

2022 第二届天府杯全国大学生数学建模竞赛论文

题 目 基于数学模型的绿色发展政策评估与探索

摘 要:

随着人类社会的发展和工业水平的提高,发展与环保似乎背道而驰。发展和环保永远会是对立面吗?“绿水青山就是金山银山”的号召告诉我们,不是的。因此,人类需要在经济增长、工业进步和环境保护之间找到一个均衡点,从非绿色经济迈向绿色经济。本文通过一系列数学模型对环保相关政策进行分析和评估,探究发展和环保之间的“纳什均衡”。

对于问题一 我们利用题目中给出的附件,首先建立了**霍尔特线性趋势预测模型**对森林面积和森林面积占比进行预测,该模型的拟合优度为 0.9999,预测结果见正文。同时考虑到线性趋势预测模型存在预测值偏高的隐患,我们又建立了**灰色 Verhulst 模型**进行预测,该模型考虑到了森林增长的饱和态,拟合优度为 0.9827,预测结果见正文。

对于问题二 我们首先从世界银行上面搜集数据,从多个角度探讨“退耕还林”政策。对国民经济和国家工业水平的影响途径。然后建立了不完全预测模型进行影响的定量评估。为了使得评估取得更好的效果,我们先后通过使用**熵权 Topsis 模型**建立综合指标、对评价量进行有界无量纲化的方式对模型进行修正。最后得到的结论为:“退耕还林”政策对国民经济影响的评分为 0.22,对工业化水平影响的评分为 0.05,对综合指标影响的评分为 0.57。

对于问题三 我们对于不同国家建立了**集合论模型**,使用正态分布对不同国家的不同指标划分集合。在经济层次上得到了低收入国家集合、中低收入国家集合、中高收入国家集合和高收入国家集合;在工业水平层次上得到了低水平工业国家集合、中等水平工业国家集合、高水平工业国家集合;在森林面积层次上,得到了低森林面积国家集合、中等森林面积国家集合、高森林面积国家集合。基于集合的划分,我们使用**雅卡尔指数矩阵**交叉评价不同集合分类指标之间的相似性。

对于问题四 我们结合问题一到问题三中挖掘到的信息和发现的结论,写成研究报告,并采用可视化的逻辑模型进行表述,层次清晰地提出了四个具体举措。

关 键 词 环境保护 预测 不完全预测模型 集合论模型 绿色经济

目录

1 准备工作	2
1.1 问题分析	2
1.2 模型假设	2
1.3 符号约定	2
2 问题一的建模与求解	3
2.1 数据的初步分析	3
2.2 基于 Holt 线性趋势预测算法的森林面积预测模型	3
2.2.1 模型的理论基础	3
2.2.2 模型的参数选择	4
2.2.3 模型的运算	4
2.3 基于灰色 Verhulst 的饱和态预测模型	5
2.3.1 模型的理论依据	5
2.3.2 模型的运算	6
3 问题二的建模与求解	7
3.1 “退耕还林”政策的影响因子分析	7
3.1.1 “退耕还林”对农业造成了严重冲击吗?	7
3.1.2 “退耕还林”客观拉动了生态林旅游业的发展	8
3.1.3 “退耕还林”通过提高环境指标间接影响工业发展	8
3.1.4 “退耕还林”客观上促进经济结构改变	9
3.1.5 其它影响因子概述	10
3.2 基于不完全预测的影响评价模型	10
3.2.1 模型的理论依据	10
3.2.2 模型的运算	10
3.3 综合指标对模型的修正	13
3.3.1 基于熵权 Topsis 模型构建的综合指标	13
3.3.2 基于综合指标对模型的修正	14
3.4 有界量对模型的修正	15
4 问题三的建模与求解	16
4.1 模型的理论依据	16
4.1.1 使用正态分布进行分类	16
4.1.2 使用雅卡尔指数评价集合相似度	16
4.2 对国家各个指标的挖掘与分类	17
4.3 不同指标间的交叉分析	17
5 研究报告书	19
6 模型分析与评价	21
参考文献	22

1 准备工作

1.1 问题分析

对于问题一 问题一要求预测 2025-2035 年的森林面积数据，应当首先绘制前一段时期的森林面积趋势曲线，根据曲线选择线性模型或非线性模型，并进行精度检验。

对于问题二 问题二要求分析和评价“退耕还林”政策对国民经济和工业水平的影响，首先要做的便是分析“退耕还林”政策以哪些途径产生影响。接下来可以借助“控制变量”的思想，假设“该政策没有出现”并进行预测，再将预测数据与实际数据比较得出结论。

对于问题三 该问首先需要做的便是将不同国家按不同指标进行分类评价，考虑到每个类别里面的国家具有互异性，唯一性和确定性，因此可以将每一类都抽象成一个集合，用离散数学的方法交叉分析集合之间的相似程度。

对于问题四 该问题需要撰写研究报告，应当充分利用前面几个小问挖掘到的信息和发现，结合正确的价值观念，提出具体的措施。为了使得逻辑上更加清晰，可以使用逻辑可视化的方法表达思路。

1.2 模型假设

- **假设 1** 附件中给出的数据与我们搜集到的数据是真实可靠的
- **假设 2** 2025-2035 年国家林木政策不会发生重大变革
- **假设 3** 不同国家的不同指标服从正态分布或近似服从正态分布

1.3 符号约定

符号	意义
$Y(t)$	某项指标的实际值
$\hat{Y}(t)$	某项指标的预测值
$\Delta X(t)$	序列 $X(t)$ 在 t 时刻的差分
α	霍尔特线性趋势预测模型水平平滑系数
β^*	霍尔特线性趋势预测模型趋势平滑系数
l_t	霍尔特线性趋势预测模型用到的中间量
b_t	霍尔特线性趋势预测模型用到的中间量
h	霍尔特线性趋势预测模型用到的中间量
R^2	模型的拟合优度，也称决定系数
$J(A, B)$	集合 A 与集合 B 的雅卡尔系数
$ A $	集合 A 的基数，正文中即 A 的元素个数

表 1: 模型的符号约定

* 注：文中出现的其他符号含义均在文中相应位置给出

2 问题一的建模与求解

2.1 数据的初步分析

附件中的数据给出了 1990-2020 年中国国内的森林面积变化数据和森林面积占土地面积百分比的数据。通过对数据的初步观察，我们发现：**森林面积随着时间增长的趋势呈现线性**。由基本的数学知识可知，线性函数的导数为常数，而在离散的数据中，我们使用差分来描述变化率。记 $\Delta X(t)$ 为离散序列 $\{X(t)\}$ 在 t 时刻的差分，则：

$$\Delta X(t) = X(t) - X(t-1)$$

下面两幅图分别展示了森林面积变化曲线和森林面积百分比变化曲线，每幅图中的绿色曲线为对于曲线的差分序列。

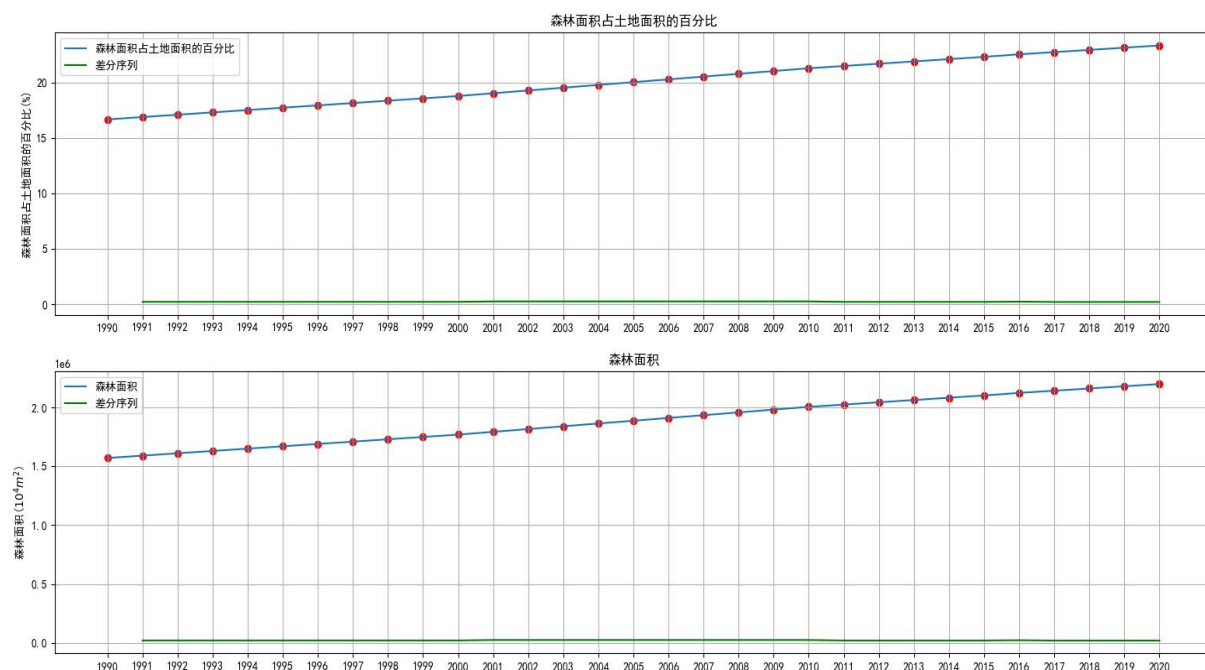


图 1: 中国近二十年森林数据曲线

观察图像，可以得到如下结论：

结论 (1) 差分序列接近一条直线，因此可以认为我国近几年森林面积的增长是**线性的**。

结论 (2) 差分序列与原数据曲线相距很远，说明差分数值与原始数据相比相对微小，即使差分序列不是严格的直线，我们也可以近似认为原曲线是线性趋势增长的。

2.2 基于 Holt 线性趋势预测算法的森林面积预测模型

2.2.1 模型的理论基础

Holt 线性趋势预测算法是一种时间序列预测模型，由 Holt 改进经典指数平滑模型而诞生。改进之后的算法相较于经典指数平滑算法，可以很好的预测**线性趋势的时间序列**。

Holt 线性趋势预测模型由**预测方程**、**水平方程**、**趋势方程**组合而成。其中**预测方程**表示为：

$$\hat{y}_{t+h|t} = l_t + hb_t$$

水平方程表示为：

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad \alpha \in (0, 1)$$

趋势方程表示为：

$$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_t \quad \beta^* \in (0, 1)$$

其中参数 α 称为模型的水平平滑参数，参数 β^* 称为模型的趋势平滑参数。这两个参数是需要建模者根据先验经验估计得到。

2.2.2 模型的参数选择

不同的 α 取值和 β^* 取值会直接影响到 Holt 线性趋势预测模型的效果，因此需要调整最适合数据的参数。我们使用 Python 程序设计语言编写算法，并通过**网格搜索算法**搜索最佳参数组合。

参数网格由 $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ 和 $\beta^* \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ 生成。筛选指标为模型对 1990-2020 年森林面积以及森林面积占比数据的拟合优度 R^2 。

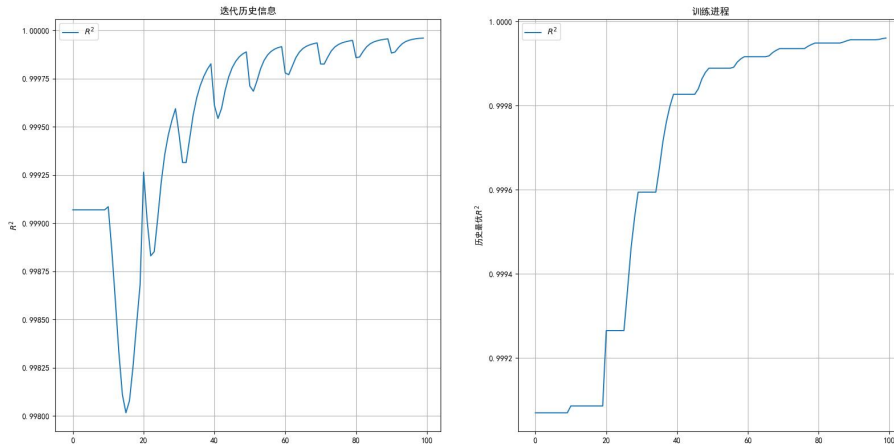


图 2: 网格搜索调整 Holt 线性趋势预测模型参数的训练进程

最终的调参结果为 $(\alpha, \beta^*) = (0.9, 0.9)$ ，对应的拟合优度达到了 $R^2 = 0.9999$ 。体现了模型效果的优越性。

2.2.3 模型的运算

根据网格搜索的最优参数组合构建模型，并向下预测，预测图像如下所示：

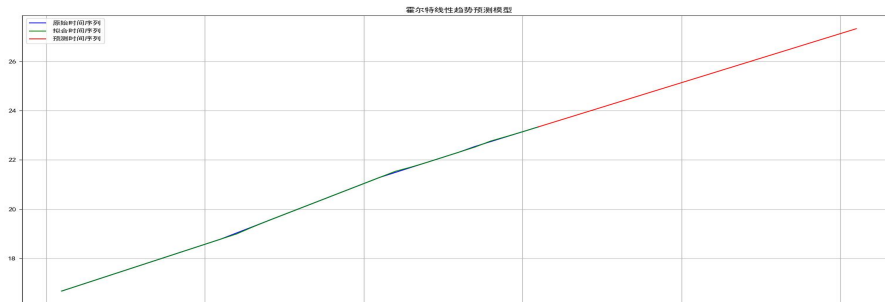


图 3: 模型的预测结果

最终的预测结果见下表：

表 2: 2025-2029 年间森林面积数据预测

年度	2025	2026	2027	2028	2029
森林面积	2387552.2	2406330.0	2425107.9	2443885.7	2462663.5
百分比	25.33%	25.53%	25.73%	25.93%	26.13%

年度	2030	2031	2032	2033	2034	2035
森林面积	2481441.3	2500219.1	2518996.9	2537774.8	2556552.6	2575330.4
百分比	26.33%	26.53%	26.73%	26.93%	27.13%	27.33%

* 其中森林面积的单位为万平方千米

2.3 基于灰色 Verhulst 的饱和态预测模型

2.3.1 模型的理论依据

Holt 线性趋势预测模型具有良好的拟合效果，但是线性趋势预测模型往往使得预测值偏高，这是由于自然界大多数增长现象往往都会随着规模增长而速率下降。为此我们引入了 **灰色 Verhulst 预测模型**作为修正。

灰色 Verhulst 预测模型是灰色系统理论中的重要预测模型之一，该预测模型的一大亮点在于，它考虑到了增长的饱和状态，使得预测曲线呈现出 S 型。

灰色 Verhulst 的建模步骤如下：

步骤一 根据原始序列分别计算累加序列。设原始序列为 $\{X^{(0)}(t)\}$ ，则其累加序列为

$$\{X^{(1)}(t)\} = \left\{ \sum_{i=1}^t X^{(0)}(i) \right\}$$

步骤二 对累加序列 $\{X^{(1)}(t)\}$ 进行平均生成

$$\{Z(t)\} = \left\{ \frac{X^{(1)}(t+1) + X^{(1)}(t)}{2} \right\} \quad 2 \leq t \leq n$$

步骤三 构造灰微分方程，称 $X^{(0)}(t) + aZ(t) = bZ^2(t)$ ，其对应的微分方程为：

$$\frac{d}{dt}X^{(1)}(t) + aX^{(1)}(t) = b(X^{(1)}(t))^2$$

步骤四 使用最小二乘法估计参数 a 和参数 b 的值

$$B = \{-Z(t), Z^2(t)\}^T \quad 2 \leq t \leq n$$

$$Y = \{X^{(0)}(t)\}^T \quad 2 \leq t \leq n$$

$$[a, b]^T = (B^T B)^{-1} B^T Y$$

步骤五 将参数带入原微分方程，构造出 Verhulst 模型的时间相应函数：

$$\hat{X}^{(1)}(k+1) = \frac{aX^{(0)}(k)}{bX^{(0)}(k) + (a - bX^{(0)}(k))e^{ak}}$$

$$\hat{X}^{(0)}(k+1) = \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k)$$

2.3.2 模型的运算

按照上述步骤构建灰色 Verhulst 模型并向后预测，做出图像如下所示：

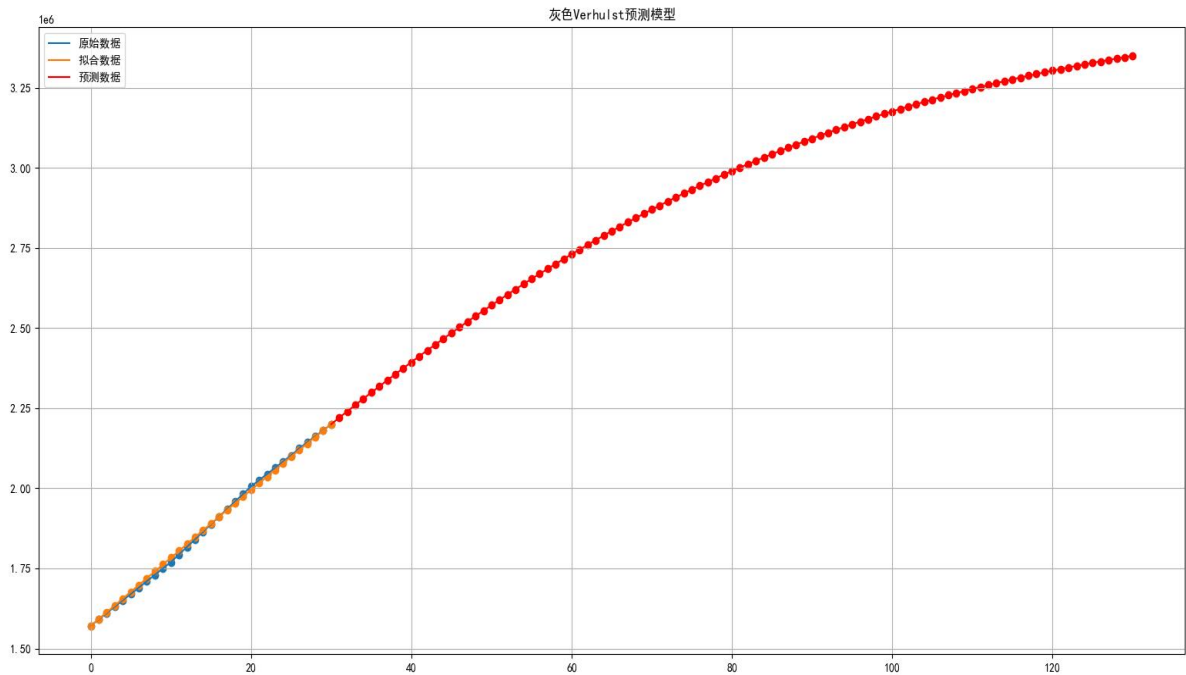


图 4: 中国近二十年森林数据曲线

经过计算，模型拟合优度 $R^2 = 0.9827$ ，平均误差率为 0.31%。并且由图可知，灰色 Verhulst 预测模型考虑到了森林面积增长的饱和状态，可以将该结果作为对 Holt 线性趋势预测模型的修正。最终的预测结果见下表：

表 3: 2025-2029 年间森林面积数据预测

年度	2025	2026	2027	2028	2029
森林面积	2393721.7	2412239.9	2430595.2	2448784.4	2466804.5
百分比	25.40%	25.59%	25.79%	25.98%	26.17%

年度	2030	2031	2032	2033	2034	2035
森林面积	2484652.7	2502326.3	2519822.8	2537139.9	2554275.3	2571227.0
百分比	26.36%	26.55%	26.74%	26.92%	27.10%	27.28%

* 其中森林面积的单位为万平方千米

3 问题二的建模与求解

3.1 “退耕还林”政策的影响因子分析

3.1.1 “退耕还林”对农业造成了严重冲击吗？

“退耕还林”政策是在 1999 年开始试点，2002 年全面推广。顾名思义，“退耕还林”政策要求适当削减农业用地，还原为林地。身为该政策的主角耕地，自然是第一影响因子。

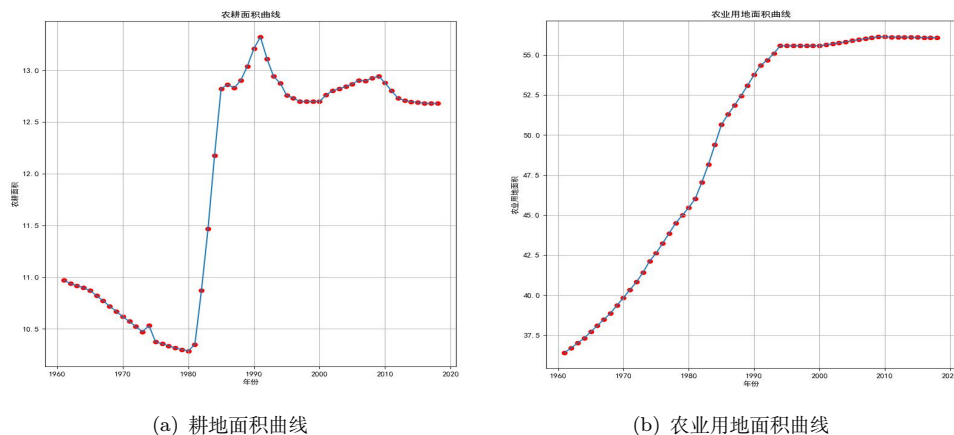


图 5: 1960-2020 年耕地和农业用地面积变化曲线

由图可知，从整体上来讲，“退耕还林”政策确实影响到了耕地面积和农业用地面积的发展趋势。那么接下来再来看我国粮食产业的变化曲线。

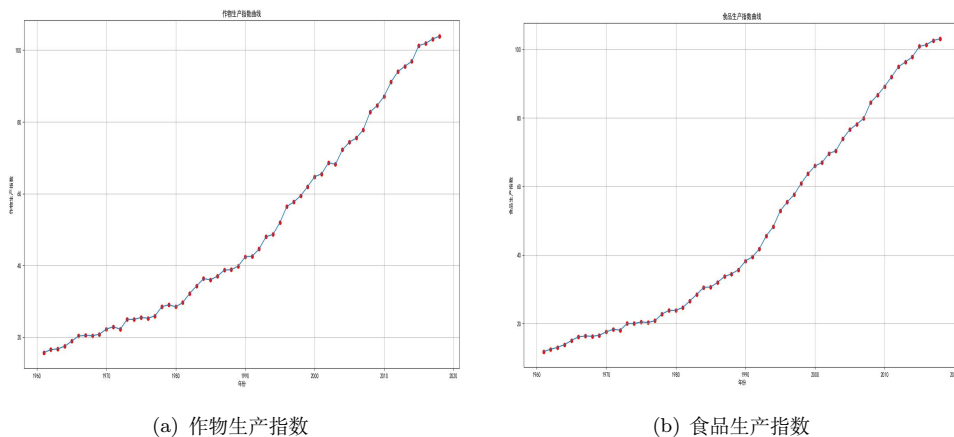


图 6: 1960-2020 年作物生产指数和食品生产指数变化曲线

我们以 1999 年作为分水岭，将数据截断，并用差分来计算序列的增长率，对差分序列进行统计性描述。1999 年之前，作物生产指数的平均增长速率为 1.3663 单位每年；1999 年之后，作物生产指数的平均增长速率上升到 2.0628 单位每年，增长了 50.98%。1999 年之前，食品生产指数的平均增长速率为 1.2179,1999 年之后，该数值上升到 2.1744，增长了 75.20%。

因此从作物生产指数和食品生产指数的角度来看，“退耕还林”政策并没有严重冲击到我国粮食产业。究其原因，我国的科技发展迅速，农业和粮食产业科技化含量增高，生产效率大大提高，成功应对了“退耕还林”政策可能带来的负面影响。

3.1.2 “退耕还林” 客观拉动了生态林旅游业的发展

绿水青山就是金山银山，山清水秀永远都是“驴友”们的向往。客观上来说，“退耕还林”政策促进了生态的恢复，有利于开发生态林旅游业，进而拉动经济增长。

由于生态林旅游业具有地域差异，因此我们选取塞罕坝林场进行案例分析。下图显示了 2003-2019 年塞罕坝林场的旅游收入和森林面积变化。

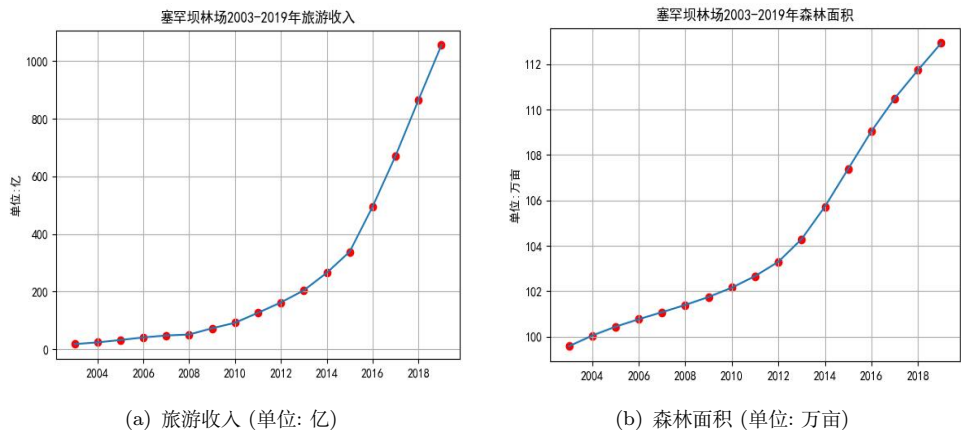


图 7: 2003-2019 年塞罕坝林场的旅游收入和森林面积变化

为了体现二者之间的关联性，分别使用线性回归模型和非线性回归模型进行回归分析。

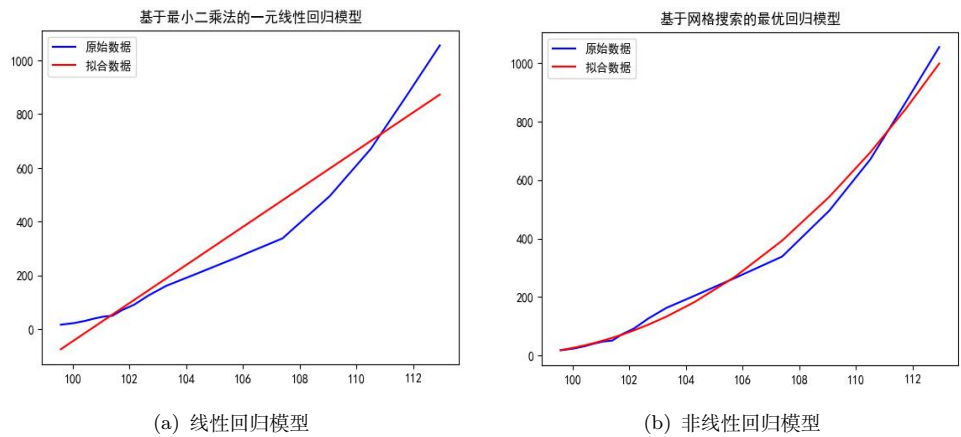


图 8: 2003-2019 年塞罕坝林场的旅游收入和森林面积回归分析结果

令变量 X 代表塞罕坝地区的森林面积， Y 代表塞罕坝林场的旅游收入。

线性模型得到的结果为 $Y = 0.0132X + 100.8575$ ，拟合优度 $R^2 = 0.9382$ ， p 值数量级为 10^{-5} 。非线性模型得到的结果为 $\sqrt{Y} = 0.0097X^2 - 91.5756$ ，拟合优度为 $R^2 = 0.9937$ ， p 值数量级为 10^{-5} 。该结果体现了森林面积和生态林旅游收入之间的高相关性。

3.1.3 “退耕还林” 通过提高环境指标间接影响工业发展

“绿水青山就是金山银山”这句话不仅仅体现了生态林旅游业的价值，这句话从根本上更新了人们对于工业发展和环境保护的认知。在绿色经济时代，环境指标是衡量国家工业化水平的重要考量，一个强大的工业体系必然能在环境保护和工业化之间找到互利共赢的平衡点。

随着社会环境保护意识的增强，越来越多的地方政府开始根据工业指标制定工业行业规范，宏观调控工业化和环境保护的平衡点。

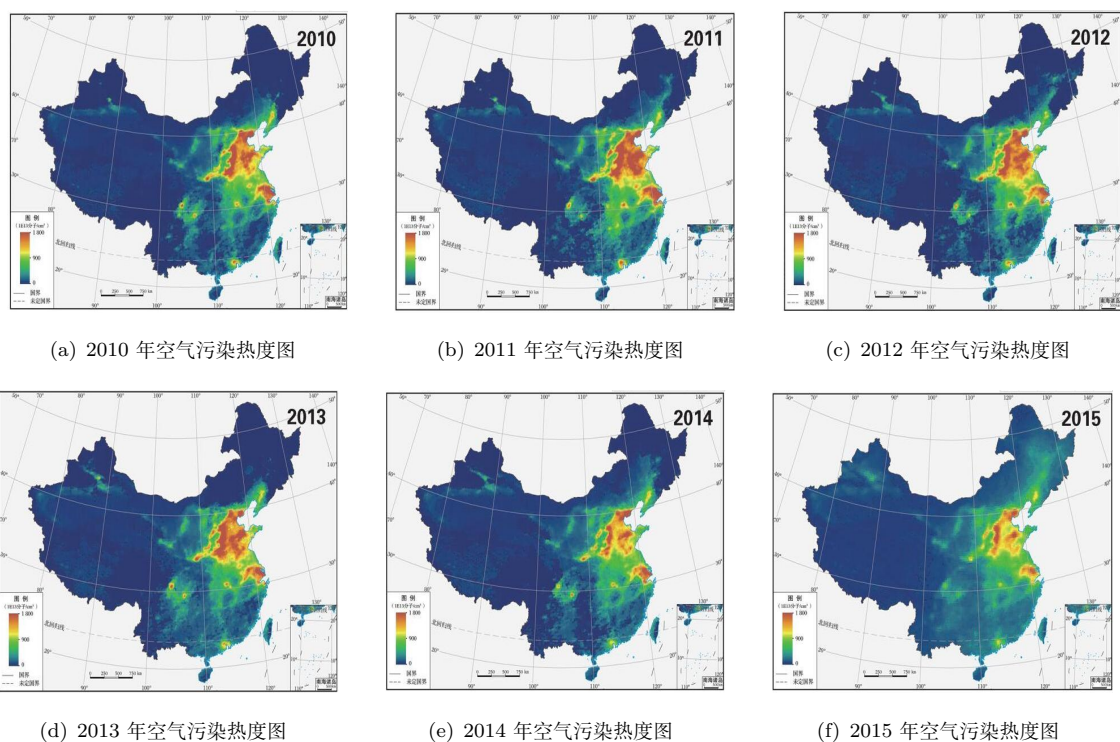


图 9: 2010-2015 年空气污染热度图 (图片选自澎湃新闻)

上面一组图片体现了我国在环境治理方面的成绩，环境污染减少将为工业发展带来更多空间，有利于拉动工业发展。

3.1.4 “退耕还林” 客观上促进经济结构改变

“退耕还林”政策客观上促进了农村人口向城市的迁移，有利于改变经济结构，客观上有利于增加就业率，提高劳动力水平从而拉动国民经济和工业化水平的发展。

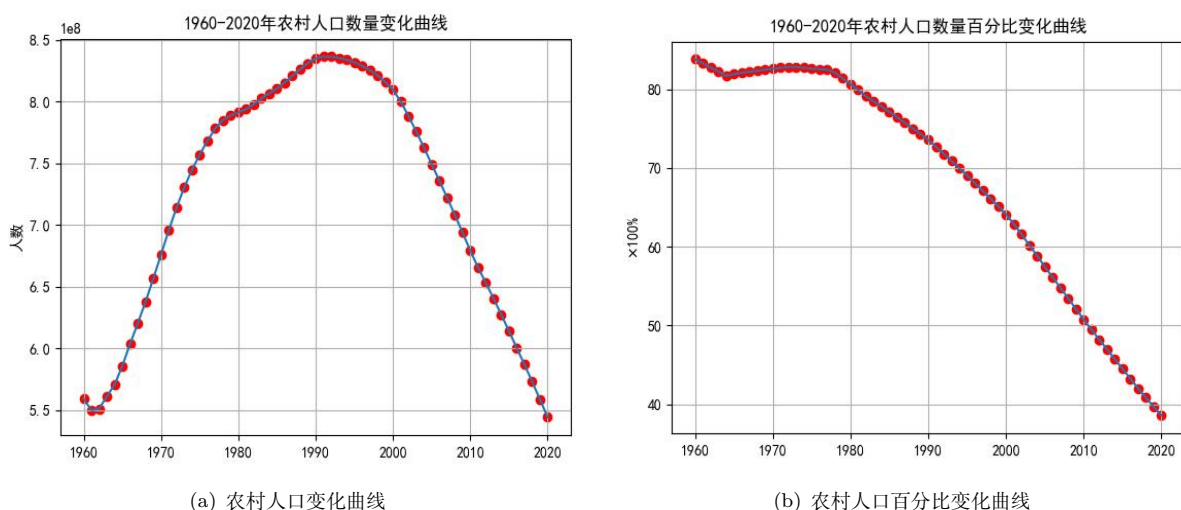


图 10: 1960-2020 年农村人口变化曲线

由图可知，在“退耕还林”政策出台的时间点，农村人口数目出现了**突变点**，同时农村人口占比开始更大幅度下降，这间接地体现了“退耕还林”对于中国城乡人口结构的改变。接下来从就业

的角度来看：

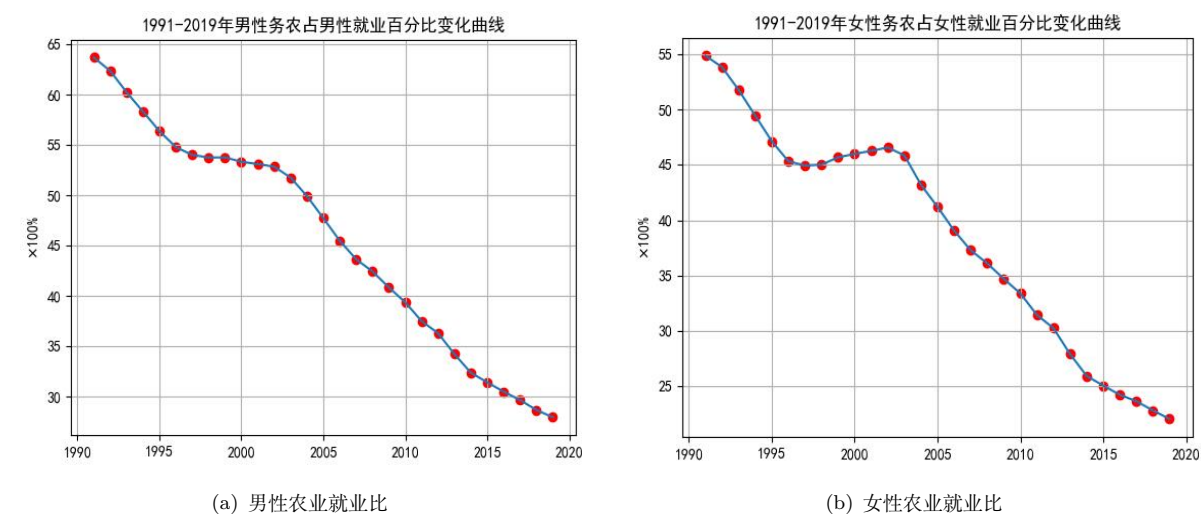


图 11: 1990-2019 年农业就业比分析

由图可知，“退耕还林”政策出台之后，男女务农人员的比例进一步下降，对应的城乡劳动力增加。对于农村而言，有利于促进脱贫致富；对于城镇而言，有利于工业化进程和国民经济的进一步发展。

3.1.5 其它影响因子概述

除了上所述的各种影响因子之外，“退耕还林”政策还通过诸多途径对国民经济发展和工业化进程产生影响。比如“退耕还林”政策推动了林业的发展，“退耕还林”政策积极响应了可持续化发展战略，“退耕还林”政策可以间接改变生物丰富度，涵养水土，带来诸多生态效益等。

因此我们认为，评价“退耕还林”政策的影响时，不能从某个指标一概而论，应当注意到各种影响的“蝴蝶效应”。

3.2 基于不完全预测的影响评价模型

3.2.1 模型的理论依据

考虑到“退耕还林”的影响具有复杂的“蝴蝶效应”，我们首先采用**不完全预测法**对“退耕还林”政策的工业化影响和国民经济影响进行评估。

不完全预测法是一种通过预测方法来评价某种影响的方法。我们认为，事物发展产生的突然变化，一定是由于外界某些客观影响因子的突变导致的。为了评估这种影响，我们使用客观影响因子突变之前的数据进行向后预测，并与实际值进行对比，对比中体现的差异便可以用来评估突变产生的影响。

3.2.2 模型的运算

步骤一：数据的预处理 我们以工业增长值和 **GDP** 作为衡量工业化和国民经济的评价指标。考虑到后面的数据操作需要数据曲线有着较高的光滑性，我们首先使用**小波分析降噪算法**对数据进行降噪处理。降噪结果如下：

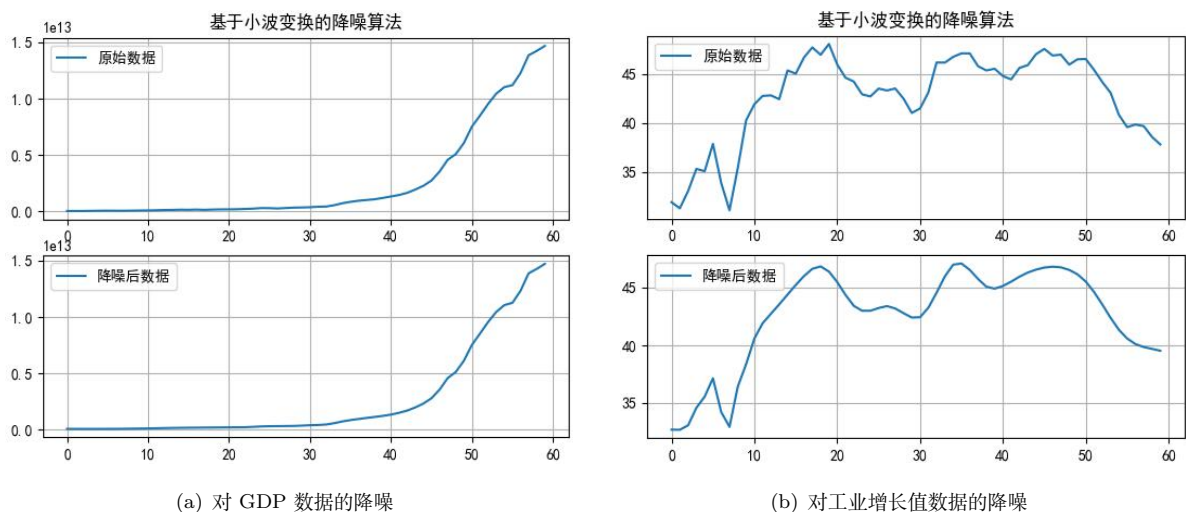


图 12: 数据降噪

步骤二：对数据进行不完全预测 我们将 1999 年作为分水岭，认为 1999 年之前的发展情况不受到“退耕还林”政策影响，1999 年之后的发展情况主要受到“退耕还林”政策的影响。

在预测模型的选择方面，我们使用 **5 阶自回归模型 (AR(5))** 进行预测。该模型的预测公式为：

$$\hat{Y}_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \beta_5 Y_{t-5}$$

自回归模型是一种经典的时间序列模型，它在预测时候考虑到了先前序列的趋势，因此我们认为该预测模型符合不完全预测模型所需要的条件。

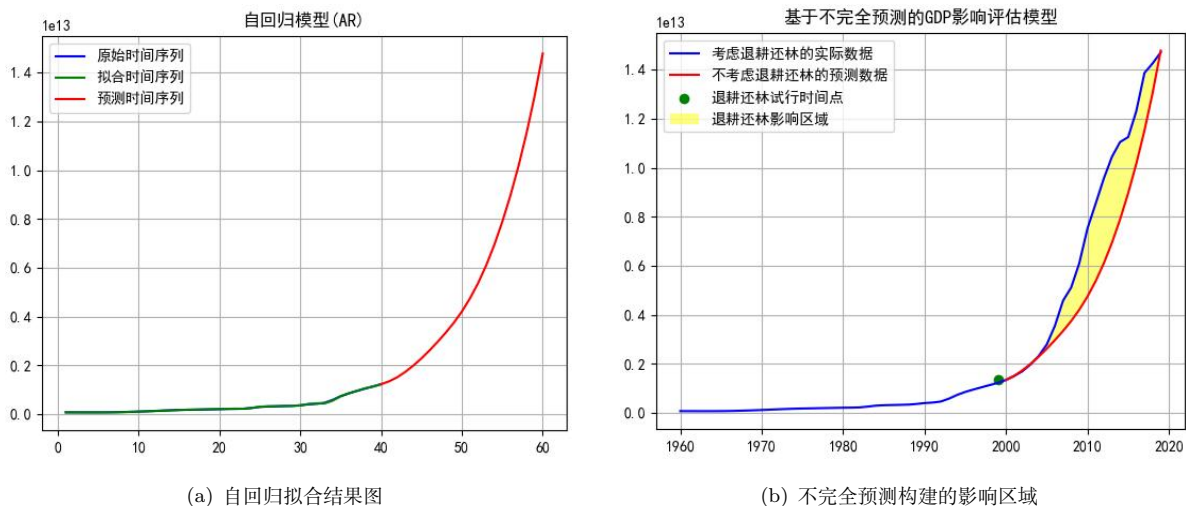


图 13: 国民经济的不完全预测评估

步骤三：插值使得数据连续化 设 $\hat{Y}(t)$ 为预测数据， $Y(t)$ 为实际数据。则二者均是离散函数，构造 $F(t) = Y(t) - \hat{Y}(t)$ ，我们使用**三次样条插值**使得函数 $F(t)$ 化为连续函数。

处理结果如下图所示：

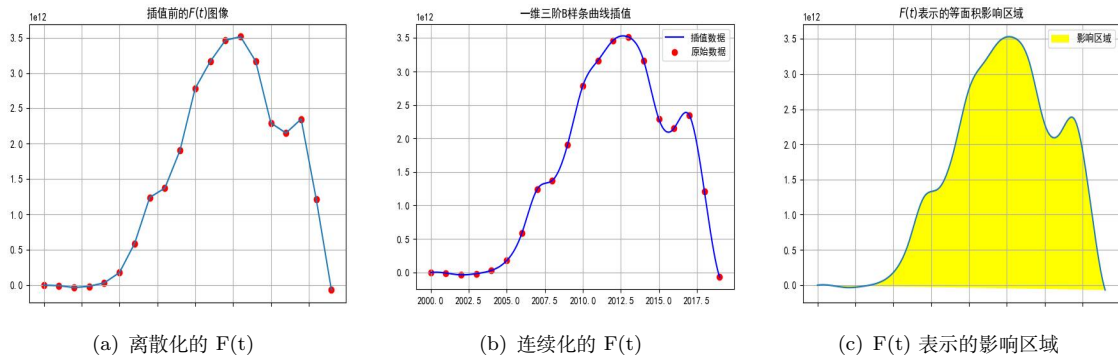


图 14: $F(t)$ 插值处理示意图

步骤四：使用黎曼和逼近 $F(t)$ 的积分 函数 $F(t)$ 在区间 $[2000, 2020]$ 上的积分可由如下公式逼近：

$$\int_{2000}^{2020} F(t)dt = \lim_{n \rightarrow \infty} \sum_{i=1}^n \left[\frac{20}{n} F\left(2000 + \frac{20i}{n}\right) \right]$$

使用 Python 迭代解得该积分值为 29346215997527.527，单位为**现价美元 × 年**。即可作为“退耕还林”政策对国民经济的影响评价量。使用同样的方法对**工业增长值**指标进行分析。

与 GDP 数据不同的是，工业增长数据随着时间变化出现了比较复杂的波动趋势，因此自回归模型不再适合对该数据进行预测。为此我们引入**霍尔特温特斯三重指数平滑预测模型**。

霍尔特温特斯三重指数平滑预测模型是一种时间序列模型，其特点是可以针对数据的**周期性、类周期性和波动性**进行指数平滑，因此我们认为该模型可以较好地拟合预测波动的工业增长量模型。

在我们的建模过程中，选取参数 $T = 10$ 作为时间序列的平滑步长值进行建模，并将模型带入到**不完全预测模型**中。

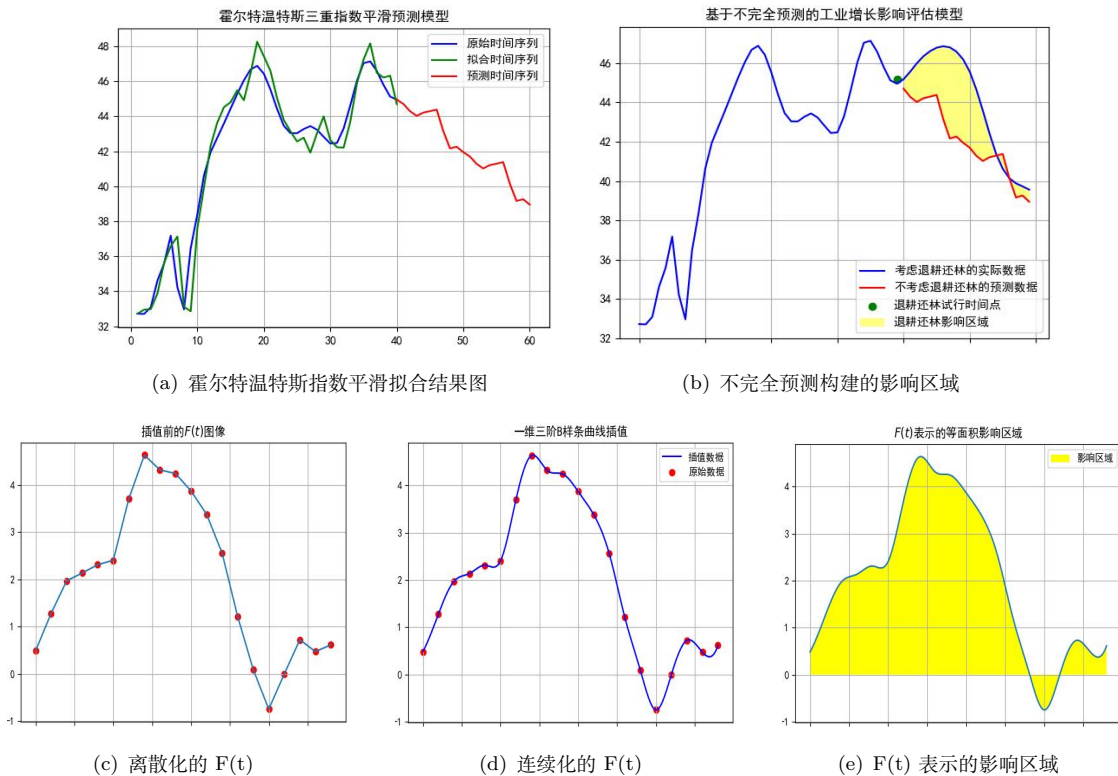


图 15: $F(t)$ 插值处理示意图

解得影响区域的积分为 39.066，单位为占 GDP 的百分比 × 年，即可作为“退耕还林”政策对工业影响的评价量。

3.3 综合指标对模型的修正

3.3.1 基于熵权 Topsis 模型构建的综合指标

如果仅仅只选取 GDP 和工业增长量来进行分析和评价，会显得过于片面。因此我们将建立多项评价指标的综合指标，相对全面地评价和分析。我们选取**人均能源使用量、森林面积、用电量、GDP、二氧化碳排放量、作物生产指数**指标，使用熵权 Topsis 模型建立工业水平和经济的综合指标。

熵权 Topsis 方法是一种客观赋权方法，该算法借鉴了信息熵理论，改进了经典的 Topsis 算法。假设有 m 个样本和 n 项指标，算法步骤如下：

步骤一：构造 m 行 n 列的评价矩阵 $\{X_{ij}\}$ ，其中 X_{ij} 代表第 i 个样本的第 j 项指标

步骤二：对评价矩阵进行**极差标准化**，负向指标和正向指标的公式分别为：

$$Z_{ij} = \begin{cases} \frac{X_j - X_{\min}}{X_{\max} - X_{\min}} \\ \frac{X_{\max} - X_j}{X_{\max} - X_{\min}} \end{cases}$$

步骤三：计算每个指标的比值，设第 j 个指标值，第 i 个样本所占的比重为 p_{ij} ，其中：

$$p_{ij} = \frac{Z_{ij}}{\sum_i Z_{ij}}$$

步骤四：计算第 j 个指标的熵权值 e_j ，公式为：

$$e_j = -\frac{1}{\ln m} \sum_i p_{ij} \ln p_{ij}$$

步骤五：计算每个指标的权重，其中 $d_j = 1 - e_j$ 称为**信息冗余值**，公式为：

$$w_j = \frac{d_j}{\sum_{j=1}^n d_j}$$

步骤六：对原始数据进行**数据归一化**，公式为：

$$Y_{ij} = \frac{X_{ij}}{\sqrt{\sum_i X_{ij}^2}}$$

步骤七：寻找**最优向量**和**最劣向量**，分别表示为 Y_+ 和 Y_- ，其中：

$$Y^+ = \{ \max_{1 \leq i \leq m} Y_{i1}, \max_{1 \leq i \leq m} Y_{i2}, \dots, \max_{1 \leq i \leq m} Y_{in} \} \quad Y^- = \{ \min_{1 \leq i \leq m} Y_{i1}, \min_{1 \leq i \leq m} Y_{i2}, \dots, \min_{1 \leq i \leq m} Y_{in} \}$$

步骤八：计算每个样本与最优向量和最劣向量的**欧氏距离**：

$$D_i^+ = \sqrt{\sum_j (Y_{ij} - Y_j^+)^2} \quad D_i^- = \sqrt{\sum_j (Y_{ij} - Y_j^-)^2}$$

步骤九：计算每个样本的综合得分：

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

由于所选指标众多，部分年份存在缺失值，我们仅选取了 1990 年-2014 年这 25 年时间跨度的相关数据进行分析。使用熵权 Topsis 模型计算得到的指标权重如下表所示：

指标名称	人均能源使用量	森林面积	用电量	GDP	二氧化碳	作物生产指数
权重	0.0514	0.1126	0.2121	0.3016	0.1983	0.1240

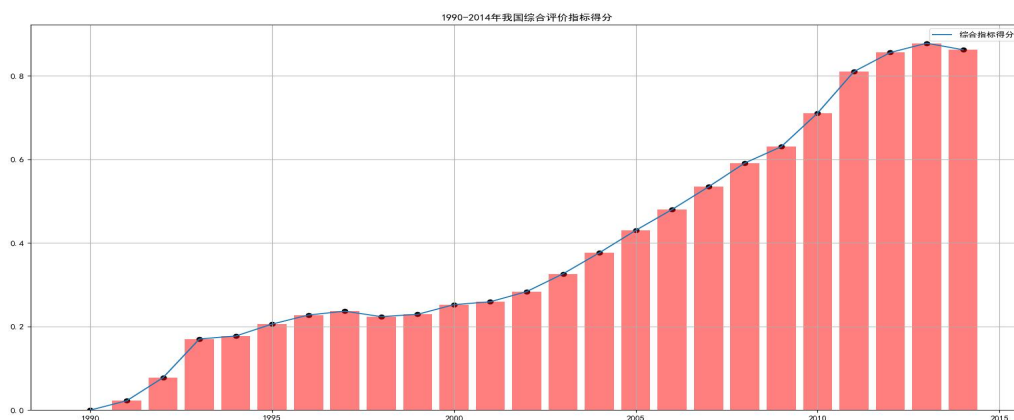
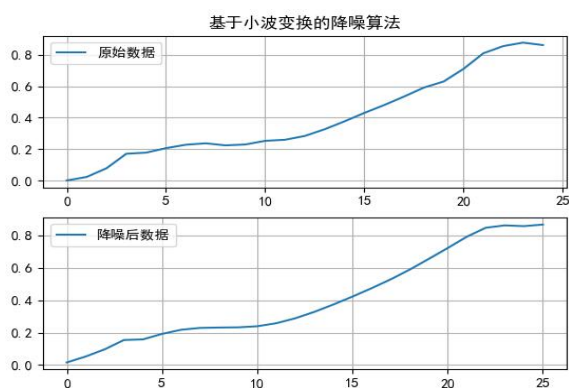


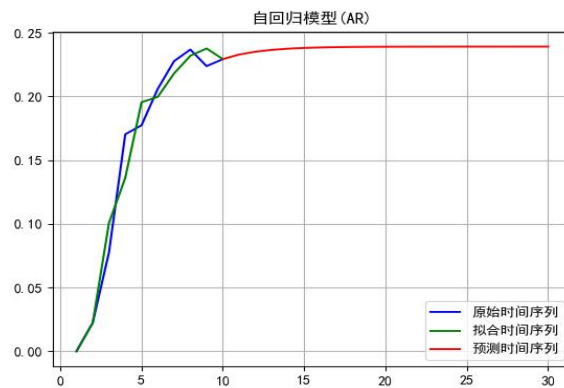
图 16: 1990-2014 年我国综合指标得分变化曲线

3.3.2 基于综合指标对模型的修正

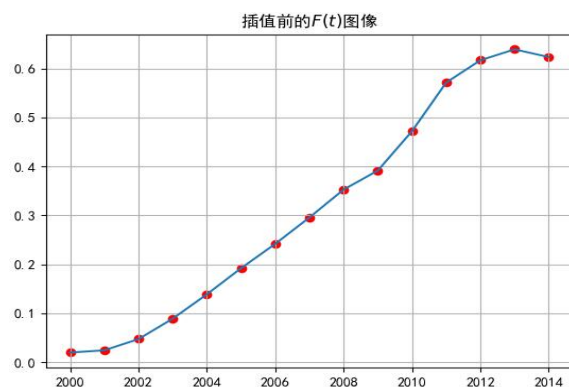
将综合指标的变化曲线带入不完全预测模型进行计算。



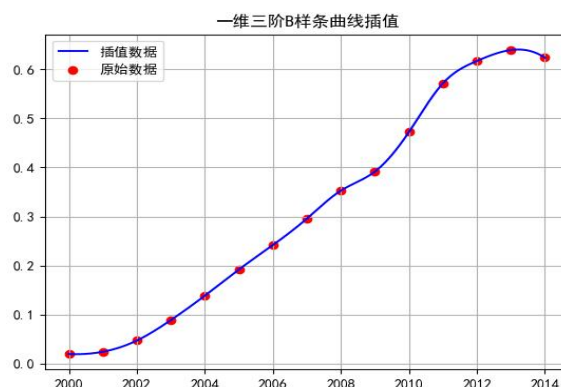
(a) 综合指标曲线的小波降噪



(b) 综合指标的不完全自回归预测



(c) 离散化的 $F(t)$



(d) 连续化的 $F(t)$

图 17: $F(t)$ 插值处理示意图

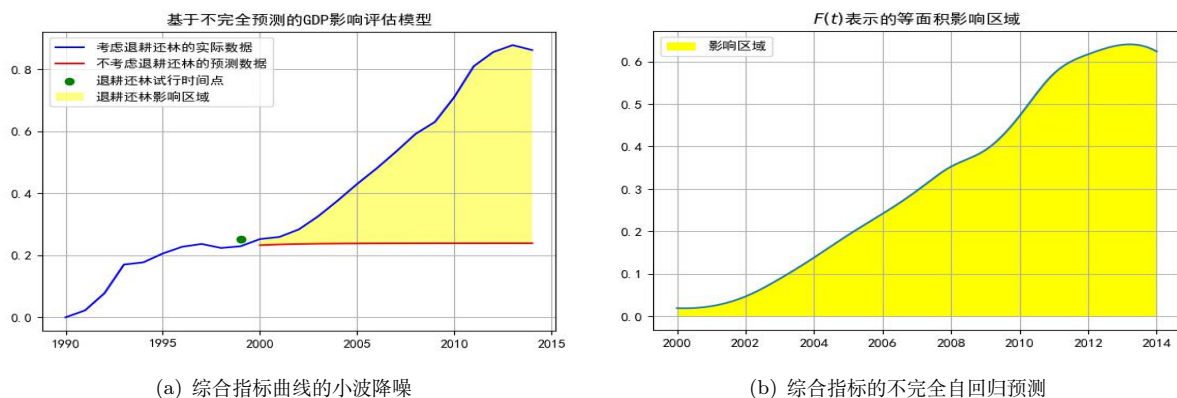


图 18: 影响区域的构建和分析

计算得到综合指标曲线的影响评价量为 4.4002。

3.4 有界量对模型的修正

在上述评价方法中，我们最后得到的统计量数量级不一，量纲不一，无法直观得出一个明显的结论。为此我们引入有界量来构造一个统一化的量化指标。设 $Y(t)$ 为各个指标的实际值，分别求解上述模型积分区间中 $Y(t)$ 的积分，并将上述求得的量纲不一的评价量与 $Y(t)$ 的积分作比即可得到一个 $(0, 1)$ 之间的有界量作为评价量。

用数学表示为：

$$\beta = \frac{\int_{T_1}^{T_2} (Y(t) - \hat{Y}(t)) dt}{\int_{T_1}^{T_2} Y(t) dt} = 1 - \frac{\int_{T_1}^{T_2} \hat{Y}(t) dt}{\int_{T_1}^{T_2} Y(t) dt}$$

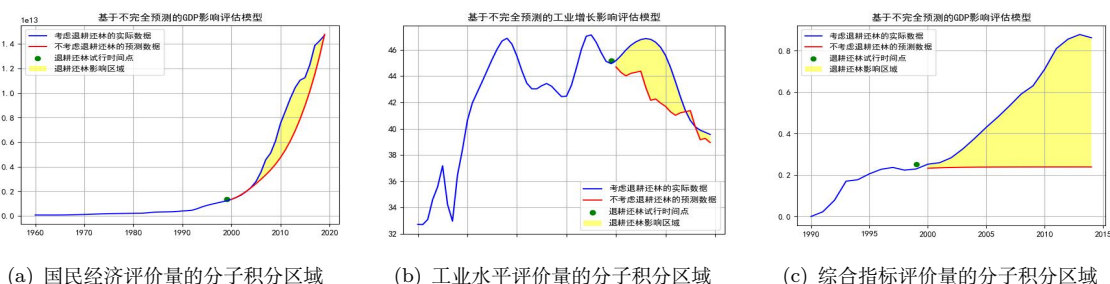


图 19: 评价量的分子积分区域

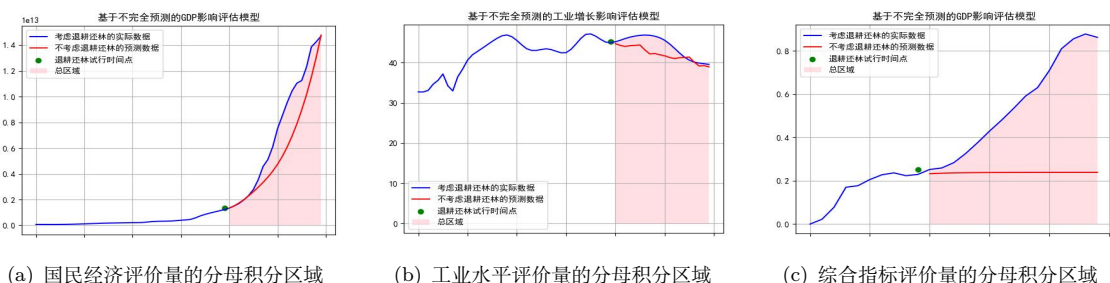


图 20: 评价量的分母积分区域

基于上面的分析，得到“退耕还林”政策对国民经济影响的有界量评分为 0.22；对工业水平影响的有界量评分为 0.05；对综合指标影响的有界量评分为 0.57。

4 问题三的建模与求解

4.1 模型的理论依据

4.1.1 使用正态分布进行分类

正态分布是一种最广泛的概率分布方式，很多事物都会服从或者近似服从正态分布。一般地，正态分布由均值和方差来决定。那么如何将正态分布应用到一维数据的分类当中呢？

我们借鉴著名的**拉依达准则**（也叫 3σ 原则）：**对于近似服从正态分布的数据，认为小于均值与三倍标准差的差的数值为异常偏低值；认为大于均值与三倍标准差的和的数值为异常偏高值。**

在模型中，我们认为：如果数值大于**均值与标准差之和**，则认为该数值处于**偏上水平**；如果数值处于**均值与标准差的波动范围之内**，那么数值处于**中等水平**；否则，数值处于**偏低水平**。

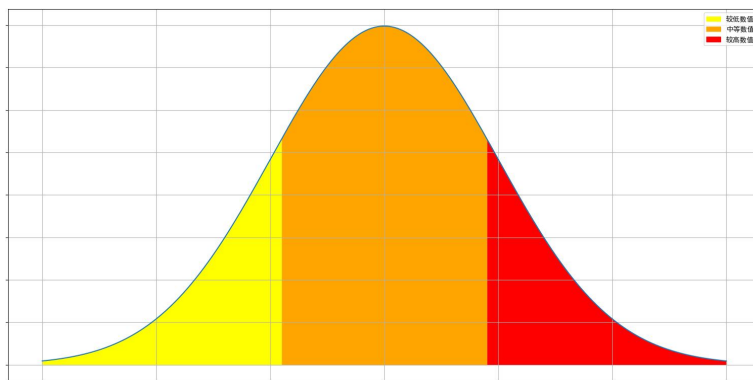


图 21: 正态分布分类示意图

4.1.2 使用雅卡尔指数评价集合相似度

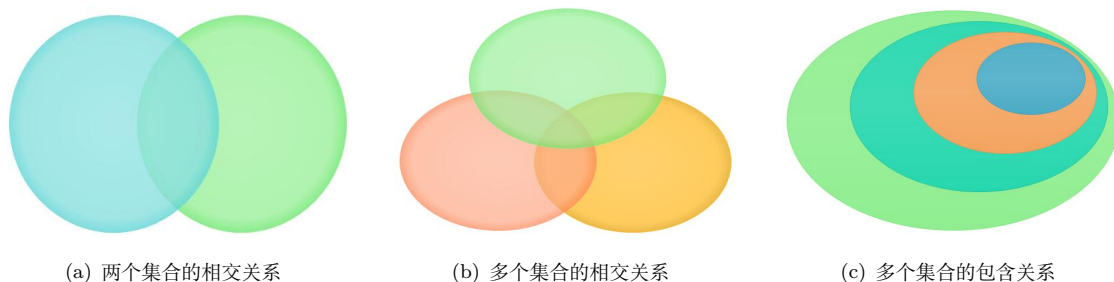
集合是数学中一种普适的、描述事物关系的数学结构，集合具有无序性、唯一性和确定性。**雅卡尔指数**又称交并比，是描述集合之间相似性的数学量。

集合 A 与集合 B 的雅卡尔指数被定义为：

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

其中 $|A|$ 表示集合 A 的基数，对于有限集合，基数便是元素个数。

雅卡尔指数的显著特点是有界，即 $J(A, B) \in [0, 1]$ 。当且仅当 A 与 B 完全相同时， $J(A, B)$ 取得最大值 1；当且仅当 A 与 B 不相交时， $J(A, B)$ 取得最小值 0。因此， $J(A, B)$ 越接近 1， A 和 B 的相似程度就越高。



(a) 两个集合的相交关系

(b) 多个集合的相交关系

(c) 多个集合的包含关系

图 22: 集合之间的关系模型示意图

在接下来的建模中，我们将对不同国家构造**集合论模型**，以对各种信息进行深度挖掘。

4.2 对国家各个指标的挖掘与分类

1. 对国家森林面积的挖掘与分类 在数据的选择方面，考虑到不同国家的土地面积各异，因此直接评价国家的森林面积是没有意义的，因此我们选择 **2020 年森林面积占土地面积百分比**作为评价森林面积的指标。

经过计算，模型分类结果如下表：

表 4: 各个国家关于森林面积百分比的集合划分

所属集合	较低水平集合	中等水平集合	较高水平集合
森林面积百分比所属区间	$\leq 0.0957\%$	$(0.0957\%, 0.5497\%)$	$\geq 0.5497\%$
国家数目	46	166	42
数目所占百分比	18%	65%	17%

2. 对国家工业化水平的挖掘与分类 在数据选取方面，我们选取 **2020 年各个国家的工业增长值数据**进行分类。经过计算，模型分类结果如下：

表 5: 各个国家关于工业化水平的集合划分

所属集合	较低水平集合	中等水平集合	较高水平集合
工业增长值所属区间	$\leq 24.7503\%$	$(24.7503\%, 26.5438\%)$	$\geq 26.5438\%$
国家数目	105	19	92
数目所占百分比	48%	9%	43%

3. 对国家国民经济的挖掘与分类 在数据选取方面，我们对世界银行官方统计出的各个国家收入的分类评级进行计数，计数结果如下：

表 6: 各个国家关于国民经济水平的集合划分

所属集合	低收入国家集合	中低收入国家集合	中高收入国家集合	高收入国家集合
国家数目	76	27	54	55
数目所占百分比	36%	13%	25%	26%

4.3 不同指标间的交叉分析

国民经济和工业化程度的交叉分析 根据上述分类，各个国家按照国民经济水平分为**低收入集合、中低收入集合、中高收入集合、高收入集合**；按照工业化水平分为**低水平集合、中水平集合、高水平集合**。

我们采用构造**雅卡尔系数矩阵**的方法描述集合之间的关联程度，即对每个集合都对另外领域的每个集合求解一次雅卡尔系数，最后构造成矩阵进行分析。

国民经济和森林面积的交叉分析 根据上述分类，各个国家按照国民经济水平分为**低收入集合、中低收入集合、中高收入集合、高收入集合**；按照森林面积分为**低水平集合、中水平集合、高水平集合**。

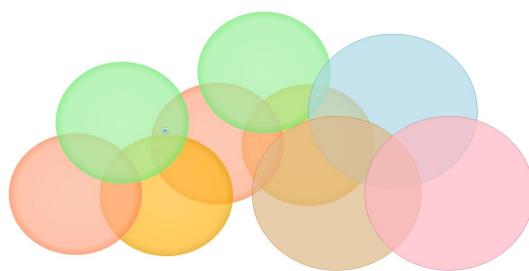


图 23: 多集合的交叉模型示意图

表 7: 国民经济和工业水平的雅卡尔系数矩阵

国民经济 \ 工业水平	较低工业水平国家集合	中等工业水平国家集合	较高工业水平国家集合
低收入国家集合	0.46	0.09	0.25
中低收入国家集合	0.37	0.11	0.33
中高收入国家集合	0.46	0.04	0.35
高收入国家集合	0.35	0.07	0.38

* 注: $J(A, B) > 0.5$ 即可认为集合之间有较高相似度

表 8: 国民经济和森林面积的雅卡尔系数矩阵

国民经济 \ 森林面积	较低森林面积国家集合	中等森林面积国家集合	较高森林面积国家集合
低收入国家集合	0.34	0.11	0.54
中低收入国家集合	0.48	0.19	0.26
中高收入国家集合	0.44	0.06	0.44
高收入国家集合	0.58	0.04	0.35

* 注: $J(A, B) > 0.5$ 即可认为集合之间有较高相似度

工业水平和森林面积的交叉分析 根据上述分类，各个国家按照工业水平分为低水平集合，中水平集合、高水平集合；按照森林面积分为低水平集合，中水平集合、高水平集合。

表 9: 工业水平和森林面积的雅卡尔系数矩阵

工业水平 \ 森林面积	较低森林面积国家集合	中等森林面积国家集合	较高森林面积国家集合
较低工业水平集合	0.61	0.11	0.12
中等工业水平集合	0.59	0.15	0.07
较高工业水平集合	0.54	0.13	0.14

* 注: $J(A, B) > 0.5$ 即可认为集合之间有较高相似度

为了更加直观体现各个雅卡尔系数矩阵的程度强弱，我们使用热度图的方法进行数据可视化。结果如下所示：(颜色越浅，集合相似度越强) 综上所述，我们可以得到诸多结论和发现，下面举了

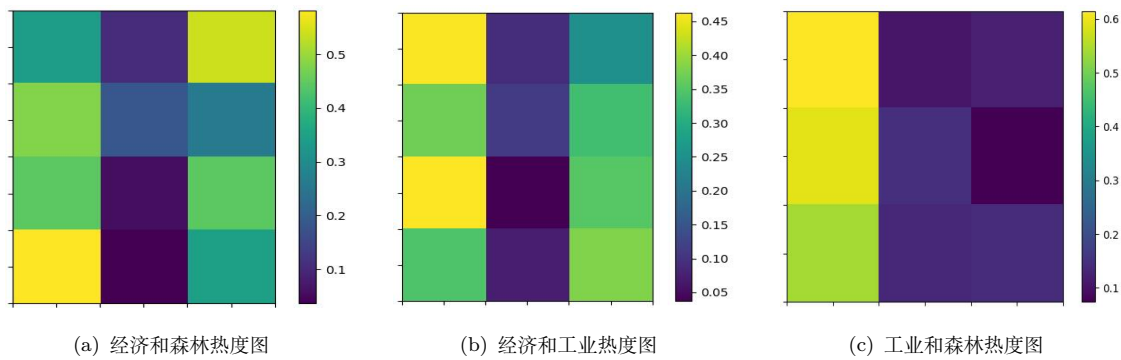


图 24: 雅卡尔矩阵的热度图可视化

一些例子：

- 1. 非高收入国家中低工业水平的国家最多，高收入国家中高工业水平的国家最多，这印证了工业与经济之间的密不可分
- 2. 低收入国家中森林面积占土地面积总比例高于高收入国家中森林面积占土地面积的总比例更容易偏低，这可能与资源利用是否充分有关
- 3. 森林面积比例较高的国家也可以拥有较高的经济和工业水平，说明森林建设和经济、工业发展并不是站在对立面的

5 研究报告书

自工业革命以来，人类社会的发展与环境保护之间的矛盾便日益凸显出来，环境因为人类社会的经济增长、工业化程度的提高而受到破坏。这一切仿佛都说明：人类社会的工业化、国民经济的发展与环保站在了对立面。

随着“绿水青山就是金山银山”口号的提出，中国、亚洲，乃至全世界的环保观念都得到了前所未有的颠覆性更新。人类社会的工业化、国民经济的发展与环境保护并没有站在对立面，二者可以相辅相成，共同促进。

从“退耕还林”中能得到什么启发？身为国际上有担当，勇于承担社会责任的大国，我国在环境保护方面始终处于积极的队伍中。“退耕还林还草”、“可持续发展战略”、“生态林产业”等等一系列政策，都使得我国环保成绩卓越，效果显著。



上面几幅图是中国近几年的环保成绩单，来自央视新闻。

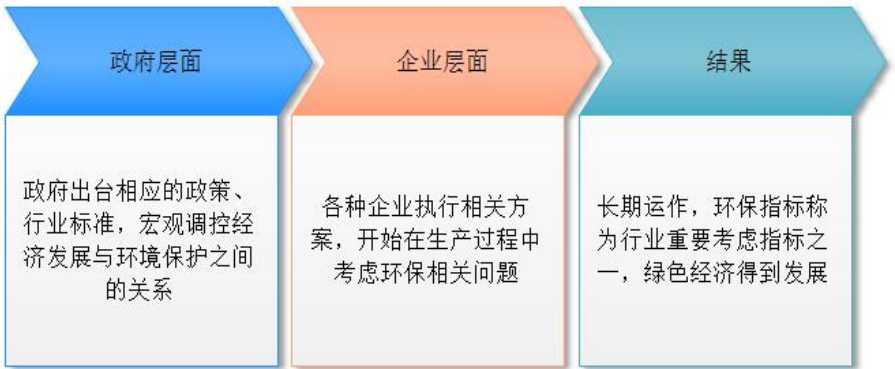
结合我们的分析，我们认为“退耕还林”等一系列政策不仅没有阻碍经济的发展和工业水平的进步，反而拉动了经济的增长，带来了诸多有益的“蝴蝶效应”。从“退耕还林”政策这个鲜活的例子来看，国家通过政策手段宏观调控，在直接层面有利于改善环境，在客观层面有利于拉动旅游业的发展，促进经济结构向绿色可持续化发展。有利于在人类社会的经济增长、工业水平提高和环境保护方面找到均衡点，真正实现“绿水青山”向“金山银山”的变现。

从不同国家和地区的各项指标中能看到什么？ 世界是一个命运共同体，每个国家和地区的环境都会深深地影响到每一个其它国家，这便是国际责任的内涵之一。分析各个国家的国民经济，工业化水平以及森林面积等指标，我们可以发现：**国民经济、工业发展和绿色环保之间没有绝对的排斥和对立**。也就是说，每个国家都可以在环境保护的前提下，在绿色的前提下发展好国民经济、发展好工业水平。

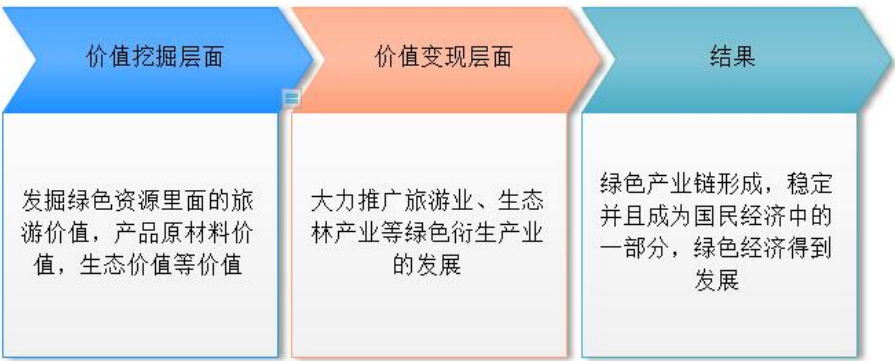
而全球国家与国家之间又是密不可分、紧密相连的，因此，国际社会应当呈现“牵一发而动全身”之势，积极响应国际“碳达峰与碳中和”、“绿色化发展”等战略目标，更新自己的发展意识，将国民经济增长、工业水平提高与环境保护从对立面放到同一侧。

应当如何去做？ 综上所述，目标是在经济增长、工业化发展和环境保护之间寻找一个平衡点。那么各个国家和地区应当如何去做呢？

1. 发挥政府的宏观调控作用 一个地区发展的大政方针都是由政府制定的，政府是发展的指向标。在环保、绿化方面，政府应当出台相应的政策以进行宏观调控。比如，政府可以通过制定行业环保标准调控工业化对环境的影响，政府可以出台战略和政策大力发展绿色经济和可持续化经济。



2. 充分挖掘绿色资源的价值 发展归根结底是为了提高人民生活水平，那么同样可以通过对绿色资源的价值挖掘获得绿色收益。森林面积大了，吸引的游客多了，便出现了生态林旅游业；水土好了，生态环境改善，生物多样性增加，那么便会带来不限于旅游业的隐性收益。同时，生态林产业、木产品工业等林业相关产业，都会随着森林的发展而得到发展。



3. 辅以科技手段 在一个国家从非绿色经济转向绿色经济的初期，各种产业必然会受到理论上

的各种消极影响。以“退耕还林”为例，理论上来说退耕还林必然会导致作物生产指数的下降，但是事实上却没有，这是为什么呢？因为科技，科学技术提高了生产效率。因此，科学技术在非绿色经济向非绿色经济的转变中，起到了缓冲、增强的作用。



4. 在思想上更新传统认知 归根结底，这点是最重要的。这里的“在思想上更新传统认知”指的是将经济发展、工业发展和环境保护从对立面放置于非对立面。我们要坚信，发展和环保之间的“纳什均衡”是可以找到的，并要承担国际责任和社会责任，坚持不懈地向可持续化发展迈进。其具体举措可以从教育、宣传方面入手。



6 模型分析与评价

对于问题一 在问题一的预测模型中，我们分别使用线性趋势预测和饱和态趋势进行预测，二者相互对比，相互修正，使得模型预测效果更好。**在模型的优点上：**拟合精度显示，两种模型的拟合效果均达到很高的水平。作为一种经典的时间序列模型，霍尔特线性趋势预测模型具有坚实的理论基础，结果更加普适，可信。**在模型的不足上：**我们并没有考虑到其他影响因子对森林面积的动态影响，比如退耕还林的程度、未来的发展政策等。

对于问题二 问题二中我们先是进行了影响因子分析，而后通过不完全预测的方式进行了影响评价。**在模型的优点方面：**影响因子分析的时候考虑到的层次全面，详略得当，不完全预测模型中先后引入了综合指标和有界评价量进行模型的修正。**在模型的不足方面：**不完全预测模型并不能绝对体现“退耕还林”的影响，没有将“退耕还林”带来的影响和其他政策带来的影响分离。

对于问题三 问题三我们使用了正态分布进行分类，构造了离散数学模型。**在模型的优点方面：**分类标准具有统计性意义，多次的交叉评价能够充分挖掘各类国家集合的特点。**在模型的不足方面：**分类完全基于国家指标服从正态分布的假设，但是没有给出正态分布的严格假设检验和修正，因此在分类效果方面会出现不同程度的偏差。

参考文献

- [1] 姜启源等,《数学模型》,北京:高等教育出版社,2004
- [2] 傅立,《灰色系统理论及其应用》:科学技术文献出版社,1992 年 10 月
- [3] <https://data.worldbank.org.cn>
- [4] 韩中庚,数学建模竞赛获奖论文精选与点评,北京:科学出版社,2007
- [5] 薛薇,SPSS 统计分析方法及应用第三版,北京:电子工业出版社,2013
- [6] 郑彤等. 环境系统数学模型 [M]. 北京:化学工业出版社-2003
- [7] 张树京等. 时间序列分析简明教程 [M]. 北京:清华大学出版社,2001

附录

附件: 程序 文件夹中包括建模用到的程序及其源代码

附件: 论文 文件夹中包括建模用到的论文及其源代码

附件: 数据 文件夹中包括建模查找到的其他数据