

## 题目: 基于多模型的环境保护与经济发展模式探究

### 摘 要

环境保护和绿色经济的发展作为我国乃至世界范围内持续共同关注的焦点领域,近年来就政策方面的研究与各国间的成果备受世界人民的关注。因此谋求一种在保证绿色经济链能够循环发展的前提下实现环境保护的做法以及与之对应的成效衡量机制,对于可持续经济链的发展至关重要。我们团队在综合了当今国际形势与附件提供的相关数据后对如下问题进行系统化的求解:

针对问题 1,我们构建了一个“基于 GM(1,1)与时间序列预测的综合序列预测体系”。分别选用 1990 年~2015 年的附件中有关中国森林面积的数据作为两预测模型的训练组;2016 年~2020 年的相关数据作为测试组。最后通过对两者预测结果的对比得出时间序列预测模型的效果较优。且成功预测出中国 2025 年的森林面积为 2293760(km<sup>2</sup>),2035 年的森林面积为 2481717(km<sup>2</sup>)。(其余结果见附件)。

针对问题 2,我们在充分考虑到中国的“退耕还林政策”对于“工业水平”与“国民经济”的作用方式后,决定首先对如上三项进行指标的量化代指:以附件中出现的“森林面积”(占国土的百分比)与“农业用地”(占国土的百分比)之比量化表示“退耕还林度”,以代表退耕还林;在充分结合附件已知数据与衡量国家“国民经济”与“工业水平”的复杂程度后,我们于“国家统计局”收集了中国 1980 年~2020 年间的“居民消费水平”(现价美元)与“居民人均可支配收入”(现价美元)相关数据,并与附件中所给出的“GDP”一并作为衡量“国民经济”的指标。随后为了消除上述三项指标间的相关性与将指标相对目标进行降维,我们团队考虑到使用主成分分析模型(PCA)对指标进行降维处理,发现结果中仅保留一项主成分的情况下来衡量“国民经济”这一目标的贡献率已达到 0.9873,并定义该主成分为“经济水平综合因子”;受此启发,我们同理以“二氧化碳排放量”、“耗电量”、“能源使用量”与“工业增加值”(现价美元)作为衡量“工业水平”的四项指标进行主成分分析,得出保留一项主成分时,针对“工业水平”的贡献率达到 0.9977,并将该主成分定义为“工业水平综合因子”因此我们决定保留“退耕还林度”、“国民经济综合因子”与“工业水平综合因子”,并进行关系的找寻;随后我们在 1980 年~2020 年的数据中将“退耕还林度”在作为母序列,以“国民经济综合因子”与“工业水平综合因子”作为子序列进行灰色关联分析,并得出其二者的灰色关联度分别为 0.6301 与 0.6272,存在较强的相关性,即退耕还林政策不会对国家工业水平与国民经济有抑制作用,且能一定程度上促进其增长。

针对问题 3,我们以附件中给出的“收入水平”的 4 个标准为依据筛选出 40 个国家作为评价方案,分别建立以 GDP 作为衡量指标的“经济水平”评价模型、基于 EWM 与 AHP 综合指标权重的 TOPSIS “工业水平”评价体系和以森林面积与森林覆盖率为指标的“森林植被面积”评价体系,并分别通过在对的评价排名中划分等级的做法最终与收入水平进行映射,一次发现了包括“前列型”、“发展型”、“稳定型”、“转换型”、“受限型”与“特殊型”六中基本国家类型,在对上述类型进行定义后还发现森林植被面积这项指标相对于前三者的关系不大。

针对问题 4,在对上述问题 3 中的模型进行简要介绍的基础上,完善了我们团队对于上述发现的建议,并将希望一并向联合国政府提出,共同希望绿色经济与环境保护活动的持续进行。

**关键词:** GM(1,1)模型; 时间序列预测; PCA; 评价体系; 灰色关联分析

## 一、问题重述

就附件中所给出的世界范围内的相关数据以及背景资料下经济、工业的发展与森林植被面积之间的矛盾，现对四个方面的题目内容进行如下重述。

(1)本题附件中给出了中国近几十年来森林面积的数据，故其要求以已知数据作为依据，对中国未来 2025-2035 年 10 年时间内森林面积的大小做出预测。

(2)本题旨在着重探究退耕还林政策实施之后，对中国的国民经济和国家工业水平的影响。题目中给出了 GDP，二氧化碳排放量，工业增加值，耗电量，能源使用量等数据，通过题目给出的或者其他自行寻找的必要数据采取一定的方法分析寻找退耕还林政策的影响。

(3)本题即通过附件中所给出的世界各国的针对环境与经济的相关数据分别进行“经济水平”、“工业水平”和“森林植被面积”的评价模型的构建，并通过所构建的模型进行各个国家间定性或定量关系的找寻。

(4)本题目的为向联合国提供一封报告书，报告书的内容包括对所建立的模型的描述与对通过该模型得到的各国之间定性或定量的发现，最后提出具体措施对上述发现进行合理建议。

## 二、问题分析

### 2.1 问题 1 的分析

本题是预测类题目，题目中给出的森林面积数据是以年份度量的时间序列，且数据是非负的，通过时间序列图可以清晰看出数据随时间呈逐渐递增的关系。为了寻找适应度更高的预测模型，我们打算分别采用灰色预测与时间序列预测两种方法进行预测，并将已知数据中 1990-2015 年的数据作为训练组，2016-2020 年的数据作为测试组。通过比较二者对于测试组预测结果与实际数据之间误差的大小，最终发现时间序列预测模型更适用于本题数据，预测结果的平均相对误差仅为 0.05%，可见极为准确。通过时间序列预测模型对 2025-2035 年中国森林面积大小的预测结果见于附件。

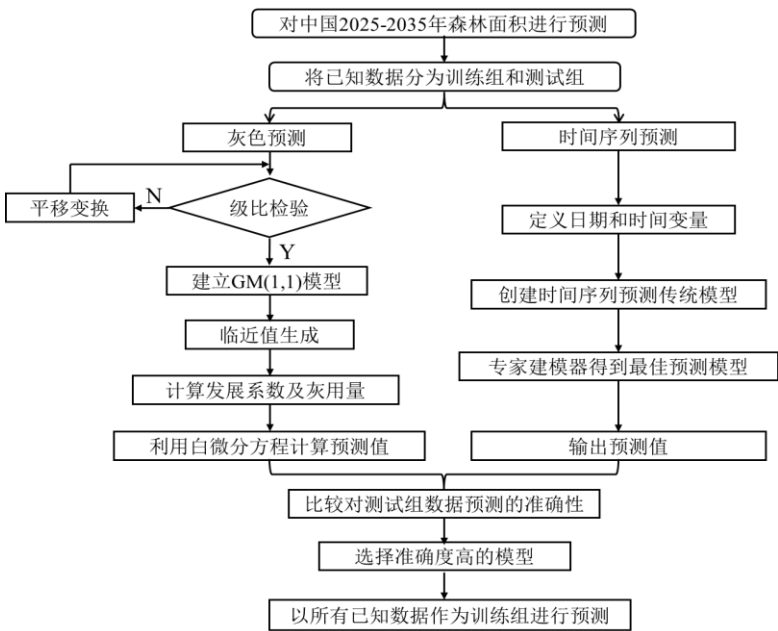


图 1-1 问题 1 分析解决流程图

## 2.2 问题 2 的分析

问题 2 中为了探究退耕还林政策对中国国家工业水平和国民经济的影响, 首先我们要寻找一种方式对 2002 年退耕还林政策正式实施之前和之后的国家工业水平和国民经济做出合理性陈述, 为此我们选择了二氧化碳排放量、耗电量、能源使用量和工业增加值作为衡量一个国家工业水平的因素, 选择了 GDP、居民消费水平、居民人均可支配收入作为衡量一个国家国民经济水平的因素。为了避免各因素之间存在一定的相关性, 从而影响到对目标的衡量, 因此我们运用主成分分析 (PCA) 模型对因素进行线性组合, 一方面能够消除因素间相关性较强的影响, 使衡量结果更加准确, 同时也进行了指标的降维, 更便于后续的分析。通过主成分分析得出的碎石图可以发现, 只需要一个主成分即可反映 99% 的原数据所包含的信息。为了寻找退耕还林政策是否会对工业水平和国民经济造成影响, 我们通过已知数据, 引入退耕还林度的概念, 并且采用灰色关联分析模型, 对上述两个主成分与退耕还林度数据的发展趋势进行分析, 最终得出退耕还林政策对国家工业水平和国民经济的影响。

## 2.3 问题 3 的分析

问题 3 的目的为使用附件中已知的各个国家的相关数据进行 3 个衡量不同方面的评价模型的构建, 最终通过所建立的评价模型对国家进行不同标准的区分, 通过对比划分为不同类型的国家间的偏重异同发现某国家分别在“经济水平”、“工业水平”和“森林植被面积”方面的优劣; 针对问题 3, 我们团队首先考虑将附件中的 9 项不同类型的数据进行定性讨论, 决定将“GDP”作为衡量“经济水平”的指标, 将“二氧化碳排放量”、“耗电量”、“能源使用量”与“工业增加值(每年的工业增加值的总额)”作为衡量“工业水平”的指标, 最后将“森林面积”与“森林覆盖率”作为衡量“森林植被面积”的指标; 考虑到上述三种评价目标中的指标数量不一, 于是我们对“经济水平”与“森林植被面积”的评价进行平均加权或单一对指标进行标准化后代指目标, 而对于“工业水平”这类具有较多评价指标的目标采用 AHP 专家定权与熵权法结合, 在充分定量考虑到数据间的变异程度的情况下, 定性考虑指标间的相对重要程度, 后将 4 个不同收入水平下的各 10 个样例国家在 2018 年的指标数据作为样本, 代入 TOPSIS 得分体系并进行排名, 最终结合上述三个评价得分体系进行分析, 将三种类型的评价结果中的每十个国家作为一类标准, 我们发现: 收入水平高的国家在经济水平与工业水平上一般较高, 但在森林植被面积上存在着 3 种标准的不唯一性, 如印度虽然作为中低收入国家, 但其经济水平与工业水平属于较高水准; 刚果金、赞比亚虽属于低收入国家, 但其森林植被面积仍然在世界范围内占比较大。

## 2.4 问题 4 的分析

问题 4 的提出即是为了向联合国介绍并展示我们对上述 3 个衡量国家各类指标的评价体系的优劣。并针对我们在问题 3 中的发现, 向当今世界上存在某些在“经济水平”、“工业水平”和“森林植被面积”间的不平衡的国家提出相应的建议, 并最后将我们的观点以报告书的形式向联合国递交。

## 三、模型假设

- 1 假设题目提供的数据均真实可靠;
- 2 假设中国在 2025~2035 年间继续实施退耕还林政策
- 3 假设 2018 年对应各国数据不受疫情形势下统计偏差的影响
- 4 假设模型在由于异常情况导致数据缺失的国家间仍可适用

## 四、定义与符号说明

符号	说明
$S^{(0)}$	中国森林面积原始数据
$JiB$	级比
$Z^{(1)}$	紧临近值生成序列
GDP	国民生产总值
entropy	信息熵
d	信息效用值
n	样本数
p	指标数
$\lambda$	特征值
c	特征向量
$y(x_0, x_i)$	灰色关联度函数
S	得分
Wei	综合指标权重
$p_{ij}$	概率矩阵

## 五、模型的建立与求解

### 5.1 针对问题 1 的模型建立与求解

我们分别建立的灰色预测模型与时间序列预测模型，对预测准确度进行比较后，选择最优模型对森林面积进行预测。

灰色系统是由华中科技大学邓聚龙教授于 1982 年提出的理论，其所衍生的灰色预测模型可以根据寻找得出的预测系统中灰色数据的内在规律，并且建立相应的微分方程模型，对数据的发展趋势进行预测，该模型对所需求的样本数据量较少，并且精度较高，被广为使用<sup>[1]</sup>。

时间序列预测是一种定量的预测方法，其利用过去的时间序列数据以及数据发展趋势的连续性，统计分析出数据在未来的发展趋势，同时还考虑到了随机因素的影响<sup>[2]</sup>，是一种较为完善，较为精确的预测模型。

#### 5.1.1 灰色预测模型的建立

Step1: 数据的检验

设中国 1990-2020 年森林面积数据为  $S^{(0)} = (S^{(0)}(1), S^{(0)}(2), \dots, S^{(0)}(n))$ 。数据只有满足准指数规律的检验，才符合 GM(1,1) 的使用前提。为此我们优先计算森林数据的级比，观察其是否都落在可容覆盖区间  $(e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}})$  中。

级比计算公式如下：

$$JiB_{(k)} = \frac{S^{(0)}(k-1)}{S^{(0)}(k)}, k = 2, 3, \dots, n$$

通过计算我们发现，森林面积数据满足级比检验要求，可以进行 GM(1,1) 模型的建立。

Step2: GM(1,1) 模型的建立

以  $S^{(0)}$  作为数据列建立灰微分方程:

$$S^{(0)}(k) + az^{(1)} = b,$$

用 OLS 法求出  $a, b$  的值之后, 得到相应的白微分方程:

$$\frac{dS^{(1)}(t)}{dt} + a S^{(1)}(t) = b$$

其解为:

$$S^{(1)}(t) = \left( S^{(0)}(1) - \frac{b}{a} \right) e^{-a(t-1)} + \frac{b}{a}, k = 1, 2, \dots, n-1$$

将  $t$  替换为  $k+1$ , 并作累减, 即可得到原始数据的预测值:

$$\hat{S}^{(1)}(k+1) = \left( S^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a}, k = 1, 2, \dots, n-1$$

$$\hat{S}^{(0)}(k+1) = \hat{S}^{(1)}(k+1) - \hat{S}^{(1)}(k), k = 1, 2, \dots, n-1$$

Step3:模型的检验

绝对残差:  $\varepsilon(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k=2, 3, \dots, n$

相对残差:  $\varepsilon_r(k) = \frac{|x^{(0)}(k) - \hat{x}^{(0)}(k)|}{x^{(0)}(k)} \times 100\%, k=2, 3, \dots, n$

平均相对残差:  $\bar{\varepsilon}_r = \frac{1}{n-1} \sum_{k=2}^n |\varepsilon_r(k)|$

### 5.1.2 时间序列预测模型的建立

在 IBM SPSS Statistics 26 软件中导入森林面积数据, 并定义日期与时间变量。创建时间序列预测传统模型, 根据专家建模器所给出的模型, 对数据进行预测。

### 5.1.3 模型预测准确度的比较与选择

我们将原始数据集中, 2016-2020 年的数据定为测试组, 两种预测方式对于测试组数据的残差检验计算结果下表, 预测结果见下图:

表 5-1 测试组预测结果对照表

预测方法	年份	绝对误差	相对误差	平均相对误差
灰色预测	2016	19960.35	0.009	0.016
	2017	26733.7	0.012	
	2018	33812.25	0.016	
	2019	41199.29	0.019	
	2020	48898.5	0.022	
时间序列预测	2016	2288.43	0.001	0.001
	2017	1716.72	0.001	
	2018	1144.68	0.001	
	2019	572.64	0	
	2020	0.6	0	

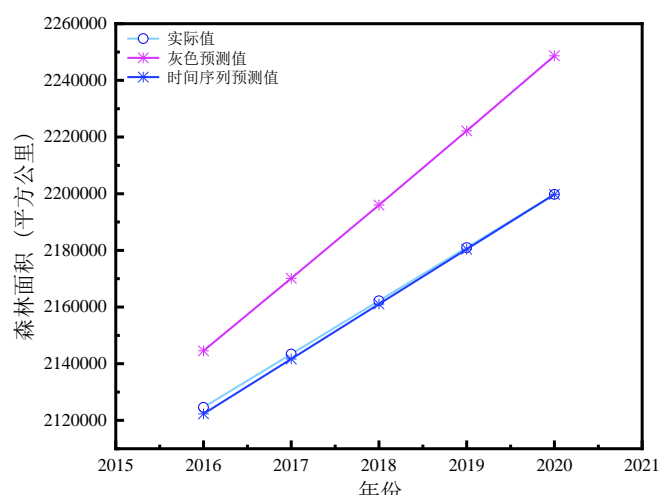


图 5-1 测试组预测结果对照图

通过以上的分析, 时间序列预测的预测结果准确度更高, 故我们选该模型进行预测。

### 5.1.4 中国 2025-2035 年森林面积的预测

将 1990-2020 的数据全都作为训练组, 用以上建立的时间序列模型进行预测, 得



到的预测结果如下表, 预测效果见下图。

表 5-2 预测结果表

年份	预测结果 (平方公里)
2025	2293760
2026	2312556
2027	2331352
2028	2350147
2029	2368943
2030	2387739
2031	2406535
2032	2425330
2033	2444126
2034	2462922
2035	2481717

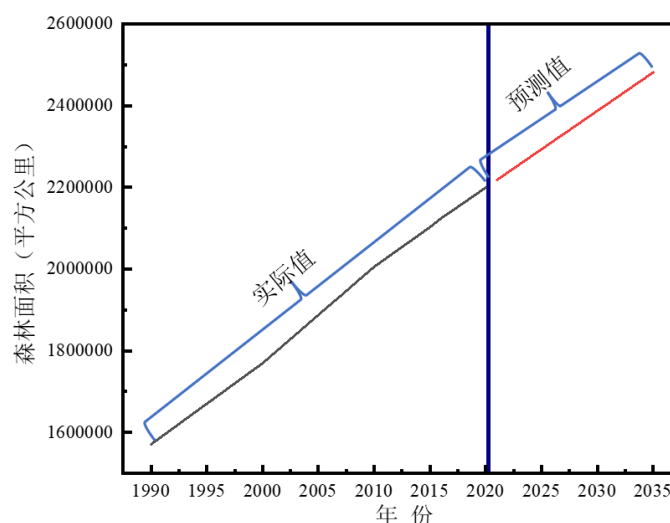


图 5-2 预测结果效果表

## 5.2 针对问题 2 的模型建立与求解

### 5.2.1 主成分分析模型

为了更加快速且准确的找出退耕还林政策对国家工业水平和国民经济的影响, 我们首先运用主成分分析模型对指标进行降维处理, 然后再通过得出的主成分对影响进行分析。

主成分分析法是一种常用的多变量分析方法, 其将原始变量线性组合成毫不相关的若干主成分来反映原始变量中的信息, 可根据个人需求选择主成分的个数, 是一种科学、便捷的多变量分析方法<sup>[3]</sup>。

Step 1: 首先分别计算工业水平和国民经济因素数据的相关系数矩阵

假设我们的数据有  $n$  个样本和  $p$  个指标, 则原始数据矩阵为:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \quad (n = 41, p = 4 \text{ 或 } 3)$$

通过公式:

$$R = \frac{\sum_{k=1}^n (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j)}{\sqrt{\sum_{k=1}^n (X_{ki} - \bar{X}_i)^2 \sum_{k=1}^n (X_{kj} - \bar{X}_j)^2}}$$

计算得出两个相关系数矩阵为:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (p = 4 \text{ 或 } 3)$$

Step2: 再用 Matlab 得到特征值  $\lambda$  和特征向量  $c$  的基础上, 运用以下公式计算得出主成分的贡献率和累计贡献率

$$\text{贡献率} = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k} \quad (i = 1, 2, \dots, p, p = 4 \text{ 或 } 3)$$

$$\text{累计贡献率} = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} (i = 1, 2, \dots, p, p = 4 \text{或} 3)$$

Step3: 通过累计贡献率碎石图确定要选择的主成分个数

### 5.2.2 灰色关联模型

灰色关联分析模型是根据数据曲线趋势的相似度来分析数列之间联系紧密程度的模型<sup>[4]</sup>, 其再非线性变量, 变量数据较少时, 相比传统数理统计模型拥有更好的适用度, 被人们广泛应用。

Step1: 对数据进行标准化, 选退耕还林度数据作为母序列 $x_0$ , 由主成分分析模型得出的主成分数据作为子序列 $x_m$ , 并计算关联系数

Step2: 定义

$$a = \min \max |x_0(k) - x_i(k)|$$

$$b = \max \max |x_0(k) - x_i(k)|$$

$$y(x_0(k), x_i(k)) = \frac{a + \rho b}{|x_0(k) - x_i(k)| + \rho b} \quad (\rho: \text{分辨系数}, i = 1, 2, \dots, m, k = 1, 2, \dots, n)$$

Step3: 定义灰色关联度函数并计算

$$y(x_0, x_i) = \frac{1}{n} \sum_{k=1}^n y(x_0(k), x_i(k))$$

### 5.2.3 模型结果与分析

相关性分析得到的各因素之间的相关系数如下图所示:

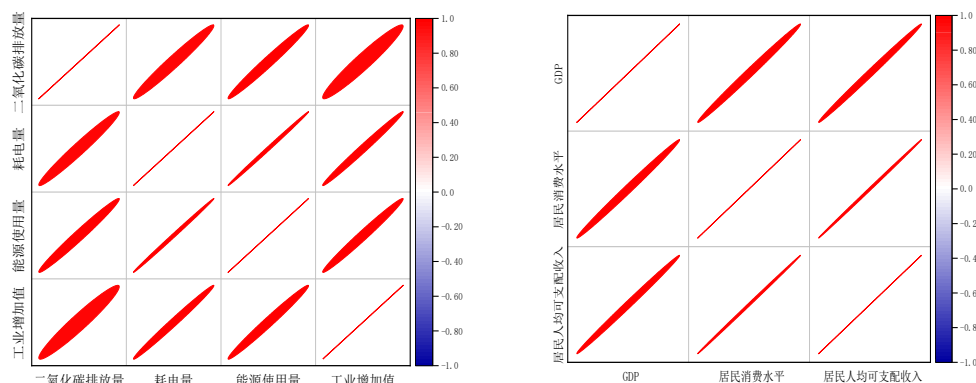


图 5-3 相关系数热力图

通过相关系数热力图我们可以看出, 因素之间存在极大的相关性, 如果直接进行对于指标的评价, 可能会造成评价结果不准确, 科学性不高的后果, 由此也可看出, 选择主成分分析模型的科学性与必要性。

主成分分析分析得到相关参数结果见下表:

表 5-3 工业水平主成分分析结果

特征向量	a1	a2	a3	a4
二氧化碳排放量 (千吨)	0.4966	0.7824	0.3638	0.0946
耗电量 (人均千瓦时)	0.5022	-0.1972	-0.4469	0.7136
能源使用量 (人均千克石油当量)	0.5024	0.0102	-0.5302	-0.6829
工业增加值	0.4987	-0.5907	0.6219	-0.1248
特征值	3.9491	0.0407	0.0091	0.001

贡献率	0.9873	0.0102	0.0023	0.0003
累计贡献率	0.9873	0.9975	0.9997	1

表 5-4 国民经济水平主成分分析结果

特征向量	$b_1$	$b_2$	$b_3$
GDP (现价美元)	0.576792	0.805242	0.137465
居民消费水平 (元)	0.57746	-0.52094	0.628615
居民人均可支配收入 (元)	0.577799	-0.2832	-0.76547
特征值	2.993162	0.006481	0.000358
贡献率	0.997721	0.00216	0.000119
累计贡献率	0.9977	0.9999	1

通过表格中累计贡献率可以看出,对于国家工业水平而言,只需要一个主因子就可以表达原始数据的 98.78% 的信息量;对于国民经济水平而言,也是只需要一个主因子就可以表达原始数据 99.77% 的信息量。同时还可以通过观察发现两个主因子对于各自的原始变量都拥有接近先等的正载荷,故我们将其解释为工业水平综合因子  $a_1$  和国民经济水平综合因子  $b_1$ 。

将得到的两个主成分的值和退耕还林度的对比图如下

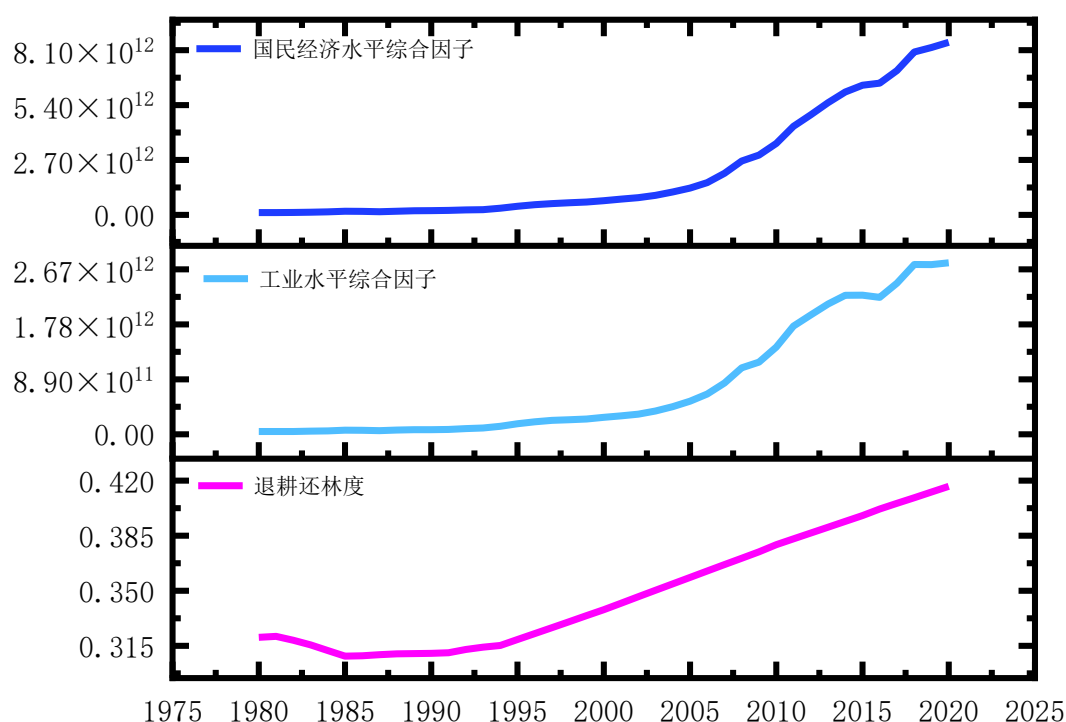


图 5-4 子序列与母序列趋势图

通过图形我们可以看出,三者均具有随年份增多的趋势,而且数据增长速度在 2002 年左右有一个迅速的提升。

运用灰色关联分析,我们得出工业水平综合因子与退耕还林度的灰色关联度为 0.6301,国民经济综合因子与退耕还林度的灰色关联度为 0.6272。由此我们可以得出,退耕还林政策对于我国工业实力和国民经济无不利影响,且具有一定的正向促进作用,在正式实行退耕还林政策之后,国家工业水平和国民经济水平的增长具有明显的小幅度提升。



### 5.3 针对问题 3 的模型建立与求解

#### 5.3.1 数据的选取与预处理

在对本问题的附件数据进行分析时我们发现, 部分国家在部分年份的相关数据存在一定程度上的缺失。且就问题 3 的目的而言, 建立评价模型的过程中若要采用不同年份的所有国家作为评价的方案, 则会在指标权重的确定上也将受到同一国家沿时间维度下的数据递增性的影响。因此首先在对评价方案的选取上, 我们考虑到已有附件数据中存在对各国收入水平的分级的基础上对数据量较为齐全的国家进行选取; 后就时间维度而言, 为避免全球疫情在近两年来的对各国经济与工业发展方面的不利影响, 我们决定采用 2018 年的相关指标数据进行模型的建立。最终在综合考虑数据完整性的情况下, 对每个收入标准下的各 10 个国家在 2018 年的相关数据作为构建后续评价模型的基础。

表 5-5 不同收入等级评价方案选择表

高收入	中高收入	中低收入	低收入
美国	中国	印度	莫桑比克
英国	巴西	蒙古	刚果金
法国	保加利亚	巴基斯坦	尼日尔
意大利	俄罗斯	印度尼西亚	苏丹
日本	苏里南	肯尼亚	埃塞俄比亚
澳大利亚	牙买加	尼日利亚	多哥
加拿大	古巴	越南	叙利亚
挪威	阿根廷	玻利维亚	也门
瑞士	马来西亚	伊朗	津巴布韦
沙特阿拉伯	土耳其	菲律宾	赞比亚

#### 5.3.2 针对经济水平的评价模型建立

经济水平作为衡量国家的经济命脉, 其影响指标较多, 但就本题而言, 我们首选采用 GDP 作为衡量各国经济水平的关键指标

Step1: 考虑到就经济水平而言, 能够侧面反映的指标过多, 且就 CPI、GDP 平减指数等硬性指标在世界范围内的获取存在极度的不完美性。而 GDP 作为衡量国家经济水平的综合指标, 对人口与国家生产总值等均具有较良好的代表意义, 因此我们采用将各样本国家的 GDP 进行标准化处理后的值作为评价各国经济水平的得分:

$$X = \begin{bmatrix} GDP_{11} \\ GDP_{21} \\ \vdots \\ GDP_{n1} \end{bmatrix} \quad n = 40$$

$$S_{i1} = \frac{GDP_{i1}}{\sqrt{\sum_{i=1}^n GDP_{i1}^2}} \quad (i = 1, 2, \dots, 40)$$

Step2: 将标准化后的得分进行排名, 并按照每 10 个国家作为一个标准进行定级, 此处仅列出每个评级中的前五个国家(此处对得分\*100 后列出, 因原数据量过小, 不易进行区分)

表 5-6 各国经济水平评价得分表

经济水平得分	国家	等级		国家	经济水平得分
78.663	美国	经济水平高	经济水平中低	巴基斯坦	1.201
53.028	中国			伊朗	1.123
19.223	日本			越南	0.936
11.071	英国			古巴	0.382
10.646	法国			肯尼亚	0.352
5.452	澳大利亚	经济水平中高	经济水平低	赞比亚	0.100
3.978	印度尼西亚			也门	0.082
3.002	沙特阿拉伯			叙利亚	0.082
2.971	土耳其			津巴布韦	0.069
2.807	瑞士			牙买加	0.060

5.3.3 针对工业水平的评价模型建立

工业水平作为决定国家发展的命脉，其主要评价指标往往较多，此处从附件中所给的 4 项指标入手，进行评价模型的构建

而层次分析法(Alytic Hierarchy Process/AHP)为美国运筹学家 T.L.Saaty 教授提出的一种多方案评价体系。即一种将评价过程进行拆分，将决策问题抽象成目标，准则和方案的三层综合评价体系，此处仅使用其将指标与指标之间构建判断矩阵并通过此进行指标权重计算的思想。

Step1：我们在初步对附件数据进行分析后决定采用“二氧化碳排放量”、“耗电量”、“能源使用量”与“工业增加值”作为衡量工业水平的 4 项指标。但就“工业增加值”一项的单位作为“占 GDP 的百分比”而言则存在一定的不合理性；即就某国家的工业增加值的占比较高，但其 GDP 总值较低，而其相对于另一 GDP 总值较高，且工业增加产值也很高的那类国家而言，其工业增加值占比存在明显优势。故附件中所给出的“工业增加值”能够一定程度上反应某国家在某一年的产业结构，却不能用于衡量工业水平的高低。

因此我们此处采用“工业增加产值”(即工业增加值的绝对数量，单位为美元)来作为衡量工业水平的指标之一。其余指标则作为极大型进行保留。

$$G = g \times GDP$$

(其中  $g$  为某国家的“工业增加值”，单位为占  $GDP$  的百分比， $GDP$  为某国家的国民生产总值)

Step2：对上述遴选出的四项衡量国家“工业水平”的指标进行判断矩阵的构建，并对其进行一致性检，(其  $\lambda_{max}$  为矩阵的最大特征值， $n$  为矩阵阶数，此时  $n=4$ )

$$CI = \frac{\lambda_{max} - n}{n - 1}$$

Step3：随后观察平均随机一致性指标表，并查找对应的 RI 代入公式

$$CI = \frac{CI}{RI}$$

表 5-6 平均随机一致性指标表

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49

可知该判断矩阵 $CR = 0.0052 < 0.1$ ,故满足一致性标准

**Step4:** 给出 AHP 重要程度对照表, 随机计算判断矩阵的指标权重, 并对其进行可视化处理

表 5-7 AHP 标度参照表

标度	含义
1	两指标同等重要
3	前者相对后者稍微重要
5	前者相对后者明显重要
7	前者相对后者强烈重要
9	前者相对后者极端重要
2,4,6,8	上述两指标间的中间值
倒数	后者相对前者的重要程度

表5-7 AHP判断矩阵表

	工业增加 产值	二氧化碳 排放量	人均耗电量	人均能源 使用量
工业增加 产值	1	2	3	5
二氧化碳 排放量	1/2	1	2	3
人均耗电量	1/3	1/5	1	2
人均能源 使用量	1/5	1/3	1/2	1

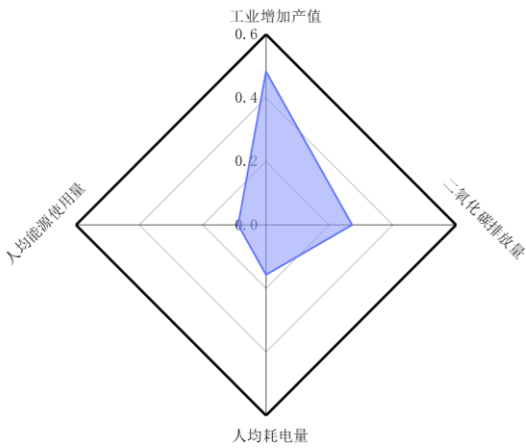


表5-8 AHP指标权重表

图5-5 权重可视化雷达图

指标	工业增加产值	二氧化碳排放量	人均耗电量	人均能源使用量
权重	0.4829	0.272	0.157	0.0881

我们在考虑到在使用 AHP 专家赋权法之后得到的上述权重之间虽然考虑到了指标的相对重要程度, 但对于各指标数据之间没有进行充分利用, 故此处我们考虑使用熵权法对所得到的指标权重进行进一步的修正, 以保证其在定性的考虑到了指标之间的重要性差异的同时, 定量的对指标数据的变异程度进行充分利用。

而熵权法(the entropy weight method EWM)作为一种能够充分挖掘数据量之间存在的变异关系以通过信息熵作为媒介的形式进行客观赋权的方法, 能够为我们在此处利用数据间的关系这一目标提供理论支撑。

**Step5:**首先构建一 40 行\*4 列的矩阵, 并对其进行标准化处理, 目的是消除掉由于不同指标之间单位的不同而产生的误差。(同 1.12 中对 GDP 的处理过程)

$$X = \begin{bmatrix} x_{11} & x_{12} & \vdots & x_{1m} \\ x_{21} & x_{22} & \vdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} \quad (n = 40 \quad m = 4)$$

$$Z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}} \quad (i = 1, 2, \dots, 40 \quad j = 1, 2, \dots, 4)$$

Step6: 随后对标准化后的矩阵进行概率矩阵 P 中每一个元素  $p_{ij}$  的计算

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & \vdots & Z_{1m} \\ Z_{21} & Z_{22} & \vdots & Z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \dots & Z_{nm} \end{bmatrix} \quad (n = 40 \quad m = 4) \quad p_{ij} = \frac{Z_{ij}}{\sum_{i=1}^n Z_{ij}}$$

Step7: 后对信息熵与信息效用值进行计算(其中 entropy 为信息熵, d 为信息效用值), 并最终计算指标熵权如下

$$entropy_j = -\frac{1}{\ln n} \sum_{i=1}^n p_{ij} \ln(p_{ij}) \quad (j = 1, 2, \dots, m)$$

$$d_j = 1 - entropy_j$$

$$wei = d_j / \sum_{j=1}^m d_j \quad (j = 1, 2, \dots, m)$$

表 5-9 熵权法指标权重表

指标	工业增加产值	二氧化碳排放量	人均耗电量	人均能源使用量
权重	0.383	0.350	0.161	0.105

随后对上述两种方法得到的指标权重进行平均汇总得到综合权重

表 5-10 综合指标权重表

指标	工业增加产值	二氧化碳排放量	人均耗电量	人均能源使用量
权重	0.433	0.311	0.159	0.097

在得到综合了指标数据信息与指标间的相对重要程度后, 我们考虑到就目标为将上述筛选所得的 40 个国家进行“工业水平”方面的排名, 以最终得到 4 个不同的标准; 于是我们考虑到使用 TOPSIS 方法对各国家的指标数据进行得分。

优劣解距离法(Technique for Order Preference by Similarity to an Ideal Solution/TOPSIS), 顾名思义, 为一种利用评价对象与最优或最劣解间所定义的距离建立相应的关系, 以定量对评价对象给出得分的方法。即认为某个评价对象以靠近最优对象或远离最劣对象为最佳, 并以此比值量化得分。但 TOPSIS 得分中默认各指标权重占比相同, 因此此处我们考虑对 TOPSIS 进行改进, 以得到将综合权重代入其中的得分体系

Step8: 沿用在熵权法定权中的标准化矩阵 Z, 并按列定义其中的最小值与最大值, 记录在如下向量中

$$Z^{max} = (Z_1^{max}, Z_2^{max}, \dots, Z_m^{max}) \quad Z^{min} = (Z_1^{min}, Z_2^{min}, \dots, Z_m^{min})$$

Step9: 分别定义第 i 个对象与最大值和最小值之间的距离, 并将综合权重记录在内, 完成 TOPSIS 赋予权重的修正。(其中 wei 为综合指标权重)

$$Dis_i^{max} = \sqrt{\sum_{j=1}^m wei_j (Z_j^{max} - z_{ij})^2} \quad Dis_i^{min} = \sqrt{\sum_{j=1}^m wei_j (Z_j^{min} - z_{ij})^2}$$

Step10: 随即计算 TOPSIS 得分, 并将得分进行排名。同理对“经济水平”的评价结果进行处理, 得到如下等级分类(即将评价得分\*100 的同时分为四个标准, 且每个标准中只保留前五项, 其余得分结果见附件)

$$S_i = \frac{D_i^{min}}{D_i^{max} + D_i^{min}}$$

表 5-11 各国工业水平评价得分表

工业水平得分	国家	等级		国家	工业水平得分
77.782	中国	工业水平高	工业水平中低	苏里南	5.369
56.411	美国			蒙古	4.116
26.419	挪威			越南	3.080
23.010	加拿大			古巴	2.626
20.850	日本			叙利亚	2.586
11.655	瑞士	工业水平中高	工业水平低	津巴布韦	1.424
10.315	意大利			赞比亚	1.264
10.031	英国			莫桑比克	1.014
9.090	伊朗			埃塞俄比亚	0.786
8.990	马来西亚			多哥	0.756

5.3.4 针对森林植被面积的评价模型建立

森林植被面积作为国家战略性储备与衡量环境建设好坏的重要参照, 其主要考虑森林的绝对占地面积与该国家森林的覆盖率。绝对的森林面积能够一定程度上决定该国家对于生态环境的改善能力, 而较高的森林覆盖率又能侧面展示该国家对于生态环境的重视程度与展现其对绿色经济发展的重要保障。

因此针对如上观点, 我们团队拟采用与在衡量“经济水平”时一样的做法, 即对指标数据进行标准化处理后进行得分的平均加权。

Step1: 经过较为系统的分析后, 我们得出将附件中的“森林面积”(平方公里)定义为某国家的“绝对森林占比”, 而将“森林面积”(占土地的百分比)作为该国家的“森林覆盖率”, 并对其二者进行标准化处理

$$A = \begin{bmatrix} a_{11} & a_{12} & \vdots & a_{1m} \\ a_{21} & a_{22} & \vdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \quad (n = 40 \quad m = 2)$$
$$Z_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^n a_{ij}^2}} \quad (i = 1, 2, \dots, 40 \quad j = 1, 2)$$

Step2:随后计算上述二者指标的加权平均得分, 最终参考对“经济水平”和“工业水平”的评价体系得分模式进行数据处理如下



表 5-12 各国森林植被面积评价得分表

森林植被面积得分	国家	等级		国家	森林植被面积得分
93.923	俄罗斯	森林面积高	森林面积较小	法国	14.447
69.590	巴西			意大利	13.997
47.221	加拿大			土耳其	13.751
41.877	美国			瑞士	13.330
41.744	苏里南			古巴	13.215
25.927	马来西亚	森林面积中高	森林面积小	苏丹	5.867
24.182	玻利维亚			英国	5.714
22.936	莫桑比克			蒙古	5.040
22.564	牙买加			伊朗	3.696
20.628	越南			肯尼亚	2.932

### 5.3.5 结果汇总

在分别对各国“经济水平”、“工业水平”和“森林植被面积”进行了系统的评价后得到了相对应的得分，并在此基础上进行了不同标准的划分。为发现各衡量标准下不同国家之间的差异，我们以国家的收入水平作为基准，后对以上三组评价标准进行横向对比，结果如下

表 5-13 不同标准下各国家横向对比表

排名	收入水平排名	经济水平排名	工业水平排名	森林植被面积排名
1	美国	美国	中国	俄罗斯
2	英国	中国	美国	巴西
3	法国	日本	挪威	加拿大
4	意大利	英国	加拿大	美国
5	日本	法国	日本	苏里南
6	挪威	印度	印度	刚果金
7	澳大利亚	意大利	俄罗斯	日本
8	加拿大	巴西	沙特阿拉伯	赞比亚
9	沙特阿拉伯	加拿大	澳大利亚	印度尼西亚
10	瑞士	俄罗斯	法国	中国
11	中国	澳大利亚	瑞士	马来西亚
12	巴西	印度尼西亚	意大利	玻利维亚
13	保加利亚	沙特阿拉伯	英国	莫桑比克
14	俄罗斯	土耳其	伊朗	牙买加
15	苏里南	瑞士	马来西亚	越南
16	牙买加	阿根廷	保加利亚	津巴布韦
17	古巴	挪威	巴西	澳大利亚
18	阿根廷	尼日利亚	土耳其	印度
19	马来西亚	马来西亚	印度尼西亚	保加利亚
20	土耳其	菲律宾	阿根廷	挪威

21	印度	巴基斯坦	苏里南	法国
22	蒙古	伊朗	蒙古	意大利
23	巴基斯坦	越南	越南	土耳其
24	印度尼西亚	古巴	古巴	瑞士
25	肯尼亚	肯尼亚	叙利亚	古巴
26	尼日利亚	埃塞俄比亚	牙买加	尼日利亚
27	越南	保加利亚	尼日利亚	菲律宾
28	玻利维亚	刚果金	巴基斯坦	多哥
29	伊朗	玻利维亚	菲律宾	埃塞俄比亚
30	菲律宾	苏丹	玻利维亚	阿根廷
31	莫桑比克	赞比亚	津巴布韦	苏丹
32	刚果金	也门	赞比亚	英国
33	尼日尔	叙利亚	莫桑比克	蒙古
34	苏丹	津巴布韦	埃塞俄比亚	伊朗
35	埃塞俄比亚	牙买加	多哥	肯尼亚
36	多哥	莫桑比克	肯尼亚	巴基斯坦
37	叙利亚	蒙古	苏丹	叙利亚
38	也门	尼日尔	也门	也门
39	津巴布韦	多哥	刚果金	尼日尔
40	赞比亚	苏里南	尼日尔	沙特阿拉伯

可见就“收入水平”处在第一梯队的国家，其对于工业水平与经济水平的评价排名也同时靠前；而对于中低收入国家的印度而言，在综合考虑其前沿领域科学技术与工业技术的发展史而言，其在分别针对经济水平与工业水平的得分排名中处在较靠前的位置也趋于合理。

且就针对收入水平对于经济水平和工业水平方面的映射而言，绝大部分收入水平较低的且处于第三或第四梯度的国家也同样在其他两个评价体系中处于较低的位次，但反观“森林覆盖面积”这一评价目标的高低排行，则相较于前两者并不难看出处于一个较为混乱的关系。

我们在存在考虑到不同国家的地理位置与相关经济水平后将依据上述 40 个国家方案中总结出了 6 种类型的国情现状来定性的作为分析某国时的衡量标准。

表 5-13 基于四项衡量标准的国家类型表

类型	定义
前列型	收入水平、经济水平与工业水平均位于世界前列，且环境方面存在一定的优势
发展型	收入水平、经济水平与工业水平较高，但往往受限于环境或依靠于环境
稳定型	收入水平、经济水平与工业水平具有一定实力，且环境方面不会受到太多牵制，可以进行小规模发展
转换型	收入水平、经济水平与工业水平均较低，但可以通过对环境的利用一定程度上获取资源与经济
受限型	收入水平、经济水平与工业水平均很低，且不能或几乎不可能利用环境资源进行产能转换
特殊型	因存在某些特异与其他地区的资源或人为活动，且能依此进行发展

结合上述各国家针对不同评价对象的分类中,“前列型”、“发展型”与“稳定型”的国家在上述包括“经济水平”、“工业水平”和“森林植被面积”上均存在发展的能力,而诸如“转换型”与“受限型”的部分国家则在上述能力中存在一定不完美性。尤其是针对部分“转换型”国家,其丰富的森林植被面积或许能成为这类国家发展的相关支柱。

#### 5.4 问题 4 的解决

尊敬的联合国政府:

我们很荣幸,也十分感激的收到了来自“天府杯”赛事方面的邀请,来向您递交一份报告书以关于我们在各国家间经济水平与环境方面的发现。

这里再次感谢天府杯方面能够给予我们一个机会向正在读这封信的您传达我们最新的发现。纵观发展,为实现环境保护与绿色经济链的循环发展,大部分国家对此均或多或少的进行了政策的提出或工作内容的宏观向导。但就近年来世界多联贸易体系的构建与工业水平的进步,多数国家为了实现发展的目的而不惜牺牲诸如国家森林面积、空气质量等环境指标来推进经济水平的进步;可也有些国家没有适应上目前经济世界的新格局,其各项相关经济水平与工业水平的衡量指标仍处在较低水平。为了进一步明确世界各国目前在经济发展与环境保护方面的关系,我们团队在通过初步探究了中国的退耕还林政策与国民经济水平和工业发展之间的关系后,成功的将其中正向的反馈推广到了对世界各国包括“经济水平”、“工业水平”与“森林植被面积”在内的评价模型的映射。以三项基本评价目标分别构建特异与其他衡量国家绿色经济与发展的评价体系。随后我们考虑到使用世界公认标准对待评价的国家进行遴选,决定以“收入水平”作为区分世界国家梯度的依据,以获得国家数据样本的原则,以美国、中国、印度、莫桑比克分别为“高收入水平”、“中高收入水平”、“中低收入水平”和“低收入水平”国家的代表,总计 40 组样本数据,来分别针对上述三个方面进行评价;随后我们团队使用了基于包括 AHP、EWM 与 TOPSIS 在内的多种算法来构建我们的评价模型,并得出如上 40 个样本国家在“经济水平”、“工业水平”与“森林植被面积”方面单独的得分,进行了排名与可视化处理。

经过上述系统而客观的对我们所采用的 40 个世界各地国家在经济与环境水平的评价之后,我们发现了世界各梯度之间存在的显而易见的规律与差异。在得到数据之后,我们团队首先依据我们的“独体评价体系群”进行了国家类型的划分,得到包括“前列型”、“发展型”、“稳定型”、“转换型”、“受限型”与“特殊型”在内的六种国家类型。于是我们发现,就处在“前列型”与“发展型”阶段的国家其工业程度、经济实力以及其在环境方面的资源均能做到基本稳定,且就对经济发展对本国环境的取代方面也能在较长时间内不会造成较为严重的危害。而处在“稳定型”位列的国家虽然能够及时的适应世界经济格局的进程,能够自身的在贸易与工业潮流中生存,但就其本国内的环境资源问题往往会在下一阶段显现出严重的反馈。针对此种状况,我们建议此类国家在政策抉择上进行改革创新,就针对自身在森林植被覆盖面积不足方面的缺陷,可以借鉴中国在 2003 年提出的“退耕还林”政策,并加大环境资源向经济水平方面的转换,以谋求真正的绿色发展;“转换型”的国家往往作为收入水平低的“第三世界”国家,其国内经济水平与工业发展基础远远落后于目前世界标准。但作为一类地理位置上又有着无可替代的“森林大国”,其在下一阶段内可以也必要的会利用其自身森林植被面积方面的优势去弥补在经济与工业水平的不足来为本国公民提供更好的生活条件。具体措施包括建立与国际接轨的碳中和交易市场,充分利用自身在环境方面的优势促进本国与相邻国家的发展,同时深度挖掘环境资源的多重价值,逐渐走

向“特殊型”国家的道路；而“受限型”的国家作为一类既在基础经济水平与工业发展基础上落后于世界平均水平，又处于较差的地理位置上，“孤立无援”的面对世界发展的激流，由于本国国民文化基础与资源或缺的限制，也很难在短时间内进入下一阶段，虽作为政府或联合国会对此类国家进行过多的干预保护，但就发展而言则没有可利用的优势可言。而基于此类国家的发展离不开世界范围内的关注，就本国而言，可以先从自身宣传与文化艺术或潜在资源的挖掘入手，寻求“受限”的突破点，并引起国际关注，助力发展；最后作为一类能够利用本国特有环境或资源来谋求发展的“特殊型”国家，其本质上作为世界某项资源或技术的提供者而言，具有较为稳定的发展前景与较为成熟的转化体系，部分国家也曾依靠此项优势进行了多次较大规模的进步，就经济与工业水平而言均处在世界前列，特殊型的国家就本国森林或环境资源存在极度的不均一性，有些国家能将森林资源作为其向特殊型转换的资本，另一些国家则完全不需要考虑环境方面的影响，依旧可以利用自身海洋、矿产资源乃地理位置方面的优势进行国民经济水平的稳步提升与发展。

以上便是我们团队在基于对各国经济与环境所建立的“独立式评价群”的发现与建议提供，而就当今全球范围内的环境保护与绿色经济的发展，也希望联合国政府能够持续关注并贡献出属于全人类的一份力量！



2022.3.21  
天府杯参赛团队

## 六、模型评价与推广

模型的评价与推广作为模型实用性的为一检验标志，对于模型的整体性要求较高，我们团队在充分考虑到上述模型的可行性与缺陷性后给出如下建议：

### 6.1 模型的优点

(1)文中的模型能够一定程度上解决了问题中所存在的疑问，能够较为清楚的给出针对不同结果的答复

(2)能够一定程度上确保对大量数据的利用程度，将数据价值进行了发挥。

(3)在评价阶段也基本满足世界格局下各国家之间的客观差异，基本没有出现偏差

### 6.2 模型的缺点

(1)与假设条件中进行了对预测年份的干预，对于预测成果的客观性存在不足

(2)评价模型建立的过程中对于数据缺失的国家没有进行过多的干预，导致评价样本咋缺少假设的情况下不一定具有普适性。

(3)就问题3中最后对问题的发现中没有充分的对得分排名与量化指标进行挖掘，可能存在更多发现。

### 6.3 模型的改进

(1)数据利用方面，可以在数据量不足的情况下对缺失的数据进行挖掘，将较为科

学的指标数据更多的介入评价。

(2)结论解释方面, 可以从由模型得出的得分排名中进行更多量化指标的拆分, 以发现更多信息。

## 七、参考文献

- [1]基于灰色系统 GM ( 1, 1) 模型的湖北省卫生人力资源需求预测研究  
 [2]管理决策方法——理论、模型与应用 河海大学出版社 吴丹 2014 年  
 [3]韩小孩,张耀辉,孙福军,王少华. 基于主成分分析的指标权重确定方法[J]. 四川兵工学报,2012,33(10):124-126.  
 [4] 刘思峰. 灰色系统理论及其应用(第五版)[M]. 北京: 科学出版社, 2010: 62.

## 八、附录

附件类型		
问题 1 数据处理与结果呈现表	sheet1	中国
	sheet2	时间
	sheet3	测试集
	sheet4	预测结果
问题 2 数据处理与结果呈现表	sheet1	工业水平综合因子
	sheet2	国民经济水平综合因子
	sheet3	退耕还林度
	sheet4	灰色关联
问题 3 数据处理与结果呈现表	sheet1	国家汇总
	sheet2	高收入
	sheet3	中高收入
	sheet4	中低收入
	sheet5	低收入
	sheet6	评价结果
	sheet7	得分表
	sheet8	全体得分
	sheet9	得分对比表
spss 时间序列输出结果	1	训练集结果
	2	最终预测结果
matlab 文件	灰色关联分析代码	