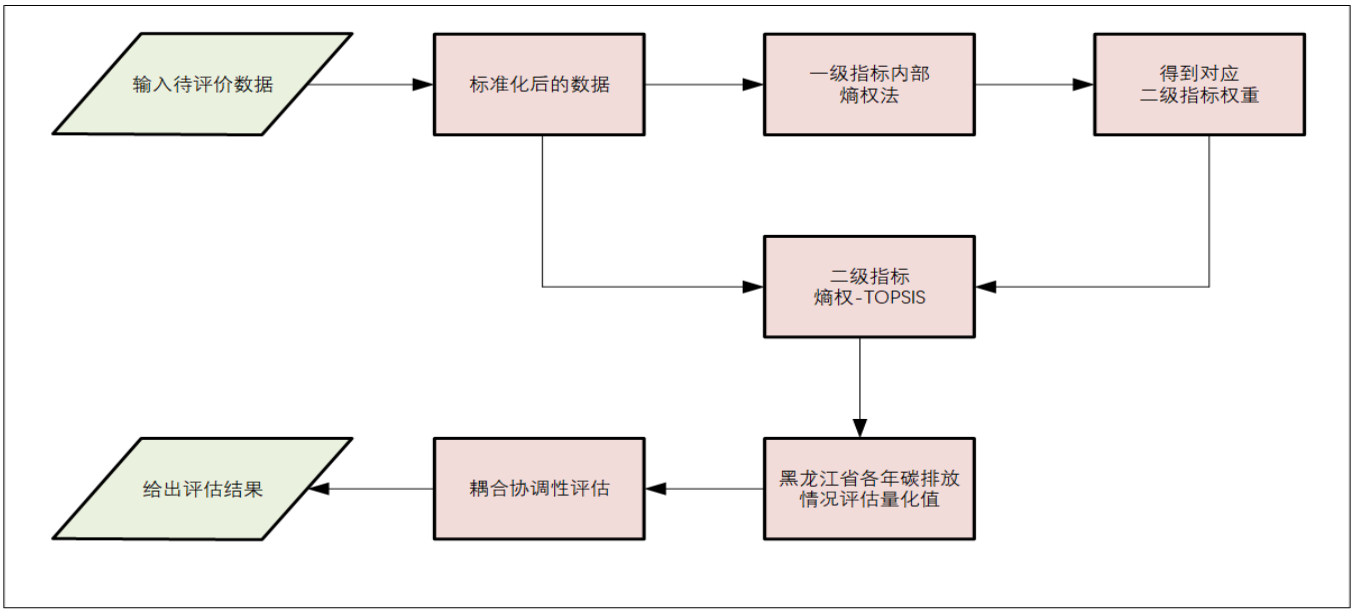


图 4 多层次碳排放评估模型原理图

（二）多层次碳排放评估模型构建

1．熵权法确定指标权重

熵权法是一种客观评价方法，不同于层次分析法等主观综合评价方法，是以各个指标的信息熵为媒介，计算各个指标相对变化程度而衡量其对整体的影响，从而决定权重的评估算法。

确定指标j权重具体步骤如下：

第一步：数据$x_{ij}^{'}$比例化处理为比重p_{ij}：

$$p_{ij}=\frac{x_{ij}^{'}}{\sum_{i=1}^{n}x_{ij}^{'}} \tag{2.1}$$

p_{ij}相对大小反映指标j的相对贡献度大小。

第二步：计算信息熵e_j：

$$e_j=-\frac{1}{ln\,n}\sum_{i=1}^{n}p_{ij}\,ln(p_{ij}) \tag{2.2}$$

e_j越大，说明指标j信息量越小。n为计算指标数目，对于第一次熵权法计算，n为某一级指标下对应二级指标数目；对于第二次熵权法计算，n=6，即一级指标数目。

第三步：计算差异系数d_j：

$$d_j = 1 - e_j \tag{2.3}$$

d_j越大，代表着该指标值的差异越大，对方案评价的影响就越大，即指标越重要。

第四步：差异系数d_j归一化得到权重w_j：

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j} \tag{2.3}$$

2.基于熵权法得到的权重进行 TOPSIS 评价

首先，基于标准化数据矩阵X'和各指标权重w_j构建加权决策矩阵V：

$$V = X' * W_{diag} = \left[v_{ij}\right]_{n\times m} \tag{2.4}$$

$$v_{ij} = x'_{ij} * w_j \tag{2.5}$$

接着，确定正负理想解：

$$V_j^+ = \max_i\left(v_{ij}\right) \tag{2.6}$$

$$V_j^- = \min_i\left(v_{ij}\right) \tag{2.7}$$

其中正理想解V_j^+代表各指标的最优值组合，负理想解V_j^-代表各个指标的最劣组合。

再计算每组样本i距离正负理想解的欧式距离测度D_i^+和D_i^-：

$$D_i^+ = \sqrt{\sum_{j=1}^{m}\left(v_{ij} - V_j^+\right)^2} \tag{2.8}$$

$$D_i^- = \sqrt{\sum_{j=1}^{m}\left(v_{ij} - V_j^-\right)^2} \tag{2.8}$$

最终综合D_i^+和D_i^-评估样本i相对接近度C_i：

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{2.9}$$

其中，D_i^+数值大小与若$D_i^+ = 0$，则为正理想解；若$D_i^- = 0$，则为负理想解。

3.耦合协调度（Coupling coordination degree）评估

本文借鉴物理学中的容量耦合（Capacity coupling）及容量耦合系数模型

（Capacity coupling coefficient model），推广得到多个系统(或要素)相互作用耦合度C的概念。

基于两次熵权-TOPSIS 方法计算得到的评估矩阵进行耦合协调性分析,逐年计算黑龙江省各个一级指标S_{ij} $(\mathrm{i}=1,2,\ldots,\mathrm{n})$的耦合度$C$，量化评估黑龙江省碳排放与区域整体发展情况关联程度。

以某一年的评估矩阵作为一个样本，计算第j年样本$[S_{1j},S_{2j},\ldots,S_{nj}]$的耦合度$C_j$：

$$C_j = n\frac{(\prod S_{ij})^{\frac{1}{n}}}{\sum S_{ij}} \tag{2.10}$$

其中S_{ij}为第j年第i个一级指标的评估值；$n=6$。且易得耦合度值$C_j \in [0,1]$。$C_j=1$时耦合度最大，内部要素之间达到良性耦合。

然而，耦合度不能完全反映区域发展与碳排放情况的协同效应，特别是在黑龙江省多个一级指标共同研究的情况下，因每个一级指标在时间尺度上均有其交错、动态和不平衡的特性，单纯依靠耦合度判别易产生评估误差。

故进而计算第j年的协调度D_j，计算公式如下：

$$T_j = \sum w_i S_{ij} \tag{2.11}$$

$$D_j = \sqrt{C_j T_j} \tag{2.12}$$

其中T_j为各一级指标的综合指数，反映碳排放与其余五个一级指标的整体协同效应；w_i为第i个一级指标的权重。本模型取每个指标平权，故(2.11)化简为：

$$T_j = \frac{\sum S_{ij}}{n} \tag{2.13}$$

实际运算中，存在极端数据$S_{ij}\approx 0$引起C_j和D_j趋于0，无法恰当反映正常得分领域的耦合情况。故引入微小噪声$\varepsilon=10^{-6}$，用$S'_{ij}=S_{ij}+\varepsilon$代替$S_{ij}$进行计算。

D_j的取值落在区间$[0,1]$中，根据D_j的取值对该年碳排放与区域整体发展协调程度进行评价，结合相关文献给出参照表如下：

表 2 D_j取值对应协调情况参照表

协调情况	严重失调	轻度失调	基本协调	优质协调
D_j取值	[0,0.3)	[0.3,0.5)	[0.5,0.8)	[0.8,1]

六、随机森林模型

随机森林模型（Random Forests，RF）通过集成多棵决策树，以双重随机性（数据自助采样、特征随机抽取）有效降低过拟合风险，提升了模型泛化能力。本文构建了多个针对指标体系中各指标进行预测的 RF 模型。

（一）数据分集初处理

1.原始训练集与测试集划分

将归一化后的数据矩阵$X'_{ij}=\{x'_{ij}\}$做初等列变换得到原始数据集$\mathcal{D}=\{(X_i,y_i)\}_{i=1}^{N}$。

其中X_i是第i个样本的指标行向量，y_i是该样本所在行的碳排放量。

将$\mathcal{D}$按行以10:1比例划分成训练集$\mathcal{D}_{train}$和测试集$\mathcal{D}_{test}$。训练集样本数$N_{train}=\left\lfloor \frac{10}{11}N \right\rfloor$。测试集样本数 $N_{\text{test}}=N-N_{\text{train}}$。

2. 留一验证（LOOCV，Leave one out cross validation）划分

LOOCV 方法从可用的数据集中保留一个数据点，根据其余数据训练模型。对每个数据点进行迭代。例如数据集中有 n 个数据点，就要重复交叉验证 n 次。

图 5 n=10 时 LOOCV 划分原理图

为充分验证模型对历史数据的泛化能力，本文采用 LOOCV 进一步划分训练集与测试集。

对每一年$t\in\{2014,2015,\ldots,2024\}$，执行 LOOCV 操作：

$$\mathcal{D}^{(t)}{}_{train}=\{(X_i,y_i)|i\neq t\} \quad (3.1)$$

$$\mathcal{D}^{(t)}{}_{test}=\{(X_i,y_i)|i=t\} \quad (3.2)$$

利用 LOOCV 操作，保留一个样本数据集，取出$\mathcal{D}^{(t)}{}_{test}$；使用$\mathcal{D}^{(t)}{}_{train}$训练模型；完成模型后，在$\mathcal{D}^{(t)}{}_{test}$中测试；如果模型在验证数据上提供了一个肯定的结果，那么继续使用当前的模型。

实际操作中，经贝叶斯超参数优化（BO 优化）完毕的某 PURE-RF 模型，用 LOOCV 划分进一步训练 11 次。其中，每次固定测试集为单一年份数据，其余年份为训练集。

（二）单一决策树（Decision stump）构建

本节阐述单个回归决策树（以第b棵树为例）详细生成步骤：

1.节点纯度计算

本文通过从第b棵树的根节点递归，逐个计算节点MSE反映节点纯度。

对于第t个节点：

$$MSE(t) = \frac{1}{N_t}\sum_{i\in t}(y_i - \bar{y_t})^2 \tag{3.3}$$

$$\bar{y_t} = \frac{1}{N_t}\sum_{i\in t} y_i \tag{3.4}$$

其中N_t为第t个节点的样本总数。

2.分裂阈值候选

为避免过拟合效应，每个节点只从全部m个特征中随机选取m_{try}个候选特征，且 $m_{try} = \lfloor \sqrt{m} \rfloor$ ，构成候选集F。仅在该子集内评估分裂点以避免特征主导效应。

对于每个特征f，将其在节点样本中的取值去重并排序，得到唯一值序列$\{x_{i_1 f}, x_{i_2 f}, \ldots, x_{i_k f}\}$（$k$为第$b$棵树的样本数）。

计算相邻值的算术平均作为候选的分裂阈值：

$$\tau_j = \frac{x_{i_j f} + x_{i_{j+1} f}}{2} \tag{3.5}$$

其中$i = 1,2,\ldots,k-1.$

3.分裂增益计算

遍历所有分裂阈值τ_j，计算分裂后左节点MSE_L，右节点MSE_R。

$$S_L = \{(X_i, y_i) | x'_{if} \le \tau_j\} \tag{3.6}$$

$$S_R = \{(X_i, y_i) | x'_{if} \ge \tau_j\} \tag{3.7}$$

$$MSE_L=\frac{1}{N_{S_L}}\sum_{i\in S_L}(y_i-\overline{y_L})^2 \quad (3.8)$$

$$MSE_R=\frac{1}{N_{S_R}}\sum_{i\in S_R}(y_i-\overline{y_R})^2 \quad (3.9)$$

进而计算指标f、分裂阈值τ的分裂增益$\Delta MSE(f,\tau)$：

$$\Delta MSE(f,\tau)=MSE(t)-\left(\frac{N_L}{N_t}MSE_L+\frac{N_R}{N_t}MSE_R\right) \quad (3.10)$$

其中N_L,N_R分别为S_L和S_R的样本个数。

从所有候选特征-阈值组合中，选择使ΔMSE取最大值的(f^*,τ^*)作为当前节点的分裂条件。

4.检查递归终止条件

对生成的左右子节点重复分裂阈值候选和分裂增益计算，逐层扩展树结构。当满足以下三个条件之一时，递归停止，记此时的深度最大节点为叶子节点。

第一，达到最大深度d_{max}

第二，分裂增益$\Delta MSE(f,\tau)\leq 0.$

第三，$N_t\leq N_0$，N_0为预设最小值。

（三）随机森林(RF)模型构建

为有效抑制单棵决策树的过拟合倾向，增强模型的泛化能力，本文采用双重随机化策略，结合代价复杂度剪枝（Cost-Complexity Pruning, CCP）优化策略构建随机森林（RF）模型，最终进行聚合输出。

1. Bootstrap 重采样

采用自助采样（Bootstrap Aggregating）从原始数据集中有放回地抽取多个子集，使每棵决策树仅基于部分样本训练，降低对特定数据的敏感度，即样本随机化。

第b棵树的训练集$\mathcal{D}_b$通过有放回抽样生成：

$$\mathcal{D}_b=\left\{\left(X_{i_b^{(1)}},y_{i_b^{(1)}}\right),\ldots,\left(X_{i_b^{(N_{\text{train}})}},y_{i_b^{(N_{\text{train}})}}\right)\right\} \quad (3.11)$$

其中 $i_b^{(k)}\in[1,N_{\text{train}}]$，其余为袋外数据 $\mathcal{D}_{OOB}^{(b)}=\mathcal{D}_{train}\backslash\mathcal{D}_b$

2.特征随机化操作

与单棵随机数生成类似，每棵树的节点分裂时，仅从随机选取的特征子空间中寻找最优分割点，进行特征随机化。

第 b 棵树从 m 维特征中随机选取 m_{try} 维作为候选集F

$$F_{\text{候}} = \left\{f_b^{(1)}, \ldots, f_b^{(m_{\text{try}})}\right\} \subseteq \{1, \ldots, m\} \quad (3.12)$$

其中m_{try}= $\lfloor \sqrt{m} \rfloor$.

3.CCP 优化

CCP 剪枝目标为：通过调整α，寻找使$R_a(T)$最小的树结构，即剪枝后模型的预测误差与复杂度达到最优平衡。

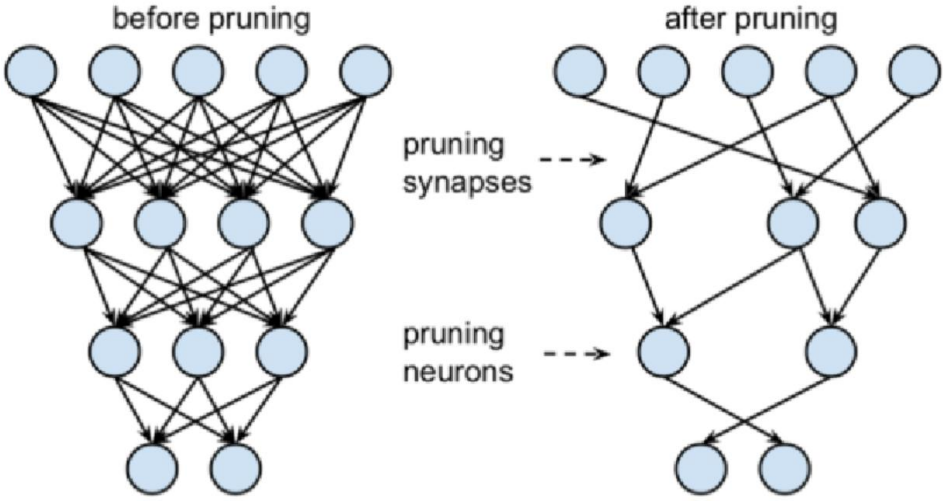


图 6 CCP 优化原理图

定义代价复杂度函数：

$$R_a(T) = R(T) + \alpha \cdot |\hat{T}| \quad (3.13)$$

其中，$R_a(T)$为子树T在训练集上的加权平方误差（叶子节点误差之和），$|\hat{T}|$为子树T的叶子节点数量（衡量模型复杂度）；α为惩罚系数（权衡误差与复杂度）。

生成子树序列自底向上遍历，逐层计算每个非叶子节点的有效α值，即α_{elf}：

$$\alpha_{elf} = \frac{R(t) - R(T_i)}{|\hat{T}| - 1} \quad (3.14)$$

其中，T_i为以节点t为根的子树，$R(t)$为剪除T_i后节点t作为叶子的误差。

α_{elf}从小到大排序，依次剪除对应子树，生成复杂度递减的候选树

$$\{T_0, T_1, \ldots, T_k\} \quad (3.15)$$

其中T_0为原始树，T_k为仅剩根节点的树。

选择使 MSE 最小的α值对应的子树作为最终模型。

4.聚合输出

进行加权平均求和，输出预测值为所有b棵树的预测值的算数均值$\hat{y}$。

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b(X) \quad (3.16)$$

其中，$T_b(X)$为测试集数据在第b棵树中的预测值。

七、多层次碳排放评估模型输出结果分析

（一）基于两次熵权法的黑龙江省相关指标权重结果与分析

1. 具体输出结果展示

第一次熵权法确定的各二级指标具体权重详见附录。第二次熵权法确定的各一级指标的权重如表 3 所示：

表 3 第二次熵权法确定的各一级指标权重

一级指标名称	权重值
经济发展	0.2295
工业发展	0.1965
农业发展	0.1410
居民生活	0.1268
人居环境	0.1883
碳排放	0.1179

2. 分析结果与相关结论

一级指标中经济发展指标对黑龙江省综合碳排放评估影响最为显著，其二级指标中 GDP 总量与人均 GDP 贡献接近，合计占比达 84.87%，表明经济增长规模与效率是所选指标中影响黑龙江省综合碳排放评估值的核心因素。

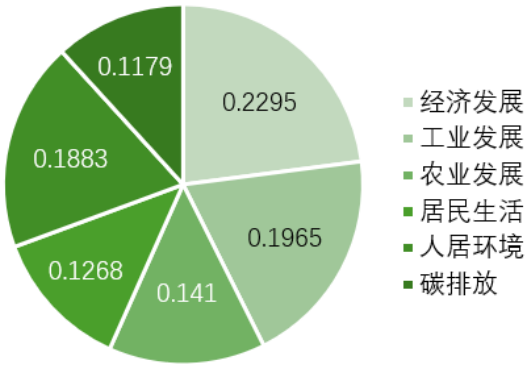


图 7 一级指标权重

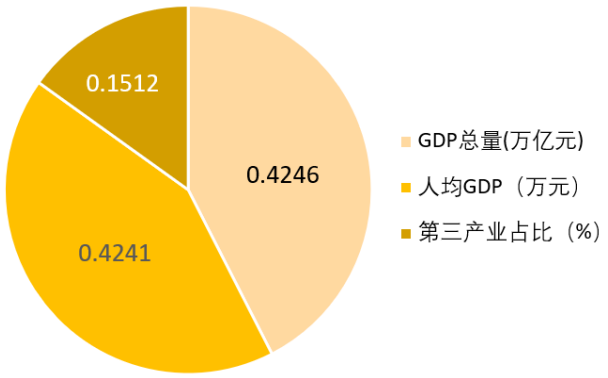


图 8 经济发展指标下各指标权重

而第三产业占比（权重 0.1512）较低权重验证了前文提出的黑龙江省整体发展与碳排放存在结构性矛盾的观点。

工业发展指标中，焦炭产量（权重 0.3192）和工业产值增加量（权重 0.2892）

的权重显著高于其他指标，凸显高能耗工业活动对碳排放存在直接影响。

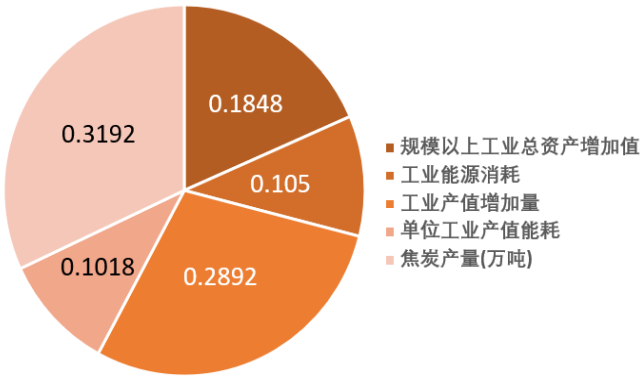


图 9 工业发展指标下各指标权重

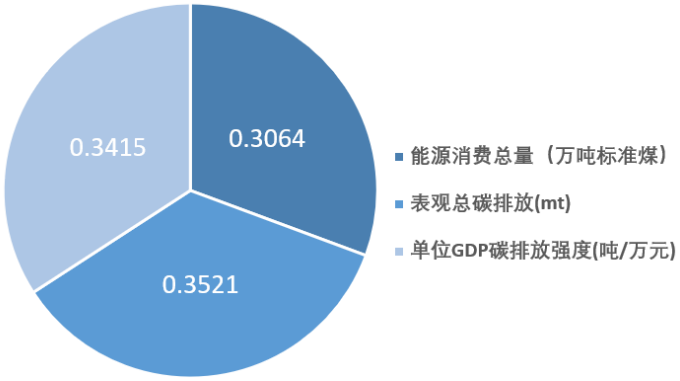


图 10 碳排放指标下各指标权重

（二）基于黑龙江省各指标数据的多层碳排放评估模型结果分析

2014－2024 年黑龙江省综合碳排放评估值呈现“先降后升再波动”的非线性特征（图 11）。

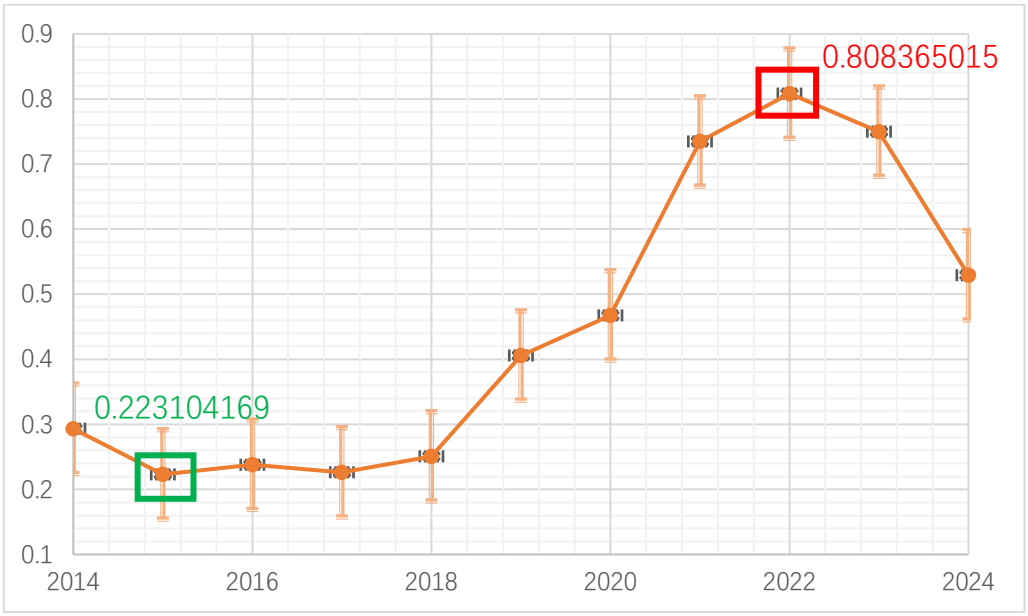


图 11　2014-2024 黑龙江省综合碳排放评估值折线图

2014 年得分为 0.293，2015－2017 年连续下降至 0.226（2017 年），年均降幅达 8.2%;2018 年起逐步回升,2022 年达到峰值 0.808,年均增长率高达 28.5%;2023－2024 年回落至 0.749 和 0.529，降幅分别为 7.6%和 29.4%。

关于黑龙江省各一级指标评估情况，根据图 12，经济发展呈现持续增长趋势， 2014 年至 2024 年年均增幅达 0.091，表明区域经济转型成效显著。工业发展呈现波动特征，2014 年达峰值 0.545 后下降至 2024 年 0.316，与产业政策调

整或技术升级周期相关。居民生活在 2021 年达到最高值 0.831，较 2014 年增长 52.3 倍，反映民生改善成效突出。碳排放整体呈下降趋势，2020 年达最低值，但 2021 年后回升至 0.445-0.482 区间，提示减排压力仍存。农业发展与人居环境虽存在短期波动，但 2024 年较基准年分别增长 11.0%和 82.5%，显示基础领域稳步优化。

数据表明，黑龙江省综合发展情况为：在经济增长与民生领域取得突破，但工业转型与碳排放控制需进一步协同。

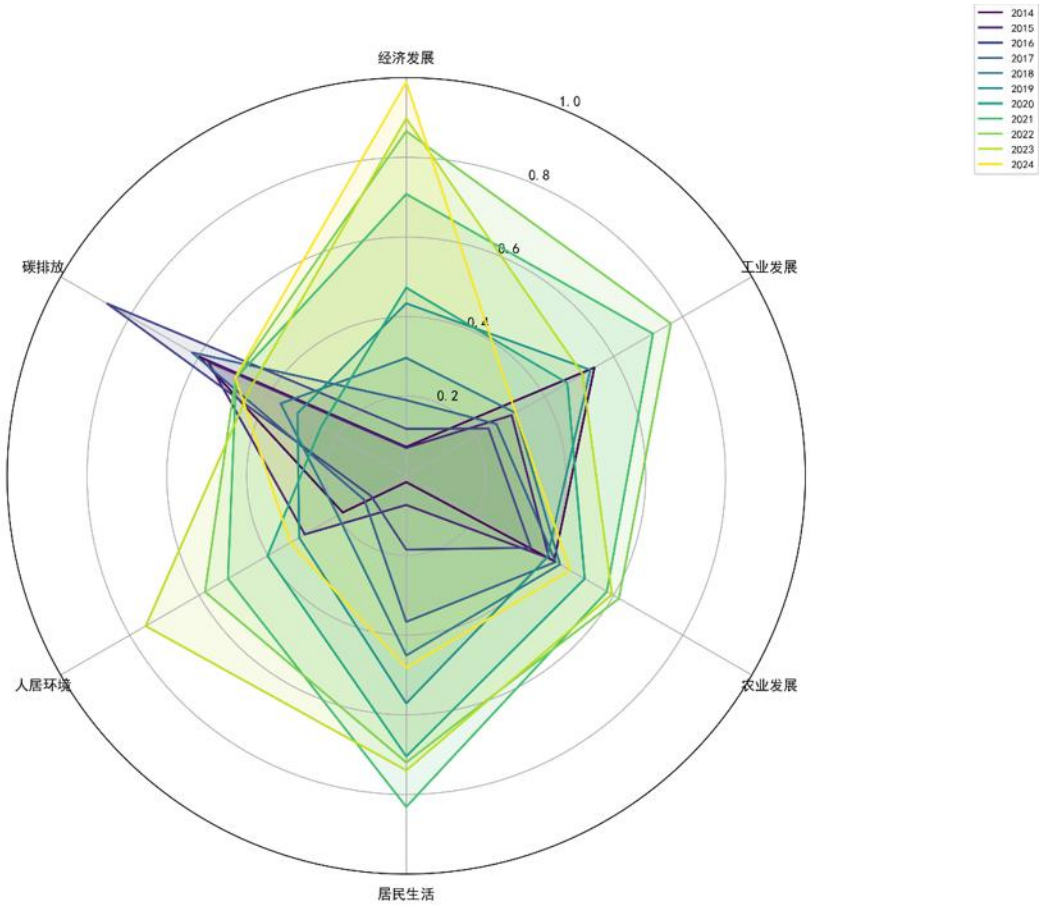


图 12 2014-2024 黑龙江省各一级指标评估雷达图

（三）黑龙江省各领域与综合碳排放耦合度、协调度评估结果分析

1. 具体输出结果展示

针对黑龙江省六项一级指标进行耦合协调度评价，相关数据如下表所示：

表 5 耦合度与协调度相关评估值

年份	耦合度	发展指数	协调度	协调等级

2014	0.578	0.2993	0.4159	轻度失调
2015	0.7529	0.2925	0.4693	轻度失调
2016	0.7471	0.3147	0.4849	轻度失调
2017	0.8831	0.3233	0.5343	基本协调
2018	0.9448	0.3163	0.5467	基本协调
2019	0.9233	0.402	0.6093	基本协调
2020	0.8199	0.438	0.5992	基本协调
2021	0.9816	0.6387	0.7918	基本协调
2022	0.9797	0.6682	0.8091	优质协调
2023	0.9742	0.6609	0.8024	优质协调
2024	0.9256	0.5071	0.6851	基本协调

2.数据分析

①关键年份的驱动因素解析

整体来看，协调等级从2014年“轻度失调”逐步优化至2022年“优质协调”（协调度0.8091）。

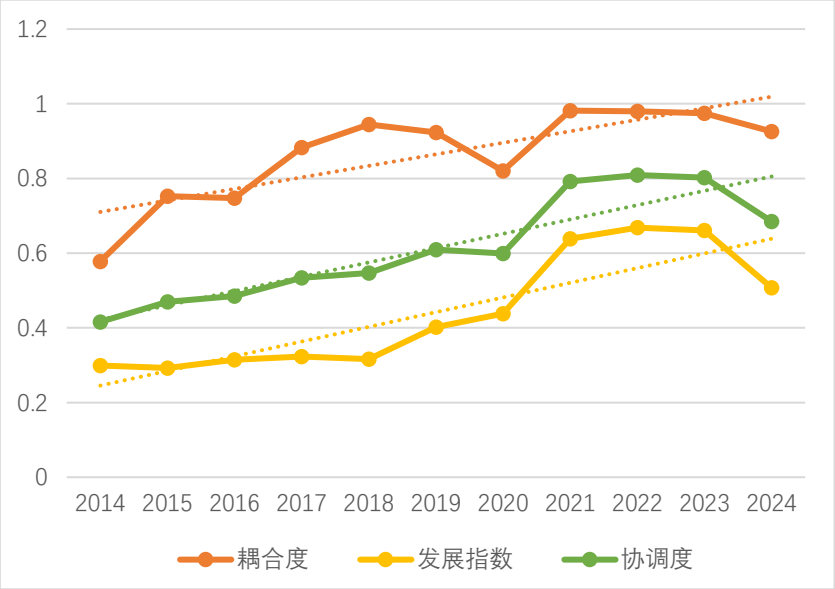


图13 2014-2024黑龙江省各领域发展与碳排放耦合度、协调度折线与趋势图

2019年得分跃升：总得分从2018年0.251增至0.406（增幅61.8%），结合权重结构与社会事实，主要驱动因素为第三产业占比（权重0.1512）的短期提

升（如旅游业或服务业扩张），但其后未能保持，导致 2020 年增速放缓至 14.9%。

2022 年峰值形成：协调度达 0.8091（优质协调）。结合指标权重与政策文件，体现了清洁生产技术推广效应（如焦化行业能效提升），叠加人均公园绿地面积（权重 0.4019）的生态修复效应以及疫情期间交通出行、生产生活减少等减排因素。

2024 年下跌：总得分骤降至 0.529（较 2023 年降幅 29.4%），协调度回落至 0.6851。数据表明，单位 GDP 碳排放强度（权重 0.3415）的恶化（如能源回弹效应）与工业能源消耗（权重 0.1050）的回升是主因，同时综合碳排放指数从 0.6609（2023 年）降至 0.5071（2024 年），反映经济增速放缓与碳排放控制的矛盾有所加剧。

②黑龙江省各领域发展与碳排放协调性演变特征

经过分析，黑龙江省各领域发展与碳排放协调性演变存在阶段性特征与驱动机制，可分为三阶段：

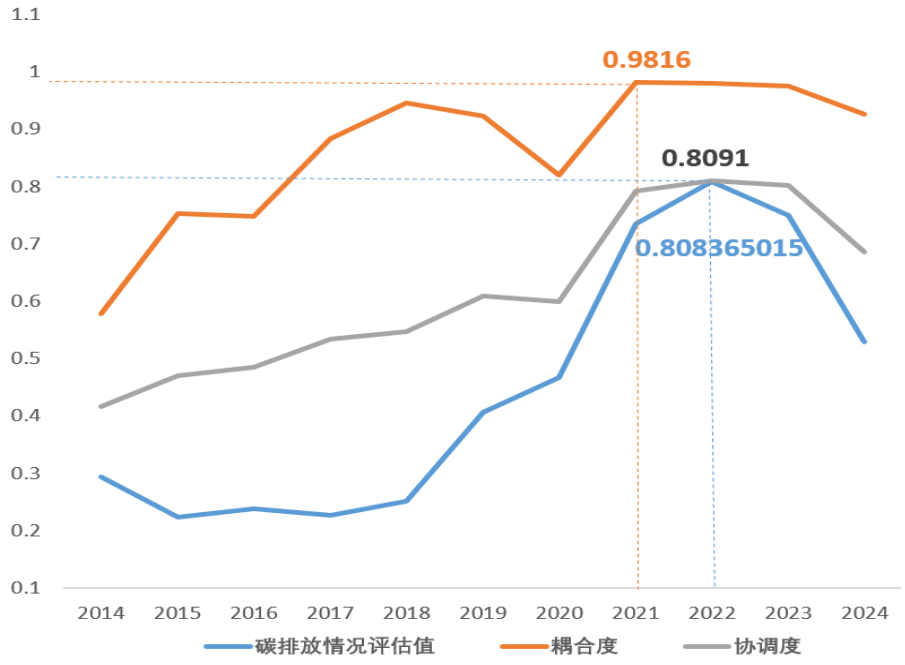


图 14 综合碳排放评估值与耦合度、协调度评估值折线图

失调缓解期（2014－2017）：协调度从 0.4159 升至 0.5343，依赖工业能效初步改善。

协调优化期（2018－2022）：2017 年协调度首次突破 0.5（0.5343），进入“基本协调”阶段，表明经济发展与碳排放控制的协同效应初显。2022 年协调度

跃升至 0.8091，耦合度与发展指数（同步增长，反映各子系统（如工业发展与碳排放强度控制）的协同效率显著提升。

波动调整期（2023－2024）：2022 年评估总得分峰值（0.808）与协调等级“优质协调”的同步出现（图 13），表明该年份经济社会综合发展与碳排放控制的平衡达到最优状态。而 2024 年协调度回落至“基本协调”但发展指数仍高于 0.5，提示尽管子系统间协同性减弱，但整体发展水平未出现断崖式下跌。

（四）多层次碳排放评估模型输出数据的特征与效应

通过对于多层次碳排放评估模型与协调度、耦合度评估的输出数据分析，发掘出以下数据特征与效应：

1.权重与得分波动的非线性响应：

高权重指标（如 GDP 总量、焦炭产量）的变动对总得分具有放大效应。例如，2022 年焦炭产量（权重 0.3192）若提升 10%，理论上可拉动总得分增加约 3.2%，但实际增幅达 9.9%，表明其与单位工业产值能耗（权重 0.1018）的协同优化产生了乘数效应。

2.协调度阈值效应：

协调度在 0.5－0.8 区间内呈现“边际收益递减”特征。例如，2017 年协调度 0.5343（基本协调）仅需发展指数提升 7.4%，而 2022 年达到 0.8091（优质协调）需发展指数增长 104.6%，说明高阶协调需多维指标的同步突破。

3.政策敏感性与系统韧性：

2024 年总得分与协调度的同步下降，反映系统对高权重工业指标（如焦炭产量）的政策调控高度敏感，而人居环境维度（权重 0.1883）的生态修复未能有效缓冲工业波动，提示需增强低碳产业的韧性布局。

八、基于贝叶斯超参数优化的随机森林模型

为突破传统网格搜索（Grid Search）和随机搜索（Random Search）的低效性，本研究采用贝叶斯优化（Bayesian Optimization, BO）对随机森林关键超参数进行自适应调参。具体流程如下：

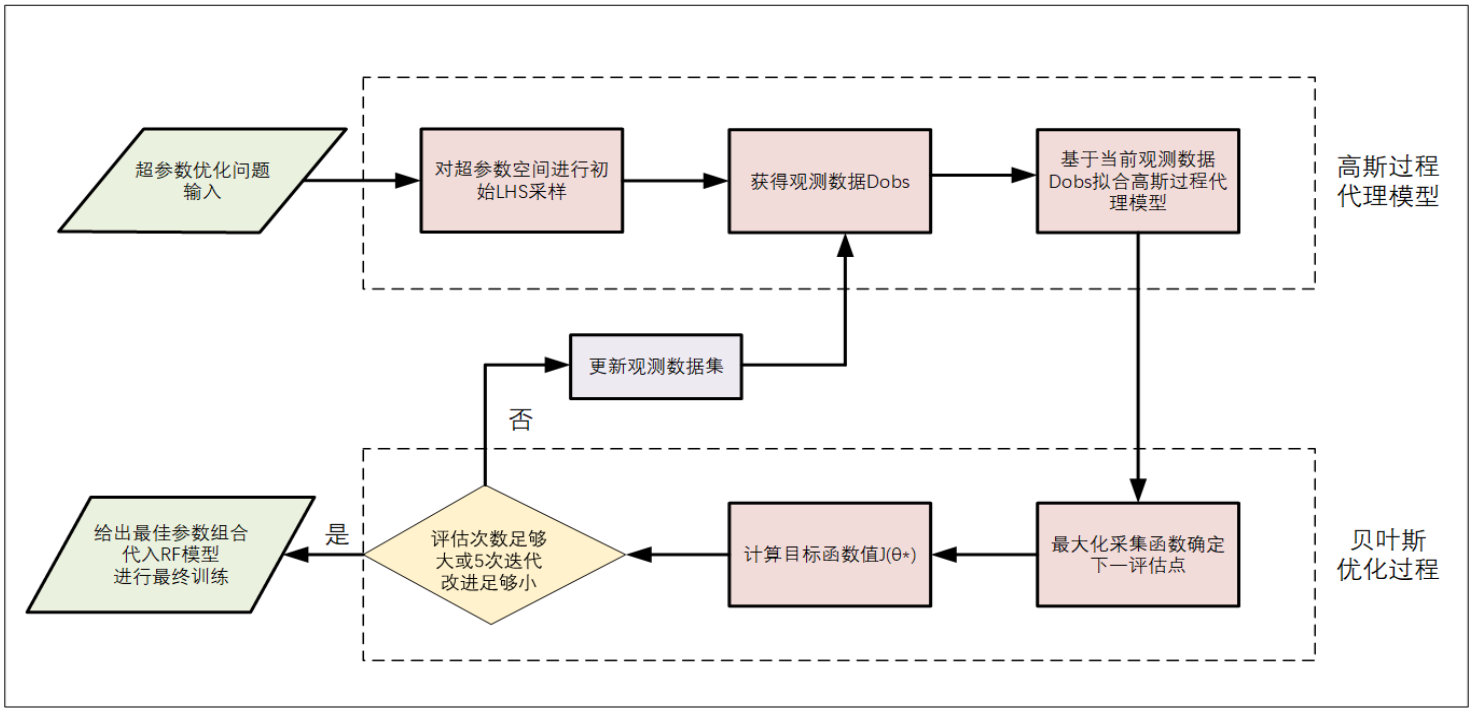


图 15 贝叶斯超参数优化结构与基本过程图

本文通过高斯过程（Gaussian Process, GP）概率代理模型与期望提升采集函数（Acquisition Function，AF）协同作用，并利用拉丁超立方采样（Latin Hypercube Sampling, LHS）初始化代理模型，在有限计算资源下平均优化训练 9 次逼近全局最优参数组合，显著提升了模型预测性能。

（一）优化目标与核心超参数

1. 优化目标函数构建

定义优化目标函数$J(\theta)$为 11年样本数据的 LOOCV 验证的平均负均方误差：

$$J(\theta) = -(1/11)\sum_{t=2014}^{2024} MSE_t(\theta) \quad (4.1)$$

单年份测试误差计算：

$$MSE_t(\theta) = \left(1/N_{test}{}^{(t)}\right)\sum_{i\in D_{test}{}^{(t)}}\left(y_i - \hat{y}_i^{\theta}\right)^2 \quad (4.2)$$

2. 核心超参数空间定义

定义超参数空间Θ：

$$\theta = \left(B, d_{max}, m_{try}, n_{min}\right) \in \Theta \quad (4.3)$$

其中，树数量$B \in [100,500]$，为自然数；最大树深度$d_{max} \in [5,30]$，为自然数；分裂特征数$m_{try} \in [1,m]$，为自然数；最小分裂样本数$n_{min} \in \{2,5,10\}$。

（二）LHS 初始化采样

LHS 采样旨在以较少的样本点实现对高维参数空间的均匀覆盖，适用于本研究样本数据相对较小（11 年数据）的实际特点。

1.参数范围划分：对每个维度d(超参数的范围）等分为N_{init}个互不重叠的区间,即采样N_{init}个点。

2.区间内随机采样：

在每个维度的每个区间内随机抽取一个点，保证每个区间仅被访问一次。

第d维第i个区间内的采样值为：

$$Nx_d^{(i)} = \frac{i-1+U(0,1)}{N}, \quad i = 1,2,\ldots, N \tag{4.4}$$

其中$U(0,1)$为均匀分布随机数。

3.维度间随机排列：

将各维度的区间索引进行独立随机排列，确保不同维度的采样位置无相关性。最终样本点坐标为各维度随机排列后的区间内采样值的组合。

（三）GP 代理模型

定义 GP 代理模型为随机过程集合，满足任意有限维边际分布服从多元正态分布，建立超参数组合$\theta = (\theta_1, \theta_2, \ldots)$与验证误差$f(\theta)$之间的隐式映射关系，即对目标函数建立带噪声观测模型:

$$f(\theta) \sim GP\big(\mu_0, k(\theta,\theta')\big) \tag{4.4}$$

$$y_i = f(\theta_i) + \epsilon_i \tag{4.5}$$

其中μ_0为均值函数$E[f(\theta)]$，$\epsilon_i \sim N(0,{\sigma_n}^2)$，$k(\theta,\theta')$为协方差函数。

选用平方指数核函数作为协方差函数以描述参数间相关性：

$$k(\theta,\theta') = \sigma_f^2 e^{-\frac{1}{2l^2}\|\theta-\theta'\|^2} + \sigma_n^2\delta(\theta,\theta') \tag{4.6}$$

其中l为特征尺度参数，σ_f^2为信号方差，σ_n^2为噪声方差。

（四）期望提升采集函数

基于当前 GP 代理模型的预测分布，本文采用期望提升函数(Expected Improvement, EI）作为 AF 以平衡探索-利用困境，选择下一个候选参数。

定义提升函数为：

$$EI(\theta) = E[max(f(\theta) - f(\theta^+), 0)] \quad (4.7)$$

其中θ^+为当前最优参数。其解析解为：

$$\mathrm{EI}(\theta) = \begin{cases} (\mu(\theta) - f\theta^+) - \xi)\Phi(Z) + \sigma(\theta)\varphi(Z), & \sigma(\theta) > 0 \\ 0, & \sigma(\theta) = 0 \end{cases} \quad (4.8)$$

其中$Z = \frac{\mu(\theta) - f(\theta^+) - \xi}{\sigma(\theta)}$，$\Phi(Z)$和$\varphi(Z)$分别为标准正态分布累积函数和概率密度函数。

（五）迭代优化算法

本文优化流程执行以下步骤：

1.初始化：在θ空间进行 LHS 采样，获取初始样本集$\{\theta_i, J(\theta_i)\}(i = 1 \rightarrow N_{init})$

2.循环迭代直至满足终止条件：

①基于当前观测数据$D_{obs} = \big(\theta_i, J(\theta_i)\big)$拟合高斯过程代理模型

②最大化采集函数确定下一评估点$\theta* = argmax_{\theta \in \Theta} EI(\theta)$

③计算目标函数值 J(θ*)并更新观测数据集$D_{obs} \leftarrow D_{obs} \cup \{\theta*, J(\theta*)\}$

3.终止条件：达到最大评估次数$N_{max} = 50$或连续 5 次迭代改进小于$\varepsilon = 10^{-4}$

（六）最优参数选择

从观测数据集中选取综合性能最优参数配置：

$$\theta* = argmin_{\theta \in D_{obs}} \frac{1}{11} \sum_{t=2014}^{2024} MSE_t(\theta) \quad (4.9)$$

将$\theta*$代入 RF 模型进行最终训练，确保 RF 模型同时满足精度与泛化能力要求。

（七）BO 优化前后绝对误差堆积比较

定义优化前绝对误差堆积K：

$$K = \sum_{i=1}^{n} (y_i - \hat{y}_i) \quad (5.1)$$

利用 Python 软件绘制得经贝叶斯超参数优化前后的历年各指标绝对误差堆积图，如图 12、13 所示。

可以明显观察到优化后K的峰值下降数十倍，初步验证了本文采取的贝叶斯超参数优化的有效性。

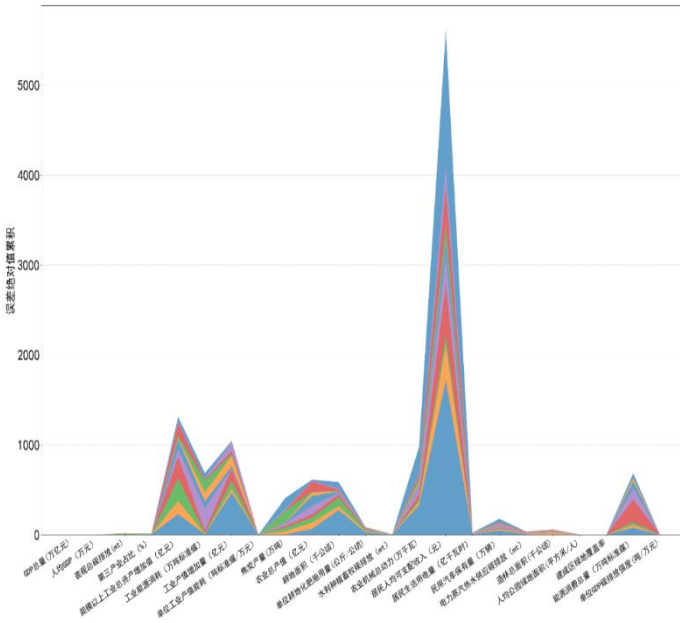


图 16 优化前绝对误差堆积

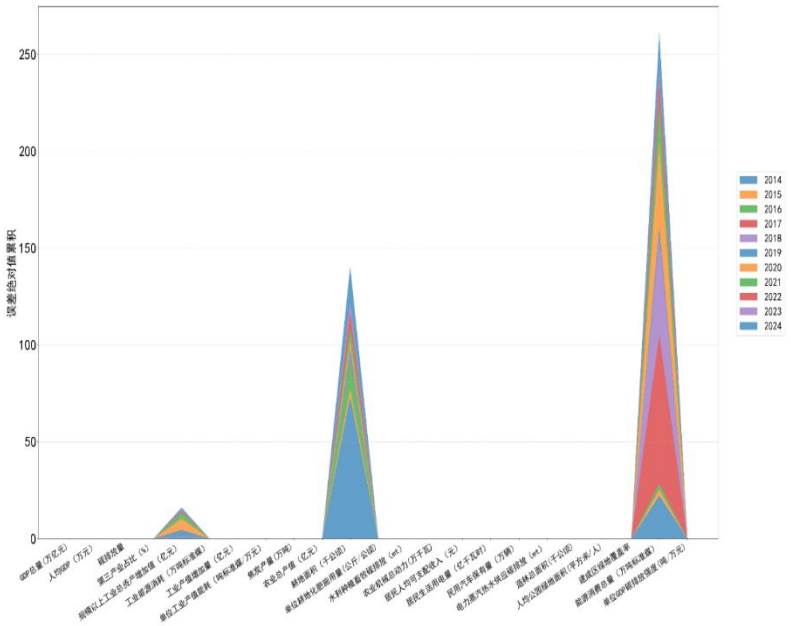


图 17 优化后绝对误差堆积

（八）经典误差指标评价 BO 优化性能

本文进一步使用决定系数（R^2）、均方误差（MSE）、均方根误差（RMSE）来评价经 BO 优化前后的 RF 模型性能。计算公式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{5.2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5.3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5.4}$$

其中，n 是预测的样本数量（预测范围内的年份数），y_i是第 i 年某指标的实际值，$\hat{y}_i$是第 i 年某指标的预测值。

本文采用未经优化的 RF 模型的默认参数如下：决策树数量（N_estimators,NE）=100，最大深度（Max_depth,MD）=10，节点在继续分裂前所需的最小样本数(Min_samples_split,MSP)=5,叶子节点所需的最小样本数（Min_samples_leaf,MLE）=3，每个决策树节点在寻找最佳分裂时允许使用的特征总数百分比（Max_features,MF）=0.5。

经计算，给出优化前后的随机森林模型针对各指标 R^2、MSE、RMSE 值：

由于指标过多,表 7 展示了部分指标的误差检测结果。完整图表详见附录 3。

表 7 优化前后 RF 模型针对各指标的 R^2、MSE、RMSE 值表（部分）

二级指标	优化前 R^2	优化后 R^2	优化前 MSE	优化后 MSE	优化前 RMSE	优化后 RMSE
GDP 总量(万亿元)	0.959705	1	0.001123	5.44E-29	0.033516	7.37564E-15
人均 GDP（万元）	0.980761	1	0.009734	1.57E-29	0.098659	3.96075E-15
碳排放量	0.905099	1	6.50439	2.54E-24	2.550371	1.59466E-12
第三产业占比（%）	0.820535	1	3.270945	9.45E-28	1.808575	3.07488E-14
规模以上工业总资产增加值（亿元）	0.947005	0.999987	21457.96	5.356775	146.4853	2.314470741
工业能源消耗（万吨标准煤）	0.787374	1	9010.207	5.41E-23	94.92211	7.35306E-12
工业产值增加量（亿元）	0.873345	1	24719.86	8.1E-22	157.2255	2.84611E-11
单位工业产值能耗（吨标准煤/万元）	0.754133	0.999976	0.014609	1.4E-06	0.120868	0.001182828
焦炭产量(万吨)	0.911397	1	3192.842	3.95E-25	56.50524	6.28324E-13
……	……	……	……	……	……	……

将优化前后的随机森林模型分别命名为RAW-RF和PURE-RF模型。可以看出，相比于 RAW-RF 模型，PURE-RF 模型针对各项模型的决定系数 R^2 大幅提高（R^2 均值从 0.8104 增至 0.9982，增长 81.7%），MSE 与 RMSE 大幅下降，充分证明了优化后 RF 模型在预测效果上存在显著提升。

结合优化后极高的 R^2（平均 0.9982）与较小的 MSE（平均 0.3222）与 RMSE（平均 0.1676），充分说明 PURE-RF 模型对城市各项指标的真实数据有很高的拟合度，对于所选的各指标有良好的预测性能。

这体现了 PURE-RF 模型具有优秀的泛化预测能力，同时兼顾准确性和稳定性，证明了本文采用 PURE-RF 模型进行黑龙江省综合碳排放预测研究是合理的。本文在附录 4 给出了预测各指标时贝叶斯优化得到的 PURE-RF 最佳参数组合。

九、总结与建议

（一）研究总结

本文通过构建“熵权-TOPSIS-耦合协调度”多层次评估模型与贝叶斯优化的随机森林预测算法（PURE-RF），系统解析了黑龙江省 2014—2024 年碳排放的动态特征及其与区域发展的协同效应。

本文发现：黑龙江省碳排放评估值呈现“先降后升再波动”的非线性趋势，2022 年因工业技术升级与生态修复协同达到优质协调（协调度 0.8091），但 2024 年受工业能耗回弹与单位 GDP 碳排放强度恶化影响回落至基本协调（0.6851），揭示了传统高耗能路径的系统脆弱性。

模型优化方面，BO 优化方法效果显著，PURE-RF 预测精度较 RAW-RF 方法有较大提升（R^2 均值从 0.8104 增增长至 0.9982，增长 81.7%），充分验证了其在处理复杂非线性数据中的强泛化能力。

（二）政策建议

为实现黑龙江省“双碳”目标，需从多维度协同发力：

1.加速工业低碳转型：针对焦炭、电力等高权重行业，推广清洁生产技术（如干熄焦工艺、余热回收），降低单位产值能耗；优化产业结构，逐步减少对重工业的依赖，培育新能源装备制造等低碳产业。例如，可借鉴德国鲁尔区转型经验，通过政策补贴引导企业技改，建立区域性碳交易试点。

2.增强系统韧性：结合黑龙江省可再生能源禀赋，优先布局风能、光伏等分布式能源网络，缓解冬季供暖能耗的季节性矛盾；强化生态修复的缓冲作用，提升建成区绿地覆盖率（权重 0.3890）与人均公园绿地面积（权重 0.4019）。建议在哈尔滨、大庆等城市试点“零碳社区”，整合绿色建筑与智慧能源系统。

3.完善动态监测与政策适配：依托 PURE-RF 模型构建碳排放实时预测平台，整合物联网传感器与卫星遥感数据，动态监测工业能耗、居民用电等关键指标；制定分阶段减排目标，例如设定 2025 年单位 GDP 碳排放强度下降 15%的约束性指标，并通过碳税、绿色信贷等经济杠杆激励企业减排。此外，需建立跨部门协同机制，统筹生态保护与经济增长，避免“运动式减碳”对区域经济的冲击

参考文献

[1]马潇颖.京津冀地区碳中和关键因素减排固碳潜力研究[D].华北电力大学(北京),2022.

[2]余碧莹,赵光普,安润颖,等.碳中和目标下中国碳排放路径研究[J].北京理工大学学报(社会科学版),2021,23(02):17-24.

[3]韩楠,罗新宇.多情景视角下京津冀碳排放达峰预测与减排潜力[J].自然资源学报,2022,37(05):1277-1288.

[4]张国兴,苏钊贤.黄河流域交通运输碳排放的影响因素分解与情景预测[J].管理评论,2020,32(12):283-294.

[5]胡鞍钢.中国实现 2030 年前碳达峰目标及主要途径[J].北京工业大学学报(社会科学版),2021,21(03):1-15.

[6]王灿,张雅欣.碳中和愿景的实现路径与政策体系[J].中国环境管理,2020,12(06):58-64.

[7]赵亚涛,南新元,贾爱迪.基于情景分析法的煤电行业碳排放峰值预测[J].环境工程,2018,36(12):177-181.

[8]崔佳旭,杨博.贝叶斯优化方法和应用综述[J].软件学报,2018,29(10):3068-3090.

[9]李亚茹,张宇来,王佳晨.面向超参数估计的贝叶斯优化方法综述[J].计算机科学,2022,49(S1):86-92.

[10]方匡南,吴见彬,朱建平,等.随机森林方法研究综述[J].统计与信息论坛,2011,26(03):32-38.

[11]Jin S S ,Kim G ,Kwag S , et al.Feasibility study of progressive Latin hypercube sampling and quasi-Monte Carlo simulation for probabilistic risk assessment[J].Geomatics, Natural Hazards and Risk,2024,15(1):

[12]Greenhouse Gases; Shandong University of Science and Technology Details Findings in Greenhouse Gases (Scenario analysis of carbon emissions' anti-driving effect on Qingdao's energy structure adjustment with an optimization model, Part I: Carbon emissions

peak ...)[J].Global Warming Focus,2018,203-.

[13]Breiman L .Random Forests.[J].Machine Learning,2001,45(1):5-32.

[14]Fredrik R ,Maxim T ,Paul M D V , et al.MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space.[J].Systematic biology,2012,61(3):539-42.

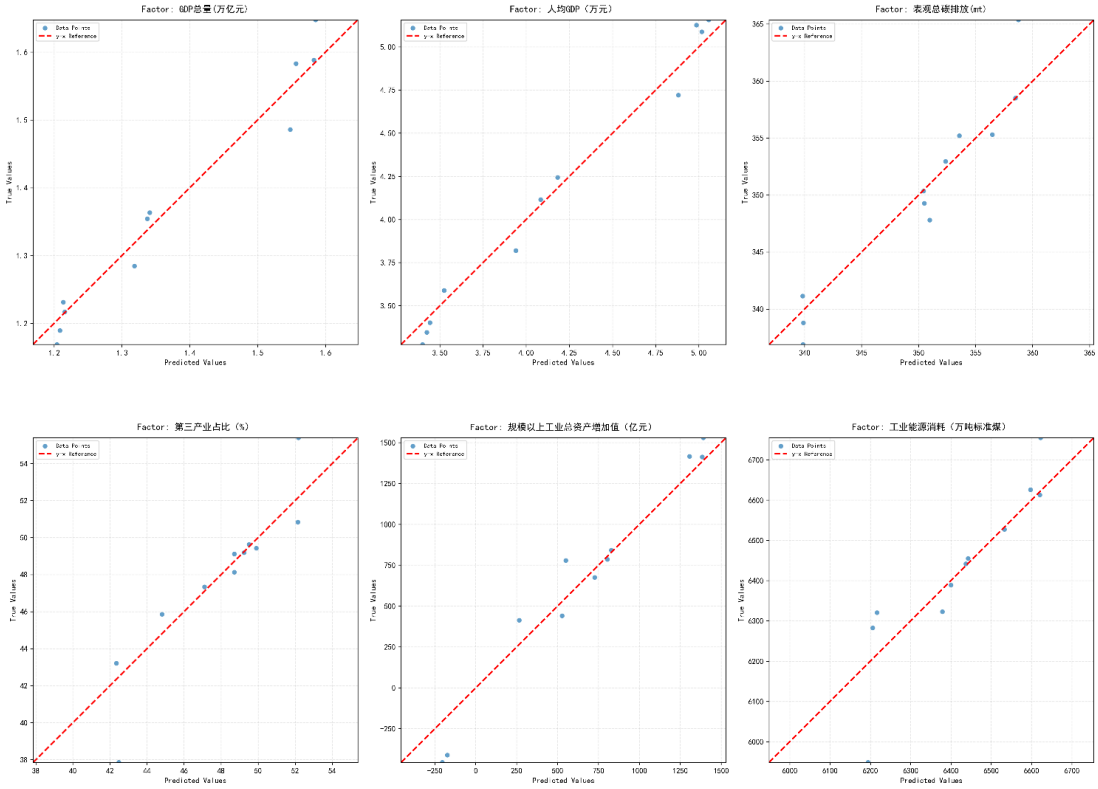
[15]Helton J ,Davis F .Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems[J].Reliability Engineering and System Safety,2003,81(1):23-69.

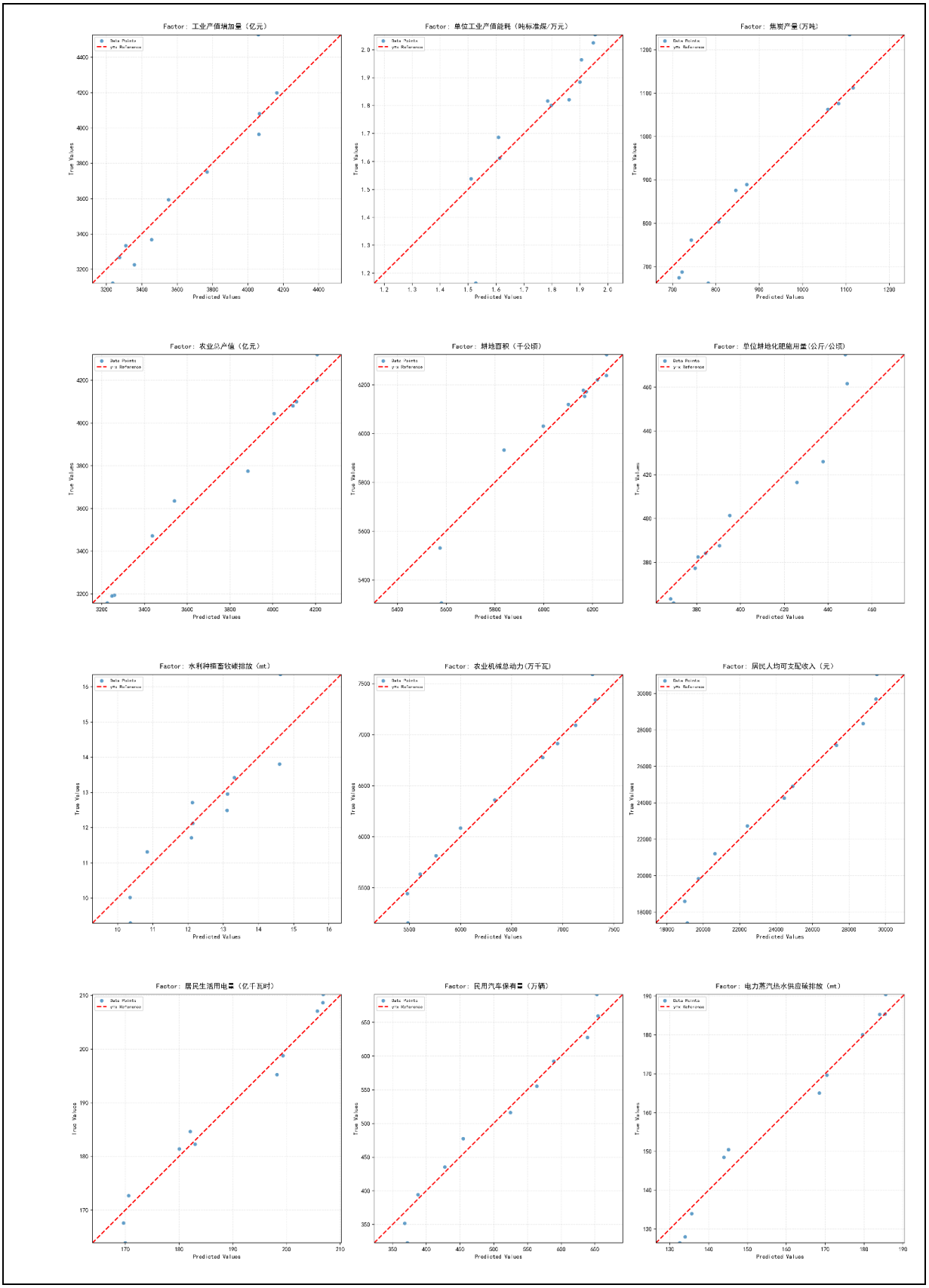
[16]Junhong H ,Fei G ,Xuanyi F , et al.Multi-factor decomposition and multi-scenario prediction decoupling analysis of China's carbon emission under dual carbon goal.[J].The Science of the total environment,2022,841156788-156788.

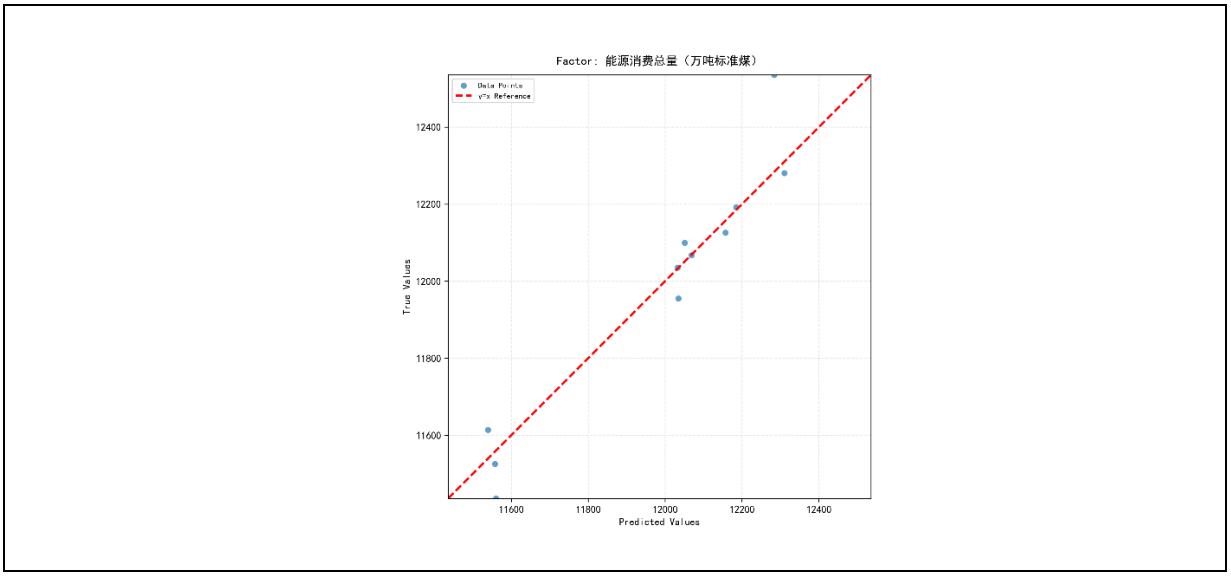
[17]Feng R ,Dinghong L .Carbon emission forecasting and scenario analysis in Guangdong Province based on optimized Fast Learning Network[J].Journal of Cleaner Production,2021,317

附录

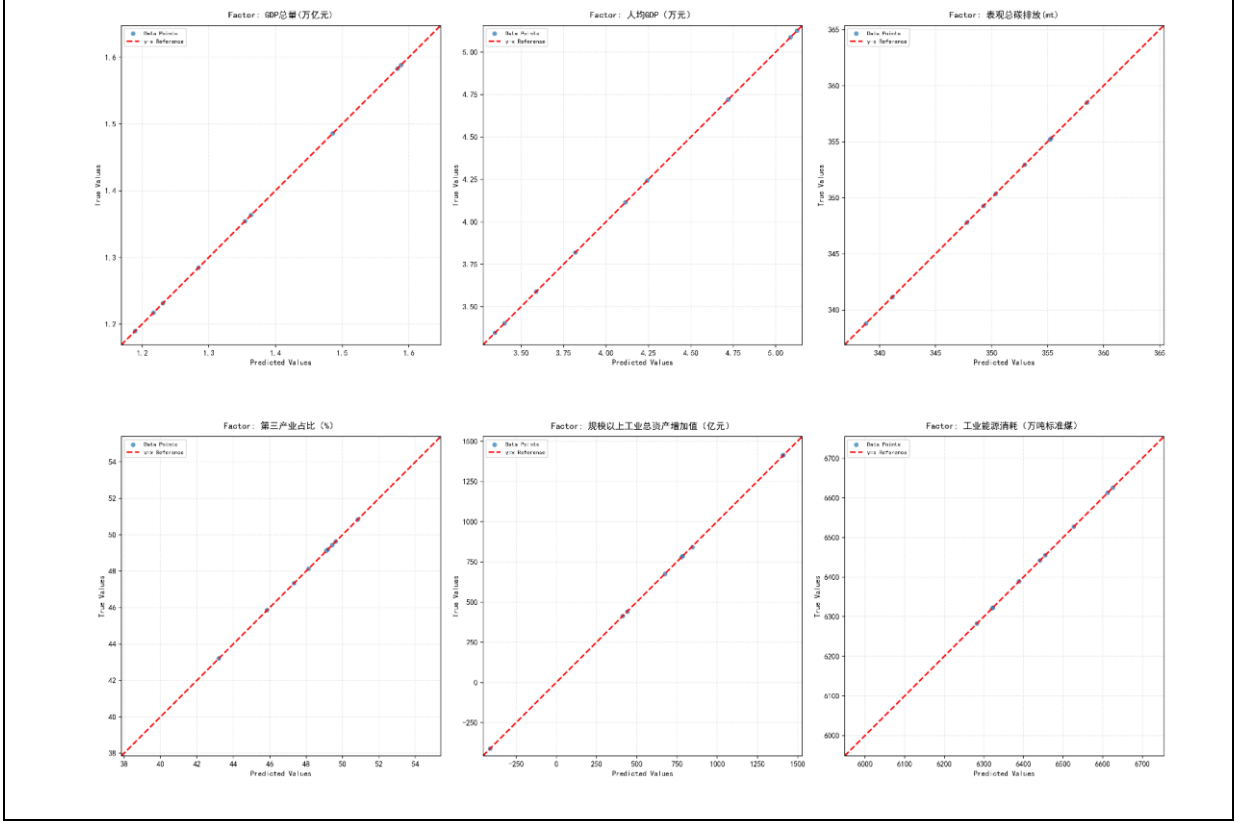
附录 1 RAW-RF 预测值与真实值相关性图

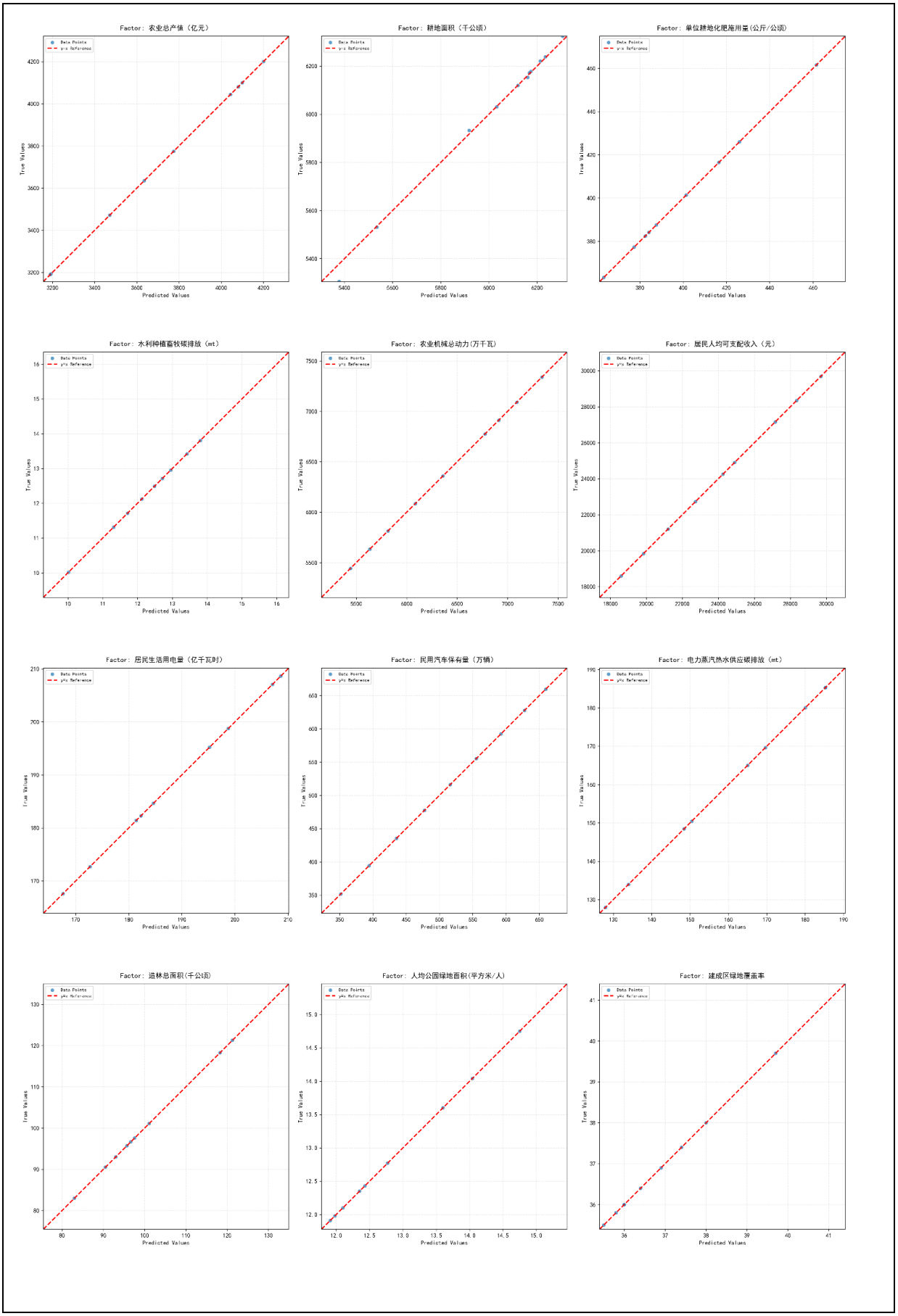


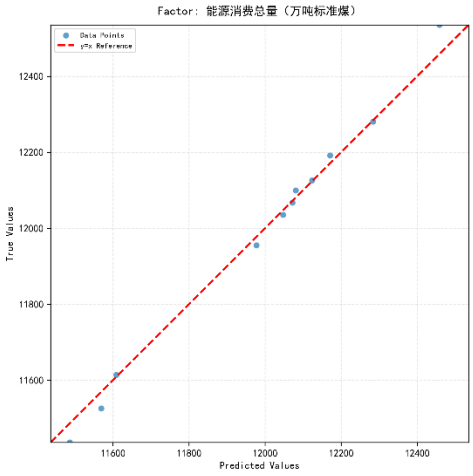

Factor: 工业产值增加量（亿元）
Factor: 单位工业产值能耗（吨标准煤/万元）
Factor: 焦炭产量(万吨)
Factor: 农业总产值（亿元）
Factor: 耕地面积（千公顷）
Factor: 单位耕地化肥施用量(公斤/公顷)
Factor: 水利种植畜牧碳排放（mt）
Factor: 农业机械总动力(万千瓦)
Factor: 居民人均可支配收入（元）
Factor: 居民生活用电量（亿千瓦时）
Factor: 民用汽车保有量（万辆）
Factor: 电力蒸汽热水供应碳排放（mt）
Data Points
y=x Reference
True Values
Predicted Values



附录 2 PURE-RF 预测值与真实值相关性图




Factor: 农业总产值（亿元）
Factor: 耕地面积（千公顷）
Factor: 单位耕地化肥施用量(公斤/公顷)
Factor: 水利种植畜牧碳排放（mt）
Factor: 农业机械总动力(万千瓦)
Factor: 居民人均可支配收入（元）
Factor: 居民生活用电量（亿千瓦时）
Factor: 民用汽车保有量（万辆）
Factor: 电力蒸汽热水供应碳排放（mt）
Factor: 造林总面积(千公顷)
Factor: 人均公园绿地面积(平方米/人)
Factor: 建成区绿地覆盖率
Data Points
y=x Reference
True Values
Predicted Values



附录 3 多层碳排放评估模型得到的黑龙江省各年综合碳排放评估值

黑龙江省各年综合碳排放评估值

年份	综合碳排放评估值
2014	0.293249
2015	0.223104
2016	0.237903
2017	0.22637
2018	0.250889
2019	0.405885
2020	0.467513
2021	0.734837
2022	0.808365
2023	0.749931
2024	0.529275

附录 4 优化后各指标最佳参数组合

各二级指标	MD	MF	MLE	MSP	NE
碳排放量	79	0.423121	1	2	443
GDP 总量(万亿元)	100	0.857133	1	2	500
人均 GDP（万元）	5	0.297767	1	2	60
第三产业占比（%）	100	1	1	2	50
规模以上工业总资产增加值（亿元）	5	0.1	1	2	304
工业能源消耗（万吨标准煤）	100	0.1	1	2	82

工业产值增加量（亿元）	100	1	1	2	500
单位工业产值能耗（吨标准煤/万元）	5	0.850493	1	2	50
焦炭产量(万吨)	5	1	1	2	50
农业总产值（亿元）	100	0.1	1	2	500
耕地面积（千公顷）	95	0.444281	1	2	116
单位耕地化肥施用量(公斤/公顷)	6	0.1	1	2	223
水利种植畜牧碳排放（mt）	33	0.1	1	2	280
农业机械总动力(万千瓦)	92	0.54682	1	2	427
居民人均可支配收入（元）	94	0.247246	1	2	135
居民生活用电量（亿千瓦时）	5	1	1	2	50
民用汽车保有量（万辆）	100	0.747088	1	2	330
电力蒸汽热水供应碳排放（mt）	89	0.373069	1	2	478
造林总面积(千公顷)	100	0.1	1	2	50
人均公园绿地面积(平方米/人)	15	0.100881	1	2	434
建成区绿地覆盖率	5	0.39884	1	2	50
能源消费总量（万吨标准煤）	5	0.266883	1	2	448
单位 GDP 碳排放强度(吨/万元)	5	1	1	2	50

附录 5 熵权法确定的各二级指标在所属一级指标内的权重

熵权法确定的各二级指标权重

一级指标	二级指标	权重值
经济发展	GDP 总量	0.4246
	人均 GDP	0.4241
	第三产业占比	0.1512
工业发展	规模以上工业总资产增加值	0.1848
	工业能源消耗	0.1050
	工业产值增加量	0.2892
	单位工业产值能耗	0.1018
	焦炭产量	0.3192
农业发展	农业总产值	0.2825
	耕地面积	0.1143
	单位耕地化肥施用量	0.2810
	水利种植畜牧碳排放	0.1323
	农业机械总动力	0.1899
居民生活	居民人均可支配收入	0.2402
	居民生活用电量	0.2467
	民用汽车保有量	0.2369

	电力蒸汽热水供应碳排放	0.2761
	造林总面积	0.2091
人居环境	人均公园绿地面积	0.4019
	建成区绿地覆盖率	0.3890
	能源消费总量	0.3064
碳排放	表观总碳排放	0.3521
	单位 GDP 碳排放强度	0.3415

致谢

值此论文完成之际，谨以最诚挚的谢意向在本项研究中所有给予帮助与辛勤付出的各位致以感激之情。

首先，衷心感谢指导老师周永春副教授在论文选题、数据分析和论文修改过程中给予的耐心指导。您严谨的学术态度和深入浅出的讲解，让我们作为大一学生在面对复杂建模问题时少走了许多弯路。

其次，感谢团队成员肖家伟提供大体框架以及论文撰写的认真工作，胡伟豪的协作支持，以及朱俊彰在数据收集与绘图整理中的认真负责。同时，感谢学校图书馆和网络资源库提供的资源支持，为研究开展提供了坚实基础。

最后，感谢我们在学业压力下仍能保持探索的热情。因学识尚浅，论文难免存在不足之处，恳请各位师长批评指正。

2025 年 4 月 23 日