

# 基于 LSTM 神经网络和 Prophet 的电力负荷组合预测模型

## 摘要

不同时期的电力系统负荷预测对提高电力系统的经济和社会效益具有很强的现实意义，而电力负荷受到众多复杂因素的影响，负荷预测变得困难。本文针对不同时段电力负荷时间序列的波动规律，运用基于 LSTM 神经网络和 Prophet 时间序列预测模型的结合方法，发挥其各自优势对电力负荷进行预测。该模型的提出对提高电力负荷的预测精度，实现精准用电调配，具有一定的参考价值。

针对问题一：因为 LSTM 模型善于记忆时间跨度长的有用信息，Prophet 模型对节假日、季节性、周期性等人为因素预测精度较高，故采用两个方法进行融合。观察到日负荷波动规律相似，周平均值的变化相似，故分拆任务，利用 LSTM 模型和 Prophet 模型分别对该地区未来 10 天日负荷的间隔 15 分钟的波动情况和每日平均值分别进行预测，再将结果加和得到最终的预测结果。利用 2021 年 8 月 15 日前的数据作为训练集，此日后的数据为模型测试集，预测精度为平均绝对百分比误差 1.2%。考虑到节假日和季节性的影响较大，本文沿用 Prophet 模型进行中期电力负荷最大值最小值进行预测分析，检验结果为平均绝对百分比误差分别为 2%与 5%，预测较准确。

针对问题二：由于影响不同的突变来源不同，我们以突变影响时长为标准，将突变点分为三类，针对不同类别的突变点采取人工检验、MK-突变点检验、滑动 T 检验等方法检测突变点、量级，并挖掘原因，最终分析出了大规模全区域停电、工业园区电网短时间故障、全区定时电网检修、节假日影响和气象突变等原因。同样鉴于 Prophet 模型对具有节假日和季节性效应的中期电力负荷预测效果较优，我们沿用该模型预测各行业未来 3 个月日负荷的最值情况，最终计算出的平均绝对误差均在 0.1 以下，预测效果较好。

最后我们综合利用上述对各行业的用电负荷分析，分析了在国家“双碳”目标下，各行业未来用电负荷的趋势，并根据已有的负荷数据结论，针对工业和商业分别提出了四条建议。

**关键词：**电力负荷预测 长短时记忆神经网络 Prophet 模型 突变点检测

# 目录

|   |    |
|---|----|
| 一、 问题重述.....  | 1  |
| 1.1 问题背景.....   | 1  |
| 1.2 问题重述.....   | 1  |
| 二、 问题分析.....  | 1  |
| 2.1 问题一(1)的分析.....  | 1  |
| 2.2 问题一(2)的分析.....  | 2  |
| 2.3 问题二(1)的分析.....  | 2  |
| 2.4 问题二(2)的分析.....  | 2  |
| 2.5 问题二(3)的分析.....  | 3  |
| 三、 模型假设.....  | 3  |
| 四、 符号说明.....  | 3  |
| 五、 模型的建立与求解.....  | 5  |
| 5.1 问题一（1）的模型建立与求解.....   | 5  |
| 5.1.1 LSTM 模型的建立.....   | 5  |
| 5.1.2 LSTM 模型预测每日波动值.....   | 7  |
| 5.1.3 Prophet 模型的建立.....  | 9  |
| 5.1.4 Prophet 模型预测每日平均值.....  | 10 |
| 5.1.5 预测精度的评估分析.....  | 12 |
| 5.2 问题一（2）的模型建立与求解.....   | 13 |
| 5.2.1 模型建立.....   | 13 |
| 5.2.2 模型求解最大最小值.....  | 13 |
| 5.2.3 模型求解最值所在时刻.....   | 14 |
| 5.2.4 预测精度的评估分析.....  | 14 |
| 5.3 问题二（1）的模型建立与求解.....   | 14 |
| 5.3.1 模型建立.....   | 14 |
| 5.3.2 模型求解.....   | 17 |
| 5.4 问题二(2)的模型建立与求解.....   | 22 |
| 5.5 问题二(3)分析与建议.....  | 22 |
| 六、 模型的评价与推广.....  | 24 |
| 6.1 模型的评价.....  | 24 |
| 6.1.1 模型的优点.....  | 24 |
| 6.1.2 模型的缺点.....  | 24 |
| 6.2 模型的改进.....  | 24 |
| 七、 参考文献.....  | 25 |
| [1] 葛娜,孙连英,石晓达,赵平.Prophet-LSTM 组合模型的销售量预测研究                                   |    |
| [J].计算机科学,2019,46(S1):446-451. ....   | 25 |
| [2]Taylor SJ, Letham B. 2017. Forecasting at scale. PeerJ Preprints 5:e3190v2 |    |
| .....   | 25 |

## 一、 问题重述

### 1.1 问题背景

电力负荷预测是电力系统规划、计划、调度等部门的基础工作，提高电力负荷的预测水平对于电网运营和社会电力保障十分重要。根据预测的时间范围不同，负荷预测大致可分为超短期、短期、中长期和长期预测。用电负荷常受到历史习惯、经济状况、气象条件、地区限制和社会事件等因素的影响，而各种因素自身的复杂性及其相互影响使电力负荷预测问题变得更加复杂，因此如何建立科学精确的负荷预测模型亟待进一步研究。

### 1.2 问题重述

本题一共给出了三个附件数据，分别为某地电力的间隔 15min 的负荷数据、该地各行业的日负荷数据和该地气象数据，本文需要据此建立数学模型解决以下问题：

- (1)地区负荷间隔 15min 的短期负荷预测与评估。
- (2)地区三个月日负荷最值及最值出现时刻的中期负荷预测与评估。
- (3)各行业负荷的突变点检测及量级和原因的分析。
- (4)各行业日负荷最值的中期负荷预测与评估。
- (5)分析“双碳”目标对各行业的用电负荷影响并提出针对性建议。

## 二、 问题分析

### 2.1 问题一(1)的分析

该任务是预测某地区电网未来十天的间隔 15 分钟的负荷数据，鉴于此问题为超短期负荷预测分析，需要选择对短期预测具有较高准确度的预测模型。由于已有的数据量较大，时间跨度较长，而负荷具有“近小远大”的原则，即与待预测日越接近的历史值对预测精度的影响越大，因此**选择的模型需要较好地处理记忆与遗忘的问题**。

日负荷数据具有很强的周期性，具体体现在：不同日的日负荷曲线其整体规律相似；同一星期类型日负荷规律相似；工作日和休息日负荷各自相似；不同年度法定节假日的规律相似。

考虑到 LSTM 模型对短期预测精准度较高，且对时间跨度长的数据能长期记忆，最大效果地保存历史有用信息的优点，本文选取**长短期记忆网络 LSTM 模型**来预测未来 10 天负荷数据去日均值后得到的波动情况；鉴于上述日负荷数据周期性强的特点，本文选

择由趋势项、季节项和节假日项组合而成的 Prophet 模型来预测未来 10 天的电力负荷平均值，它能很好地处理节假日因素和季节性趋势对时间序列的影响。将上述两种模型的预测结果相加即可得到对未来 10 天间隔 15 分钟的负荷预测。

## 2.2 问题一(2)的分析

该任务是预测该地电网未来三个月日负荷的最大最小值及相应的时刻。本文将最大值与对应的时刻这两个问题分别处理。由于日负荷数据具有很强的日周期性、周周期性、月周期性、年周期性，且受到节假日的影响较为明显，本文依然采用 Prophet 模型进行建模预测。鉴于该预测为中期负荷预测，气象的预测精度得不到保证，增加了模型的复杂程度，因此不考虑温度、风力等气象参数。

## 2.3 问题二(1)的分析

题目要求挖掘各行业日负荷突变点的时间，量级，并分析可能原因。根据对绘制出的原始数据图像的直观分析，本文考虑将突变点按突变的影响时长分为三类突变点，针对不同类型的突变点采取不同的检测方法，并结合天气、节假日、停电等因素的影响进行突变原因分析。

第一类：影响单个观察值的离群突变点。突变的主要原因可能是大规模停电、极端天气短时间影响。由于该突变原因概率极小且易于观察，可以考虑直接绘图找到显著离群点，对比该点对应的日期、天气等相关信息进行原因分析。

第二类：影响一段时间观察值的离群值。造成的突变原因可能是疫情封城、天气或温度不稳定、夏日空调负荷加重、疫情解封恢复生产。对于第二类突变点，若突变值较小，则可以在第一类突变点检测时一并筛出；若较显著且可在较短期内体现，可以对数据进行一定的平滑处理和压缩处理后，使用不同的突变点检测算法，再人工检测结果，综合分析后得出。

第三类：突变在局部显著水平不高，并且持续时间较长。例如由于疫情影响，商业 2020 年一年内日负荷均偏低。此种突变用第二种方法不能很好地检测出来，本文采用以年为周期的同比分析，将三年的数据绘制在同一张图里进行对比。

## 2.4 问题二(2)的分析

该任务是预测各行业电网未来三个月日负荷的最大最小值，同样由于 Prophet 模型对具有周期性和易受节假日影响的电力日负荷数据具有很好的预测效果，本文依然采用该模型进行建模预测。

## 2.5 问题二(3)的分析

为了研究“双碳”目标对各行业未来用电负荷的影响，本文考虑将 2020 年 9 月国家“双碳”目标的提出作为时间分界点，对比分析该时刻前后各行业用电负荷的数据。考虑到大工业、非普工业和普通工业都属于工业的范畴，具有相似的特点，本文将其合并为工业，根据工业和商业的行业特点提出了针对性建议。

## 三、 模型假设

1. 假设该地中短期内不发生严重的社会事件，如经济危机、气象灾害、地质灾害。
2. 假设该地中期内季度的气象整体变化趋势与往年相同
3. 假设该地发电情况正常，无停电或电网故障。
4. 假设时间序列各个点随机独立，且具有相同的连续分布。

## 四、 符号说明

表 1：符号说明

| 符号            | 说明           | 单位 |
|---------------|--------------|----|
| $W_f$         | 遗忘门的权重矩阵     |    |
| $f_i$         | 遗忘门          |    |
| $i_t$         | 输入门          |    |
| $o_t$         | 输出门          |    |
| $b_f$         | 输入门的偏置项      |    |
| $c_t$         | $t$ 时刻细胞单元状态 |    |
| $\tilde{c}_t$ | 当前输入的单元状态    |    |
| $X_{std}$     | 标准化后的数据      |    |
| $X_{scaled}$  | 归一化后的数据      |    |

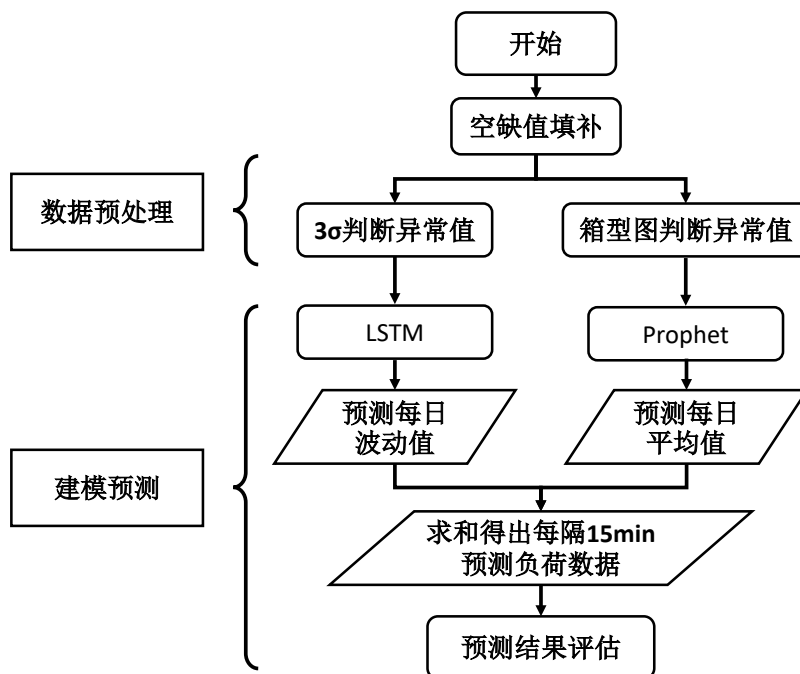
---

|               |  |
|---------------|--|
| $y_i$         | 实际值  |
| $\tilde{y}_i$ | 预测值  |
| $S_k$         | 对于具有 $n$ 个样本量的时间序列 $X_1, X_2, \dots, X_n$<br>构造出的秩序列 |
| $UF_k$        | 假设时间序列随机独立，定义的检测统计量                                  |
| $E(S_k)$      | 累积数 $S_k$ 的均值  |
| $Var(S_k)$    | 累积数 $S_k$ 的方差  |
| $UB_k$        | 时间序列 $X$ 逆序 $X_n, X_{n-1}, \dots, X_1$ 的检测统计量        |
| $t$           | 突变显著性系数的统计量  |
| $s$           | 遵从自由度 $\nu = n_1 + n_2 - 2$ 的 $t$ 分布                 |
| $K_t$         | 秩统计量中最大的值  |

---

## 五、模型的建立与求解

### 5.1 问题一（1）的模型建立与求解



#### 5.1.1 LSTM 模型的建立

长短时记忆网络(Long short term memory network, LSTM)是一种改进之后的循环神经网络 (recurrent neural network, RNN)。每次迭代在 RNN 的基础上增加了 3 个门 (gate)的逻辑控制单元：输入门、遗忘门和输出门，通过门单元的逻辑控制即时状态的更新、旧状态数据的保留程度及当前数据输出，克服了 RNN 因权重影响过大容易产生梯度消失和爆炸问题的缺点，使得网络可以更快地收敛到合适的值，从而更精确地进行短期电力负荷预测。

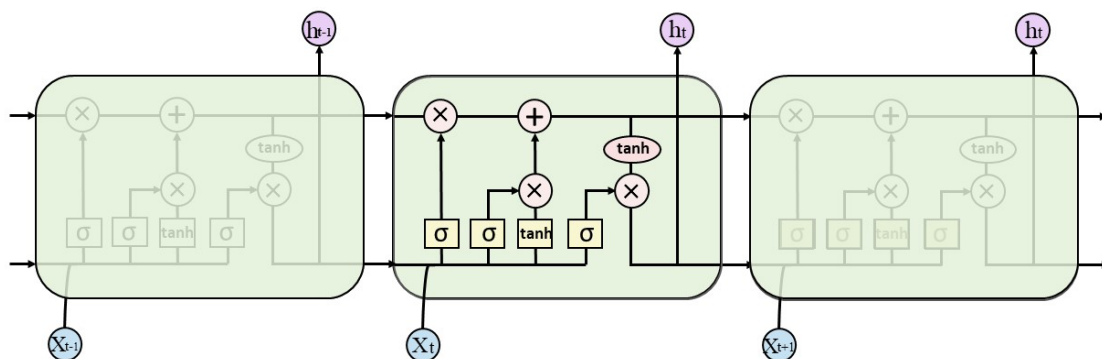


图 1. LSTM 模型示意图

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

$$\begin{aligned} \tilde{c}_t &= \tanh(W_c[h_{t-1}, x_t] + b_c) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c} \end{aligned} \quad (3)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t \odot \tanh(c_t) \quad (5)$$

式(1)为遗忘门:

$W_f$ 表示遗忘门的权重矩阵;  $[h_{t-1}, x_t]$ 表示将上一个状态向量和当前结点向量连接成一个更长的向量, 两者的点积决定旧数据和当前数据的保留程度;  $b_f$ 是输入门的偏置项,  $\sigma$ 为 sigmoid 函数。经过遗忘门后越接近 1 的向量越容易被保留, 越接近 0 的向量越容易被丢弃, 因此遗忘门决定了上一时刻的单元状态有多少被保留到了当前时刻。

式(2)为输入门:

$W_i$ 为输入门的权重矩阵;  $b_i$ 是输入门的偏置项, 输入门决定了当前时刻的输入有多少被保存到单元状态中。

式(3)表示细胞单元:

当前时刻的单元状态  $c_t$ , 是由上一次的单元状态  $c_{t-1}$  按元素乘以遗忘门  $f_t$ , 再用当前输入的单元状态  $\tilde{c}_t$  按元素乘以输入门  $i_t$ , 再将两个积加和产生。把 LSTM 关于当前的记忆  $\tilde{c}_t$  和长期的记忆  $c_t$  组合在一起, 形成了新的单元状态  $c_t$ 。通过遗忘门的控制, 它可以保存时间较为久远的信息, 通过输入门的控制, 它又可以丢弃一些当前无关紧要的内容。

式(4)表示输出门, 计算输出门的输出值, 式(5)最终得到 LSTM 输出。



### 5.1.2 LSTM 模型预测每日波动值

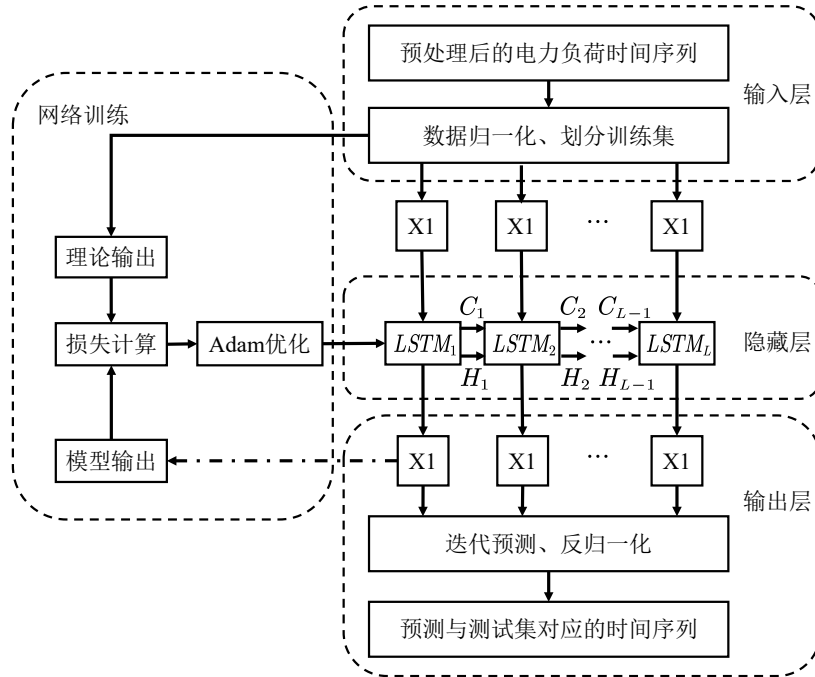
#### 数据预处理:

通过比对时间序列的长度可知,数据中存在空缺值,运用 Pandas 库中的 `resample` 方法将时间序列间隔 15min 重采样后用中位数填补空缺值,将有断点的时间序列连续化。

为了得到电力负荷数据在每一天的波动情况,本文将每天测得的负荷数据减去该天的功率的平均值,获得了在一条水平线上下波动的负荷数据。

为了防止异常值对建立预测模型的影响,本文利用正态分布的  $3\sigma$  原则,找出了 472 个负荷数据急剧波动的异常点。先使用中位数填充异常点,再用滑动平均数给其赋值,使异常点的数值与所在滑动窗口内的数值连续。

#### 模型训练:



#### (1) 数据归一化

使用 sklearn 库中的 `MinMaxScale` 方法将数据归一化,映射到(0,1)区间:

$$X_{std} = \frac{X - X.\min(axis=0)}{X.\max(axis=0) - X.\min(axis=0)} \quad (6)$$

$$X_{scaled} = X_{std} \cdot (\max - \min) + \min$$

#### (2) 分割数据集

调用 sklearn 库的 `train_test_split` 函数将原始数据分割为训练集和测试集，将 `shuffle` 参数设为 `False` 防止重洗打乱破坏时间序列的前后连续性。

### （3）定义网络

通过 keras 库中的 `Sequential` 创建顺序类的实例，通过 `add` 方法创建了 LSTM 循环层，指定时间步长数为训练集的时间序列长度，设定特征为单变量，再创建用于输出预测的完全连接层 `Dense`，鉴于 `sigmoid` 函数可以压缩数据且幅度不变，便于求导，适合前向传播，本文选取 `sigmoid` 作为激活函数。

### （4）编译网络

使用 `model.compile` 配置训练方法，首先选择损失函数，由于 MSE 的函数曲线光滑、连续、处处可导，便于使用梯度下降算法，因此本文指定损失函数为 `mse`，

$$MSE = \frac{1}{n} \sum_{i=1}^m (y_i - \tilde{y}_i)^2 \quad (7)$$

其次选择优化器，为了优化 LSTM 模型参数的更新速度，需要动态调节每个参数的学习率，本文比较了众多基于梯度的优化算法发现，自适应估计(Adaptive Moment Estimation,Adam)算法是训练神经网络的较好选择，它利用梯度的一阶和二阶矩估计动态调整参数的学习率，使得参数更新更为平稳，在实际应用中表现更优。

### （5）训练网络

使用 `model.fit` 执行训练过程，采用反向传播算法进行训练，同样由于时间序列的连续性不对数据进行打乱。通过不断调整参数，发现选取训练集样本数量 `batch_size` 为 64，定义训练总轮数 `epoch` 为 18，选定验证集比例 `validation_split` 为 0.2 的时候模型的预测效果较好。

### （6）评估网络

通过 `model.predict` 方法对测试集进行预测，选取 `R2_score` 作为评价指标，计算出决定系数（拟合优度）`R`：

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

该式得到的拟合优度为 0.9673，说明模型拟合效果很精确。

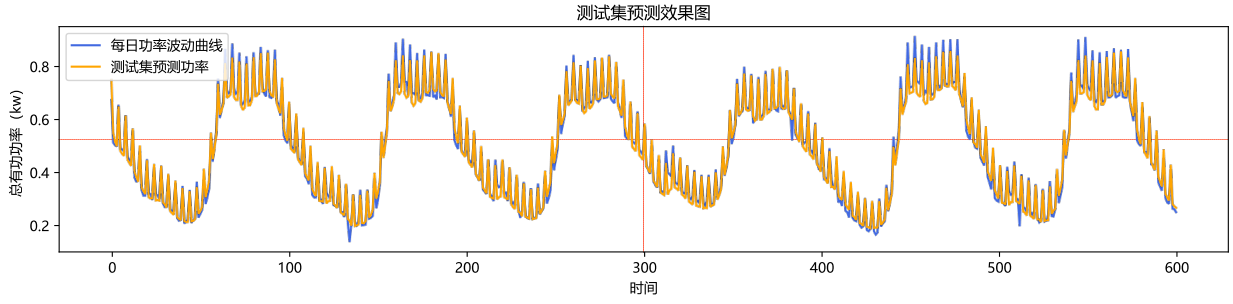
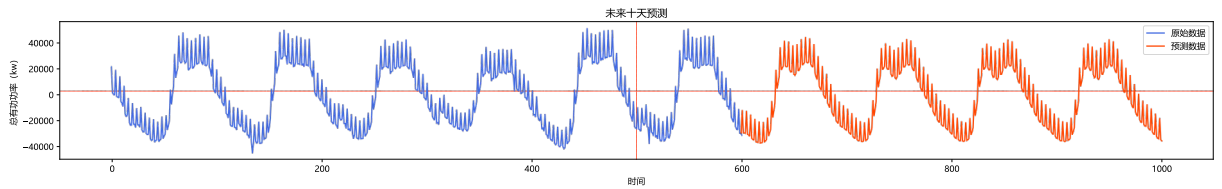


图 2. LSTM 模型预测每日负荷波动预测效果对比图

### (7) 进行预测

将训练出的模型对该地区电网未来 10 天间隔 15 分钟的负荷每日波动情况进行预测，运用 `inverse_transform` 反归一化后得到对每日负荷波动情况的预测值如下图所示，具体预测值见附件未来十天日功率波动预测值.csv。



## 5.1.3 Prophet 模型的建立

Prophet 模型是一种基于加性模型(additive model)的由趋势项(trend)，季节项(seasonality)和节假日因素(holidays)组成的可分解的时间序列模型，非线性趋势与年、周、日的规律性及节假日效应吻合，适用于具有强烈季节性影响和拥有数个季节的历史数据的时间序列。Prophet 很善于应对数据缺失和趋势变化，能很好地处理异常值，预测结果准确而快速。其对时间序列的分解如下：

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (9)$$

$g(t)$ 为趋势函数，表示非周期性变化的值， $s(t)$ 表示每周、每季度和每年的周期性变化因素， $h(t)$ 表示在可能不规律的时间表上发生的假期的影响，误差项 $\epsilon_t$ 代表模型的误差项并假设其符合正态分布，表示模型未预测到的波动。

趋势项 $g(t)$ 是整个 Prophet 模型的核心组件。Prophet 模型实现了两个趋势模型，其一为分段逻辑回归模型，它将传统的 logistic 模型的响应函数添加了一些参数后变为

$$g(t) = \frac{C(t)}{1 + e^{-(1-m(t))k(t)}} \quad (10)$$

其中 $C$ 为承载力， $k$ 为增长速率， $m$ 为偏移参数，三者都是随时间变化的函数。其二为分段线性趋势模型。

周期项 $s(t)$ 是用傅里叶级数(Fourier series)近似表达周期性分量，具体表达式如下：

$$s(t) = \sum_{n=1}^N \left( a_n \cos\left(\frac{2\pi nt}{T}\right) + b_n \sin\left(\frac{2\pi nt}{T}\right) \right) \quad (11)$$

其中 $T$ 表示某个固定的周期， $2n$ 表示期望在模型中使用周期的个数。

$h(t)$ 是节假日项，将每个节假日设为独立模型并分别设置不同的虚拟变量

$$\begin{aligned} h(t) &= \sum_{i=1}^L K_i 1(t \in D_i) \\ Z(t) &= [1(t \in D_1), \dots, 1(t \in D_L)] \\ h(t) &= Z(t)\kappa, \kappa \sim \text{Normal}(0, \nu) \end{aligned} \quad (12)$$

其中， $\kappa_i$ 表示窗口期中节假日对预测值的影响； $D_i$ 表示第 $i$ 个虚拟变量，若时间 $t$ 属于虚拟变量， $D_i$ 的值为1，否则为0； $i$ 表示节假日， $D_i$ 表示窗口期中包含的时间 $t$ 。

#### 5.1.4 Prophet 模型预测每日平均值

##### 数据预处理：

采用滑动平均为原始数据做平滑处理并且为滑动所产生首尾两边的空值进行填充。将原始数据减去滑动后的数据得到每15min 负荷的波动值，考虑到原始数据不一定是服从正态分布的，因此采取 $3\sigma$ 原则除去异常值可能带来较大误差，而箱型图判断异常值是以四分位数和四分位距为基础，具有一定的耐抗性，且能除去较多异常值，即使将其除去，Prophet 模型也能很好地处理数据有部分缺失值的问题。综上所述，本文采用箱型图除去异常值的方法，根据波动值绘制出箱型图，将内限以外位置的点表示的数据作为异常值。

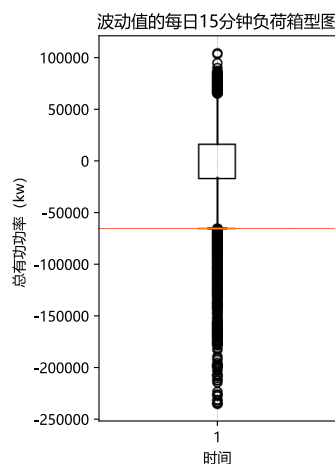


图 3. 波动值的每日 15 分钟负荷箱型图

由于疫情的影响，2018 年 9 月以前的数据整体高于疫情后的数据，这会对现在后疫情时代的电力负荷预测产生误差，因此将其舍弃。

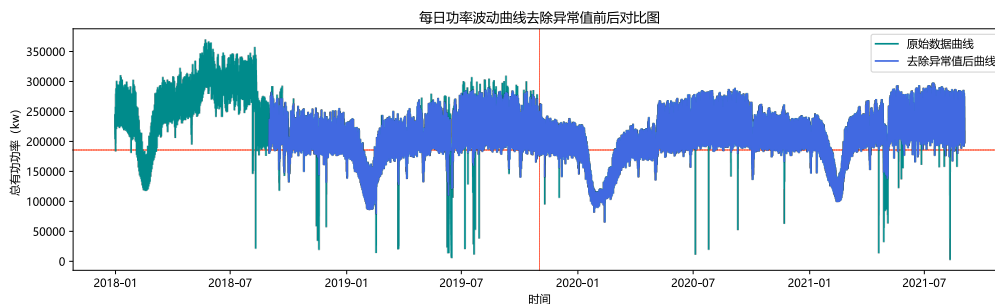


图 4. 每日功率波动曲线去除异常值前后对比图

将处理后的数据按天进行重采样得到日平均功率，如下图所示

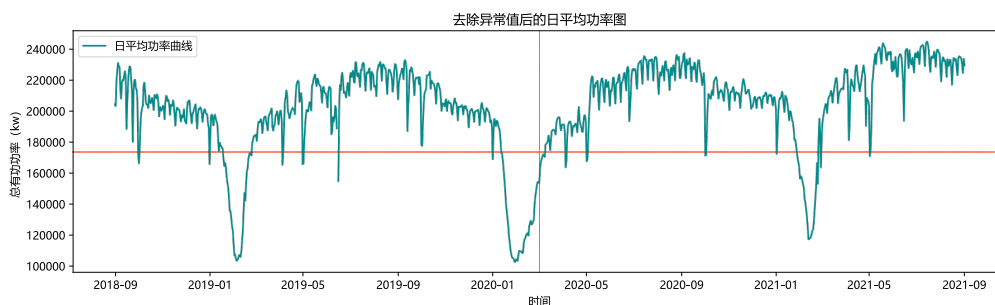


图 5. 去除异常值后的日平均功率图

## 模型训练:

首先实例化 Prophet 对象，由于电力负荷这一时间序列具有年周期性和周周期性，因此本文设置周规律拟合和日规律拟合参数，考虑到电力负荷值可能随季节周期性突变，因此本文按照经验设置季度周期性突变的灵活度为 0.2，最后设置置信区间宽度为 0.8 以检验预测效果。将中国所有的节假日添加到 Prophet 模型中，开始模型训练，预测值保存在附件 未来十天日平均功率预测值.csv 中，最后结果预测如下图所示，（浅蓝色代表

置信区间，黑点代表预测值，蓝线为原始电力负荷的时间序列。)

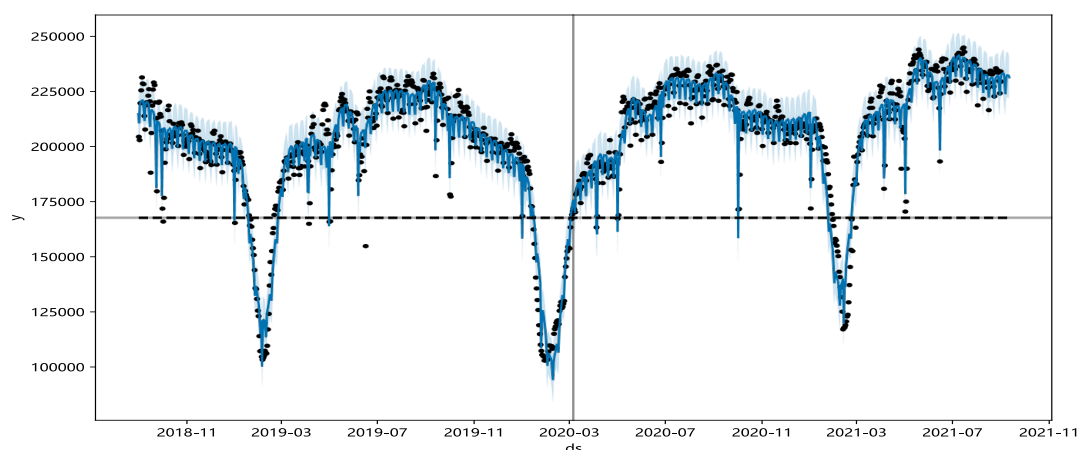


图 6. Prophet 模型对每日功率平均值的预测

### 5.1.5 预测精度的评估分析

将上述两个模型的预测结果相加，得到未来 10 天间隔 15 分钟的电力负荷预测结果，将预测结果保存到附件 未来十天功率预测值.csv 中。

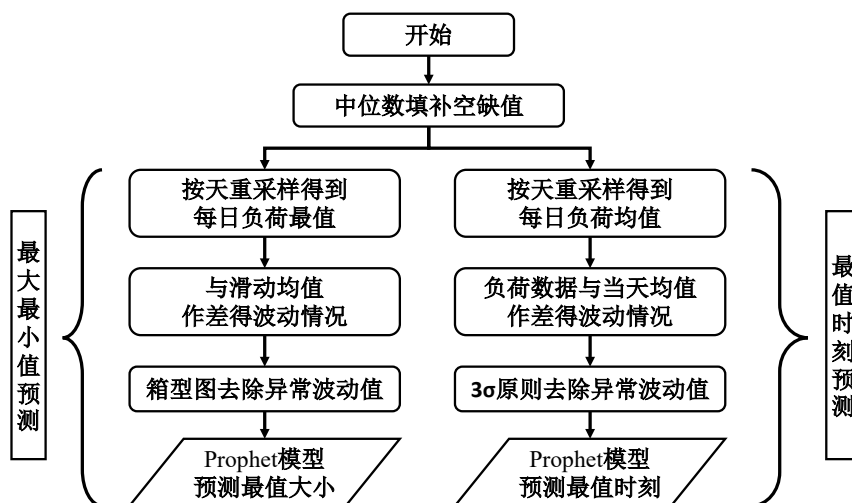
通过划分测试集，将预测的数据与真实数据比对，选取 8 月 16 号到 8 月 31 号的数据来分析预测精度，选取的评价误差指标为拟合优度(R)(公式(8))、平均绝对百分比误差(MAPE)。、平均绝对误差(MAE)。

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{actual(t) - forecast(t)}{actual(t)} \right| \times 100\% \quad (13)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |forecast(t) - actual(t)| \quad (14)$$

计算得到平均绝对误差为 2956，平均绝对百分比误差仅为 1.2%，说明预测效果非常好。

## 5.2 问题一（2）的模型建立与求解



### 5.2.1 模型建立

由于 Prophet 模型对数据中的节假日效应和趋势变化点具有出色的适应能力，因此本文仍然选择 Prophet 模型对该地区电网未来三个月日负荷的最大值和最小值及最值对应的时间进行预测，该模型的原理已在上文给出，此处不再赘述。

### 5.2.2 模型求解最大最小值

用中位数填充间隔 15 分钟的日负荷数据这一时间序列的空缺值，然后将按天重采样，取每天的最大值和最小值。将重采样得到的数据减去滑动窗口大小为 15 的滑动平均值求出波动图后，发现异常值较多，本文利用箱型图的方法将除去异常值，Prophet 模型会使用插值的方法自动填充空值。

用 Prophet 模型预测经上述操作得到的时间序列，参数与上一问的设置大致相同，为了防止过拟合的现象，本文将季度周期性突变的灵活度设置为 0.05，得到的对未来三个月日负荷的最大最小值预测结果保存在附件 *各行业未来三个月日负荷最大值最小值预测.csv* 中，Prophet 对电力负荷时间序列的趋势项、周周期性、年周期性、节假日因素的分解如下：

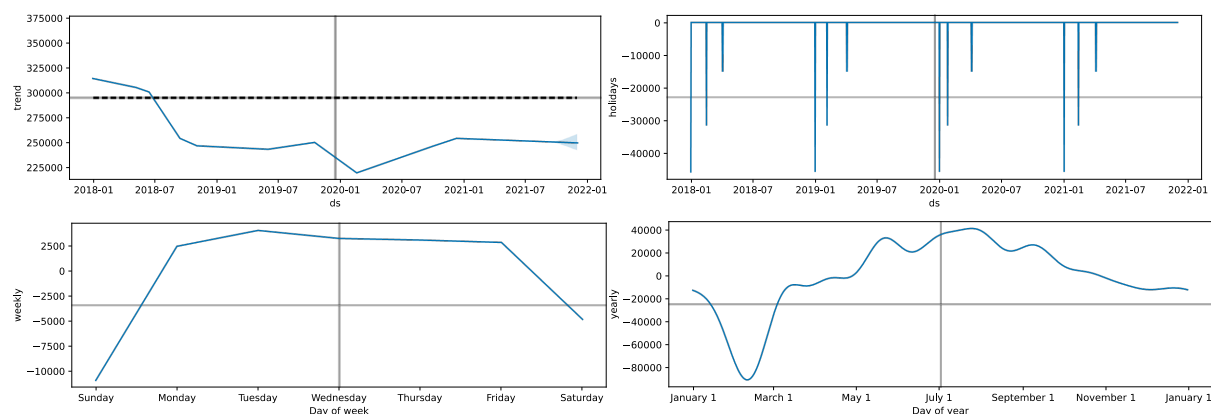


图 7：Prophet 模型对电力负荷日最大值的周期项和节假日分解

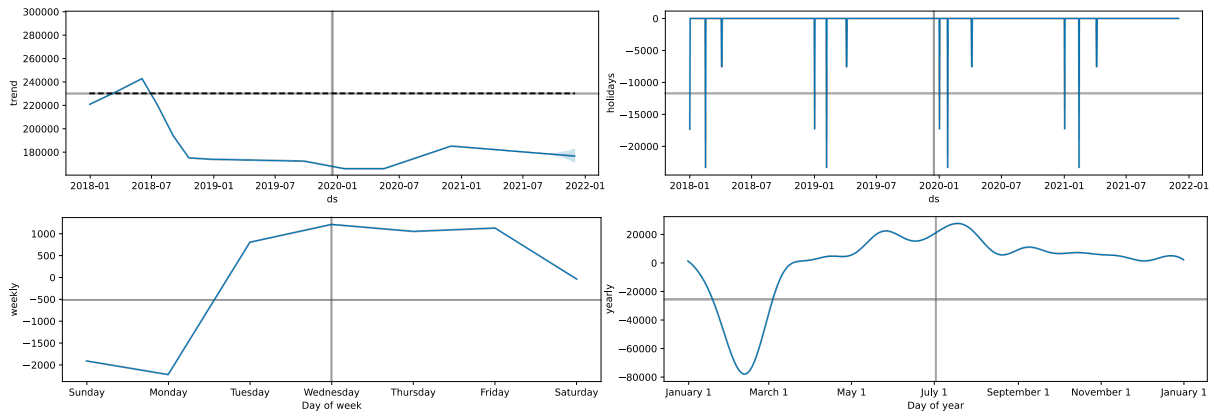


图 8: Prophet 模型对电力负荷日最小值的周期项和节假日分解

### 5.2.3 模型求解最值所在时刻

数据预处理的方法与 5.1.1 相同，考虑到数据具有天周期性，本文设置天周期性拟合参数，考虑到电力负荷与节假日因素密切相关，因此添加中国所有节假日进入模型，另外根据经验和拟合预测的结果调整其他参数，得到的预测结果保存在附件。。。中。

### 5.2.4 预测精度的评估分析

通过划分训练集和测试集，选取 2021 年 6-8 月的真实数据与预测得到的数据进行比对，计算最大值与最小值的平均绝对误差百分比(MAPE)分别为 0.02 与 0.05，说明模型预测效果较为准确。

## 5.3 问题二（1）的模型建立与求解

根据突变点按突变的影响时长分为三类突变点。

第一类：影响单个观察值的离群突变点。

例如加性离群值 (Addictive)，瞬时 (Transient) 离群值。

第二类：影响一段时间观察值的离群值。

例如水平变动 (Level shift) 离群值 (如疫情封城期间)，创新 (innovational) 离群值 (如天气、温度不稳定)，季节加性 (Seasonal addictive) 离群值 (如夏日空调负荷加重)，局部趋势 (Local trend) 离群值 (如疫情解封后缓慢上升)。

第三类：突变在局部显著水平不高，并且持续时间较长

### 5.3.1 模型建立

对于第二类突变点，本文主要采取 MK-突变点检测模型及滑动 T 检测模型，并将突变点数据结果与 Pettitt、SNHT、BuishandU 模型的结果进行比较，综合得出突变点。(设



定显著性水平为 0.05)

#### 主模型一：MK-突变点检测模型

Man-Kendall 计算简便并且可以指出突变点开始时间，适合在大量独立分布的样本内寻找突变点。

$$S_k = \sum_{i=1}^k R_i \quad (K=2, 3, \dots, n) \quad (15)$$

$$UF_k = \frac{S_k - E(S_k)}{\sqrt{Var(S_k)}} \quad (k=1, 2, \dots, n) \quad (16)$$

$$E(S_k) = n(n+1)/4 \quad (17)$$

$$Var(S_k) = n(n-1)(2n+5)/72 \quad (18)$$

(16) 式以  $S_k$  表示对于具有  $n$  个样本量的时间序列  $X_1, X_2, \dots, X_n$  构造出的秩序列，式中： $R_i$  表示  $X_i$  大于  $X_j$  ( $1 \leq j \leq i$ ) 的累积数。

(17) 式中，在时间序列随机独立的假设下，定义统计量  $UF_k$ 。其中：当  $k=1$  时， $UF_1=0$ 。 $E(S_k)$ ， $Var(S_k)$  是累积数  $S_k$  的均值和方差，

(18) (19) 式表示，在  $X_1, X_2, \dots, X_n$  相互独立，且有相同连续分布时， $UF_k$  及  $E(S_k)$  的表达式。本题假设时间序列每个观察值相互独立且有相同的连续分布。

按时间序列  $X$  逆序  $X_n, X_{n-1}, \dots, X_1$ ，再重复上述过程，同时使  $UB_k = -UF_k$ ， $k=n, n-1, \dots, 1$ ，且当  $k=1$  时， $UB_1=0$ 。分析  $UB_k$  和  $UF_k$  曲线图，如果  $UB_k$  和  $UF_k$  两条曲线出现交点，且交点在显著性 0.05 的临界线之间，那么交点对应的时刻便是负荷突变开始的时间。

#### 主模型二：滑动 T 检测模型

滑动 T 检验是通过考察两组样本平均值的差来检查是否有显著突变。基本思想是，将同一个时间序列的子列问题，转化为来自两个定长时间序列，判断均值有无显著差异的问题来检验。

本题中，不断在用电负荷时间序列中取连续两段取定长子列，如果两段子列的均值差异超过一定的显著水平 0.05 则可以认为两个子列中的数据点有突变发生。在整个时

间序列中，取定滑动窗口，取遍每一个定长子列，便可得到整个序列突变点的预测。

对于具有  $n$  个样本量的时间序列  $X$ ，设置某一时刻  $t$  为基准点，此点前后两段子序列  $x_1$  和  $x_2$  的样本分别为  $n_1, n_2$ ，两端子序列的平均值为  $\bar{x}_1, \bar{x}_2$  方差为  $s_1, s_2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (19)$$

$$s = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \quad (20)$$

(20) 式中，定义统计量  $t$ ，此式遵从自由度  $\nu = n_1 + n_2 - 2$  的  $t$  分布。将  $t$  随时间变化的图绘出， $t > 0$  表示增长， $t < 0$  即下降，超过显著区间 (0.05) 则表示有较大突变。

辅模型一：标准正态检验 (SNHT)

原理：利用正态分布公式计算，设定 0.05% 之外的为显著突变。

辅模型二：Pettitt

原理：类似 MK 模型构造秩序列，但是直接利用此序列进行检验。

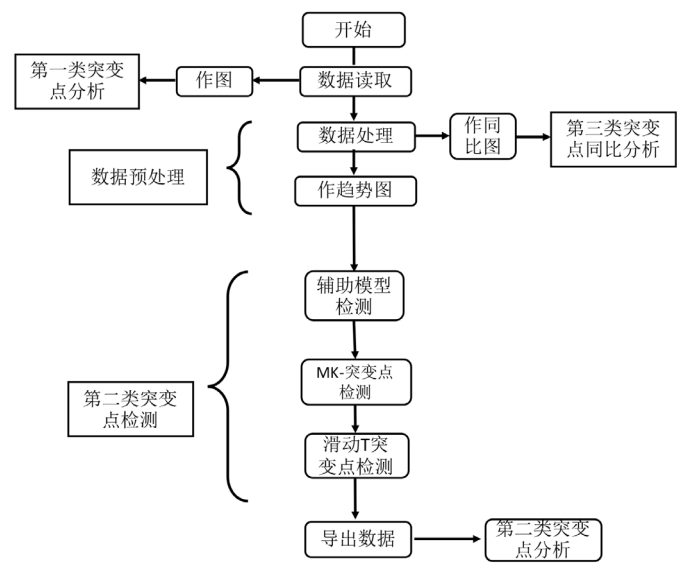
$$U_t = U_{t-1} + V_t, \quad t \in [2, N] \quad (21)$$

$$V_t = \sum_{j=1}^N \text{sgn}(x_t - x_j) \quad (22)$$

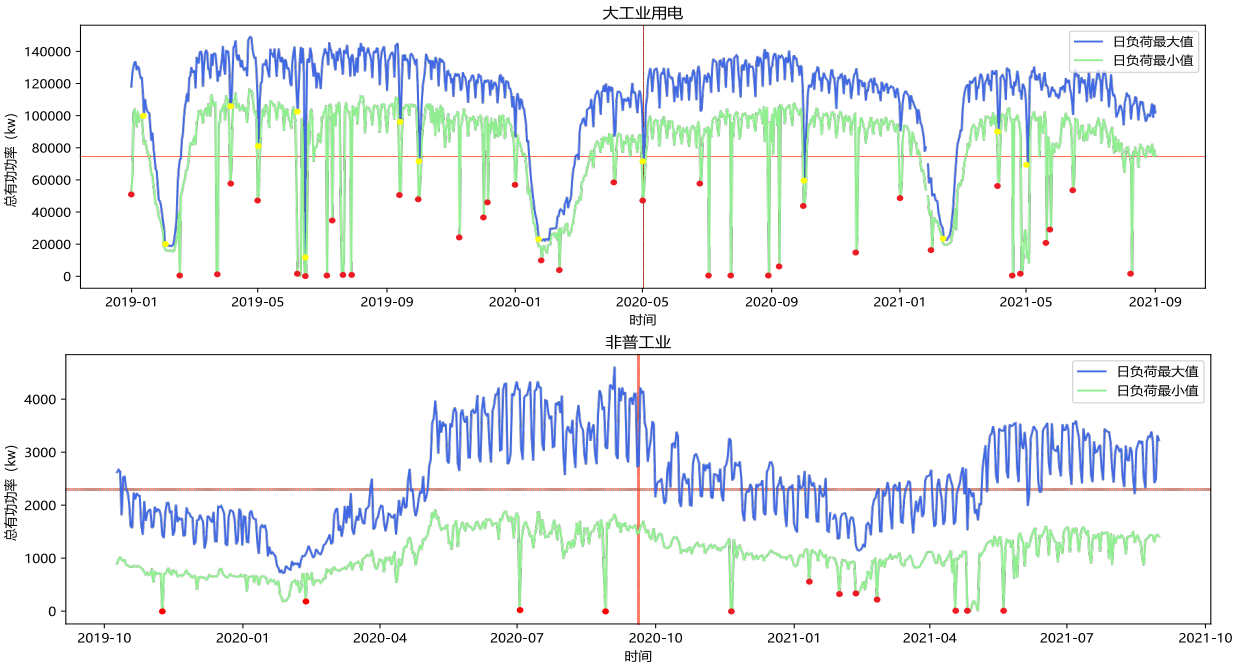
$$U_1 = V_1 \quad (23)$$

式 (22) (23) (24) 表示对于包含  $N$  个样本的时间序列  $X$ ，计算统计量  $U_k$ ；取  $U_t$  中绝对值最大的值  $K_t$ ，该点为最显著的突变点，并计算  $K_t$  对应的统计量  $P$ ，如果  $P \leq 0.05$ ，则认为此突变点显著，此点对应的时刻便是突变开始的时间。

5.3.2 模型求解



对于第一类的突变点，由于数据影响时间较短，异常值处理可能将突变点略去，而平滑处理可能影响突变的显著性并且对检测产生较大误差，所以直接采用原始数据绘图分析。首先选出突变时间极小的点，然后考虑到第二种方法也无法检测到突变值较小的点，本文在图中也一并选取这种点加以分析。



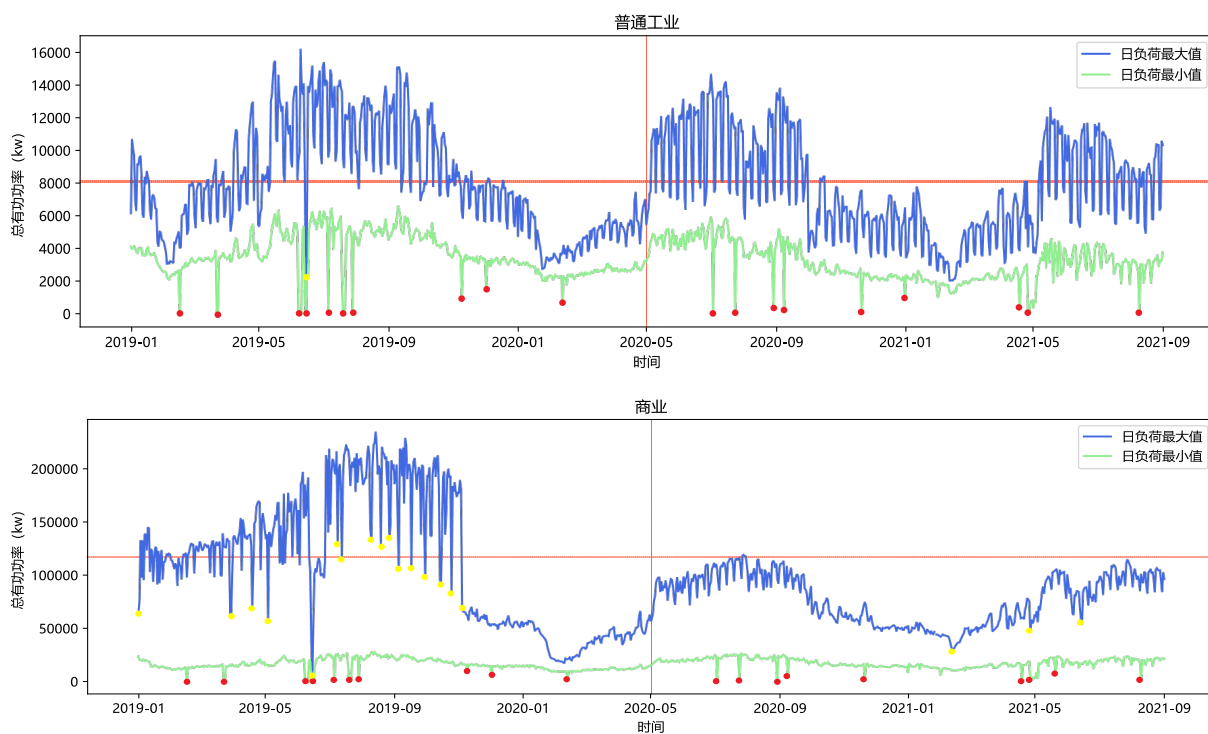


图 9：各行业原始数据图

第一类突变点的可能原因：（下降类）

1. 人为因素：① 大规模全区域停电。如 2019 年 6 月 15 日，各行业最大最小值均降为 0 附近。

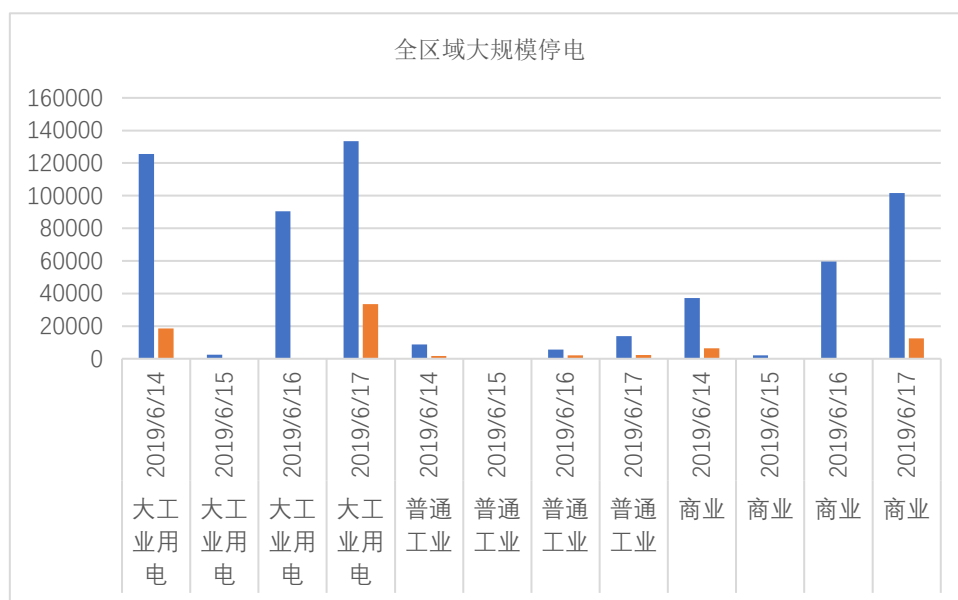


图 10. 第一类突变点原因分析

- ②该地短时间停电，如 2019 年 11 月 9 日及 2019 年 3 月 23 日，最低用电量突变。

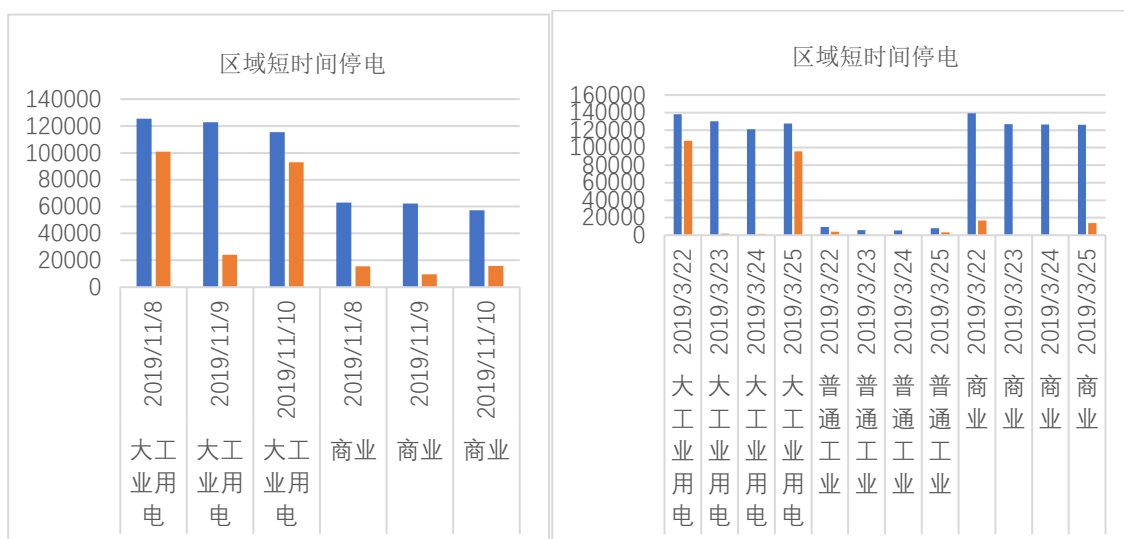


图 11. 第一类突变点原因分析

③工业园区电网短时间故障（只影响大工业单日最小值）

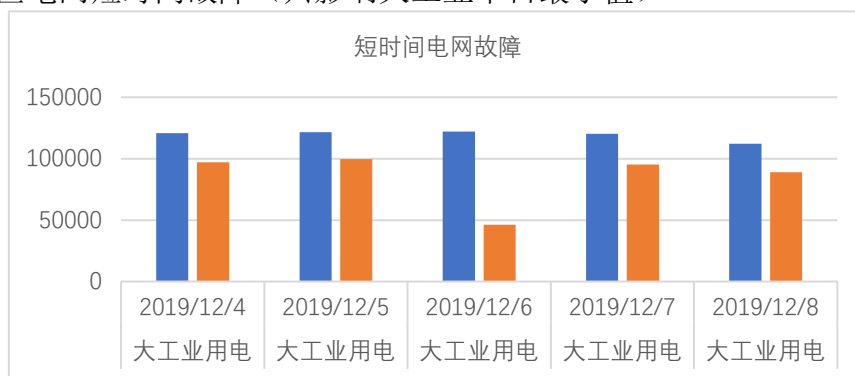
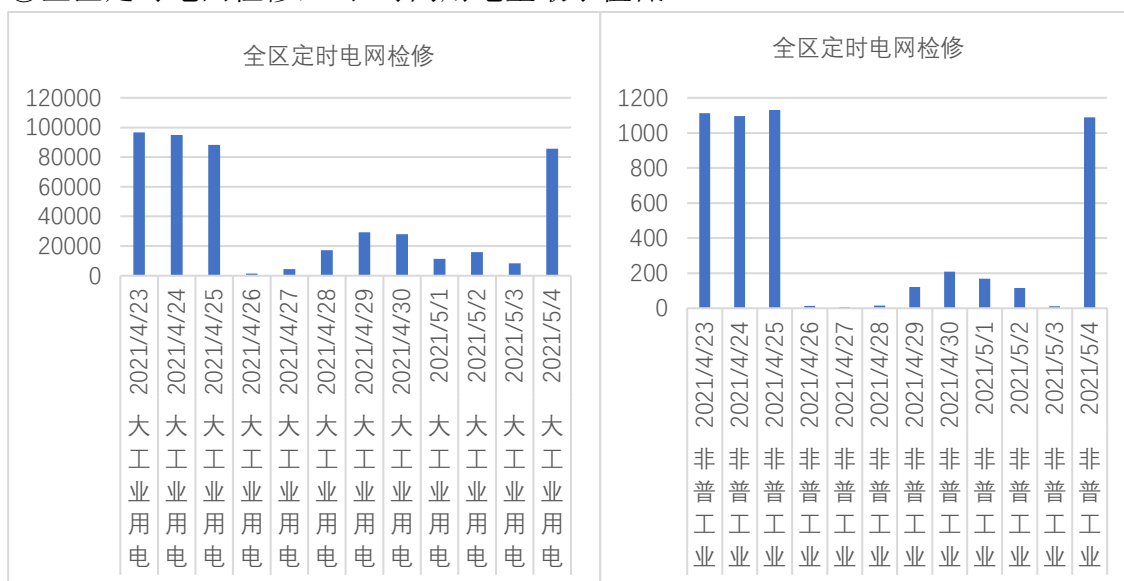


图 12. 第一类突变点原因分析

④全区定时电网检修，（长时间用电量最小值低）



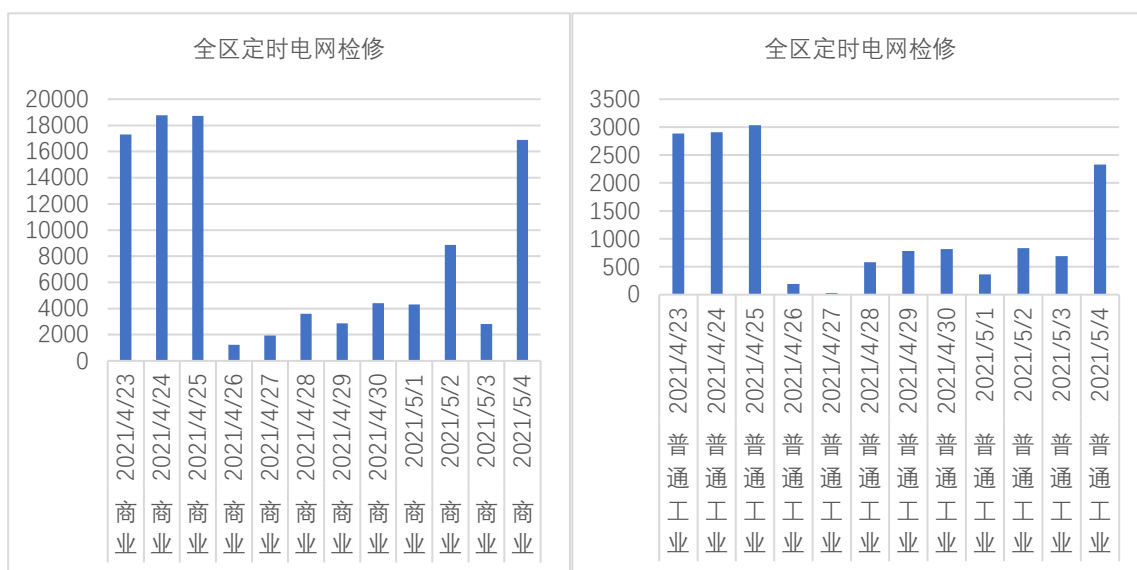
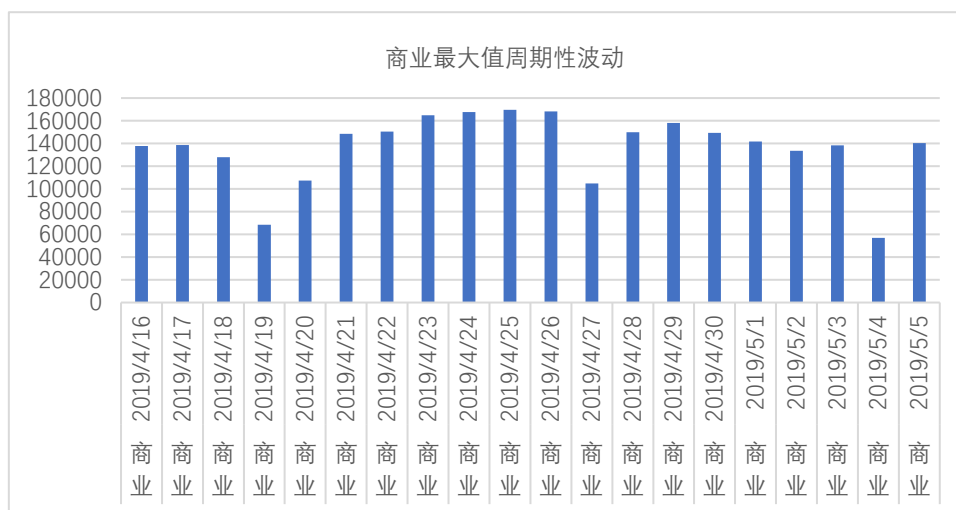


图 13. 第一类突变点原因分析

⑤ 节假日及周期因素（商业在周一显著下降，节假日后一天显著下降）



2. 自然因素：① 此前天气有雨累计超过 3 天，（大工业最大值）

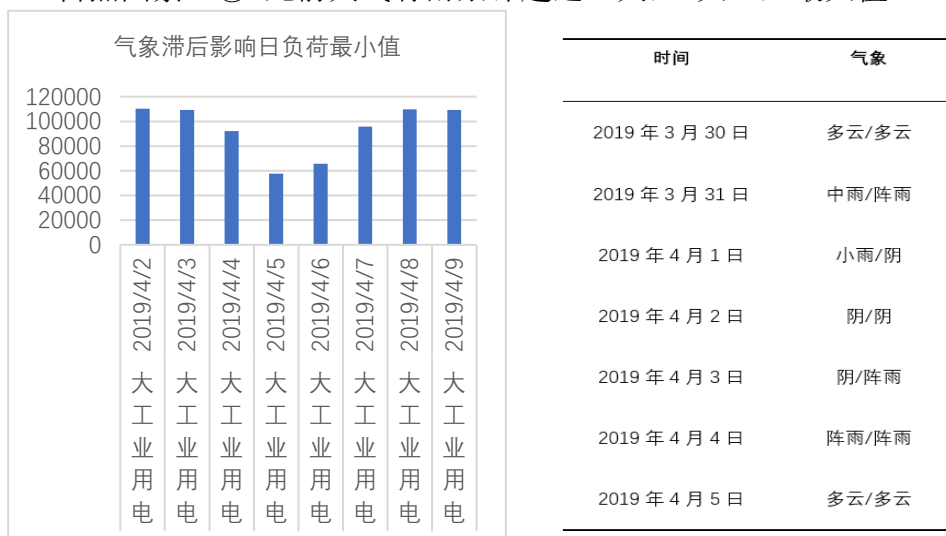


图 14. 第一类突变点原因分析

对于第二类突变点

数据预处理：

第一步，数据读取及数据处理。

将所有数据以时间为索引，分解为大工业，非普工业和商业，并分别存在最大最小值的数组里。采用类似于 5.1.15.1.2 的方法进行数据预处理。

第二步，简易画图，用于检验模型的正确性。

使用 matplotlib 将经过上述处理的，每个行业的最大最小值数据绘制在同一张图中。（以普通工业为例，如下）

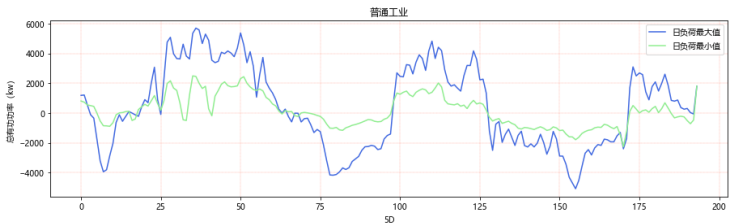


图 15. 普通工业平滑处理后趋势图

第三步，采用主模型和辅助模型进行第二类突变点检测。（选取置信区间 0.05）

第四步，画出滑动 T 的检测图，导出数据。（以商业为例，检测图如下）

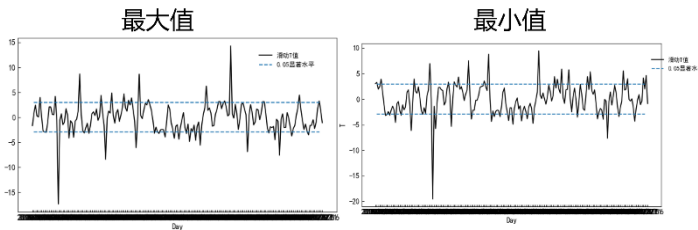


图 16. 商业滑动 T 检测图

对于第二类突变点，大致有以下原因：

1. 社会因素：①节假日(如国庆节、春节等长假)，对各行业影响不同，如对大工业而言，春节期间有阶段性的显著下降及后期恢复，商业变化则较缓和。国庆节 7 天假期间，大工业用电前三天用电较少，普通工业 7 天整体较少，商业用电量与平时相比未明显改变。
2. 自然因素：①极端天气影响 如台风登陆等

表格 1

| 时间              | 气象     | 风力及风向（白天）   | 风力及风向（夜晚）   |
|-----------------|--------|-------------|-------------|
| 2018 年 9 月 14 日 | 多云/多云  | 无持续风向 1-2 级 | 无持续风向 1-2 级 |
| 2018 年 9 月 15 日 | 晴/多云   | 东北风 3-4 级   | 东北风 3-4 级   |
| 2018 年 9 月 16 日 | 暴雨/大雨  | 东风 8-9 级    | 东风 8-9 级    |
| 2018 年 9 月 17 日 | 大雨/阵雨  | 东南风 3-4 级   | 东南风 3-4 级   |
| 2018 年 9 月 18 日 | 雷阵雨/多云 | 无持续风向 1-2 级 | 无持续风向 1-2 级 |

第二类突变点量级检测：

运用上述滑动 T 方法，导出 T 的值。若 T 值大于显著水平 0.05，则将点导出。T 值若为正，则证明是增长类突变，为负则是下降类突变。T 的值越大证明突变量级越大，反之亦然。

对于第三类突变点，在数据处理后，画出三年的同比图，进行第三类突变点分析。

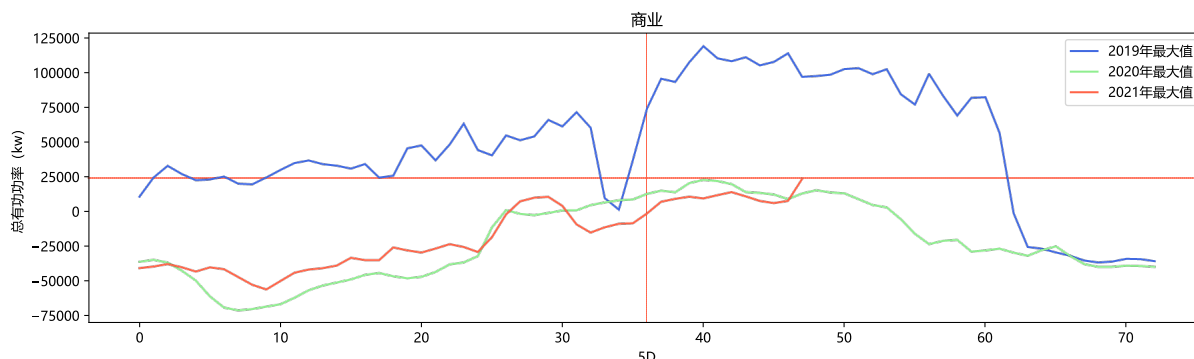


图 17. 商业三年同比图

工业类的年差异不大，但商业显然受到疫情影响，持续低迷到 5 月。

#### 5.4 问题二(2)的模型建立与求解

这里采用类似于 5.2.2 的方法对各行业未来三个月日负荷最大值最小值进行预测，由于数据不存在空缺值因此不需要填充，考虑到各行业的电力负荷具有年周期性与周周期性本文设置其对应的拟合参数得到了预测结果(下图以大工业最大最小值的预测结果为例)。浅蓝色代表置信区间，黑点代表预测值，蓝线为原始电力负荷的时间序列。

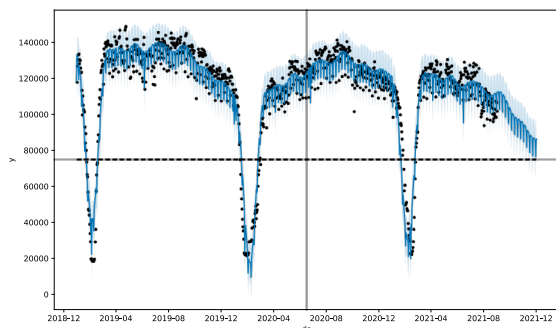


图 18. 大工业行业日负荷最大值的预测

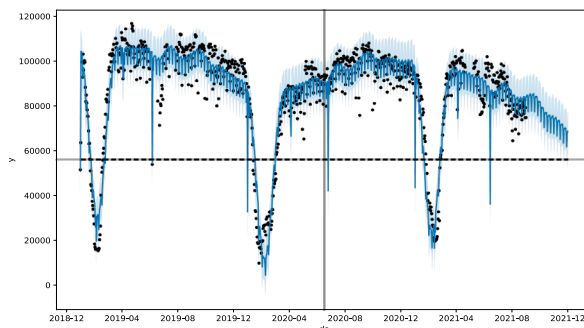


图 19. 大工业行业日负荷最小值的预测

对商业采用 Prophet 模型进行预测发现结果出现负值，预测效果出现异常，说明数据有问题，需要手动筛选。通过将 2019 年 12 月以前的数据剔除得到了正常的预测结果。预测精度分析：

通过划分训练集和测试集，将预测结果与真实数据进行比对，计算出大工业、非普工业、普通工业和商业最大值的平均绝对百分比误差(MAPE)分别为 0.046, 0.056, 0.078, 0.064，最小值的 MAPE 分别为 0.080, 0.099, 0.118, 0.084，说明预测结果较为准确。

#### 5.5 问题二(3)分析与建议

2020 年 9 月 22 日，中国在第 75 届联合国大会上正式提出 2030 年实现碳达峰、2060 年实现碳中和的目标，展现了中国主动承担应对全球气候变化责任的大国担当。随着“双碳”目标的提出，电力作为清洁高效地二次能源，将在支撑社会经济发展，服务民生等各行各业发挥更加重要的作用。“十四五”是中国迈向基本实现社会主义现代化的关



键时期，电力系统的转型与电力供需平衡对保障中国能源安全意义重大，以下为本文对“双碳”目标对各行业用电负荷的影响进行分析并提出针对性建议：

### （一）工业

通过对以往电力数据的分析发现：①整体波动：大工业有较强的年周期波动性，除2020年疫情期间停工停产导致用电量较低外，其他时间负荷基本稳定；普通工业和非普工业两者均呈现出夏季用电负荷最高，冬季较低等特点，峰谷同期，变化趋势基本相同。②“双碳”目标的影响：2020年9月“双碳”目标提出后，对于电力负荷最大值而言，2021年大工业用电负荷略有降低，但总体上仍无较大变化；普通工业和非普工业电力负荷最大值降低幅度略高于大工业用电，但仍跌幅不大。对于电力负荷最小值而言，大工业用电、普通工业与非普工业均略有降低，其中以普通工业降低幅度最为明显。

因此本文预测：大工业用电、非普工业和普通工业的电力负荷在“双碳”政策下仍将保持总体稳定、略有降低，其中普通工业受到“双碳”政策的影响最大。而我国大工业用电负荷占有所有行业用电量近半，因此大工业能否率先实现碳达峰是2030年达峰目标实现的关键。因此，本文重点关注大工业，对工业用电提出以下建议：

#### 1. 优化能源结构。

由于我国目前对能源需求仍在增长期，化石能源以及高耗能产业占比过高，为了达到碳达峰的目标，需要在“十四五”和“十五五”期间做好能源结构转型的准备。

对于自发电的工业园区（如大部分大工业、普通工业），充分使用清洁能源、可再生能源替代燃料，减少对煤等碳排放量高的能源的依赖。并且由突变点数据可知，阴雨天气对大工业用电负荷有较大影响，清洁能源如太阳能、风能等的使用可以在晴日储存一定电量，弥补在阴雨天的用电亏损。对于利用电网的其余工业产业（如非普工业），提倡其选择利用清洁能源的地区电网，从源头减少碳排放。

#### 2. 成立独立供电区以减小故障影响，完善综合梯级利用等服务。

由突变点数据可知，园区电网故障对大工业用电负荷影响较大，而园区内能源消耗大，用户需求大，可以增配电网，成立独立供电区，减小供电不足甚至停电对生产造成的影响。另外，独立电网的使用可逐渐带动新能源的应用，将分布式能源充分整合，智能化分配及调度，削峰填谷。因此，独立电网利于充分利用生产过程中产生的热量，实现电热转化等回收利用，实现能源效率的提高。用户方面，尝试错峰用电，共同改进使电网更加平稳可靠运行

#### 3. 引进先进的技术，节能的同时保障能源供应。

可以完善工业物联网，引进更先进的技术（例如利用机器学习实现自动检测故障、自动维修等）。使用先进的节能科技，实现产业升级，提升能源利用率，达到节约电能的目的。

#### 4. 发展低碳经济及完善碳交易市场体系。

从原料供应环节降低碳排放，提高采购供应链中绿色原料的比例，并在运输环节减少碳足迹，从而实现低碳经济的发展。政府方面，建设全国碳市场，收紧配额总量，在一级市场逐步提升“拍卖”形式的有偿分配比例，适时引入碳税作为碳市场的补充。

### （二）商业

分析：

通过对以往电力数据的分析可以发现：①商业用电负荷受到疫情的影响显著，②“双碳”政策提出后，2021年整体电力负荷略低于2020年，说明商业的节电意识有所提高。随着碳达峰、碳中和的目标推进，本文预测商业的用电负荷会继续缓慢降低。综上本文给出以下针对性建议：

#### 1. 提高用户节电意识。提倡商场、酒店、写字楼等商业场所在用电高峰阶段自觉减少使

用或停用大功率用电设备和非必要的照明灯具，做到公共照明区域“人走灯灭”；合理设置空调温度，合理减少空调使用时间和频率；不使用电器时关闭电源，减少待机耗电。

2. **积极推广低碳产品和节能型电器。**改造与更新老旧设备，如用 LED 等、智能新风系统等节能设备对商业大楼进行二次改造，推动绿色转型。
3. **加快商业用电数字化建设。**童工大数据、云计算等人工智能技术为商业建筑赋能。例如安装电能自动监测系统，对用电量进行智能化节能管理，使用户实时了解用电明细，便于用户合理规划用电情况。
4. **负碳吸收。**利用碳汇和负碳技术实现碳中和目标，例如在商业建筑中建植物墙，不仅美观而且有利于碳的吸收。亦可在商业建筑中采用碳捕集技术实现碳的吸收。

## 六、 模型的评价与推广

### 6.1 模型的评价

#### 6.1.1 模型的优点

- (a) 第一问采用 LSTM 和 Prophet 模型组合预测，综合了两者的优点，预测精度较高。
- (b) 中期负荷预测综合考虑了节假日、年趋势、季节性、周和日的周期性等因素，考虑较为全面。
- (c) 针对不同类型突变点采用不同方法处理，能更好地识别出突变点并分别进行原因分析。

#### 6.1.2 模型的缺点

- (a) 在短期预测负荷中未考虑天气因素的影响。
- (b) 未结合不同行业的特点进行预测分析，而只是全部套用同一模型。
- (c) 突变点检测模型仍需结合人工筛查，有部分突变点模型难以检测出或检测不准确。

### 6.2 模型的改进

- (a) 将天气因素量化分析，用皮尔逊相关系数研究对负荷的影响较大的天气特征并据此进行建模分析。
- (b) 可考虑将 LSTM 和 Prophet 组合模型同样应用到各行业电力负荷的分析，通过调整两者预测结果的加权系数得到最准确的预测结果
- (c) 如果能结合气象部门对未来天气的预测数据可更好地对电力负荷进行预测。

## 七、 参考文献

- [1] 葛娜, 孙连英, 石晓达, 赵平. Prophet-LSTM 组合模型的销售量预测研究[J]. 计算机科学, 2019, 46 (S1) : 446-451.
- [2] Taylor SJ, Letham B. 2017. Forecasting at scale. PeerJ Preprints 5:e3190v2