

For office use only

T1 _____

T2 _____

T3 _____

T4 _____

Team Control Number

86765

Problem Chosen

B

For office use only

F1 _____

F2 _____

F3 _____

F4 _____

2018

MCM/ICM

Summary Sheet

Your title

Summary

However, this causes a couple of warnings and bad boxes in the compilation process and the figure gets moved down and placed not after the "esempio" box, but after even the next paragraph, and it also creates huge blank spaces... quite weird.

However, this causes a couple of warnings and bad boxes in the compilation process and the figure gets moved down and placed not after the "esempio" box, but after even the next paragraph, and it also creates huge blank spaces... quite weird.

However, this causes a couple of warnings and bad boxes in the compilation process and the figure gets moved down and placed not after the "esempio" box, but after even the next paragraph, and it also creates huge blank spaces... quite weird.

key words : complex network, entropy

MEMO

FROM: Team 86765 , MCM C

To: The group of Governors

Date: February 13, 2019

SUBJECT: Goals for the inter state energy compact

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Sincerely yours
ICM E Team 86765

Contents

1	Introduction	1
1.1	Restatement of the Problem	1
1.2	Our Work	1
2	Assumptions and Notations	1
2.1	Assumptios	2
2.2	Notations	2
3	Data Preprocessing	2
3.1	Default Value Processing	2
3.2	Data Normalization and Classification	3
4	Model Construction	3
4.1	Model I: Construct Energy Profile	3
4.1.1	Calculate Profile Weight	3
4.1.2	Energy Evolving	5
4.2	Model II	7
4.2.1	Interpret Similarities And Differences	7
4.2.2	Analysis Influential Infactor	8
4.3	Model III	9
4.3.1	Producing and Planning Goals	9
5	Model Extension and Sensitivity Analysis	10
5.1	Sensitivity	10
5.2	Model Extension	11
6	Strengths and Weaknesses	11
6.1	Strengths	12
6.2	Weaknesses	12
7	Conclusion	13
	References	14
	Appendices for Code and Data	14

1 Introduction

1.1 Restatement of the Problem

随着能源的日益紧缺和环境污染的加重,人们开始考虑使用清洁能源,能源结构发生转型,即非清洁能源的使用开始减少。同时结合地域、科技等优势条件,发展相对有利的清洁能源来满足能源需求。

因此能够准确的构建能源画像,对能源使用的现状进行描述和分析,描述能源的分布情况和使用情况显得尤为重要。另外,根据能源画像能够确定出能源使用的最佳情况,并且分析各大州的异同点,包括对不同大州自身而言更有利的清洁能源甚至某些缺陷。最终能够根据能源画像的对比和差异实现对未来时间能源的使用做出合理的规划。

1.2 Our Work

首先建立画像模型,使用多级标签描述四大州的能源画像,标签能够直观的描述和分析能源是否清洁、数量、种类和用途等;之后确定画像模型中标签权重和计算标签得分,以此来描述四大州能源演变情况、使用差异和衡量标准。基于上述内容,对各个标签下的能源的使用趋势进行预测,得到 2025-2050 年能源的使用情况;进而确定 2025-2050 年间能源使用的总目标。

We will proceed as follows for the sake of tackling these problems:

1. 根据提供的数据,建立画像模型,设计多级标签构建四个州的能源画像,描述能源使用情况与分布情况。
2. 对画像模型中的标签赋予权重并计算得分,描述 1960-2009 年内四个州的能源演变情况,并根据得分量化四大州之间的清洁、可再生能源使用最佳情况及使用的差异,最终从经济、人口角度分析产生差异的原因。
3. 同样基于画像模型,在没有政策约束的条件下,综合考虑不同标签下能源的得分差距和历史分布情况,模拟 2025-2050 年内能源演变情况;根据模拟情况对 2025-2050 年间的能源使用情况列出规划目标。

2 Assumptions and Notations

2.1 Assumptios

为了能够合理的简化问题，我们在模型建立的过程中提出如下假设：

1. 随着环境污染的日益严重，清洁能源的使用逐渐进入人们视野。因此假设清洁能源比非清洁能源更加重要 [1]，而且本文也重点讨论清洁能源。
2. 题目中数据表的负数表示两个含义 [2]，表示入不敷出或者对以往数据的更新。为简化计算，本文假设负数的含义为能源入不敷出的情况。
3. 假设各州在未来时间内各州的气候、人口、经济状况不会产生严重变化。使我们的模型能够准确的应用于普遍情况；而对于极端变化的情况本文不予讨论。

2.2 Notations

Here are the variables and their meanings in our paper.

Symbol	Meaning
η	能源种类的权重
P	人口数量
μ	能源使用的增长率
T_i	未来 i 年各大州能源使用的目标
$ \vec{d} $	各州能源使用量总差异
x	能源种类

3 Data Preprocessing

在数据中存在较多的冗余数据、缺失数据和一些未记录的数据，为了提高模型的有效性和结论的准确性，首先对数据进行预处理。

3.1 Default Value Processing

(1) 删除冗余数据。因题目中网址 [3] 指出了数据间转换关系，因具有相同的 MSNcode 的数据可相互转换。因为能源总热量相比总数量、总金额能更好的描述能源画像，因此我们只选择数据类型为热量的数据建立模型，删除其余类型的数据。

(2) 为了减少缺失数据对建模造成的精度影响，对缺失 70% 数据以上的属性进行删除。而对于并非严重缺失得到数据，我们使用拟合的方法进行填充，假设拟合数据产生的误差可忽略不计。

3.2 Data Normalization and Classification

(1) 为了消除数据量纲对结果的影响，我们对数据进行标准化处理，标准化形式为：

$$x_{new} = (x - x_{min}) / (x_{max} - x_{min}) \quad (1)$$

x 表示同一类型能源的取值， x_{max} 和 x_{min} 表示相同能源的最大值和最小值。

(2) 最终，对处理后的数据分为清洁能源和非清洁能源两类。清洁能源的含义为：能被直接使用和不排放污染气体的能源；非清洁能源的定义为：不可再生且污染环境的能源 [4]。最终，清洁能源包括水电、风能、太阳能、地热能、核能和乙醇；非清洁能源包括石油、天然气、煤。

(3) 而考虑到未来主要以消耗清洁能源为主，本文也将重点分析清洁能源，因此，保留所有清洁能源属性的数据。缺失数据仍以拟合的方式进行填充。

4 Model Construction

本文首先建立能源画像模型，以此为基础，计算画像下四大州能源的得分情况，基于得分情况描述能源使用情况和确定清洁能源最佳使用情况。最终对 2025-2050 年间能源使用的情况进行预测，根据预测情况规划目标。基于能源画像与预测情况，便可对政府提出能源使用的建议。

4.1 Model I: Construct Energy Profile

为了能够全面描述能源画像，需要尽可能减少冗余数据。因此，一级标签对能源的划分标准为：是够清洁；二级标签在一级标签基础上进一步划分为光能、风能、化学能等各种能源；而三级标签在二级标签的基础上将能源划分为不同的来源和用途。画像模型中的多级标签实现了对能源的层次划分和减少冗余，同时也能够全面的描述能源画像。最终，建立的画像模型中的多级标签依赖如图 1 所示。

在建立起能源画像之后，本问计算了标签权重与能源得分，详细描述了各大州的能源结构与使用情况，指出各大州主要消耗的能源与需求增长率，并且分析了各大州生产清洁能源的主要来源。

4.1.1 Calculate Profile Weight

(1) 对于能够过获得数据的标签，如图 1 中的二级标签和三级标签，使用系数变异法计算其权重。系数变异法的定义为标准差 σ 与均值 $E(x)$ 的比值，

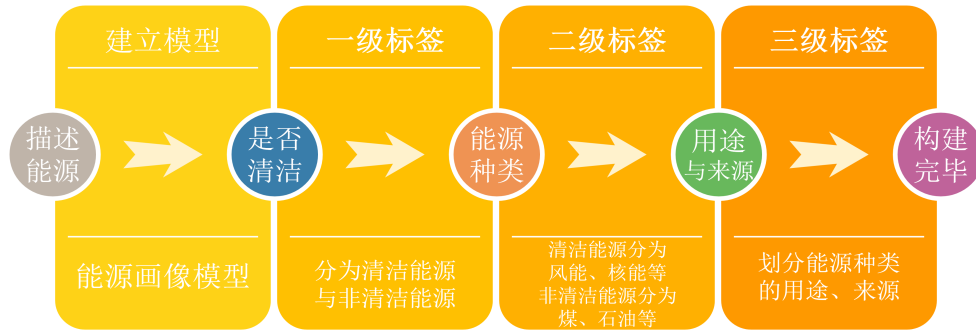


Figure 1: 标签分级示意图

能够反映出能源的使用状况和稳定状况。

- 为能更详细描述能源，使用变异系数法从能源生产和消耗两个角度分别计算了各大洲的权重。记能源生产的权重为 \mathcal{W} ，能源消耗的权重为 \mathcal{Y} 。
- 而能源的最终目的为使用，由哪个部门使用并无区别，因此假设能源的各种用途权重均相等。

由此，计算得到的四大州各个能源的消耗权重 \mathcal{W} 如表 1 所示。

Table 1: 能源消耗权重

States	BMTCB	EMTCB	ESTCB	GETCB	CLTCB	NGTCB	PMTCB
AZ	0.672	1.41	0.611	0.14	0.708	0.4	0.395
CA	0.291	1.464	0.35	0.751	0.228	0.148	0.191
NM	0.425	0.749	0.519	0.679	0.533	0.151	0.233
TX	0.381	1.885	0.502	0.711	0.755	0.126	0.349

生产能源的权重 \mathcal{Y} 如表 2 所示，定义在二级标签中，同一能源的权重为生产权重与消耗权重的均值。

Table 2: 能源生产的权重

States	HYTCB	NUETB	SOEGB	WYTCB
AZ	0.385	0.326	0.819	0.1
CA	0.298	0.824	0.518	0.616
NM	0.647	0	0	0.525
TX	0.372	0.367	0.581	1.305

(2) 对于抽象的概念，如图 1 中一级标签内的清洁能源与非清洁能源，使

用层次分析法确定其权重。最终得到的清洁能源的权重 $\eta_1 = 0.75$ ，非清洁能源的权重为 $\eta_2 = 0.25$ 。

在得到各标签之间的权重之后，以权重乘以数据得到能源使用情况的得分，至此，能源画像模型构建完毕。之后根据得分情况对各州的能源画像进行分析和表述，分析能源使用情况。

4.1.2 Energy Evolving

如图 2 所示，对消耗的总能源而言，TX 和 CA 州的能源使用量最高，由能源消耗的权重 \mathcal{W} 可知能源消耗以电能为主，分析数据可知能源主要用于工业和商业部门。同样，TX 州和 CA 州对清洁能源的使用比例也相对较高，而清洁能源消耗的权重可知主要来源为风电、水电和地热能。

而对于 NM 州而言，清洁能源的利用率也较为良好。而详细分析 NM 州能源生产的权重 \mathcal{Y} 可以了解到虽然对核能和太阳能的利用较为落后，但是对风电、水电能源的利用较为良好；分析 \mathcal{W} 可知，NM 州清洁能源利用率低于 CA, TX 州的原因对于非清洁能源煤和天然气的使用较为频繁，应增加风电、水电的投入以减少煤和天然气的使用。

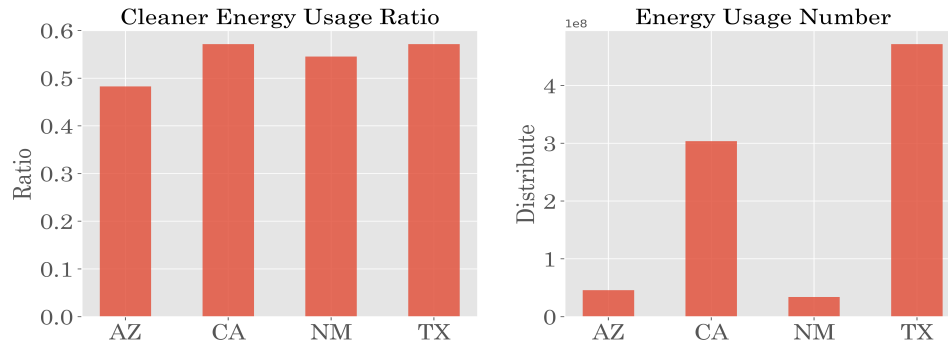


Figure 2: 能源使用示意图

如图 3 所示，随着清洁环保的概念开始普及，各州的清洁能源的使用量都在上升，只是因各州情况不同而导致使用量也不相同，TX 州和 CA 州的使用量明显高于 AZ 州与 NM 州。

为了消除能源使用基数的影响，我们定义各州在 49 年内清洁能源使用的增长率 μ :

$$\mu = \exp(10^{-6}(\sum_{i=1961}^{2009} C_i - C_{i-1})/49) \quad (2)$$

其中, C_i 表示第 i 年的清洁能源使用量。计算得到四大州的增长率为: AZ:0.69, CA:0.76, NM:0.93, TX:0.61。表明 NM 州在清洁能源利用中有较大发挥空间。

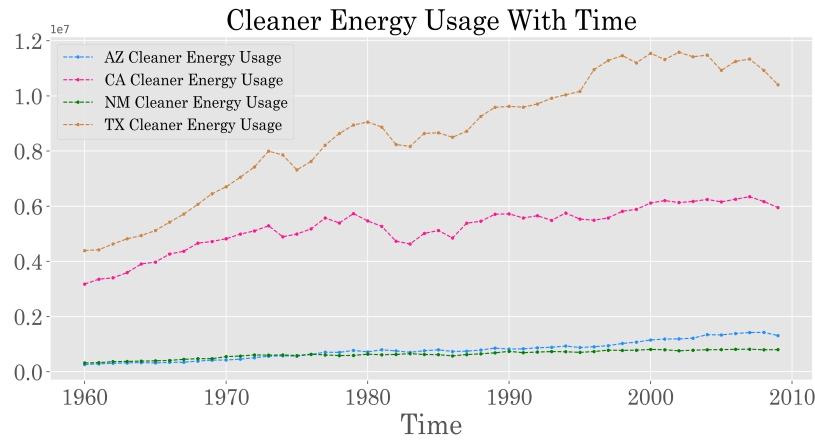


Figure 3: 清洁能源使用对比图

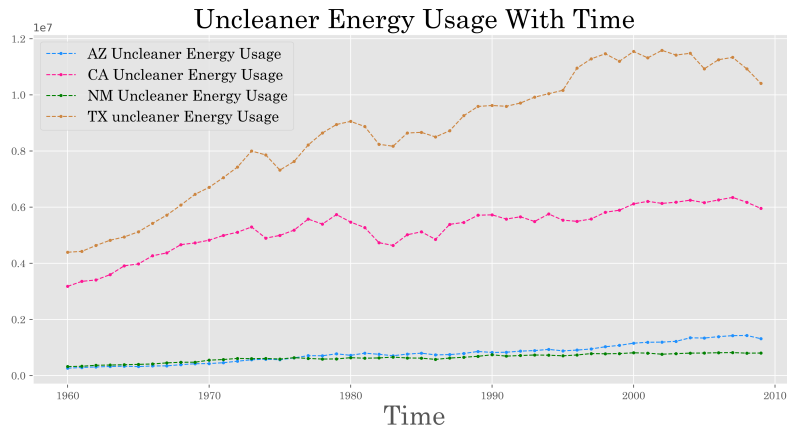


Figure 4: 非清洁能源使用对比图

如图 4 所示，因各大州对能源总量需求不同的原因，各大州非清洁能源的消耗总量情况与清洁能源类似，CA 州与 TX 州的消耗量明显高于 AZ 州与 NM 州。同样根据公式 2，计算得到的四大州对非清洁能源使用的增长率，AZ:0.76，CA:0.72，NM:0.73，TX:0.64。因此可知，四大州对非清洁能源需求的增长量几乎相同。但是 AZ 州对非清洁能源的需求增长量最高，应增加清洁能源的使用以减少环境污染。结合 \mathcal{W} 可知，AZ 州在非清洁能源的使用中，煤、石油、天然气的使用量明显高于其他州，且主要以消耗天然气为主。

而对于具体能源的描述，因电力为新时代的主要能源，本文仅以清洁能源的电能 (包括风电、水电、太阳能发电求和) 为例进行描述分析，其余结果见附录。

由图 5 所示，电力需求随时间变化而逐渐增多。TX 州和 CA 州对于电力

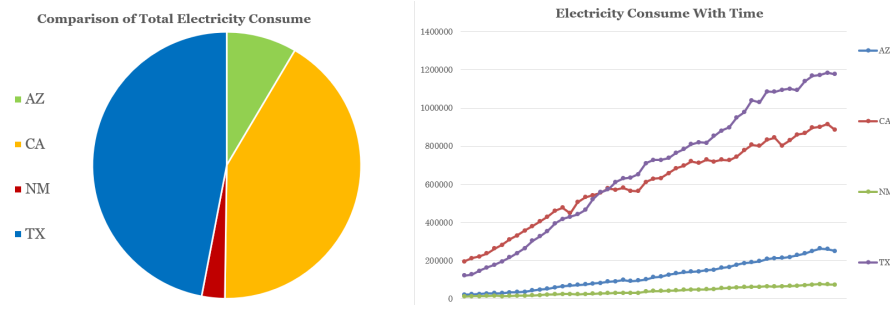


Figure 5: 电力能源使用对比图

能源消耗多于 AZ 州和 NM 州，且结合 \mathcal{Y} 可知 TX 州主要以风力发电来满足电能的需求；而 AZ 州的风力发电较为薄弱，但以太阳能发电来满足电能的需求。同样根据公式 2 计算电力的需求的增长率，得到以下结果：AZ:0.81, CA:0.82, NM:0.89, TX:0.61。结果显示 TX 州的增长率最低，NM 州增长率最高结合 \mathcal{Y} 可知以水力发电为主，而 CA 州则以核能发电为主。

4.2 Model II

根据 Model I 中确定的标签权重，计算四大州清洁能源的得分情况。根据得分情况来确定衡量标准与量化差异，最后从经济与人口的角度分析产生差异的原因。

为分析四大州能源差异和相似的原因，本文从 GDP 和人口数量的角度进行解释。首先，对于 GDP 而言，使用同公式 (4) 的方法计算各大洲的经济总量；对于人口数量而言，对人口数量进行累积求和是不合理的行为，因此选取一个区域人口最多的数量作为总人口数量。

4.2.1 Interpret Similarities And Differences

记 X 表示能源种类 x 对应的数据， t 为时间 (单位为年)，以 y 表示某一能源的得分：

$$y = \sum_{t=1960}^{2009} \eta \cdot X_t \quad (3)$$

本文从能源的生产和消耗两方面描述四大州之间的清洁能源的差异与相似点。以欧式距离量化差异的数值：

$$d(y_m, y_n) = \sqrt{\sum_{i=1}^k (y_m - y_n)^2} \quad (4)$$

式中以 y_m 和 y_n 表示归一化后不同大州在能源上的得分。

为了分析 GDP 与人口数量对清洁能源消耗造成的差异，列出清洁能源总量消耗、GDP 总量 (单位：百万美元) 与人口数量 (单位：千人)，如表 3 所示。

Table 3: 清洁能源消耗总量

States	BMTCB	EMTCB	ESTCB	GETCB	GDP	Population
AZ	439778.88	145211.63	3726334.64	1735.15	3941831	6587.653
CA	2280690.63	953603.01	10507760.24	2900691.44	33604222	36887.615
NM	149190.82	25116.71	1037200.37	7830.82	1375712	2007.315
TX	1493127.04	554510.53	16969582.3	15110.98	18967518	24770.651

分析差异：记不同能源的得分最高者为 y_{max}^i ， i 表示能源的种类。将得分最高者评价为最佳能源“用户”。在清洁能源的使用总量中，CA 州的能源评分最高，是最佳的能源使用者。而对于具体的清洁能源的最佳用户，AZ 州对地热能的使用明显较差，CA 州对地热能、生物质能、地热能的使用较为良好，NM 州对生物质能利用明显较差，TX 州对电能的使用较为广泛。

通过分析表 3 可得，表明 CA 州相对而言需要强大的燃料资源，政府部门应该加强粮食作物的种植和收购来增产乙醇能源，同时减少化石燃料等非清洁能源的使用；对于地热能而言，NM 州和 AZ 州区域内阳光照射时间短，政府部门应该减少地热能的投入，转为对需求较大的能源的投入，如电力系统等，形如：结合 NM 州电能生产与消耗权重，NM 州应加大对水电的投入，减少对核电的支出。

4.2.2 Analysis Influential Infactor

分析影响：记四大州能源总量为 G ，则 $\vec{G} = (AZ_{sum}, CA_{sum}, NM_{sum}, TX_{sum}) = (4313060.3, 16642745.32, 1219338.72, 19032330.85)$ 。同理，能够得到 GDP 总量 \vec{D} 和人口总量 \vec{P} 。

计算得到 \vec{D} 和 \vec{G} 的 pearsonr 的相关性系数为 0.84，同理得到 \vec{P} 和 \vec{G} 的 pearsonr 的相关性系数为 0.91，均表现出较强的相关性。因此得到如下的结论：能源的消耗与区域的人口数量、经济发展程度密切相关，一个区域人口数量越多，经济越发达，则消耗的能源越多。因此，当一个区域的人口数量和经济达到一定程度时，政府部门应开始着重思考投入部分经济实力用于建设能源设备，提高能源利用率和生产率，甚至以购买的形式进口能源；此外还应该大力发展清洁能源，减少能源大量消耗带来环境污染等不良后果。

4.3 Model III

为了预测 2025-2050 年的能源使用情况并做出规划目标，本文从各大洲能源使用和消耗历史情况与各大洲之间的差异两个角度出发，建立相应模型并进行求解。

4.3.1 Producing and Planning Goals

对于历史情况，设置预测步长为 s ，即使用历史 s 年的数据预测下一年的数据，而预测方的法为使用神经网络进行回归，记预测每年的得分为 T_1 。

除此之外，还需要考虑各州之间存在的差异 \vec{d} ， \vec{d} 的各个分量为各个能源差异，则能源总消耗为 $|\vec{d}|$ 。

- 考虑到现实生活中从投入能源生产、测试、运维需要较长时间，设置差异值 $|\vec{d}|$ 每 e 年更新一次 ($e = 10$)。
- 考虑到各大州对能源的使用情况不同，NM, AZ 州和 TX, CA 州的差异较大，而考虑较大的差异也显得不切实际。因此设 TX 州和 CA 州的差异值为 $|\vec{d}_1|$ ，设 NM 州和 AZ 州的差异值为 $|\vec{d}_2|$ 。

根据能源的历史消耗情况，假设能源消耗随时间变化服从指数分布；而能源的使用在人口、地理、气候、政治等因素影响下存在演变过程，设演变因子为 α ，因此设演变过程中的差异产生的影响为 T_2 ：

$$T_2 = \alpha \cdot |\vec{d}| \cdot \sqrt[3]{t} \quad (5)$$

综合考虑自身的历史演变过程和大州之间的差异，得到最终的调节方程，而求解调节方程在每年的数值即为大州的最终规划目标。

记能源使用总目标 T ，而能源的需求会随人口变化和经济发展而逐渐增长并呈现不规则的增长，因此计算出总目标 T 的取值范围，当某年的能源使用情况取值属于此区间时，认为该能源的使用情况达到了目标。规定 T 的取值范围是： $T \in [T_1, T_1 + T_2]$ 。

得到的结果如图 6 所示，因为对目标的规划过程中采取了弹性措施，目标值落在阴影部分内即认为达到了使用目标。图中显示出 TX 州具有较大的提升空间，和 4.1.2 小节得到的结论相同。

本文将重点分析清洁能源的使用，求解未来 25 年对清洁能源的消耗目标。而未来间会以消耗清洁能源为主，因此适当的对演变因子 α 进行增大以扩大对清洁能源的使用，结果如图 7 所示。

在图 7 中，能够看出 NM 州需要对清洁能源的使用具有较大的上升空间，结论仍与 4.1.2 小节相同。结合 NM 州能源画像，应加大对水力发电的投入，也可以考虑对可能发电等新技术的研发，以增加清洁能源的使用量缓解能源

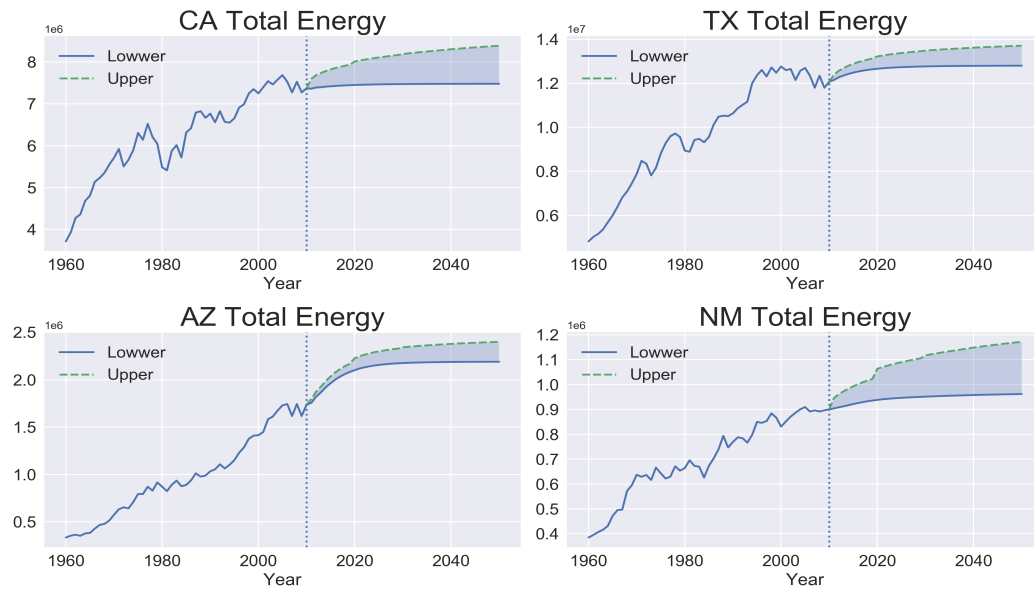


Figure 6: 未来 25 年能源消耗目标

压力。而对于其他州而言，也应保持本州的清洁能源的优势，填充不足，为日益增长的能源需求提供充足的支持，同时减少非清洁能源的使用。

5 Model Extension and Sensitivity Analysis

首先分析模型对时间、演变参数的变换后的稳定性进行了分析，在得到模型有足够的稳定性后对模型进行推广与扩展。

5.1 Sensitivity

为了检验模型的正确性，我们对模型进行灵敏性分析。未来能源消耗的目标规划对一个州的发展至关重要。因此重点对大州未来能源的发展进行灵敏性分析。

首先，我们对州未来发展的能源约束中使用的差异更新步长 e 进行检验。我们设置 e 分别取值为 5,10,15，计算 TX 州的能源总消耗量，以分析时间因素对能源消耗总量的影响。

如图 8 所示，随着步长 e 的变化，TX 州在 2025-2050 年消耗总能源的数量并没有明显的波动。结果表明了能源使用的差异并非在十年左右的短时间内可以弥补，因此对于大州的发展计划而言，应该逐步发展基础设施和有利资源，而不应该突然的对能源投入大量成本，否则只会带来经济的负面影响，实现短时间内的超越是不切实际的行为。

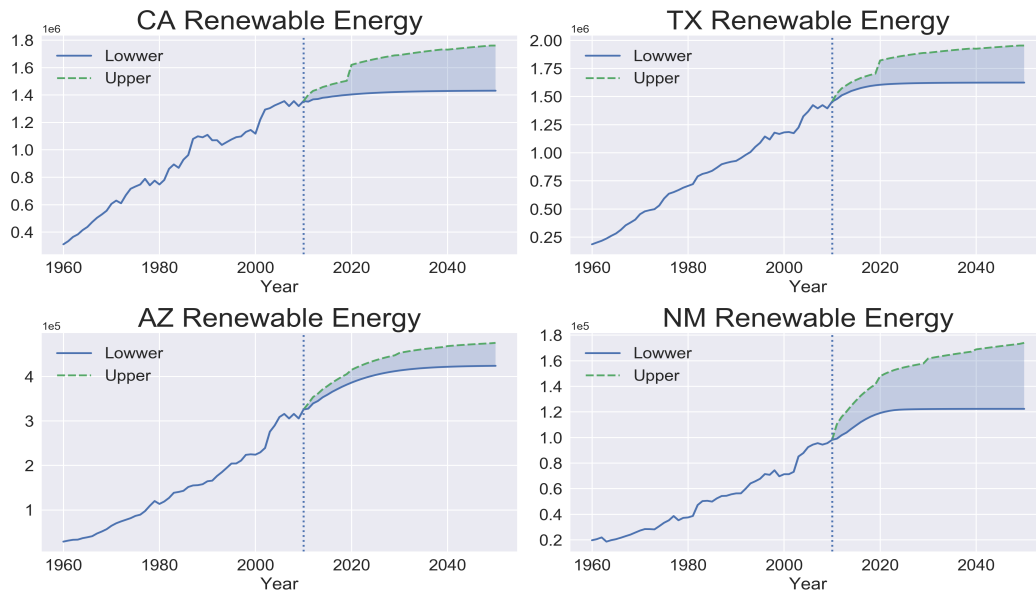


Figure 7: 未来 25 年清洁能源消耗目标

其次，还应该考虑演变因子的影响，在演变因子 α 取值不同的情况下对 TX 州 2025-2050 年能源总消耗量进行预测，得到的结果如图 9 所示。

在图 9 中，我们发现随 α 的逐步增大，未来能源的消耗量也在逐步增大，因演变因子 α 是综合考虑地理、气候、人口、经济等各种指标的综合因子，所以 α 的取值还应考虑实际情况进行确定。总而言之，从图中能够分析出 α 从正面影响着能源使用量，当确定出演变因子 α 后，便可以确定出 2025-2050 年间的能源使用的目标情况，证明了我们模型的稳定性。

5.2 Model Extension

模型能够扩展到需要构建画像，描述结构分析异同的背景，如工业画像、经济画像和气候画像等情况。

以气候画像为例，根据一个区域的气候数据，我们能够对气候构建多级画像并计算得分，对区域的气候进行详细可量化的描述。根据得分情况分析区域之间的异同，便能够对政府部门提出合理化的建议。与此同时，还能够考虑干预政策、差异情况、历史情况对区域未来年间的气候情况进行预测，而预测数值可以代表政府部门干预气候的规划目标。

6 Strengths and Weaknesses

Based on the modeling process, we make some comments on our model as listed below.

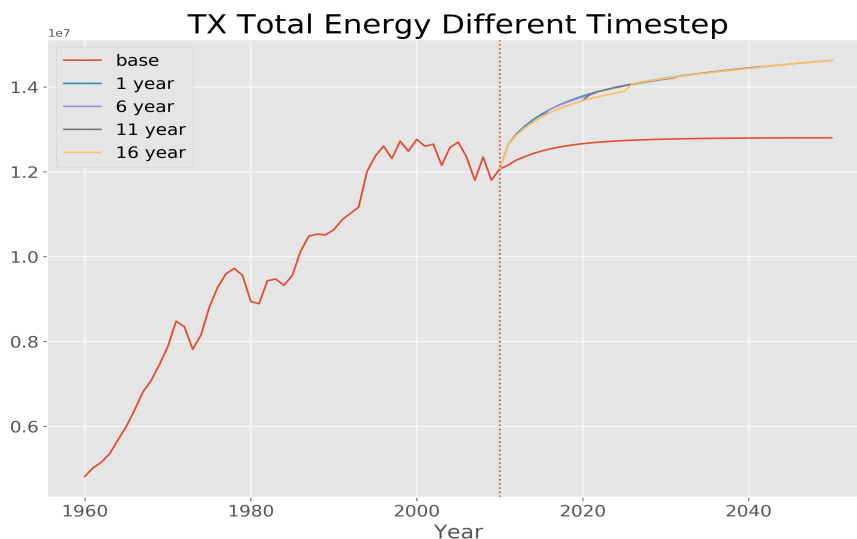


Figure 8: 对 e 的灵敏性检验

6.1 Strengths

- 构建了较为全面的能源，基于画像利用数据对画像对大州进行描述，量化的结果更容易分析异同点。
- 简洁明了的分析了能源使用量与人口、经济的相关性，以此对政府部门提出合理化建议。
- 设置弹性区间规划了未来时间内能源的使用情况，目标具有弹性限度更容易调整。
- 经过灵敏性分析，证明了我们的模型足够稳定。

6.2 Weaknesses

- 在数据处理的过程中，删除了较多的冗余数据和缺失数据，这样处理会使得模型精度降低。如果想要改进模型，应该分析冗余数据之间的关联性。
- 应该搜寻更多的数据，如四大州 2025-2050 年中的能源基础设施情况、蕴含能源情况的数据，来提升画像的全面性与准确性。

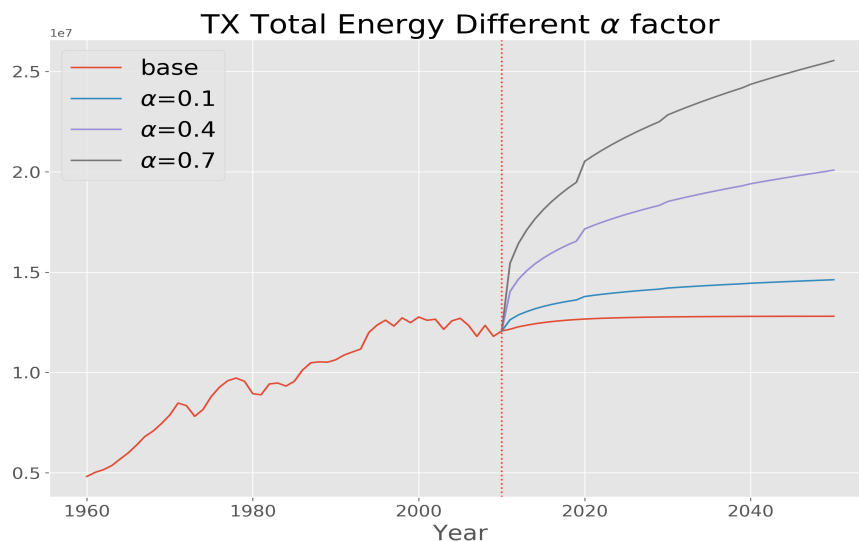


Figure 9: 对差异步长 e 的灵敏性检验

7 Conclusion

我们建立了能源画像模型，使用不同的方法对标签确定权重，量化描述的各大州的能源分布和使用情况。并且计算了能源使用和人口、经济的相关性系数，得到了较强的相关性。之后对未来时间能源的使用情况进行了合理的预测，设置弹性区间保证结果的合理性，对应调整政策的可调度性。最终对模型的稳定性进行了检验，分别从优缺点两个方面讨论了我们的模型。

References

- [1] I. Dincer, “Renewable energy and sustainable development: a crucial review,” *Renewable and Sustainable Energy Reviews*, vol. 4, no. 2, pp. 157 – 175, 2000.
- [2] “State energy data system 2016 production technical notes,” https://www.eia.gov/state/seds/sep_prod/Prod_technotes.pdf.
- [3] “United states - seds - u.s. energy information administration (eia),” <https://www.eia.gov/state/seds/seds-technical-notes-complete.php>.
- [4] M. S. Guney and Y. Tepe, “Classification and assessment of energy storage systems,” *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 1187 – 1197, 2017.

Appendices

Here is Code we used in our model, which python is the main development language, relying on the third-party library numpy.

Appendices A: Dynamic Program to solve Problem 1

```
import numpy as np
import pandas as pd
from collections import Counter

def step1(df):
    keep_typeOfData_list = list('PDRX')
    df_rest = [df.Year]
    for i in range(1, len(df.columns)):
        column = df.iloc[:, i]
        if column.name[-1] in keep_typeOfData_list:
            df_rest.append(column)
```

```

df_rest = pd.concat(df_rest, axis=1)
return df_rest

def step2(df, threshold=.7):
    df_rest = [df.Year]
    for i in range(1, len(df.columns)):
        column = df.iloc[:, i]
        if sum(column == 0) < threshold: # 閾值判断
            df_rest.append(column)
    df_rest = pd.concat(df_rest, axis=1)
    return df_rest

def step3(df, df_raw):
    renewable = ['BM', 'EL', 'EM', 'EN', 'ES', 'GE', 'HY', 'NU',
                 'RE', 'RO', 'SO', 'WD', 'WS', 'WW', 'WY']

    df_add = []
    for i in range(1, len(df_raw.columns)):
        column = df_raw.iloc[:, i]
        name = column.name
        if name[:2] in renewable and name not in df.columns:
            df_add.append(column)
    df_add = pd.concat(df_add, axis=1)
    df = pd.concat([df, df_add], axis=1)
    return df

if __name__ == "__main__":
    df_AZ = pd.read_excel('./data/data_set.xlsx',
                          sheet_name='AZ')
    df_CA = pd.read_excel('./data/data_set.xlsx',
                          sheet_name='CA')
    df_NM = pd.read_excel('./data/data_set.xlsx',
                          sheet_name='NM')
    df_TX = pd.read_excel('./data/data_set.xlsx',
                          sheet_name='TX')
    df_rest_AZ = step3(step2(step1(df_AZ)), df_AZ)

```

```
df_rest_CA = step3(step2(step1(df_CA)), df_CA)
df_rest_NM = step3(step2(step1(df_NM)), df_NM)
df_rest_TX = step3(step2(step1(df_TX)), df_TX)
```

Appendices B: Maximum Network flow to solve Problem 2
