

2020 年第五届“数维杯”大学生 数学建模竞赛论文

基于综合评价体系的可持续发展研究

摘 要

本文通过对结合现在相关统计数据 and 新冠病毒的大环境,探究分析各省的全面建成小康社会的总体情况,通过运用 matlab、spss 软件和主成分分析法、时间序列方法,构建可持续发展模型,模拟退火模型进行分析,提出相关建议。

对于问题一,我们分析总结了过去 20 年期间我国全面建成小康社会的成功经验,据题意考虑到对于可持续发展的影响可以分为资源和人文方面,确定二氧化碳浓度、森林面积、总人数变化、水资源总量为四个主要的影响可持续发展的指标,通过 SPSS 运用时间序列方法对各指标未来 20 年变化的进行预测,根据分析结果对未来 20 年的可持续发展提出发展思路。

对于问题二,我们首先我们先确定了人均 GDP、高中入学率、恩格尔系数、人均居住面积和平均寿命等指标,然后建立分层次、多维度的指标体系进行综合评价,通过 MATLAB 采用的是主成分分析法对各省的全年建成小康社会总体情况的排序,进而根据这些进行合理评价。

对于问题三,本文到考虑疫情影响,以湖北省为例,确定了自然人口变化、老龄化人口、城乡居民收入、资源禀赋和环境承受力这五个可大致表示新型冠状病毒肺炎疫情对可持续发展产生影响的因素,通过这五个指标建立短期及长远的经济社会可持续发展模型。后通过 SPSS 用主成分分析法得出相关因素的表达式,并计算结果分析决定广东省未来的发展方向。

对于问题四,我们在问题二的基础上,将问题二中提到的指标作为二级指标,然后将 GDP 增长率和资源禀赋作为一级指标,运用聚类分析法将二级指标聚合成几个一级指标,然后对一级指标进行分析,建立模拟退火模型得出每个省的发展程度,得出各省的分配比例,可根据各省所占比例进行合作。

关键词: 时间序列; 主成分分析; 可持续发展; 模拟退火模型

目录

一、问题重述.....	1
二、问题分析.....	1
2.1 问题 1 的分析.....	1
2.2 问题 2 的分析.....	1
2.3 问题 3 的分析.....	1
2.4 问题 4 的分析.....	2
三、模型假设.....	2
四、定义与符号说明.....	2
五、模型的建立与求解.....	2
5.1 问题 1 的模型建立与求解.....	2
5.1.1 总结全面建成小康成功经验.....	2
5.1.2 问题 1 模型的建立.....	3
5.1.3 问题 1 模型的求解.....	4
5.1.4 结果.....	6
5.2 问题 2 的模型建立与求解.....	6
5.2.1 问题 2 模型的建立.....	6
5.2.2 问题 2 模型的求解.....	7
5.2.3 结果.....	8
5.3 问题 3 的模型建立与求解.....	9
5.3.1 问题 3 模型的建立.....	9
5.3.2 问题 3 模型的求解.....	10
5.3.3 结果.....	11
5.4 问题 4 的模型建立与求解.....	11
5.4.1 聚类分析.....	12
5.4.2 模拟退火模型.....	13
六、模型的评价及优化.....	15
6.1 误差分析.....	15
6.1.1 针对于时间序列预测的误差分析.....	15
6.1.2 针对主成分分析方法的误差分析.....	17
6.2 模型的优点.....	18
6.3 模型的缺点.....	18
6.4 模型的推广.....	18
参考文献.....	19
附录.....	20

一、问题重述

2020 年是我党全面建成小康社会的决战之年，是第 100 个奋斗目标的实现的关键之年，“小康”是中国人自古以来的梦想，也是中国共产党带领中华民族走向复兴的阶段性目标。当前，小康社会建设已到了最后的冲刺和收官阶段，人民生活水平即将从小康水平步入全面建成小康社会的胜利道路，是一个中华儿女民族发展史上的一个新的里程碑。但是在 2020 年初期由于新冠病毒的蔓延，从武汉到全国，对我国的发展造成了重大的影响。目前，新冠病毒正在更大规模的蔓延全世界，以致多我国的经济内外部发展造成了重大的损失，为此需要结合数学建模方法对全面建成小康社会进程的问题进行合理的优化建模和解决。

现需要我们团队解决以下问题：

- 1.总结归纳过去 20 年期间我们国家全面建成小康社会过程中的成功经验，并需结合全球气候变暖、生态环境、人口压力倍增等发展现状来提出对我们未来 20 年的可持续发展思路。
- 2.结合国家统计局年鉴、地方统计年鉴等官方给出的数据，总结归纳诸如我国各省全面建成小康社会完成难度系数、完成比率及未来可持续发展能力等在内的多重指标对各省全面建成小康社会总体情况进行合理的评价与排序。
- 3.对于目前世界处于变化的大变局及新型冠状病毒疫情全球性蔓延的大环境，需要建立以某个省为代表的短期及长远的经济社会可持续发展模型。其中，模型中需要充分考虑诸如人口变化、老龄化趋势、城乡居民收入、资源禀赋及生态环境承载力等关键性因素。
- 4.提供一个各省之间的合作计划。其中，在这计划需要充分考虑各省份经济社会发展状况及资源禀赋，而且要能对合作双方产生积极的促进作用和总体效率的提升。

二、问题分析

本文首先通过对总结过去全面建成小康社会经验，运用时间序列方法预测相关因素对未来 20 年可持续发展的影响，结合现在相关统计数据 and 新冠病毒的大环境，用主成分分析法分析各省的全面建成小康社会的总体情况并进行评价排序，构建可持续发展模型，最后对各省之间的规划问题采用模拟退火模型进行分配，提出相关建议。

2.1 问题 1 的分析

针对问题一，首先查找文献资料对我国过去 20 年全面建成小康社会的成功经验进行总结，然后结合过去 20 年的发展影响因素和题意，考虑影响可持续发展的主要资源和人文因素，确定全球气候变暖、生态环境、人口压力倍增、自然资源为主要因素，并分析影响因素之间的内部关系，最终选定二氧化碳浓度、森林面积、总人数变化、水资源总量为各个方面的影响指标，统计过去 20 年的数据，运用时间序列方法对未来 20 年的各指标进行预测，对比结果并提出未来 20 年的可持续发展发展思路

2.2 问题 2 的分析

针对问题二，我们考虑到综合评价，通过运用的是主成分分析法对各省的全面建成小康社会的总体情况进行合理的评价与排序。首先根据题意对完成难度系数、完成比率及未来可持续发展能力，我们先确定了人均 GDP、高中入学率、恩格尔系数、人均居住面积和平均寿命等指标，然后建立分层次、多维度的指标体系进行综合评价，并利用 MATLAB 得出最后各省的全年建成小康社会总体情况的排序，进而根据这些进行合理评价。

2.3 问题 3 的分析

针对问题三，我们考虑疫情严重性的影响，主要以湖北省为例，确定了自然人口变

化、老龄化人口、城乡居民收入、资源禀赋和环境承受力这五个可大致表示新型冠状病毒肺炎疫情对可持续发展产生影响的因素，而后可以建立短期及长远的经济社会可持续发展模型，通过数据统计、spss 再次进行主成分分析得出相关因素的表达式，并计算结果分析决定广东省未来的发展方向。

2.4 问题 4 的分析

针对问题四，我们结合问题二，在问题二的基础上，将问题二中提到的指标作为二级指标，然后来确定 GDP 增长率和资源禀赋作为一级指标，通过运用聚类分析的方法将二级指标聚合成几个一级指标，然后进行一级指标分析，最后可以构建模拟退火模型，利用 matlab 计算得出每个省的发展程度及分配结果，根据结果进行对比分析，利用分配比例得出各省之间按照分配率的合作计划。

三、模型假设

- 1、假设所给出的指标和数据能够很好的代表各个省份全面建成小康社会的发展程度和总体情况。
- 2、假设从文中所给的统计数据信息真实可靠。
- 3、假设我国可持续发展、全面建成小康社会只受文中相关因素的影响。
- 4、假设各相关因素之间彼此独立，不会相互影响。

四、定义与符号说明

符号定义	符号说明
a_{ij}	第 i 个评价对象的第 j 个指标的取值
μ_j	第 j 个指标的样本均值
s_j	第 j 个指标的样本标准差
R	相关系数矩阵
b_j	第 j 个主成分的信息贡献率
$r(q_j, x_i)$	q_j, x_i 之间的相关系数
S	初始解状态
T	初始温度
$C(S)$	评价函数

五、模型的建立与求解

5.1 问题 1 的模型建立与求解

5.1.1 总结全面建成小康社会成功经验

20 年以来全面建成小康社会从从总体到全面、从低水平到高水平、从量的积累到质的飞跃的发展过程。如今我们已经全面建成了小康社会，回顾历史，有对成功和经验都值得去总结，这对我们国家未来进行可持续发展有重大的意义。参考文献内容，我们可以总结出全面建成小康社会成功的原因共有以下五个点：

(一) 始终坚持党的集中统一领导。改革开放我国信息化、农村现代化等都发生了巨大的变化，我国的经济可以保持高速发展，经济总量跃升至世界第二，且取得全面建成小康社会的成绩，创造了世界经济奇迹，关键成功点都在于有中国共产党的正确领导。

(二) 始终坚持以人民为中心的发展思想

长期以来，坚持发展依靠人民，在任何时候，我们发展的主体都是人民，所以当小康社会进程推进过程中可以最大限度地激发了人民群众的能动性、创造性，使之发展能够顺利进行。

(三) 始终坚持以经济建设为中心

国家构建了“五位一体”的总布局，在这过程，以经济建设为中心的重大方针没有动摇，是全面建成小康社会的整体重要布局构建。

(四) 始终坚持循序渐进的目标导向

从刚开始的“中国式的现代化”的“小康”目标设立，后从总体小康到全面小康的调整，再从小康社会和社会主义现代化国家阶段性目标的调整，无一不体现我国建设小康社会过程中始终坚持遵循发展规律、尊重历史经验、坚持一切从实际出发的发展规律。

(五) 始终坚持改革开放的根本途径

改革开放从 1978 年开始，与小康社会一同发展，正因为是改革开放推动着小康社会建设持续向前发展。

5.1.2 问题 1 模型的建立

首先，据题意考虑到对于可持续发展的影响可以分为资源方面和人文方面，从题目中的全球气候变暖、生态环境、人口压力倍增、自然资源等因素中各方面选取较确定重要的指标进行数据处理，最后确定二氧化碳浓度、森林面积、总人数变化、水资源总量为四个主要的影响可持续发展的指标。

根据国家统计局数据，选取 2000 年到 2019 年过去 20 年的数据进行筛选统计，如图：

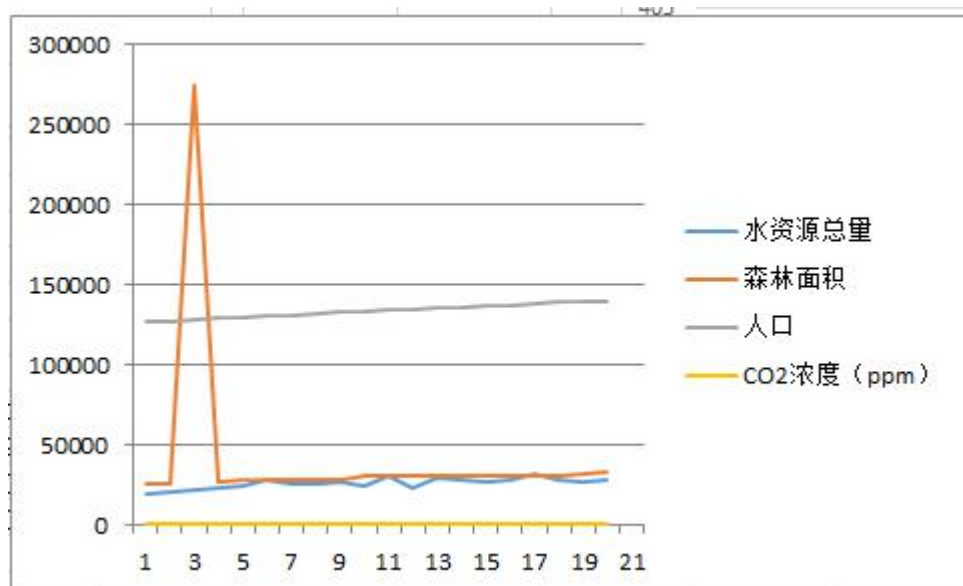


图 1-1 可持续发展的相关指标变化图

其中横坐标为时间，纵坐标为对应的变化量。

根据以上数据统计，；利用 spss 进行时间序列预测，对各指标未来 20 年变化的进行预测计算

取得 $n=1$ ，拟合 ARMA (2n, 2n-1) (即 ARMA (2,1)) 模型

(1) $p=3$ ，拟合 AR (p) 模型，设所要拟合出模型为

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varphi_3 X_{t-3} + a_t$$

(2) 用最小乘法拟合出系数 $\varphi_1, \varphi_2, \varphi_3$,

(2) 估计 ARMA (2,1) 模型 $X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} = a_t - \theta_1 a_{t-1}$ 参数的初始值。

对于 ARMA (2,1) 模型，我们有：

$$I_j - \theta_1 I_{j-1} = 0, j > 2$$

于是

$$I_3 - \theta_1 I_2 = 0 \Rightarrow \theta_1 = \frac{I_3}{I_2}$$

注意：以 AR (3) 中的 I_1, I_2, I_3 代替 ARMA (2,1) 中的 I_1, I_2, I_3 是一种近似代替。通过这种方法求得的 θ_1 的绝对值若大于 1，则取其倒数作为初始值，以满足可逆性条件。

知道了 I_1, I_2, I_3 及 θ_1 ，再用下式来确定 ARMA (2,1) 模型中的 φ_1, φ_2 ,

$$\varphi_1 = I_1 + \theta_1; \varphi_2 = \theta_2 - \theta_1 I_1 + I_2$$

(3) 以 (2) 中得到的 $\varphi_1, \varphi_2, \theta_1$ 为初始值，利用非线性最小二乘法得到 $\varphi_1, \varphi_2, \theta_1$ 的终值和置信区间，并且求出残差平方和 (RSS)。

3. $n=n+1$ ，拟合 ARMA (2n, 2n-1) 模型，基本步骤和 2 中相似

4. 用 F 准则检验模型的适用性。若 F 检验显著，则转入第 2 步。若 F 检验不显著，转入第 5 步。

对于 ARMA 模型的适用性检验其实就是对 a_t 的独立性检验。检验 a_t 的独立性的一个简便而有效的办法就是你和更高阶的模型，若更高阶的模型残差平方和有明显的减少，就意味着现有的模型 a_t 不是独立的，因而模型不适用，若更高阶的残差平方和有明显减少，同时更高阶模型中的附加参数的值也很小（其执行区间包含 0），则可以认为该模型是使用的。具体的检验准则为：

$$F = \frac{\frac{A_1 - A_0}{A_0}}{\frac{N - \gamma}{N - \gamma}} \sim F(s, N - \gamma)$$

其中 $\gamma = n_2 + m_2, s = n_2 + m_2 - (n_1 + m_1)$

若是这样得到的 F 值超过由 F 分布表所得的在 5% 置信水平上的 $F(s, N - \gamma)$ 值，那么由 ARMA (n_1, m_1) 模型改变为 ARMA (n_2, m_2) 时，残差平方和的改善是显著的，因而由 ARMA (n_1, m_1) 的适用性假设，F 值低于查表所得的值，就可以认为在该置信水平上这个模型的适用性。

5.1. 问题 1 模型的求解

利用 SPSS 时间序列求得预测 2020 年至 2039 年的结果为如下图：



图 1-2 未来 20 年森林变化面积

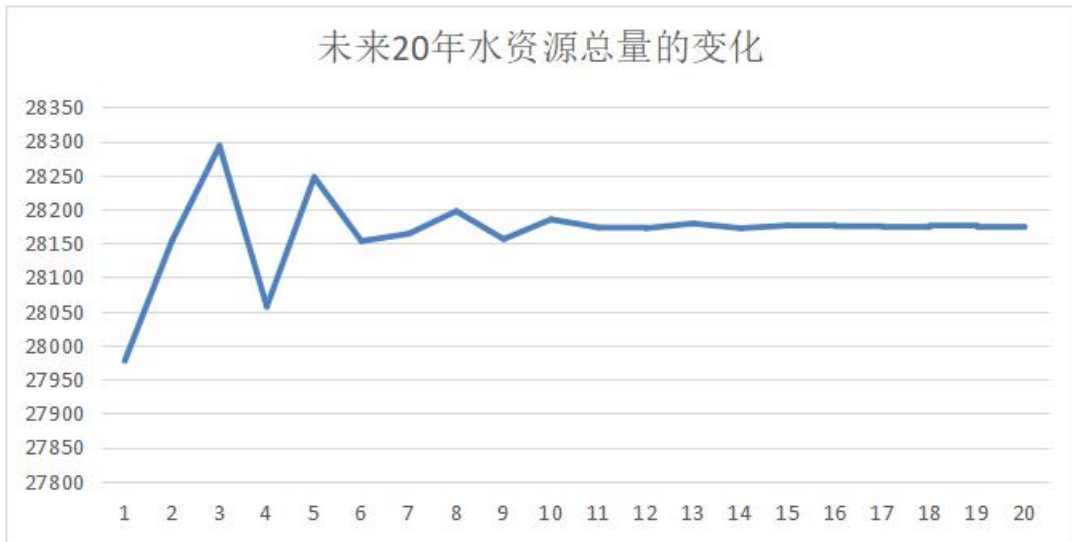


图 1-3 未来 20 年水资源的总量变化

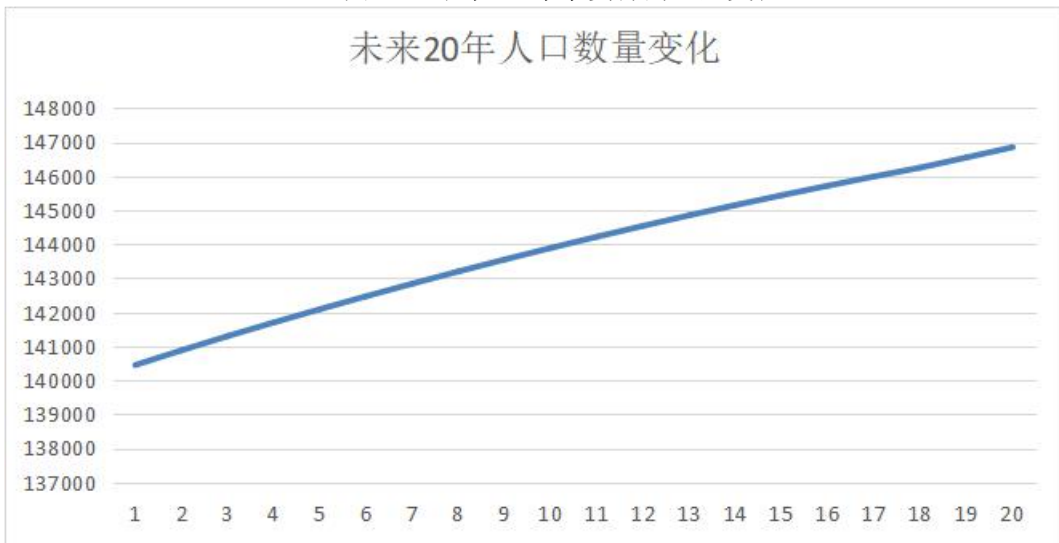


图 1-4 未来 20 年人口的数量变化

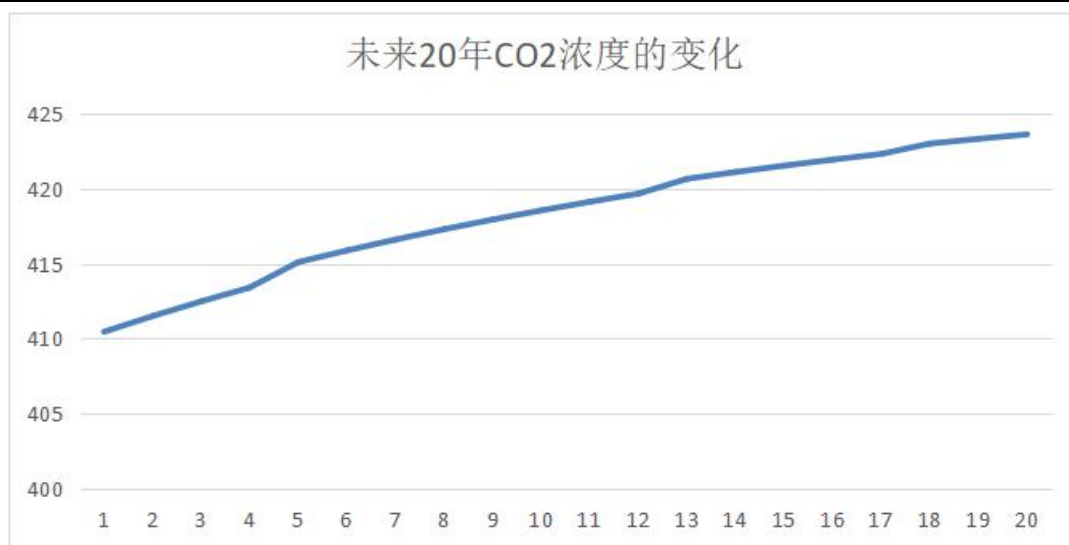


图 1-5 未来 20 年二氧化碳的浓度变化

5.1.4 结果

根据以上图示的预测结果，对未来 20 年可持续发展提以下发展思路：

从图像上看，未来 20 年对于生态方面变化并不明显，我们可以多种植树林，增加氧气的含量，创建更加美好的生态环境供人类居住，对于水资源总量变化也相对平稳，但是我们同样也不可以浪费水资源，增强对水资源，淡水资源的保护，增强我们淡水资源转化技术，而人口和二氧化碳方面，历年来会逐步上升，二氧化碳会引起温室效应，使得温室效应更为严重，同时人口过分增长，人口过分的增加会影响国家的经济发展，应当控制人口和降低二氧化碳含量的增加，开展新型的工业技术，减少汽车尾气的排放等措施。

5.2 问题 2 的模型建立与求解

5.2.1 问题 2 模型的建立

这一问题的目的是为了对各省的全面建成小康社会的总体情况进行合理的评价和排序，这一问题属于指标评价题。首先我们确定了人均 GDP、高中入学率、恩格尔系数、人均居住面积和平均寿命，然后应用主成分分析法对问题进行分析。分析步骤如下：

1、对原始数据进行标准化处理

假设进行主成分分析的指标变量有 m 个：，共有 n 个评价对象，第 i 个评价对象的第 j 个指标的取值为 a_{ij} 。将各指标值 a_{ij} 转换成标准化指标 \tilde{a}_{ij} ，

$$\tilde{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j}, (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$$

中 $\mu_j = \frac{1}{n} \sum_{i=1}^n a_{ij}, s_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_{ij} - \mu_j)^2}, (j = 1, 2, \dots, m)$ ，即 μ_j, s_j 为第 j 个指标的样本均值和样本标准差。对应地，称

$$\tilde{x}_j = \frac{x_j - \mu_j}{s_j}, (j = 1, 2, \dots, m)$$

为标准化指标变量。

2、计算相关系数矩阵 R

$$R = (r_{ij})_{m \times m}$$

河北	3
上海	4
新疆	5
内蒙古	6
重庆	7
湖南	8
西藏	9
青海	10
安徽	11
辽宁	12
贵州	13
云南	14
河南	15
浙江	16
山西	17
湖北	18
宁夏	19
福建	20
山东	21
陕西	22
江苏	23
广东	24
江西	25
四川	26
海南	27
甘肃	28
天津	29
吉林	30

5.2.3 结果

通过上面的表格，我们得出了中国 30 个主要省份的全面建成小康社会总体情况的排序。可以看出黑龙江的总体情况最好，这与黑龙江省一直实施精准脱贫、助力省内全面振兴发展等政策是离不开的。为实现全面建成小康社会，摆脱贫困是一项十分重要的工作，黑龙江省在经过 5 年的努力，精准扶贫工作取得了一定的成果，相对于中国其他省份，黑龙江的成功尤为显著。其次情况较好的为北京，北京作为中国的首都，是全国的政治中心、经济中心、文化中心、国际交往中心、科技教育中心。北京的发展对周边地区的发展具有影响和带动作用，率先实现全面小康社会对中国其他城市的发展，具有极强的示范作用和借鉴作用。其他省份在全面建成小康社会的情况也有一定进步，但是受一些地方因素影响，进程稍微较慢，距离全面建成小康社会仍有较大差距，面临这一问题不容忽视。其中最为落后的是吉林省，在 2012 年进入全面建成阶段后，十年振兴政策消化期结束，新常态下东北经济遇冷，吉林省实现程度滑落至全国平均水平以下，如期实现全面建成小康社会的更高目标面临极大挑战。

5.3 问题 3 的模型建立与求解

5.3.1 问题 3 模型的建立

在问题 3 中, 针对世界处于百年未有的大变局及新型冠状病毒疫情全球性蔓延的大环境, 我们建立了以湖北省为代表的短期及长远的经济社会可持续发展模型。其中考虑了人口变化、人口老龄化趋势、城乡居民收入、资源禀赋和生态环境承载力这五个关键性的因素。采用主成分分析法和主成分回归分析的方法进行求解。

主成分分析 (Principal Component Analysis, PCA), 是一种统计方法。通过正交变换将一组可能存在相关性的变量转换为一组线性不相关的变量, 转换后的这组变量叫主成分。

主成分分析首先是由 K.皮尔森 (Karl Pearson) 对非随机变量引入的, 尔后 H.霍特林将此方法推广到随机向量的情形。信息的大小通常用离差平方和或方差来衡量。

主成分分析的主要目的是希望使用较少的变量去解释原来资料中的大部分的变异, 将我们手上许多相关性很高的变量转化成彼此相互独立或不相关的变量。通常是选用比原始变量个数少, 且新变量能解释大部分资料中变异的几个新变量即所谓的主成分, 且以解释资料的综合性指标。综上所述, 主成分分析法实际上是一种降维方法。

假设 X_1, X_2, \dots, X_n 表示以 x_1, x_2, \dots, x_n 为样本观测值的随机变量, 如果能够找到 c_1, c_2, \dots, c_n 能够使

$$\text{Var}(c_1 X_1 + c_2 X_2 + \dots + c_n X_n)$$

的值达到最大, 因为方差表示了数据差异的程度, 则表明我们设置的这 n 个变量的最大变异。基于种种不确定因素, 需要在式加上一部分限制, 否则权值就有可能选择无穷大则加权式就失去了意义。

规定

$$c_1^2 + c_2^2 + \dots + c_n^2 = 1$$

约束条件为式, 求出式的最优解。而求出的解为 p —维空间的一个响亮单位, 代表了一个“方向”, 即主成分方向。

假设 Y_i 表示第 i 个主成分, 且 $i = 1, 2, \dots, n$, 则

$$\begin{cases} Y_1 = c_{11}X_1 + c_{12}X_2 + \dots + c_{1n}X_n, \\ Y_2 = c_{21}X_1 + c_{22}X_2 + \dots + c_{2n}X_n, \\ \vdots \\ Y_n = c_{n1}X_1 + c_{n2}X_2 + \dots + c_{nn}X_n, \end{cases}$$

其中, 对于 i 个主成分, 均有 $c_{i1}^2 + c_{i2}^2 + \dots + c_{in}^2 = 1$ 且 $[c_{11}, c_{12}, \dots, c_{1n}]$ 使得 $\text{Var}(Y_1)$ 的值达到最大, $[c_{21}, c_{22}, \dots, c_{2n}]$ 不仅垂直于 $[c_{11}, c_{12}, \dots, c_{1n}]$ 且使得 $\text{Var}(Y_2)$ 的值达到最大, $[c_{31}, c_{32}, \dots, c_{3n}]$ 同时垂直于 $[c_{21}, c_{22}, \dots, c_{2n}]$ 、 $[c_{11}, c_{12}, \dots, c_{1n}]$, 且使 $\text{Var}(Y_3)$ 的值达到最大, 以此类推可得 n 个主成分。如果使用计算机来对主成分的个数进行确定将会非常方便。

注意事项如下:

主成分分析最大的问题在于量纲变化, 若各自变量的单位不一致, 且量纲发生变化, 结果将会发生不同。回归分析使不存在这种情况的, 实际中应用主成分分析解决问题时, 可以先把变量的数据标准化, 再使用协方差矩阵或相关系数矩阵进行分析;

使方差达到最大的主成分分析不用转轴;

主成分的保留。使用相关系数求主成分时，可以将特征值小于 1 的主成分放弃掉（SPSS 软件的默认值）；

主成分的目的是降维，减少变量的个数，所以一般选用少量的主成分，且选用的这一部分能够解释变异的绝大多数累计贡献率即可。

接下来进行特征值因子筛选，假设存在 n 个指标变量 x_1, x_2, \dots, x_n 则它在第 i 次实验中的取值为

$$a_{i1}, a_{i2}, \dots, a_{in}, i = 1, 2, \dots, m$$

则其矩阵形式为

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

矩阵 \mathbf{A} 成为设计阵。

实际问题中，解决问题式中的系数采用矩阵 $\mathbf{A}^T \mathbf{A}$ 的特征向量。

将矩阵 $\mathbf{A}^T \mathbf{A}$ 的特征值进行由大到小的次序排列之后，对这一部分特征值进行筛选。删除一部分特征值之和小于整体特征值之和 $\sum \lambda_i$ 的 15%，即余下的特征值所占比重超过 85%。

实际解决问题时，单纯考虑累计贡献率可能不够，需要选择主成分对原始变量的贡献值，使用相关系数的平方和便是，假设选用的主成分为 q_1, q_2, \dots, q_k ，则他们对原变量 x_i 的贡献值为

$$\rho_i = \sum_{j=1}^k r^2(q_j, x_i)$$

其中， $r(q_j, x_i)$ 表示 q_j, x_i 之间的相关系数

主成分回归分析是克服了最小二乘（LS）估计在数据矩阵 \mathbf{A} 存在多重共线性时，表现出来的不稳定性而提出来的。

主成分回归分析采用将原先的回归自变量变换到另一组变量，即主成分，选择其中的一部分重要的作为新的变量，丢弃一部分不重要的自变量，实际上是为了达到降维的目的，接着再使用最小二乘法对选取主成分后的模型参数进行估计，最后再变换回原先的模型求出参数的估计。

5.3.2 问题 3 模型的求解

基于上述模型，我们利用 python 编程软件对结果进行求解。最后得出如下图所示的一个热力图。

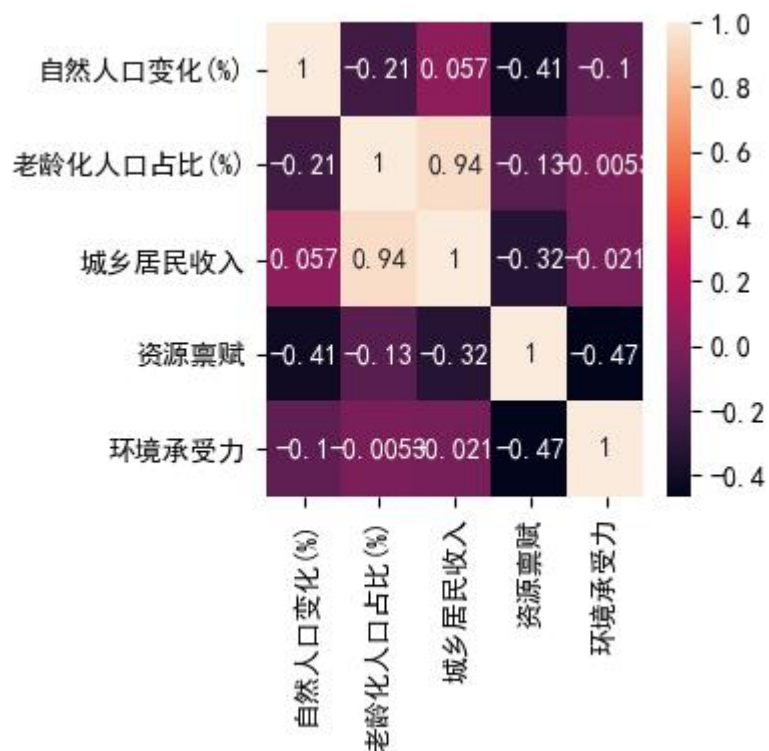


图 3-1 湖北省可持续发展情况热力图

从上图可以看出，在基于湖北省的研究基础上，得出了自然人口变化、老龄化人口比例、城乡居民收入、资源禀赋和环境承受力自己的关系

5.3.3 结果

总体上来说，以正统经济理论为基础，强调通过内部化未来激励生态服务提供的环境要素论，与坚持生态属性约束，强调规模、公平的经济要素论，处于生态和经济协调这个问题的两端，在它们中间，即在生态约束的条件下，经济激励的方法应该是在目前情况下协调经济和生态问题的有效率的途径。

可持续发展涉及到生态、经济、伦理、习惯、文化、价值等领域，但基础是生态环境的可持续性。而经济活动方式对我国可持续性构成了首要威胁。因此，经济往往成为实施可持续发展的核心问题。

在 2020 年 1 月，新型冠状病毒肺炎疫情发生后，中国政府迅速采取一系列有力措施阻止疫情蔓延，与疫情相关的科研持续推进并已经取得了一定进展。但当前仍处于疫情的高发期，现在要对疫情对中国经济的影响进行完全准确的评估仍较为困难。但政府采取的持续控制疫情的措施有助于尽快控制疫情，防范其危害的进一步扩大。疫情冲击着交通运输、住宿餐饮、旅游等服务业首当其冲，成为经济增长最大的拖累因素。

虽然遭受疫情影响，但是我国展现出了大国风范，积极采取各项有效措施，很好的控制了疫情的进一步蔓延。也能为我国实现经济社会可持续发展这一目标。

5.4 问题 4 的模型建立与求解

题目要求我们提供一个合作计划，需要考虑到各省份经济社会发展状况及资源禀赋。我们团队考虑在问题二的基础上，将问题二中提到的指标作为二级指标，再将 GDP 增长率和资源禀赋作为一级指标，运用聚类分析的方法将二级指标聚合成几个一级指标，然后对一级指标进行分析，建立模拟退火模型得出每个省的发展程度和分配结果，得出各省之间的按比例的合作计划。

5.4.1 聚类分析

因为首先经济社会发展状况及资源禀赋涉及到多方面，在此问中可以作为一级指标进行分析，可以通过结合问题二中提到的指标综合考虑，问题二中的指标作为二级指标，运用聚类分析法将二级指标聚合成一级指标，采用模拟退火的算法建立可持相关模型，综合分析得出每个省的发展程度和分配结果。

采用 SPSS 对第二问中的人均 GDP、高中入学率、恩格尔系数、人均居住面积和平均寿命、城镇人口比例、未来可持续发展的指标。在题目中，考虑到各个因素对经济社会发展状况、资源禀赋的影响，我们采用 R 型聚类分析法对附件数据按社会发展和资源禀赋相关性进行分组。

- (1) 先用 spss 对数据进行预处理，选择近十年的人均 GDP、高中入学率、恩格尔系数、人均居住面积、平均寿命、水资源、森林面积、二氧化碳浓度、人口数量数据进行筛选。
- (2) 定义一个衡量数据点之间相似度的函数
- (3) 进行聚类、分类

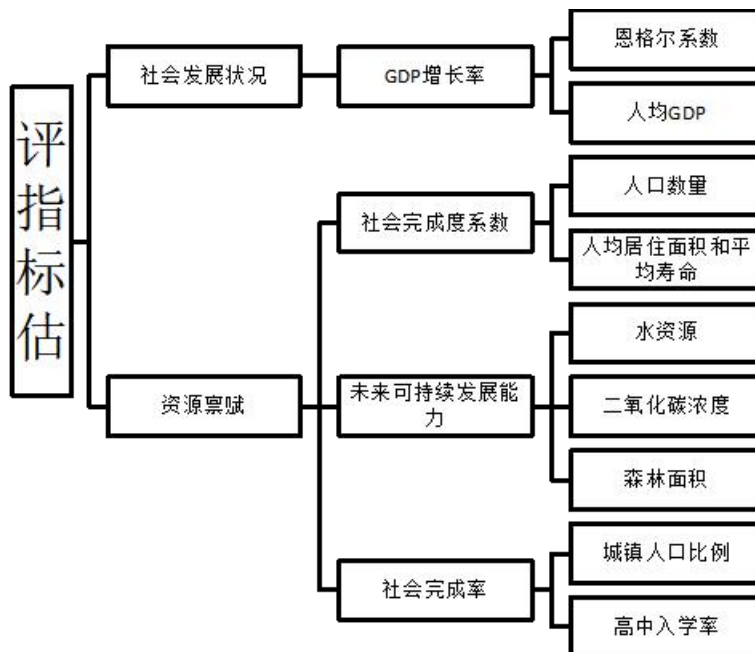


图 4-1 评价指标图

- (4) 最后分组聚类成经济社会发展状况、资源禀赋两大方面，如图所示：

前解。

⑥如果满足终止条件则输出当前解作为最优解，结束程序。终止条件通常取为连续若干个新解都没有被接受时终止算法。

⑦ T 逐渐减少，且 $T \rightarrow 0$ ，然后转第2步。

模拟退火的算法流程图如下：

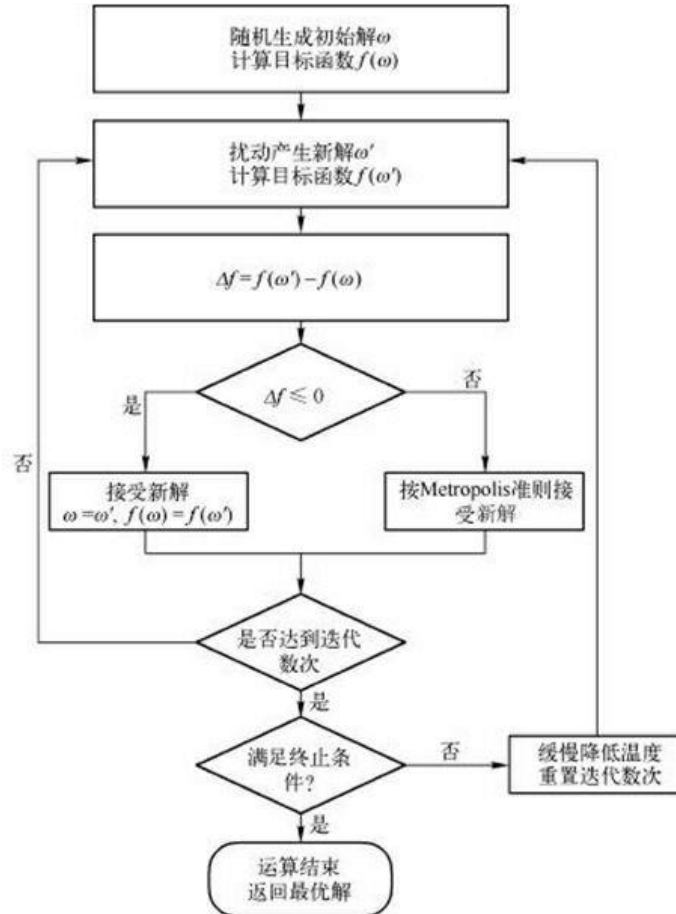


图 4-4 模拟退火算法流程图

模拟退火算法新解的产生和接受可分为如下四个步骤：

第一步是由一个产生函数从当前解产生一个位于解空间的新解；

第二步是计算与新解所对应的目标函数差。

第三步是判断新解是否被接受,判断的依据是一个接受准则，常用的接受准则是 *Metropolis* 准则: 若 $\Delta t' < 0$ 则接受 S' 作为新的当前解 S ，否则以概率 $\exp(-\Delta t/T)$ 接受 S' 作为新的当前解 S 。

第四步是当新解被确定接受时，用新解代替当前解，这只需将当前解中对应于产生新解时的变换部分予以实现，同时修正目标函数值即可。

根据 *Metropolis* 准则，粒子在温度 T 时趋于平衡的概率为 $\exp(-\Delta E/kT)$ ，其中 E 为温度 cT 时的内能， ΔE 为其改变数, k 为 *Boltzmann* 常数 *Metropolis* 准则常表示为：

$$p = \begin{cases} 1 \\ \exp\left(-\frac{E(x_{\text{new}}) - E(x_{\text{old}})}{T}\right) \end{cases}$$

Metropolis 准则表明，在温度为 T 时，出现能量差为 dE 的降温的概率为 $p(dE)$ ，表示为： $p(dE) = \exp(dE/kT)$ 。其中 k 是一个常数， \exp 表示自然指数，且 $dE < 0$ 。所以 p

和 T 正相关。这条公式就表示：温度越高，出现一次能量差为 dE 的降温的概率就越大；温度越低，则出现降温的概率就越小。又由于 dE 总是小于 0（因为退火的过程是温度逐渐下降的过程），因此 $dE/kT < 0$ ，所以 $p(dE)$ 的函数取值范围是 $(0,1)$ 。随着温度 T 的降低， $p(dE)$ 会逐渐降低。

通过建立模拟退火模型得出每个省的发展程度和分配结果，如图：

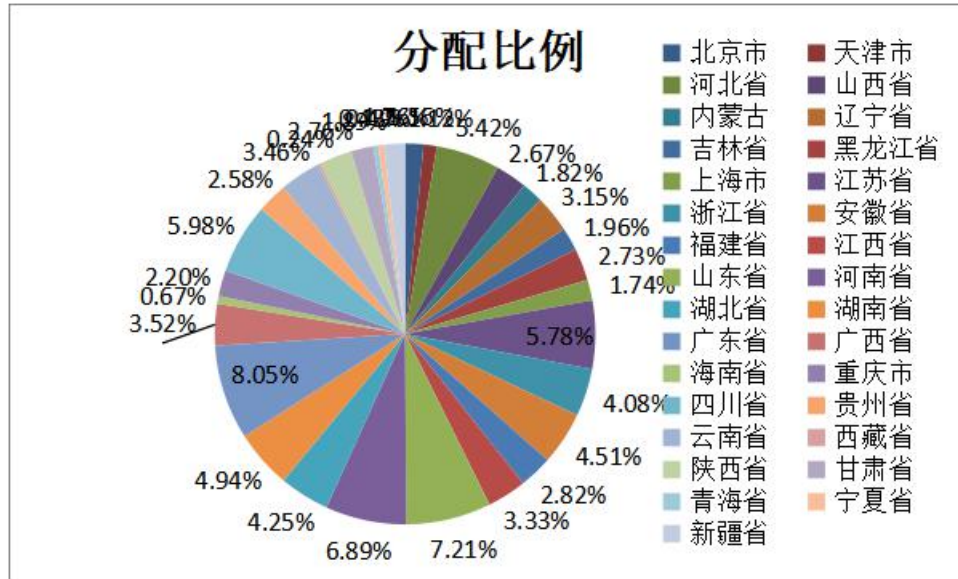


图 11 各省合作分配比例图

我们综合且充分考虑了各省的经济社会发展状况、资源禀赋，得到上述综合的各省发展程度分配比例图，按照图示进行各省的合作，提高各省之间相互合作双方产生积极的促进作用和总体效率。

六、模型的评价及优化

6.1 误差分析

6.1.1 针对于时间序列预测的误差分析

利用时间序列进行预测，预测值和国家统计局的真实值进行比较，做出各指标的相对误差图，如下：



图 12 水资源相对误差图

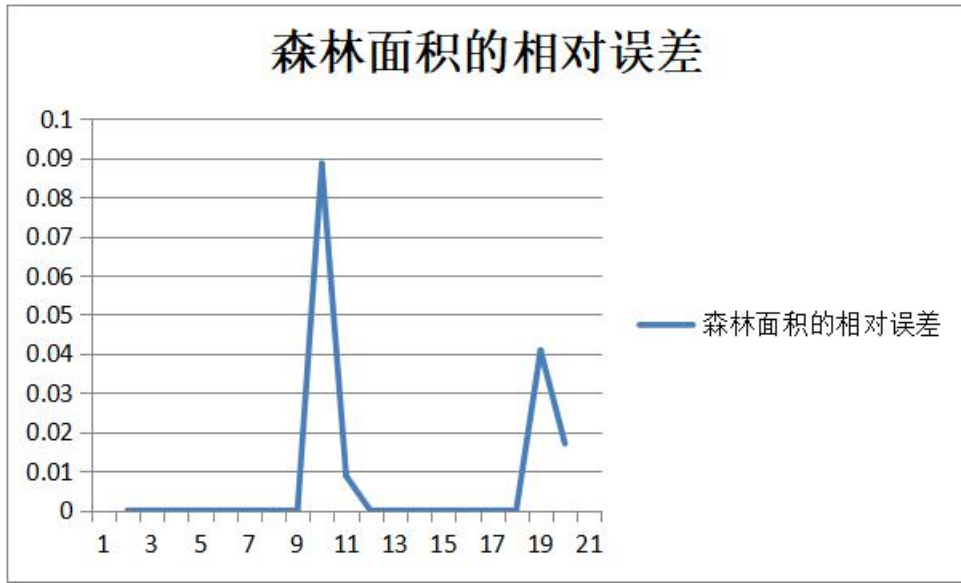


图 13 森林面积的相对误差图

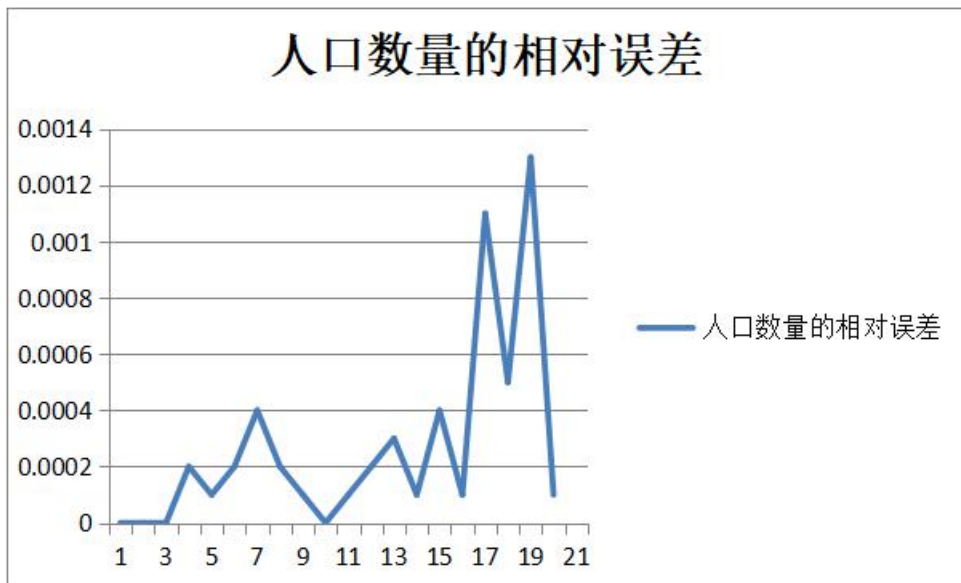


图 14 人口数量的相对误差图

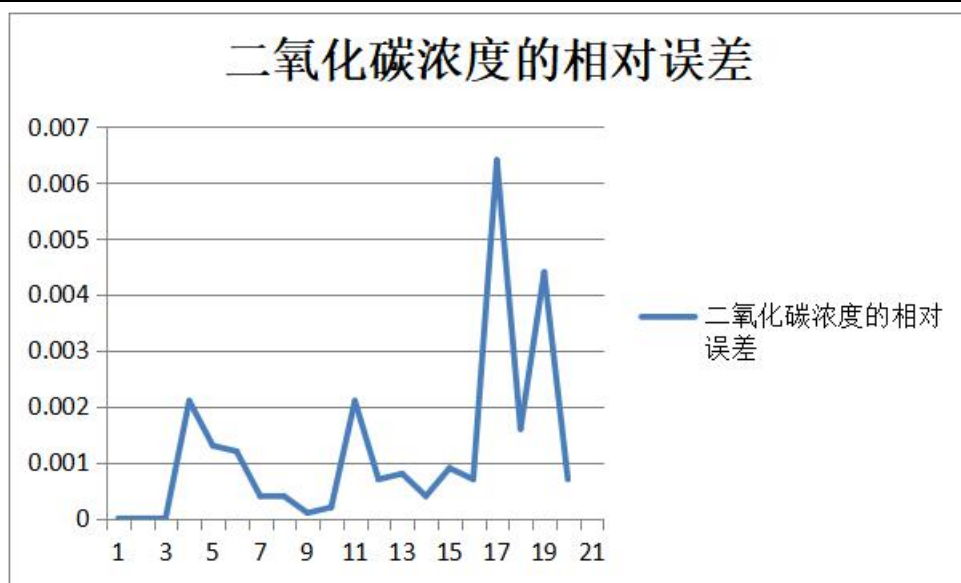


图 15 二氧化碳的相对误差图

从上述图中可以看出，虽然曲线变化比较大，但是纵坐标的数值量比较小，所以预测值和真实值的相对误差并不大，证明采用时间预测是具有一定的可行性，比较能够准确的预测未来 20 年的水资源、森林面积、人口数量、二氧化碳浓度，从而制定比较合乎发展的可持续发展思路参考。

但是由于可持续发展在资源方面只考虑了水资源，并没与考虑石油、天然气、铁矿等有限资源，在生态环境方面也只考虑了二氧化碳、森林面积，并没有考虑到自然承受能力等指标，也没余考虑科技因素所以该预测还是存在一定的误差，需要尽可能考虑多种因素使可持续发展的思路得以保障。

6.1.2 针对主成分分析方法的误差分析

单纯进行主成分分析法，单纯的进行因素表达式求解和排列，并没有考虑到所因素之间的相互联系、相互影响。我们可以进行模型的优化在主成分分析法的基础上建立回归模型，建立各因素之间的联系来使结果更为准确。

先对每个指标进行细分，各自建立主成分分析，

$$F_1 = 0.316X_1^* + 0.352X_2^* + 0.356X_3^* + 0.242X_4^* + 0.360X_5^* + 0.367X_7^* + 0.262X_8^*$$

$$F_2 = 0.368X_1^* + 0.398X_2^* + 0.268X_3^* + 0.423X_4^* + 0.256X_5^* + 0.354X_7^* + 0.354X_8^*$$

$$F_3 = 0.301X_1^* + 0.312X_2^* + 0.365X_3^* + 0.246X_4^* + 0.349X_5^* + 0.667X_7^* + 0.257X_8^*$$

F1、F2、F3 分别表示我国各省全面建成小康社会完成难度系数、完成比率及未来可持续发展能力的三个指标。

在主成分分析的基础上建立回归模型，可以利用多元回归分析建立一个因素与多个因素相联系，得到以下权重占比表达式：

$$Y = 0.536F_1 + 0.648F_2 + 0.395F_3$$

然后将各省数据带式中重新建成情况系数，并标准化如下

表 2 各省标准化系数表

北京	天津	河北	河南	山西	山东	黑龙江	吉林
.936	.911	.699	.755	.601	.845	0.599	0.594
辽宁	浙江	福建	广东	内蒙古	江苏	广西	海南
0.622	.758	.701	.945	0.602	.845	0.609	
四川	重庆	贵州	云南	陕西	甘肃	青海	宁夏

.736	0.866	0.598	0.611	0.635	0.566	0.544	0.531
湖南	湖北	新疆	西藏	安徽	江西		
0.697	.711	0.564	0.522	0.688	0.614		

根据表中数据，可以分析排序，结果比只单纯的进行主层析分析考虑更为合理。

6.2 模型的优点

- 1、主成分分析法较好的解决的综合评价的问题，能够使得出来的结果更合理准确。
- 2、模拟退火算法的局部搜索能力强，可以简单运用此模型通过 GDP 增长率、资源禀赋指标进行各省的规划分配，运行时间较短、便捷的解决规划问题。
- 3、本文建立的模型能够与实际紧密联系，结合实际情况对问题进行求解，使得模型具有很好的通用性和推广性。
- 4、模型是计算使用了专业是数学模型，可行度较高。
- 5、本文对模型中涉及到的众多影响因素进行了量化分析，使得文章具有说服力。

6.3 模型的缺点

- 1、在编程中，没有加入一些适当的约束，导致了最终的运算结果出现误差，最后，我们采用人工的方法进行了较好的弥补。
- 2、在模型计算中，由于有些结果的小数点过多，我们取了预计值作为近似值来计算，这与实际情况会有一点出入。
- 3、模型中使用的数据不够全面，只是采用了主要指标对模型做的简单处理和分析。

6.4 模型的推广

本文的可采用时间序列、主层次分析、持续发展模型、模拟退火模型可以运用到各个省份中，也可以进行两个省份之间的对比，可以在对各省经济问题等方面进行简单的综合评估，且本文中使用的模型和求解都比较简单易懂，运行时间也比较短，预测值和综合评价都具有具有一定的合理性。但是由于选取的相关因素较为少，模型的结果具有一定的欠缺性，所以我们的模型如何多方面建立联系，综合的考虑多因素的影响是需要进行改进的地方。

参考文献

- [1]国家统计局, <http://www.stats.gov.cn/>, 2020.6.4
- [2]韩保江, 邹一南, *中国小康社会建设 40 年: 历程、经验与展望[J], 经济研究参考, 2020, 4, 102-117
- [3]杨海民, 潘志松, 白玮, 时间序列预测方法综述[J], 计算机科学, 2019,46 (1), 21-28
- [4]张鹏. 基于主成分分析的综合评价研究[D].南京理工大学,2004.
- [5]韩小雷. 粒子群—模拟退火融合算法及其在函数优化中的应用[D].武汉理工大学,2008.
- [6]张建萍. 基于计算智能技术的聚类分析研究与应用[D].山东师范大学,2014.

附录 1: 问题 1 中时间序列的 matlab 处理

```

时间序列
clear;clc;close all;
A=xlsread('AA.xlsX');%万通地产股票历史数据 19to20 年数据
N = length(A);
H = lbqtest(A)%原始序列纯随机性检验 =1 不是随机, =0 纯随机

t=1:N;
plot(t,A,'k*'); %原始序列时序图
axis([1 303 1 8]);
title('原始序列时序图');

D = 0; %初始化差分阶数
dA = A;
if adfstest(A) ~= 1 %adfstest=0 表示非平稳序列, adfstest=1 表示平稳序列
    dA = diff(A); D = 1;
    while adfstest(dA) ~= 1 %判断一次差分后是否平稳,如果时间序列不平稳就再次进行下一次差分,直到平稳为止
        dA = diff(dA); D = D+1;
    end
end
adfstest(dA) %adfstest=1 表示差分后的时间序列达到平稳
D %输出最后达到平稳所需要经过的差分次数
figure
nd=length(dA);%差分后的数据长度(会随着 D 增大而减小)
t=1:nd;
plot(t,dA,'r*-');%D 阶差分后时序图
title('2 阶差分后时序图');
axis([1 nd -0.25 0.2]);

%定 p\q 值
figure
subplot(121);autocorr(dA) %画出自相关图,图中上下两条横线分别表示自相关系数的上下界,超出边界的部分表示存在相关关系。
[ACF,lagsacf,boundsacf] = autocorr(dA); %ACF 为各阶的相关系数, lagsacf 为滞后阶数
subplot(122);parcorr(dA) %画出偏自相关图
[PACF,lagspacf,boundspacf] = parcorr(dA); %PACF 为各阶的偏自相关系数, lagspacf 为滞后阶数

k = 0;%初始化的试探模型的个数;
for i=0:2 %p 最大为 2
    for j=0:2 %q 最大为 2
        if i==0 && j==0
            continue
        elseif i==0
            ToEstMd=arima('MALags',1:j,'Constant',0);%指定模型的结构为 MA 模型
        elseif j==0
            ToEstMd=arima('ARLags',1:i,'Constant',0);%指定模型的结构为 AR 模型
        else
            ToEstMd=arima('ARLags',1:i,'MALags',1:j,'Constant',0);%指定模型的结构为 ARMA 模型
        end
    end
end

```

```

        k=k+1; p(k)=i; q(k)=j;
        [EstMdl,EstParamCov,logL,info]=estimate(ToEstMd,dA); %模型拟合,将假设的模型和数据拟
合起来;
        numParams = sum(any(EstParamCov)); %计算拟合参数的个数
        %compute Akaike and Bayesian Information Criteria
        [aic(k),bic(k)]=aicbic(logL,numParams,nd); %求取每个模型对应的 AIC, BIC 值, 值取到最
小时对应的模型最优
    end
end
fprintf('p,q,AIC,BIC 的对应值如下\n %f'); %显示计算结果
check=[p',q',aic',bic']
[minrowaic,minaic]=min(check(:,3)) %输出 AIC 最小值所在行和相应的最小值
[minrowbic,minbic]=min(check(:,4)) %输出 BIC 最小值所在行和相应的最小值
r=input('输入阶数 p=');m=input('输入阶数 q=');

M = armax(dA,[r,m]); %模型参数估计
dAs =predict(M,dA); %返回预测结果
res = dA - dAs;%计算残差向量
h = lbqtest(res) %进行模型适应性检验, h=0 表示接受假设,即模型适合,1 表示拒绝假设
if h == 0
    if r == 0 %即 p=0
        ToEstMd=arima('MALags',1:m,'Constant',0); %指定模型的结构为 MA 模型
    elseif m==0 %即 q=0
        ToEstMd=arima('ARLags',1:r,'Constant',0); %指定模型的结构为 AR 模型
    else %即 p,q 均不为 0
        ToEstMd=arima('ARLags',1:i,'MALags',1:j,'Constant',0); %指定模型的结构为 ARMA 模型
    end
    np=1;%预测未来 1 天
    [EstMd,EstParamCov,logL,info]=estimate(ToEstMd,dA); %模型拟合
    dx_Forecast=forecast(EstMd,np,'Y0',dA); %计算 1 步预测值
end

%求一阶差分的还原值
% x_Forecast=zeros(np+1,1);%预测值,第一个值是原始数据的最后一个值
% x_Forecast(1)=A(end); %x(k+1)-x(k)=dx_Forecast
% for tp=1:np
%     x_Forecast(tp+1)=x_Forecast(tp)+dx_Forecast(tp); %求 x 的预测值
% end
% x_Forecast

%求二阶差分的还原值
x_Forecast=zeros(np+2,1);%x_Forecast 预测值,前两个值是原始数据的后两个值
x_Forecast(1)=A(end-1); x_Forecast(2)=A(end);
for tp=1:np %x(k+2)-2*x(k+1)+x(k)=dx_Forecast
    x_Forecast(tp+2)=-x_Forecast(tp)+2*x_Forecast(tp+1)+dx_Forecast(tp); %求 x 的预测值
end
x_Forecast

%画图:
figure
plot(A,'bo');
hold on
plot(N+1:N+np,x_Forecast(3:np+2),'ro','LineWidth',2);

```

```
legend('实际值','预测值')
```

附录 2: 问题 2 中 matlab 数据处理排序

```
clc
clear all
A=xlsread('1.xlsx');
%得到的数据矩阵的行数和列数
a=size(A,1);
b=size(A,2);
%数据的标准化处理:得到标准化后的矩阵 SA
for i=1:b
    SA(:,i)=(A(:,i)-mean(A(:,i)))/std(A(:,i));
end
%计算系数矩阵:CM
CM=corrcoef(SA);
%计算 CM 的特征值和特征向量
[V,D]=eig(CM);
%将特征值按降序排列到 DS 中
for j=1:b
    DS(j,1)=D(b+1-j,b+1-j);
end
%计算贡献率
for i=1:b
    DS(i,2)=DS(i,1)/sum(DS(:,1));%单个贡献率
    DS(i,3)=sum(DS(1:i,1))/sum(DS(:,1));%累计贡献率
end
%假定主成分的信息保留率
T=0.9;
for k=1:b
    if DS(k,3) >= T
        com_num=k;
        break;
    end
end
%提取主成分的特征向量
for j=1:com_num
    PV(:,j)=V(:,b+1-j);
end
%计算主成分得分
new_score=SA*PV;
for i=1:a
    total_score(i,1)=sum(new_score(i,:));
    total_score(i,2)=i;
end
%强主成分得分与总分放到同一个矩阵中
result_report=[new_score,total_score];
%按总分降序排列
result_report=sortrows(result_report,-4);
%输出结果
disp('特征值、贡献率、累计贡献率: ')
DS
disp('信息保留率 T 对应的主成分数与特征向量: ')
com_num
PV
```



```
disp('主成分得分及排序')
```

```
result_report
```

附录 3: 问题 3 中 matlab 主成分分析数据处理

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA

data = pd.read_csv('./data.csv')
data_ = data.iloc[:,1:]
# 求变量之间的相关系数
corr = data_.corr(method='pearson')
# 画出变量相关性热力图
plt.rcParams['font.sans-serif']=['SimHei'] #用来正常显示中文标签
plt.rcParams['axes.unicode_minus']=False #用来正常显示负号
f,ax = plt.subplots(figsize=(4,4))
sns.heatmap(corr,annot=True)
plt.show()
# 经过分析发现相关系数矩阵中 corr 存在 0.94 这么大的数值,
# 这说明这些因素中存在相关性比较强的指标, 这种指标取一个就行
# 多个显得多余(进行变量压缩降维)

# 数据的量纲不一致, 需要对数据进行标准化再进行主成分分析
from sklearn.preprocessing import scale

# 进行中心标准化
data_scaled = scale(data_)
#whiten 进行白化处理可以减少数据之间的相关性
pca = PCA(n_components=5,whiten=True).fit(data_scaled)
"""
pca.explained_variance_:经主成分之后得到的方差值
pca.explained_variance_ratio_:经主成分之后得到的方差值所占的比例
pca.components_:主成分向量
```

附录 4: 问题 4 中模拟退火模型 matlab

模拟退火算法代码

```
sol_new2=30;% (1) 解空间 (初始解)
sol_new1=2-sol_new2^2;
sol_current1 = sol_new1;
sol_best1 = sol_new1;
sol_current2 = sol_new2;
sol_best2 = sol_new2;
sol_current3 = sol_new3;
sol_best3 = sol_new3;
E_current = inf;
E_best = inf;
rand('state',sum(clock));%初始化随机数发生器
t=90;%初始温度
tf=89.9;%结束温度
a = 0.99;%温度下降比例
```

```

while t>=tf% (7) 结束条件
    for r=1:1000 %退火次数

        %产生随机扰动 (3) 新解的产生
        sol_new2=sol_new2+rand*0.2;
        sol_new1=2-sol_new2^2;

        %检查是否满足约束
        if sol_new1^2-sol_new2<=0    &&    -sol_new1-sol_new2^2+2==0    &&    sol_new1>=0
&&sol_new2>=0
            else
                sol_new2=rand*2;
                sol_new1=2-sol_new2^2;
                continue;
            end
        %退火过程
        E_new=sol_new1^2*sin(2*(sol_new2+sol_new3))/g+h/tan(sol_new 2+ sol_new 3);
        % (2) 目标函数
        if E_new<E_current% (5) 接受准则
            E_current=E_new;
            sol_current1=sol_new1;
            sol_current2=sol_new2;
            if E_new<E_best
                %把冷却过程中最好的解保存下来
                E_best=E_new;
                sol_best1=sol_new1;
                sol_best2=sol_new2;
            end
        else
            if rand<exp(-(E_new-E_current)/t)% (4) 代价函数差
                E_current=E_new;
                sol_current1=sol_new1;
                sol_current2=sol_new2;
            else
                sol_new1=sol_current1;
                sol_new2=sol_current2;
            end
        end
        end
        plot(r,E_best,'*')
        hold on
    end
    t=t*a;% (6) 降温
end

disp('最优解为: ')
disp(sol_best1)
disp(sol_best2)
disp('目标表达式的最小值等于: ')
disp(E_best)

```