

## 题目

第一部分：

第二部分：

第三部分：

## 论文大概

## 解题框架

## 假设

1A

1B

1C

1D

2A

605个特征中提取出20个

# 题目

2018 MCM C 题：生产能源

背景：能源生产和使用是任何经济的重要组成部分。在美国，能源政策的许多方面都分散下放到州的水平。此外，不同州的地理区域和工业影响能源的使用和生产。在1970年，12个在美国西方的州形成西部州际能源协议（WIEC），其任务重点是促进这些州之间的合作，促进发展和核能技术管理。州际契约是一种在两个或多个州之间的契约安排。在这些州就具体的政策问题达成一致意见，就某一特定的区域或州事项采取一套标准或相互合作。问题：在美国与墨西哥的边境，有四个州—美国加利福尼亚州（CA），亚利桑那州（AZ），美国新墨西哥州（NM）和得克萨斯州（TX），这是希望到形成一个关注增加使用清洁、可再生能源的来源的实际的新能源契约。你的团队需要帮助这四个州的州长对这些数据进行分析 and 建模，然后通知他们制定一系列发展他们的州际能源契约的目标。该附件数据文件“ProblemCData.xlsx”提供的第一工作簿（“seseds”）是50年中四个州的能源生产和消费数据的605个变量，伴随一些人口和经济信息。在这个数据集中，605个变量的名称被定义在第二工作簿（“msncodes”）中。

## 第一部分：

- A. 使用所提供的数据，为这四个州分别创建一个能量概况。
- B. 建立一个模型来描述从1960到2009年这四个州的能量是如何演化的。分析和解释模型的结果，以解决四个州使用清洁和可再生能源的问题，用州长们很容易理解的方式，帮助他们了解四个州之间的异同。你的讨论包括可能影响异同的因素（例如地理，工业、人口和气候）。
- C. 确定在2009年这四个州中哪一个州有“最佳”使用清洁和可再生能源的概况。解释你的标准和选择。
- D. 基于这些州使用能源的历史演变，以及你对所建立的州概况之间差异的理解，如你所定义的那样，在每个州长办公室没有任何政策改变的情况下，分别预测2025和2050年每个州的能源概况，。

## 第二部分：

- A. 基于你对这四个州的比较，用你的“最佳”概况的标准和你的预测，来确定2025和2050年的可再生能源使用的目标，并将它们作为新的四州能量契约的目标。
- B. 确定并讨论四个州可能采取的至少三项行动，以实现他们能源紧缩契约目标。

## 第三部分：

准备一份一页总结 2009 年州概况的备忘录给州长们，包括在没有任何政策 改变时你对能源使用的预测，以及你建议的能源契约采纳的目标。

# 论文大概

建立CAFE，这是一种用于状态能量描述（EP，多方面构造）的表征，分析，预测和评估的新颖框架。

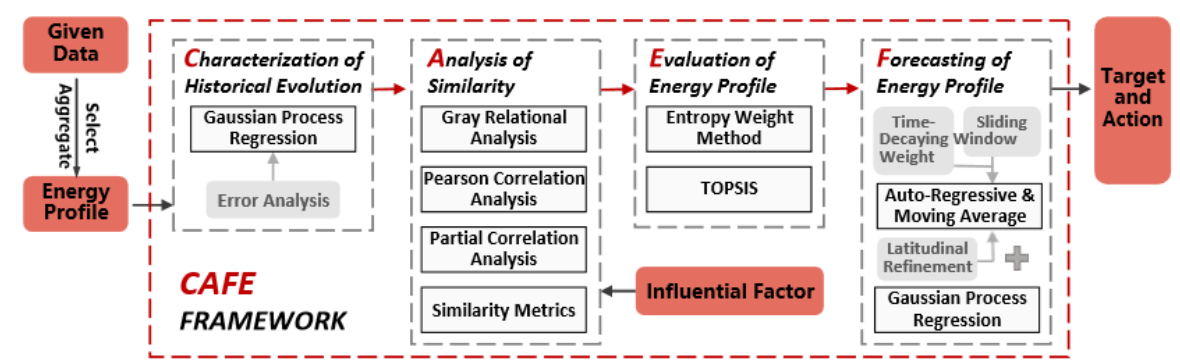
[A、B题]我们利用**高斯过程回归（GPR）模型**来描述1960年至2009年每个州不同EP时间序列的基本演变趋势，强烈波动和随机噪声水平。

[C题]我们采用灰色关联分析和Kendall秩来衡量各EP之间的相似性分别在价值和趋势上确定状态，并使用Pearson相关系数和偏关系系数来揭示相似性的外部影响因素。我们考虑了7个有关可再生能源的标准，通过熵权法赋予它们不同的重要性，将它们整合到基于TOPSIS方法的最终评分中，并最终判断出拥有“最佳”配置文件的CA。

[D、2A题]开拓性地为2025年和2050年的EP预测制定了ARMA-GPR混合模型，该模型自适应地适合于短期和长期预测。我们还提出了一种滑动窗口机制，并结合了更多的信息预测功能，采用了“超前”优化方法。然后，我们确定2025年和2050年可再生能源使用的目标和方法。我们发现了可再生能源消费（RC）初始化的优势，并因此对目标设定问题进行了多目标优化。

# 解题框架

自己定义一种CAFE模型，框架如下：



- 表征：我们为每个状态创建一个能量分布图，将其整合为20个典型项目，并建立高斯过程回归（GPR）模型以显示能量分布图的演变。
- 分析：利用相关分析来确定各州之间能源分布的相似性，其潜在原因包括地理，气候，工业等。
- 评估：我们根据**理想的标准**，采用与理想解决方案相似的偏好顺序技术（TOPSIS）来形成综合标准。
- 预测：我们新颖地提出了一种ARMA-GPR模型，该模型自适应地适合短期和长期能量分布预测。该模型不仅通过分配随时间变化的权重集成了GPR和ARMA，而且还融合了各州之间的相似性。
- 目标与行动：我们利用多目标程序设计来确定2025年和2050年四个州的可再生能源消费的不均衡目标，并根据我们的预测提供刺激亚利桑那州，新墨西哥州和德克萨斯州使用可再生能源的行动。

# 假设

辟除天灾人祸

# 1A

我们将能源从宏观上分为可再生能源（电力，乙醇燃料，核燃料，木材等）和不可再生能源（石油/汽油/石油，煤炭，天然气）。我们提取了反映能源不同方面的全球信息，主要是**进口，出口，生产，消费和支出的内部信息**。基于内部信息添加人均消费、人均支出、可再生能源熵等**概念和指标**并列出关系式，参与对总体能源概况进行评估。

分类：根据能源是否可再生、供应、使用分类，各个类别中又对指标进行分类熵，体积，比率和人均容积。

## 1B

模型：模型高斯过程回归（GPR）模型（稳定、适应噪音）拟合时间序列

对于数据集  $D: (X, Y)$ , 令  $f(x^i) = y_i$ , 从而得到向量  $f = [f(x_1), f(x_2), \dots, f(x_n)]$ , 将所需要预测的  $x_*$  的集合定义为  $X_*$ , 对应的预测值为  $f^*$ , 根据贝叶斯公式有:

$$p(f^* | f) = \frac{p(f | f^*)p(f^*)}{p(f)} = \frac{p(f, f^*)}{p(f)}$$

高斯回归首先要计算数据集中样本之间的联合概率分布,  $f \sim N(\mu, K)$ ,  $\mu$  为  $f(x_1), f(x_2), \dots, f(x_n)$  的均值所组成的向量,  $K$  为其协方差矩阵, 再根据需要预测的  $f^*$  的先验概率分布  $f^* \sim N(\mu^*, K^*)$  与  $f \sim N(\mu, K)$ , 来计算出  $f^*$  的后验概率分布。

其中共有两个核心问题: (1) 如何计算和方差矩阵 (2) 如何具体如何计算  $f^*$  的概率分布。

$$y(x) = f(x) + s(x)$$

高斯回归首先要计算数据集中样本之间的联合概率分布 (包含), 再根据需要预测的的先验概率分布, 再计算出的后验概率分布。

其中  $f$  是数据中的真实值, 协方差矩阵由高斯函数计算, 计算过程中引入数据的随机噪音变量, 再一次计算概率分布。

在可视化时, 该拟合效果中包含有**置信空间**, 使结果更合理。

将演变量化 = 各个州各个指标的预测 (用体趋势、噪音、波动)

## 1C

分层次:

- 1、纵向分析: 各个州的各个指标
  - 2、横向分析 (1个州): 分析能量演化的相似性, 灰色关联分析 (GRA)
- 【3】计算灰色关联系数

(3) 计算灰色关联系数:

$$\xi_i(k) = \frac{\min_t |x_0(t) - x_i(t)| + \rho \max_t |x_0(t) - x_i(t)|}{|x_0(k) - x_i(k)| + \rho \max_t |x_0(t) - x_i(t)|}$$

为比较数列  $x_i$  对参考数列  $x_0$  在第  $k$  个指标上的关联系数, 其中  $\rho \in [0, 1]$  为分辨系数。

其中, 称  $\min_t |x_0(t) - x_i(t)|$ 、 $\max_t |x_0(t) - x_i(t)|$  分别为两级最小差及两级最大差。

一般来讲, 分辨系数  $\rho$  越大, 分辨率越大;  $\rho$  越小, 分辨率越小。<http://www.cnblogs.com/sk18192449347>

这是一个比较复杂的公式, 给出的代码可以直接运行出来, 可以先不管这个公式。

关联系数->加权关联度->评价分析

考虑到地理, 工业、人口和气候等因素, 皮尔逊相关系数 (PCC) 和部分关系系数 (PRC) 研究相关性。

新“最佳”评估标准（对社会影响而言）：新标准包括：RCPC（人均可再生消费量），IRRC（可再生消费增加率）和REGR（可再生支出占GDP的比率）。（熵权重方法分配权重）

使用了TOPSIS2方法，该方法的核心是通过最小化所有条件均达到最佳状态的“虚拟最优”距离来直接识别最佳，最终计算各个州的得分。

## 1D

---

模型：ARMA（短时间段预测）和前述的高斯过程回归模型对比

数据:上题目中的指标和时间

预测之余还有敏感性分析和模型的优化（调参）

## 2A

---

模型：多目标规划，整合多个目标的评估，设计一个目标函数。

关联到上述指标的计算值，计算出各个指标是否偏斜等情况，分析其在各个州的可行性指定目标方程

## 605个特征中提取出20个

---