

基于非平稳广义极值分布模型的电力预测

1 数据介绍

原数据：2015 年 1 月 1 日至 2018 年 6 月 3 日用电量数据，2015 年 12 月 22 日前为 5 分钟采一个值，从 2015 年 12 月 23 日起，每 1 分钟采取一个值。

数据处理：取每天用电量的最大值，我们非常关心用电量过载的问题，具体来说哪一天会出现用电高峰，用电量能够达到多少，所以取每天用电量的最大值。

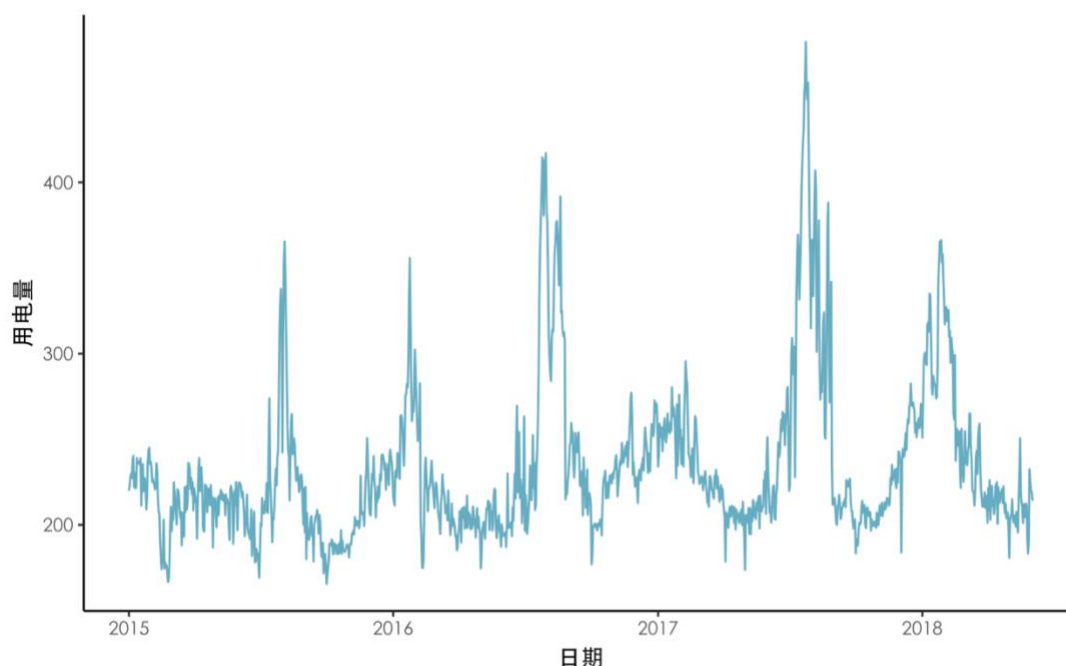


图1 用电量日极值时间序列

下表展示了 2015-2017 年用电量的均值，最大用电负荷，及对应日期，用电量超过 300，400 的天数。可以看到，2015-2017 用电平均水平增加，最大用电负荷也不断上升，超过 300，400 的天数也明显增加。而且，达到最大用电负荷的日期在不断提前。故不能将上一年的用电水平看做是本年度的用电水平，或直接从上一年用电负荷最大的日期得到本年度的用电负荷最大的日期。另外，可以注意到 2016 和 2018 年冬天的用电量明显高于 2015 和 2017 年（见 2.4 的分析）

年份	2015	2016	2017
均值	214.3	236.9	245.0
最大用电负荷	365.5	417.2	482.0
最大用电日期	2015-08-04	2016-07-30	2017-07-24
超过 300 天数	9	37	41
超过 400 天数	0	5	11

表 1

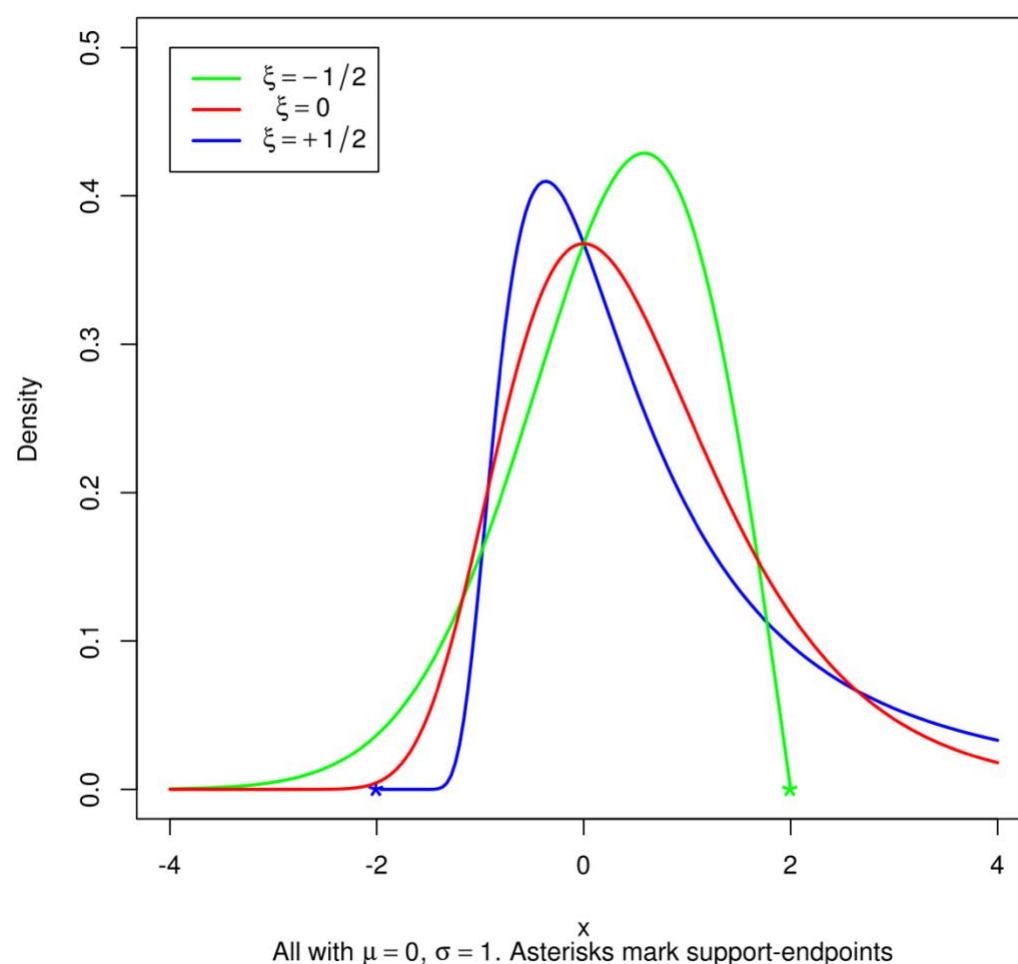
2 模型方法

2.1 广义极值模型 (GEV):

$$\text{参数: } \begin{cases} \mu \in R - \text{位置} \\ \sigma > 0 - \text{尺度} \\ \xi \in R - \text{形状} \end{cases}$$

$$\text{概率密度函数: } \frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)}, \quad t(x) = \begin{cases} \left(1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ e^{-\frac{x-\mu}{\sigma}} & \xi = 0 \end{cases}$$

Generalized extreme value densities



2.2 数据分布特征

下面是分年度用电量频数分布直方图, 可以看到用电量分布大致符合极值分布。

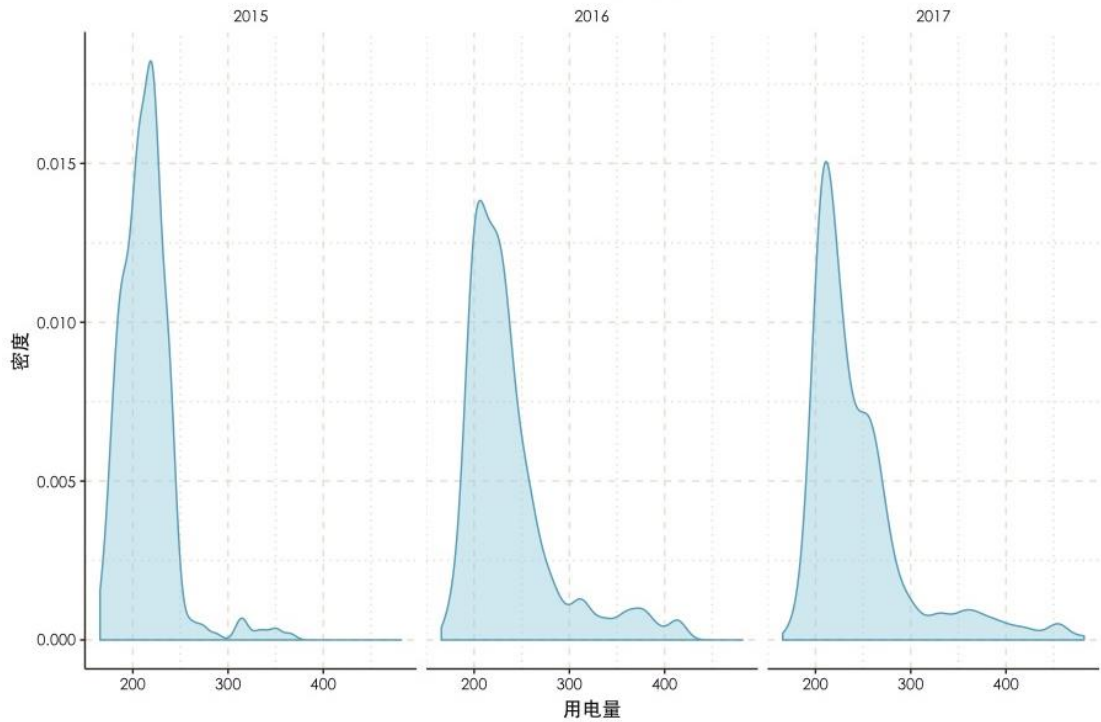


图2 用电量频率分布密度曲线图

2.3 平稳 GEV 模型

GEV 参数是常数，其他变量变化都不会影响分布的三个参数的取值。

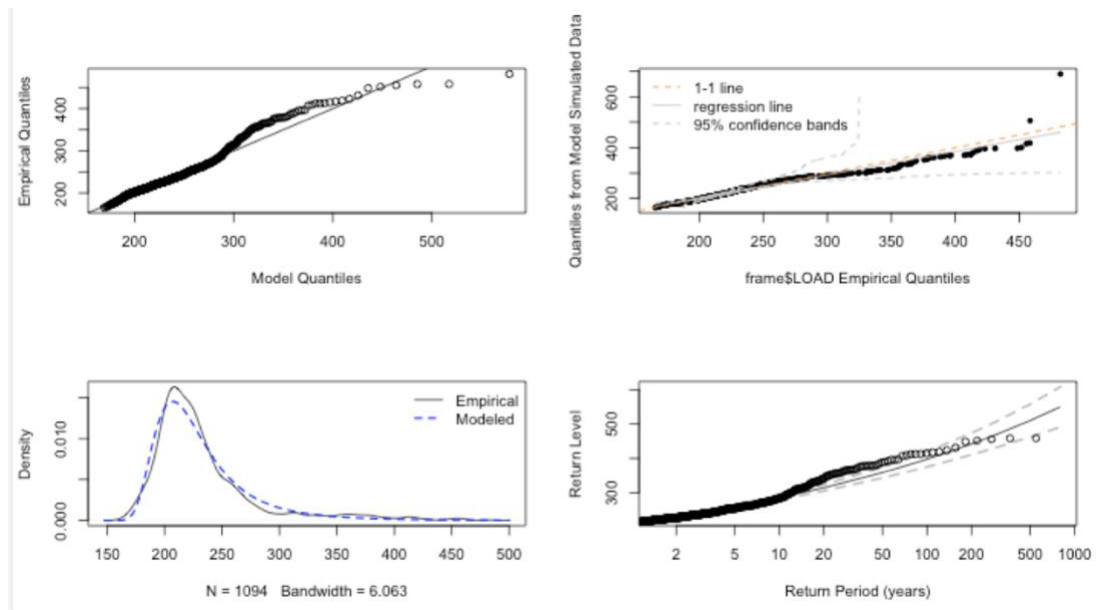


图3 平稳 GEV 模型拟合效果

实证分析：

1. 尾部偏差较大
2. 基本是符合 GEV 分布的
3. 可能是非平稳的

2.4 非平稳 GEV

μ, σ 的水平依赖于其他变量，采用极大似然估计（MLE）方法估计模型参数。

1. μ, σ 可能与温度有关。

实际数据（2015-2017）表明用电量水平和温度（由于日温差稳定，这里使用的是当天最高温度）有很强的关系。随着温度升高，用电量先下降再上升，并且高温对于用电量的影响明显大于低温。这种关系并不适合用二次函数拟合，因为是不对称的，这里我们使用了 4 次函数，并且，通过似然比检验，当天最高温度（1-4 次幂）都是显著的。

注意到第 1 部分 2016，2018 年用电量水平很高，对应到温度，这两年冬天最高气温出现了零下的情况，其中 2016 年最低温度更低，出现在 1 月 23 日和 1 月 24 日，而 2018 年在 10 度以下用电量，与前 3 年有明显上升。

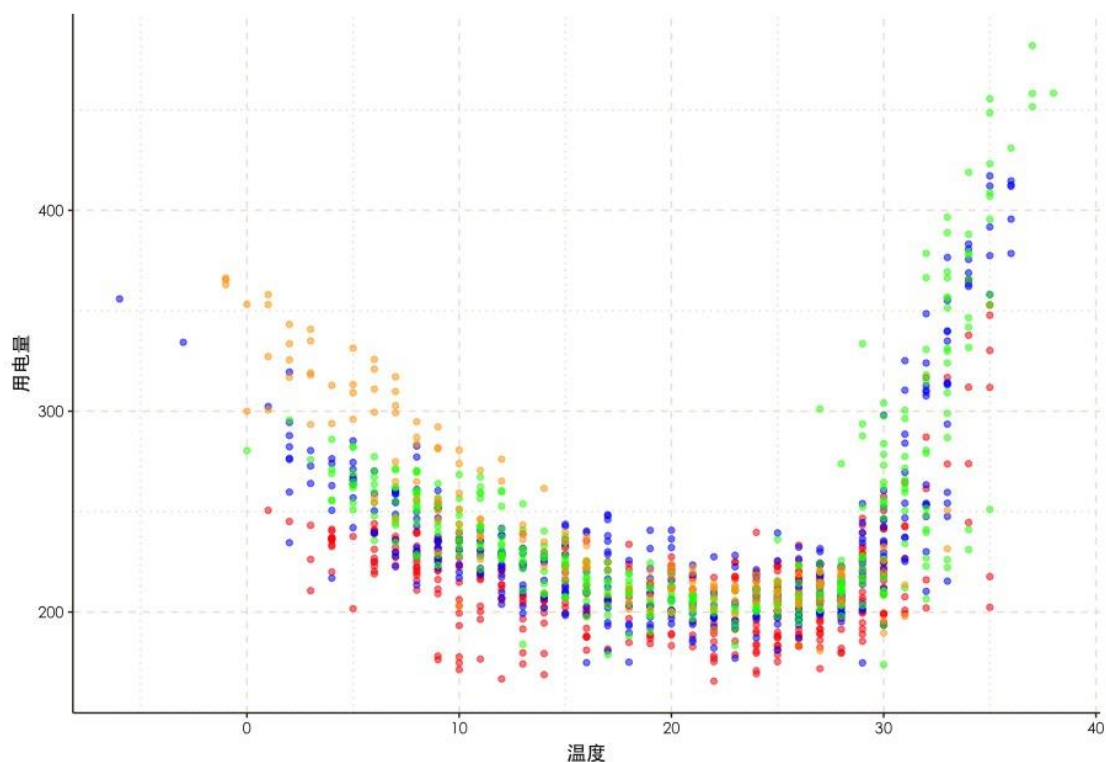


图4 最高温度，用电量散点图(红色-2015，蓝色-2016，绿色-2017)

实际数据（2015-2017）也表明用电量方差和温度有很强的关系。随着温度升高，用电量波动水平先下降再上升，并且高温对于用电量的影响明显大于低温。与用电量水平和温度拟合同理，我们也使用了当天最高温度（1-4 次幂）。

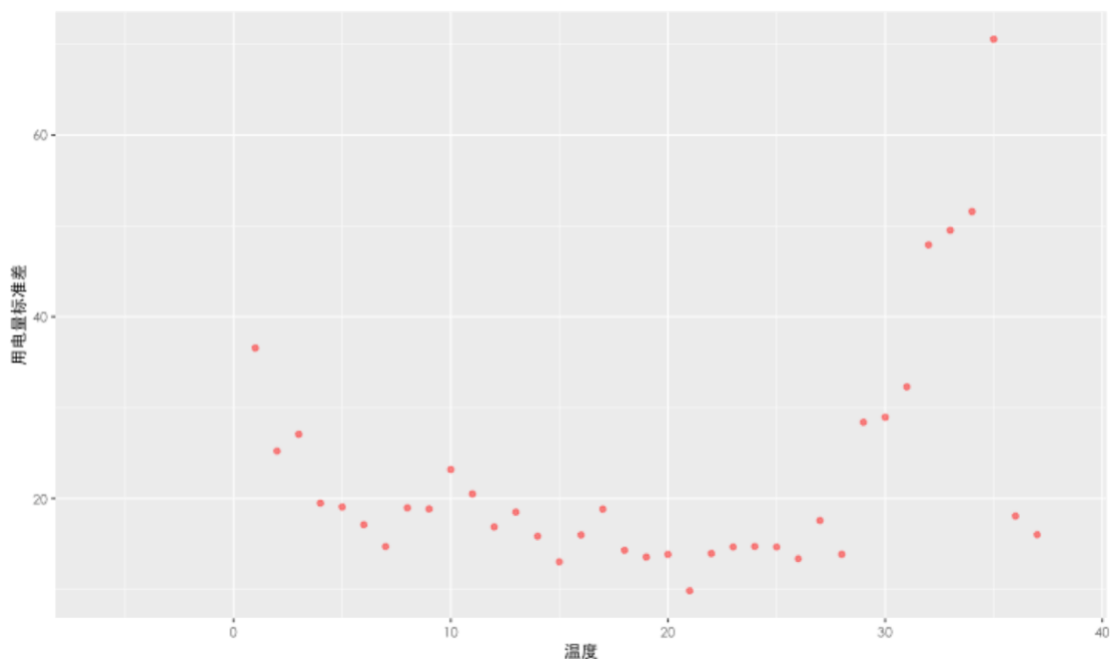


图5 温度，用电量标准差散点图

2. μ 与前一天用电量有关，并且由图 1，每年的用电量有增加的趋势。

对于趋势的刻画有两种方法（见第 3 部分），一种是将年份作为趋势项（-2015），另一种是同时加入前一天用电量和一年前的用电量。在长，短期预测中，这部分变量的选择受到限制，比如提前 7 天的预测，不能使用前一天用电量作为解释变量，具体见表 1。

3 模型

我们采用广义极值模型，实现了 1 天，7 天，以及 1 年的预测。具体变量见下表，均通过似然比检验。

提前时间	变量 (μ)
1 天 (a)	当天最高温度 (1-4 次幂)，滞后 1, 2 天用电量，趋势 (年份)
7 天	当天最高温度 (1-4 次幂)，滞后 7 天用电量，趋势 (年份)
1 天 (b)	当天最高温度 (1-4 次幂)，滞后 1 天，1 年用电量

表 2

注： σ 的解释变量皆为当天最高温度 (1-4 次幂)。

具体流程

- 先通过预测年份前两年的数据使用极大似然估计方法(MLE)估计模型参数，可以实现滚动预测与数据更新（每年都可以更新数据）。
- 将当年的数据（假设温度已知）对应的变量带入模型，使用估计出来的模型参数，计算得到 GEV 模型的参数 μ, σ, ξ 直接使用估计得到的值（认为是常数）
- 得到了广义极值分布之后 $GEV(\mu, \sigma, \xi)$ ，将分布的众数作为预测值。
众数的表达式为：

$$\begin{cases} \mu + \frac{\sigma((1+\xi)^{-\xi} - 1)}{\xi}, & \xi \neq 0 \\ \mu, & \xi = 0 \end{cases}$$

4 模型预测效果

这里我们主要采用了三个维度的标准来评价模型预测效果：相对误差，最高气温预测误差，相对误差在 5%，10% 以内的天数占比。

➤ 采用相对误差作为评估预测效果的指标：

$$\text{相对误差} = |\text{预测值} - \text{真实值}| / \text{真实值}$$

三个模型预测效果（2017）如下表：

提前时间	均值	最大值
1 天 (A)	3.8%	26.8%
7 天	5.5%	32.0%
1 天 (B)	4.2%	24.5%

表 3

三个模型预测效果（2018）如下表：

提前时间	均值	最大值
1 天 (A)	4.4%	16.7%
7 天	5.8%	25.5%
1 天 (B)	5.2%	16.9%

表 4

1 天和 7 天预测相对误差均值都可以控制在 5% 左右。从相对误差均值和最大值的角度考虑，1 天 (a) 的预测效果比 1 天 (b) 略好一些。

注：由于 18 年的数据不完整，预测效果不是全年的统计。建议以 2017 年为准。1 年预测虽然预测效果还可以，但是 1 年的温度难以获得。

➤ 最高气温预测误差（2017）

提前时间	实际日期	实际负荷	预测负荷	相对误差
1 天 (A)	2017-07-24	482.04	454.50	5.7%
7 天	2017-07-24	482.04	406.90	15.5%
1 天 (B)	2017-07-24	482.04	468.26	2.9%

提前时间	预测日期	实际负荷	预测负荷	相对误差
1 天 (A)	2017-07-23	458.35	465.48	1.6%
7 天	2017-07-23	458.35	434.58	5.2%

1 天 (B)	2017-07-23	458.35	486.13	6.1%
---------	------------	--------	--------	------

表 5

表 5 中上面的表格展示了 2017 年实际达到最大负荷的日期，实际负荷，预测负荷，和相对误差，下面的表格展示了 2017 年预测达到最大负荷的日期，实际负荷，预测负荷，和相对误差。实际达到最大负荷的日期为 7 月 24 号，而预测达到最大负荷的日期为 7 月 23 号。1 天预测的相对误差都比较小，1 天 (a) 在预测达到最大负荷的日期的相对误差较小为 5.6% 相比 1 天 (b) 的 6.1%。而 1 天 (b) 在实际达到最大负荷的日期的相对误差较小为 2.9% 相比 1 天 (a) 的 4.2%。

➤ 2017 预测相对误差 5%，10% 天数占比

提前时间	相对误差 5% 之内	相对误差 10% 之内
1 天 (A)	77.0%	91.2%
7 天	60.5%	85.2%
1 天 (B)	71.2%	92.1%

表 6

表 6 中 1 天 (a) 预测误差在 5% 内的比例为 72.9%，略高于 1 天 (b) 的 71.2%，而相对误差 10% 之内比例 1 天 (b) 的比例为 92.1% 高于 1 天 (a) 90.1%。7 天预测误差 60.5% 可以控制在 5% 以内，85.2% 可以控制在 10% 以内。

5 温度敏感性

实际操作中，温度项一般采用预测温度，这就使温度本身带有了不确定性。我们在 2017 年最高气温上加入了一定随机性 ($N(0, sd)$)，通过模拟得到模型预测效果和温度随机性 (sd) 的关系。（由于运算速度限制，只进行了 100 次模拟）。

- ✓ 1 天预测的平均误差基本可以控制在 4.5%，最大相对误差控制在 30%。考虑到实际天气预报温度 sd 在 3 内比较合理。

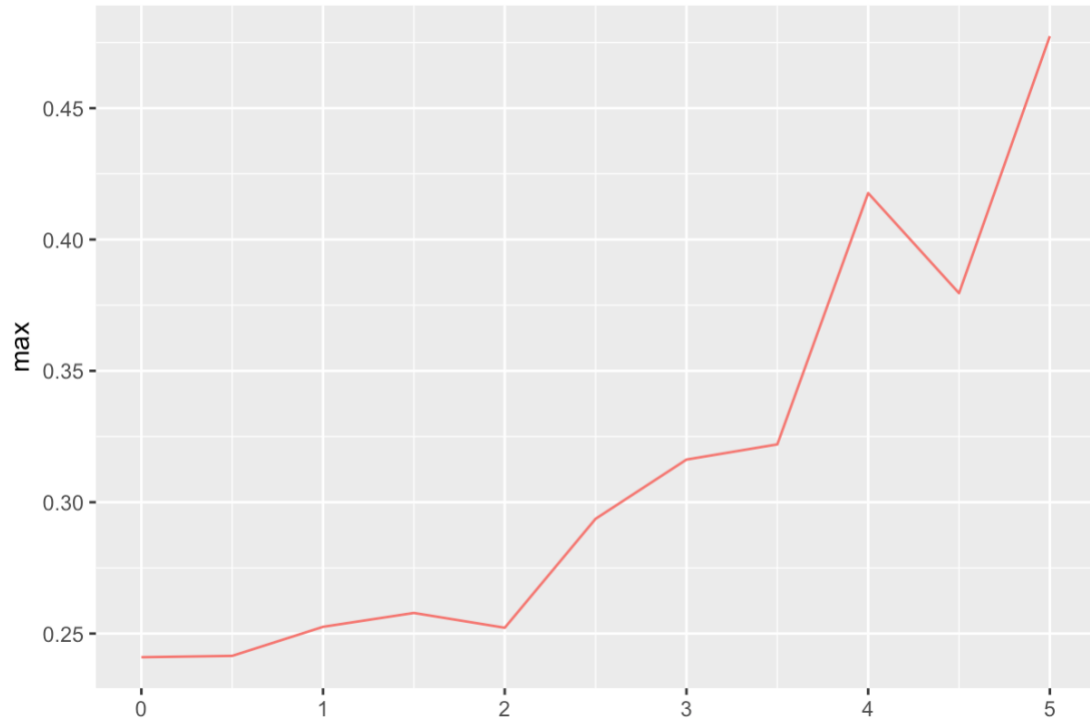


图6 相对误差最大值和温度随机性(sd)的关系

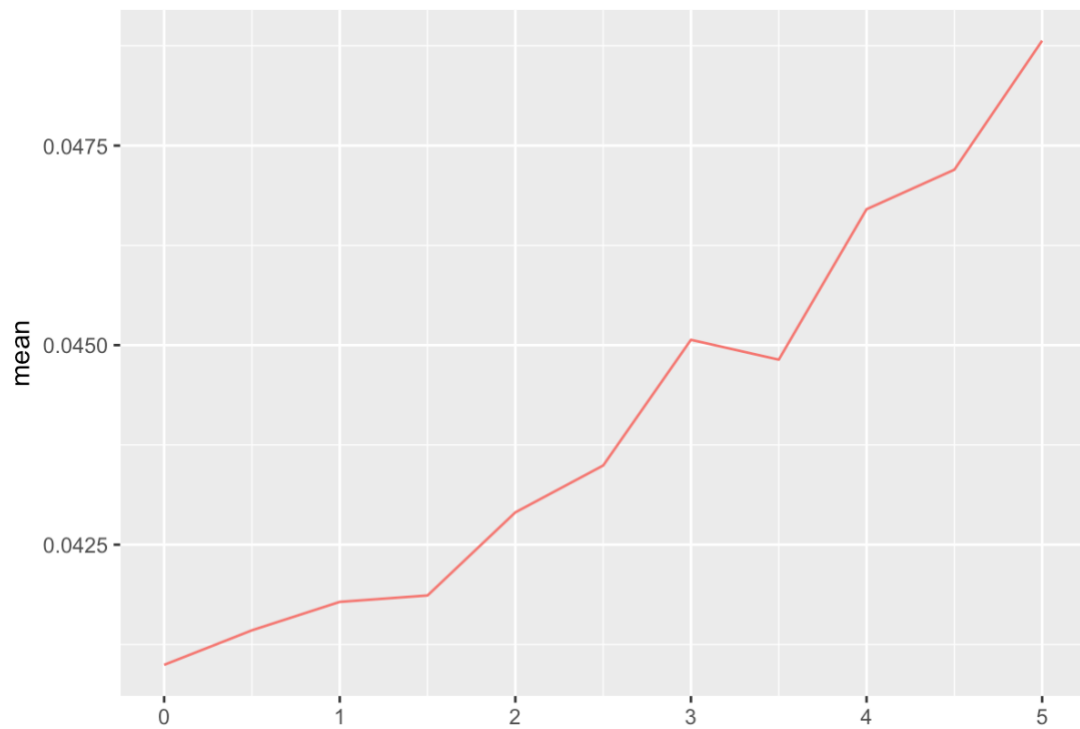


图 7 相对误差均值和温度随机性(sd)的关系

- ✓ 7 天预测的平均误差基本可以控制在 6%，最大相对误差控制在 40%，考虑到实际天气预报温度 sd 在 3 内比较合理。

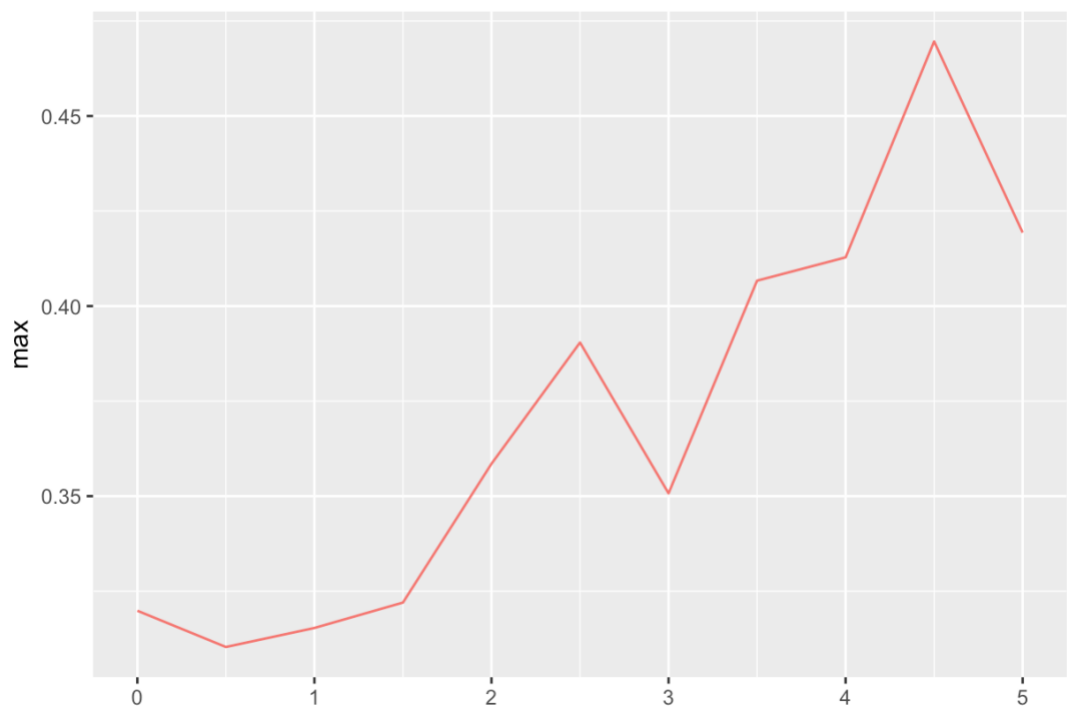


图 8 相对误差最大值和温度随机性(sd)的关系

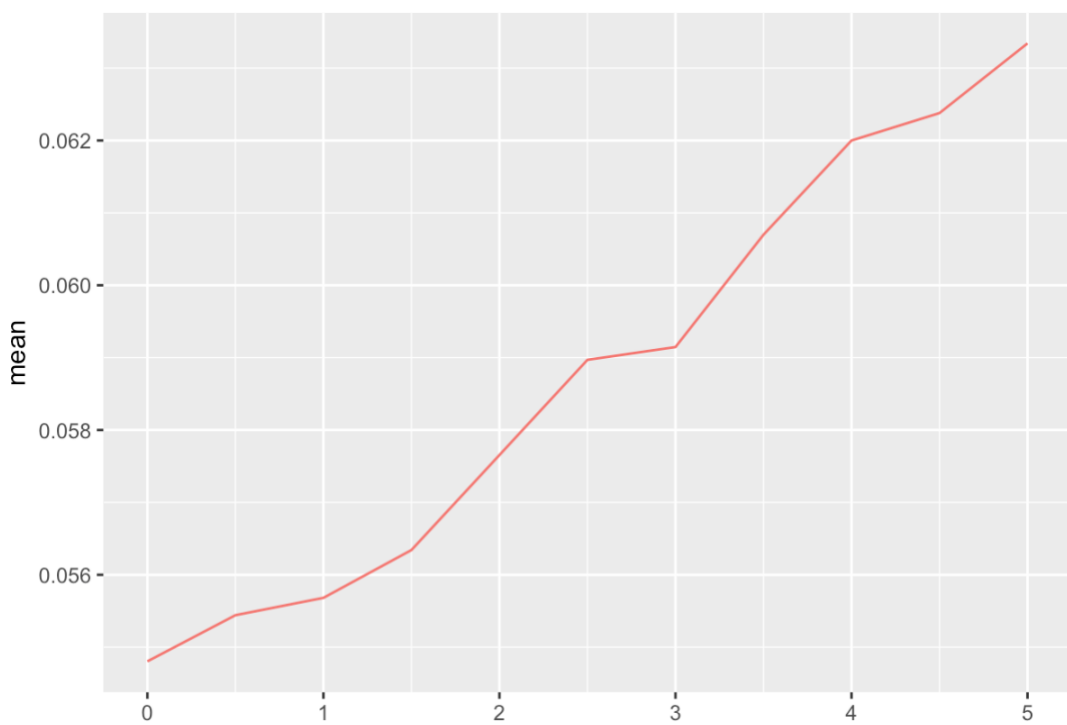


图 9 相对误差均值和温度随机性(sd)的关系