

带修正预测的 SVR 模型

0.前言

在我们的报告中，结构组织逻辑如下：

1. 第一部分介绍了该课题使用的背景知识以及实际的工程应用背景，指出了我们实验主要使用到的时间序列预测模型 SVR 和 RNN，以及要针对的问题：水文预报中的预测失准问题。
2. 第二部分介绍了我们针对工程应用问题的抽象、建模和理论基础，以及提出了主要使用的修正预测算法，我们认为通过历史数据提取异常值并构造相应分布来训练修正模型，可以帮助我们解决一些预测情况下的预测失准问题。
3. 第三部分则是我们主要进行的实验与实验结果。
 - 1) 第三部分第一节介绍了我们的数据集及尺寸，其中特别指出了在预测模型和修正模型中使用的数据集尺寸不同，具体来说我们使用不同的参数对时间序列数据进行采样，这使得我们后面要用一定的手段来使得两个模型的数据集的尺寸相同。
 - 2) 第三部分的剩余 4 节主要介绍了我们的 4 个实验，
 - a) 第 1 个实验比较了采用不同特征来训练预测模型性能的优劣，综合考虑训练速度与预测结果，最终我们选择只使用水位高度 Z 来训练预测模型；
 - b) 第 2 个实验我们通过定性分析一些预测中的异常水位值与其相邻若干天的水位值，得到了“欠优点”的特征以及“欠优点”和“异常点”的关系，我们后面的实验即是在以“欠优点”代替“异常点”的合理假设上进行的；
 - c) 第 3 个实验我们比较了通过不同的准则选择“异常点”来训练修正模型的修正效果，为了便于理解与叙述，我们首先提出了修正模型训练算法，分别应用 3-sigma 原则和基于 bootstrap 的置信区间估计的方法训练修正模型，并设置对照试验来检验我们修正模型的性能，同时我们还指出可以直接根据测试集上的预测结果选取数据，作为评判各种准则训练的修正模型的依据（换言之这种方式为我们提供了评判模型的手段）；
 - d) 在第 4 个实验中，通过我们给出的 4 个指标 N 、MAE、MRE 和 RMRE，我们对各个准则确定的修正模型进行了比较与评价，并给出了一些研究结论。

1.背景

1.1 SVR

支持向量回归(SVR)：可以容忍模型输出 $f(x)$ 和真实标签 y 之间带有 ϵ 偏差的回归模型。即：

$$\ell_{\epsilon}(z) = \max\{0, |z| - \epsilon\}$$

仿照SVM寻找间隔最大的超平面思想，SVR问题可表示为下面一般形式：

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \ell_{\epsilon}(f(x_i) - y_i)$$

利用凸优化理论可求得SVR解的一般形式：

$$f(x) = \sum_{i=1}^m (\hat{\alpha}_i - \alpha_i) \kappa(x, x_i) + b$$

其中 $\kappa(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ 为核函数。

我们经常用高斯核，也称RBF核作为核函数。

$$\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \sigma > 0$$

1.2 RNN

循环神经网络(Recurrent Neutral Network)：专门用来处理时序问题，核心思想是将 $t-1$ 时间片的隐结点输出作为当前 t 时间片的输入。一般形式为：

$$h_t = \sigma(x_t \times \omega_{x_t} + h_{t-1} \times \omega_{h_t} + b)$$

其中 x_t 为当前时间的输入， h_{t-1} 是上一时间隐结点的输出。

1.3 课题背景

在预测任务中，使用机器学习的方法可以对一些指标进行较为准确的预测，尤其当指标满足一定的概率分布时，我们的预测具有较强的鲁棒性，但是在某些时间段下部分指标会出现偏移概率分布的情况，具体表现为在某一时刻下出现了明显的增加或者下降。

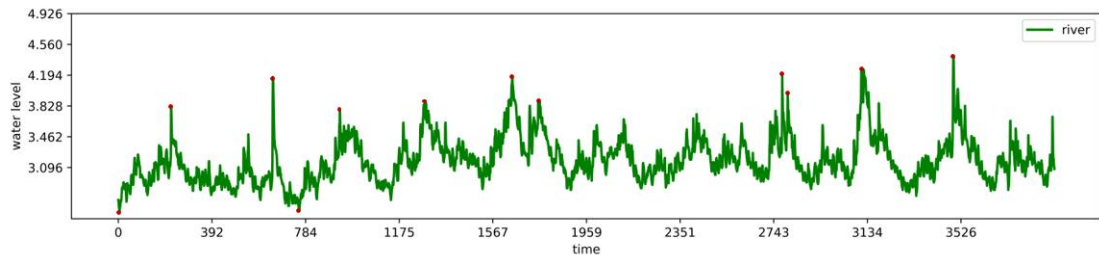


图 1.1 水位高度观测序列中的异常值

比如在上图中红点位置就是作为在整个观测周期中与其它值偏移较大的全局异常值，当然

在某一段观测周期内，部分观测点也可以认为是**局部异常值**。

类似地，在实际的水文预测场景中，预测模型对某些特征的预测具有滞后性，如在大部分时间点模型都能做出较好的预测，但是也存在部分时间点某个指标的观测值明显异于过去一段时间，此时传统方法会利用一些工程上的公式来修正预测结果，而我们使用机器学习的方法训练出一个修正模型来对于这些预测偏差的数据进行修正。

2. 模型与算法

2.1 模型

假设在观测周期： $T = (t_1, t_2, \dots, t_n)$ 内某时刻的观测值 $\mathbf{x}_i = (x_1, x_2, \dots, x_d, y_i)$ 中， x_k 为相关特征， y_i 为真实标签。真实标签集合 $\mathbf{y} = (y_1, y_2, \dots, y_n)$ 总体服从参数为 φ 的某一分布 D_φ ，即 $\mathbf{y} \sim D_\varphi$ ，同时存在某些时间点 $T' = (t_p, \dots, t_q, \dots, t_r)$ ($|T'| \ll |T|$)下的真实标签 $\mathbf{y}' = (y_p, \dots, y_q, \dots, y_r)$ 服从不同参数 φ' 的同一分布 $D_{\varphi'}$ ，即 $\mathbf{y}' \sim D_{\varphi'}$ ，那么我们称 D 为全局分布， D' 为局部分布。

考虑在全局分布 D 下对真实标签 \mathbf{y} 的预测 $\hat{\mathbf{y}} = f(\mathbf{x}) = f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ ，如果分布 D 是满足可加性的，且 $\hat{\mathbf{y}}$ 与 \mathbf{y}' 满足独立同分布，那么随机变量： $\hat{\mathbf{y}}_{T'} + (-\mathbf{y}') \sim D_{\varphi - \varphi'} (T' \subseteq T)$ ，其中 $\mathbf{y}_{T'}$ 表示对于分布 D 在时刻 T' 的预测标签，我们称 $D_{\varphi - \varphi'}$ 为局部的修正分布。

我们的主要思想就是通过在修正分布 D 下训练我们的修正模型，我们可以将异常值进行修正，使其服从全局分布 D 。

2.2 主要算法

算法 1：修正预测算法

Input: 训练集 $Train = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$;

测试集 $Test = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_k, y_k)\}$;

判定准则criterion.

Output: 测试集上的修正值 $\tilde{\mathbf{y}}_{\text{final}} = \{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_k\}$

1: 在训练集上通过判定准则criterion 筛选出不属于分布 D_φ 的全局异常值 $\mathbf{y}' = (y_p, \dots, y_q, \dots, y_r)$ ，记录对应观测点 $T' = (t_p, \dots, t_q, \dots, t_r)$;

2: 在全局分布 D_φ 下训练预测模型 M_1 ，并在训练集 $Train$ 和测试集 $Test$ 上得到预测值 $\hat{\mathbf{y}}_{\text{train}}$ 和 $\hat{\mathbf{y}}_{\text{test}}$;

3: 选取 M_1 在观测点 T' 下对应的预测值 $\hat{\mathbf{y}}_{T'}$ ，得到差分标签 $\tilde{\mathbf{y}}_{T'} = \hat{\mathbf{y}}_{T'} - \hat{\mathbf{y}}_{\text{train}}$;

4: 令输入特征为 $\mathbf{x}_i' = (x_1, x_2, \dots, x_d, y_i)$ ，预测标签 $\mathbf{y}_i' = \tilde{y}_i$ ，在修正分布 $D_{\varphi - \varphi'}$ 下训练修正模型 M_2 ;

5: 在测试集 $Test$ 上得到修正模型的修正值 $\hat{\mathbf{y}}_{\text{modify}}$;

6: 对于测试集上不属于分布 D_φ 的预测值 $\hat{\mathbf{y}}_{\text{test}}$ ，对其修正为 $\tilde{\mathbf{y}}_{\text{final}} = \hat{\mathbf{y}}_{\text{test}} + \hat{\mathbf{y}}_{\text{modify}}$ 。

7: **return** $\tilde{\mathbf{y}}_{\text{final}}$

3. 实验与结果

3.1 实验准备

1. 数据集

河流流量和降雨数据集，数据集形状为 3922×33，包含 3918 条不同观测时间下某条河的水位高度数据(4 条数据缺失)、沿河上下游 29 个站点降雨观测站的降雨量数据和 2 个闸的相关数据。

2. 字段说明

字段	TM	Z	DRP(id)	TGTQ,TDZ
字段名	观测时间	水位高度	降雨数据	闸数据
备注	格式为年/月/日 时:分	某条河的水位高度	29 个降雨观测站的降雨量数据，id 为雨量站的编号。	闸的相关数据

3. 训练集和测试集

1. 预测模型：RNN 或 SVR

每 7 天为一个采样序列，以 Z 或以 DRP(id)+Z+TGTQ+TDZ 作为特征，预测第 8 天的水位高度 Z。训练集和测试集按照 80%：20%进行划分。在使用多特征预测时，我们要对所有特征进行归一化处理，下文同。

采样后的形状：

特 征	DRP+Z		Y		Z		Y	
	X				X			
	2D	3D	2D	3D	2D	3D	2D	3D
训 练 集	(3115,210)	(3115,7,30)	(3115,1)	(3115,1,1)	(3115,7)	(3115,7,1)	(3115,1)	(3115,1,1)
测 试 集	(774,210)	(774,7,30)	(774,1)	(774,1,1)	(774,7)	(774,7,1)	(774,1)	(774,1,1)

2. 修正模型：SVR

以 DRP(id),Z,TGTQ,TQZ 作为特征，删除观测周期内超过 80%降雨量都是 0 值的观测站点(S13,S15,S19,S20,S23,S25,S28 共 7 个站点)，以异常值出现当天及前 3 天数据作为一个采样点，预测值为当天的**差分标签** $\hat{y}_{T'}$ 。差分标签的定义如下：

$$\hat{y}_{T'} = \hat{y}_{T'} - \hat{y}_{\text{train}}$$

采样后的形状：

特征	X	Y
训练集	(3119,100)	(3119,1)
测试集	(778, 100)	(778,1)

4. 异常值的确定(criterion)

- 1) 3σ 原则：对于服从正态分布 $N(\mu,\sigma)$ 的随机变量 X ，数值分布在 $(\mu-\sigma,\mu+\sigma)$ 中的概率为 0.6826，数值分布在 $(\mu-2\sigma,\mu+2\sigma)$ 中的概率为 0.9544 数值分布在

$(\mu - 3\sigma, \mu + 3\sigma)$ 中的概率为 0.9974，其取值几乎全部覆盖在区间 $(\mu - 3\sigma, \mu + 3\sigma)$ 中，落在该区间外的值都视为异常值。使用该原则前要满足如下两个条件：

1. 样本数据满足或近似满足正态分布；2. 测量次数充分大。

- 2) Bootstrap：基于 Bootstrap 方法的置信区间估计使用重采样技术生成多个数据集来估计统计量，根据获得的一系列预测结果来计算置信区间。设预测结果为 X ，方差为 σ ，置信度为 p ，对应标准正态分布的分位数为 z （特别地当 $p = 0.05$ 时 $z = 1.96$ ），那么区间上下界可以表示为： $(X - \sigma z, X + \sigma z)$ ，落在该置信区间外的点我们有理由认为该点为异常点。使用 Bootstrap 进行区间估计也要满足两个条件：

1. 重采样的每个样本独立同分布；2. 预测结果的分布近似为正态分布。

但是即使条件 2 不满足，使用 Bootstrap 方法生成的预测结果仍然可以提供一种对预测结果的不确定性的估计，并且也可以采取不同的分位数来对置信区间进行构造。

- 3) MRE：MRE 准则基于实际预测的结果对异常值进行判定，可用于没有大量统计数据的当前预测结果的修正，在水文预测中 MRE 大于 5~10% 区间可近似认为是当前预测模型下可被认为是异常值，我们对这些值使用修正模型进行修正。

3.2 实验一：选取所有指标或 Z 作为特征训练预测模型性能的比较

训练集上的训练速度比较：使用 Z 作为特征训练时，在相同隐藏层的条件下，epoch=100 时收敛，收敛时 MSE 在 $1e-3$ 左右，而使用所有指标作为特征训练时，epoch=800 时收敛，收敛时 MSE 在 $5e-3$ 左右，可见使用 Z 作为特征训练可以更快达到收敛且效果更好。

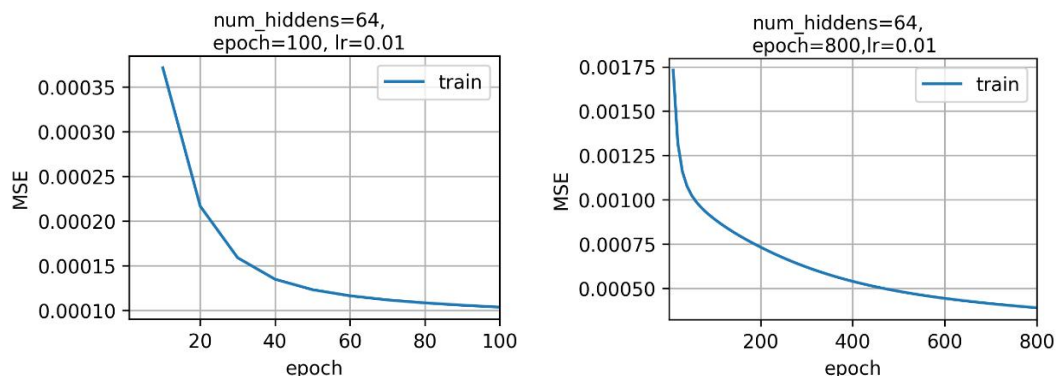


图 3.1 Z 和 DRP+Z 训练过程的比较

预测效果比较：总体上说使用 Z 作为特征训练和使用所有指标作为特征训练都能对测试集上的结果进行还不错的拟合，但也存在一定区别：

1. 仅使用 Z 进行训练时预测值要比实际值普遍高出一部分；而使用所有指标进行训练时预测曲线没有前者那么平滑；
2. 在大部分值上所有指标的拟合程度更好，但是 Z 对于一些特殊点的预测比使用所有指标要好。

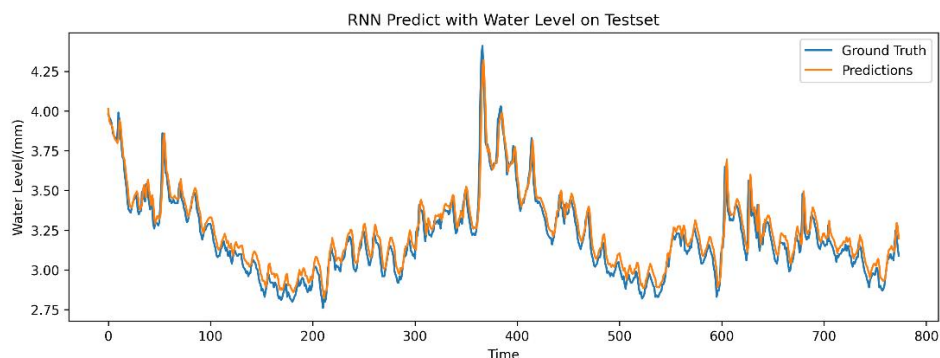


图 3.2 使用 Z 的预测结果

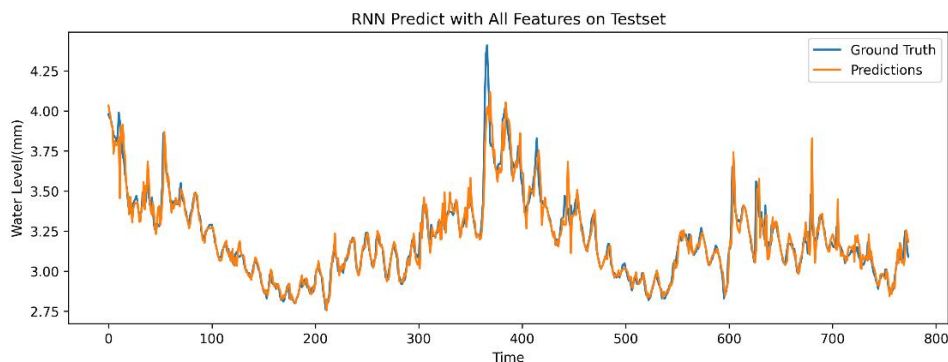


图 3.3 使用所有指标的预测结果

为了进一步验证我们的结论，我们计算了在测试集上预测的 MAE 和 MRE，结果显示如下：

表 3.1 使用两种特征训练后模型在训练集上的表现

Features	Z	ALL
MAE	0.0587	0.0434
MRE	0.0183	0.0135

通过计算 MAE 和 MRE 发现二者预测表现相当，考虑到使用 Z 更简单，所以后续的实验都只使用 Z 作为特征训练预测模型。

3.3 实验二：水文预测中异常值的特性探究

由于各种模型对一些异常值的预测效果欠佳，我们考虑对**测试集上预测效果欠佳的部分数据与其在同一时间范围内的观测数据进行绘图**，在这里我们作如下定义：

- 1) 训练集中 MRE 在 5~10%以上的预测点称为“欠优点”；
- 2) 同一时间范围我们取前后 3 天内的数据。

如图 3.4 所示，我们展示了一些欠优点前后三天（包含当天）水位高度的变化，其中标号为 3 所示的柱形表示对当天预测的 MRE，折线图代表当日的水位高度。我们可以观察到如下几个显著的特点：

- 1) “欠优点”靠近或就是近几日水位高度最高点；
- 2) “欠优点”前后预测的效果并不佳；
- 3) 即使前后预测效果都不错，也可能出现“欠优点”，且此时对“欠优点”的预测效果更加不准确。

因此我们可以从以下几个角度来寻找“欠优点”：

- 1) 从数值上寻找，“欠优点”就是显著异于周围水位高度值的点；
- 2) 从预测效果上看，“欠优点”就是效果不佳的点。

虽然“欠优点”并不代表该点的值就对应的时水文预测中的异常现象，但是可以反映出一部分的异常情况，并且可以在一定程度上反映这些点对于某一分布的偏移情况，**为了问题的简化，我们可以考虑通过寻找“欠优点”来代表“异常点”。**

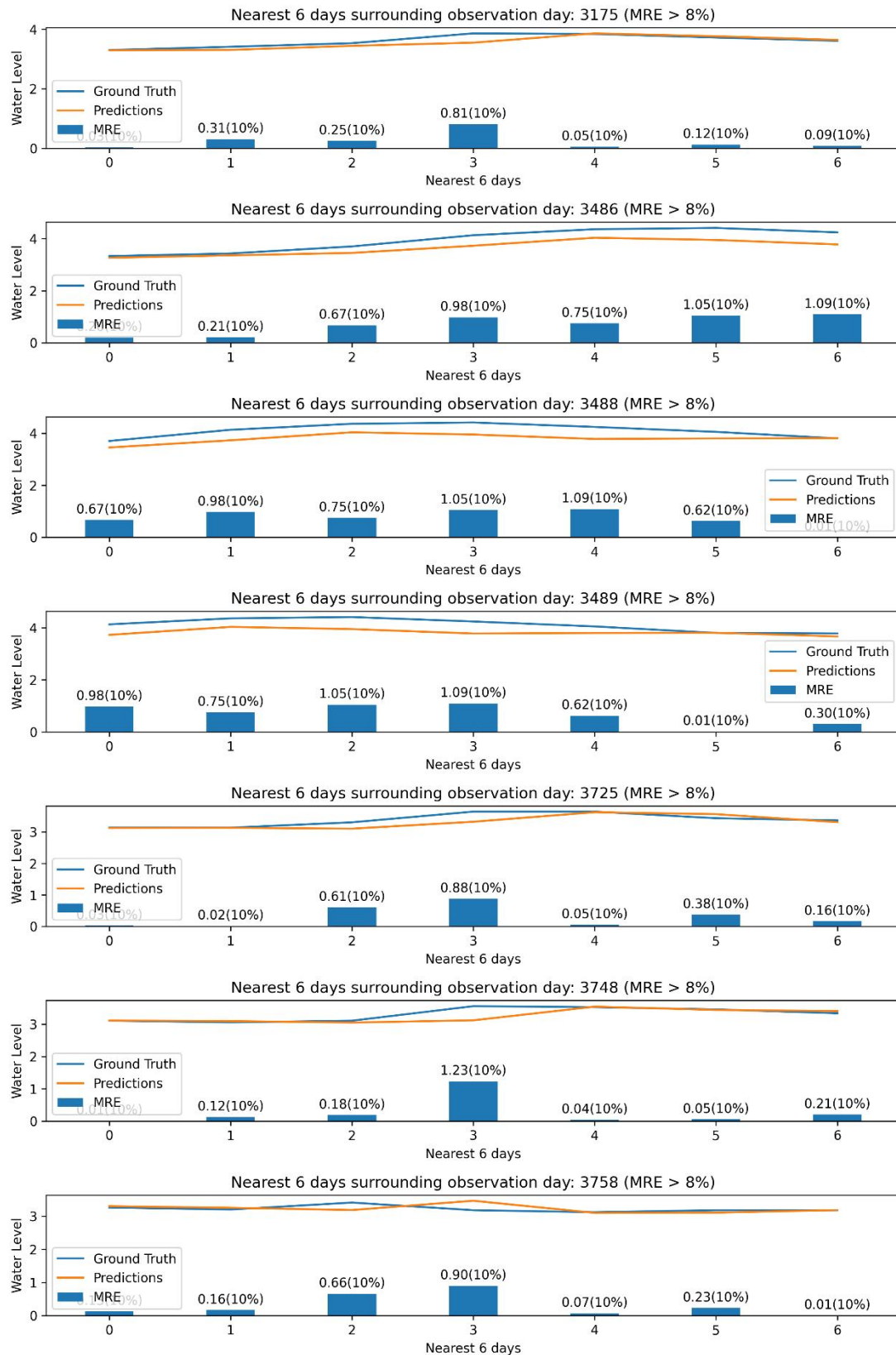


图 3.4 测试集上预测不佳日周边值的分布

3.4 实验三：训练修正模型的数据选择

如果要使用修正模型对一些“异常点”进行预测值的修正，我们就得从训练集上选取部分反映异常现象的“异常点”作为修正模型的训练数据，从而能够使模型对各种预测情况具有良好的泛化能力，并进行准确预测。

为了进行对照，我们首先使用训练集上所有的数据作为训练数据来训练修正模型，同时为了反映我们之前提到的“欠优点”的显著特征，我们使用我们选取的数据以及其前 3 天的数据作为特征一同作为训练数据参与训练，以反映“欠优点”同其周围点的联系。

3.3.1 修正模型训练算法

算法 2：修正模型训练算法

Input: 预测模型 $Model_p$; 预测模型选取的观测周期 n_p .

训练集 $Train_p = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$. $\mathbf{x}_i \in R^d$ (\mathbf{x}_i^0 equals y_i)

修正模型选取的观测周期 n_l .

判定准则 $criterion$.

Output: 修正模型 $Model_m$

1: 修正模型特征 $U_F = \emptyset$, 修正模型标签 $U_L = \emptyset$, 训练集上的预测值 $Y = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$

2: **for** $i = n_l$ **to** n :

3: $U_L = U_L \cup \{\mathbf{x}_i^0\}$, $U_F = U_F \cup \{\mathbf{x}_{i-n_l}, \dots, \mathbf{x}_i\}$

4: $U_F \leftarrow U_F[n_p - n_l:n]$, $U_L \leftarrow U_L[n_p - n_l:n]$ //选取 U_F, U_L 中第 $n_p - n_l$ 到第 n 个元素组成新的集合:

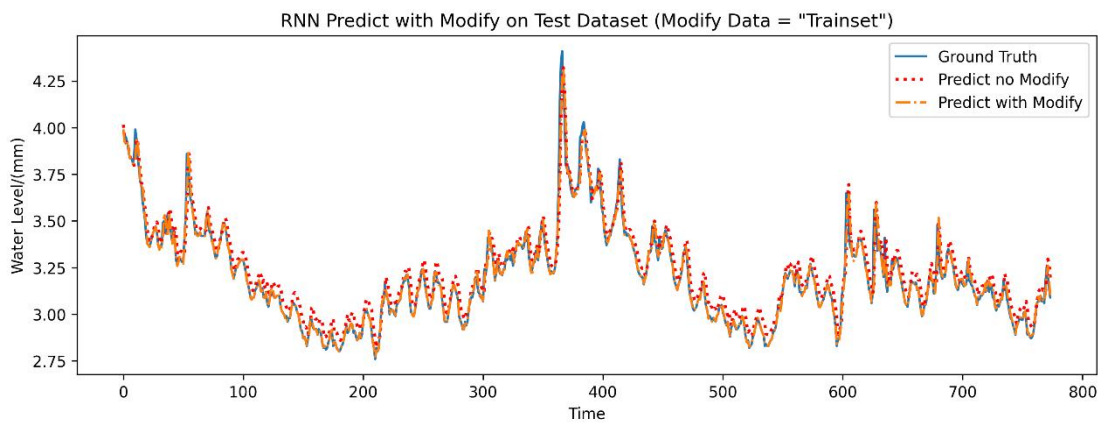
5: $Y_{diff} \leftarrow U_L[n_p - n_l:n] - Y$ //计算差分标签

6: $U_F \leftarrow criterion(U_F)$, $U_L \leftarrow criterion(U_L)$ //使用判定准则选取进入模型训练的特征和标签

7: 使用 U_F, U_L 训练修正模型 $Model_m$

8: **return** $Model_m$

3.3.2 对照试验:使用全部训练集数据训练修正模型



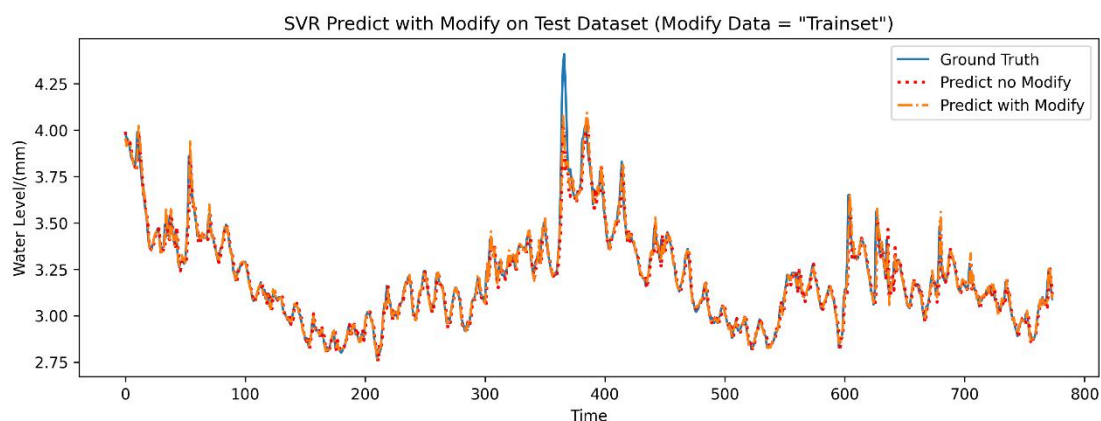


图 3.5 使用整个训练集作为训练数据的修正模型修正结果

观察结果我们可以发现，无论预测模型采用 RNN 或者是 SVR，当我们使用整个训练集作为训练数据来训练修正模型时，修正的效果很小，虽然修正后的曲线较修正前整体上升了一部分，但与原来差别不大，因此**我们应该利用一些准则来确定一些反映异常情况的、有代表性的数据参与我们修正模型的训练。**

3.3.3 使用 3-sigma 原则确定异常值

我们计算训练集上水位高度 Z 的均值 μ_Z 和标准差 σ_Z ，凡是位于区间 $[\mu_Z - 3\sigma_Z, \mu_Z + 3\sigma_Z]$ 外的数据点，我们均认为是异常值，利用这些异常值来训练修正模型，修正后测试集上预测的结果如下：

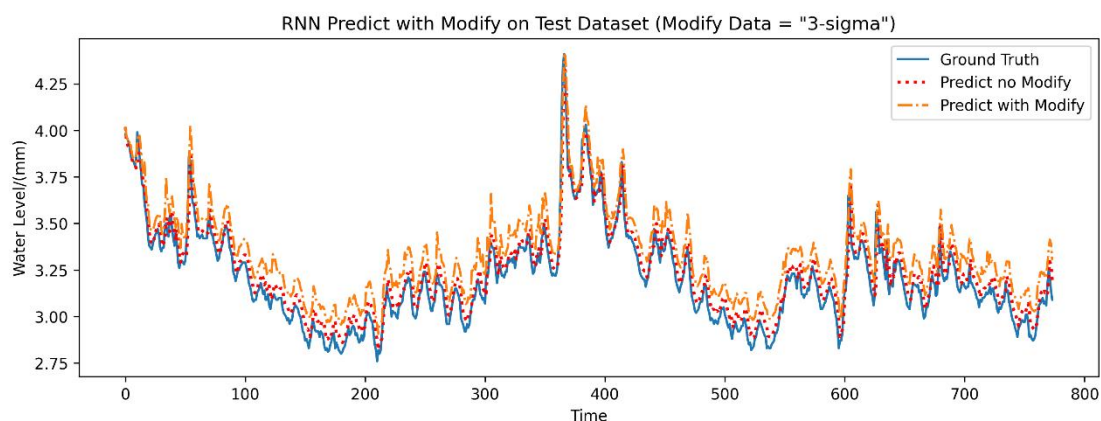


图 3.6 使用 3sigma 原则确定训练数据的修正模型修正结果

比较修正前后与实际观测到的水位数据曲线，发现对于部分“异常值”经过修正后得到了很好的预测，最显著的就是对 Time=380 左右峰值的预测，但是我们也发现修正后使得对于一些原来预测良好的值造成了偏移。

3.3.4 使用 bootstrap 得到置信区间确定异常值

通过多次的重采样并训练我们可以得到多次的训练结果，将这些训练结果样本构造统计量进行分析从而得到置信区间，落在置信区间外的值即为异常值。确定置信区间我们又有两

种方法，一种是基于正态分布，利用标准正态分布分位值 Z_p 进行计算。

当使用第一种方法时，我们计算多次预测结果在每个时间点上预测的均值 \bar{X} 和标准差 std ，给定置信度 p 求得标准正态分布下的分位值 Z_p ，置信区间上下界 L, U 通过下式给出：

$$\begin{cases} U = \bar{X} + \text{std} \times Z_p \\ L = \bar{X} - \text{std} \times Z_p \end{cases}$$

如果使用第二种方法，则只需给定上分位点百分比 u 和下分位点百分比 l 选取每个时间点对应的上下分位点，置信区间上下界 L, U 通过下式给出：

$$\begin{cases} U = X_u, X_u \text{ 是预测结果 } X \text{ 的上 } u\% \text{ 分位点} \\ L = X_l, X_l \text{ 是预测结果 } X \text{ 的下 } l\% \text{ 分位点} \end{cases}$$

比较使用上述两种方法训练修正模型的效果如下：

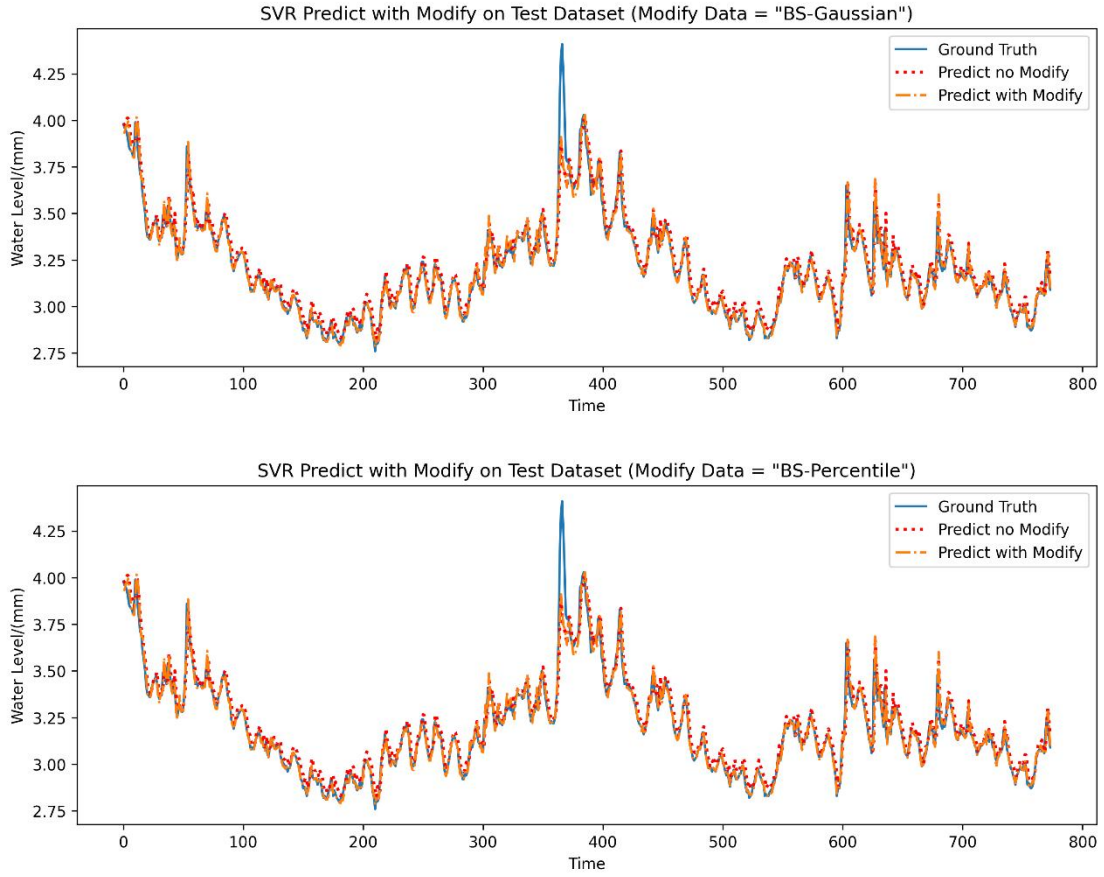


图 3.7 使用 bootstrap 两种方式得到置信区间确定训练数据的修正模型修正结果

两种方式的修正效果相差不大，但相比于修正前的情况在一些局部点上预测效果更好，经过修正后大部分值都比较贴近真实曲线，即使对一些“极端值”的修正效果不好（如 $\text{Time}=380$ ）。

进一步观察通过两种方式在训练集上确定的异常值分布情况如下：其中灰色部分为置信区间，绿色曲线为实际值，黄色点则代表了训练集上确定的异常点，两种方式确定的异常点分布相差不大，所以训练出的修正模型效果也较为接近。

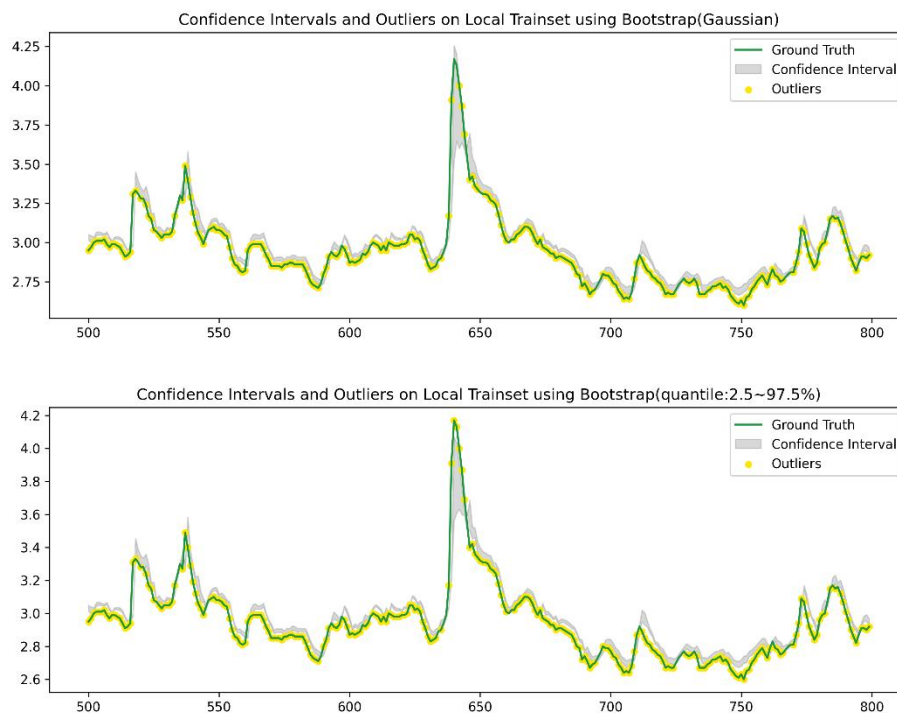


图 3.8 使用 bootstrap 两种方式在训练集上确定异常点的分布情况

3.3.5 使用训练集上预测不佳的数据作为异常值

我们还可以直接根据我们的预测结果，认为预测不佳的值即为我们的异常值，选取训练集上 $MRE > 0.05$ 的数据作为异常值参与训练，结果如下：

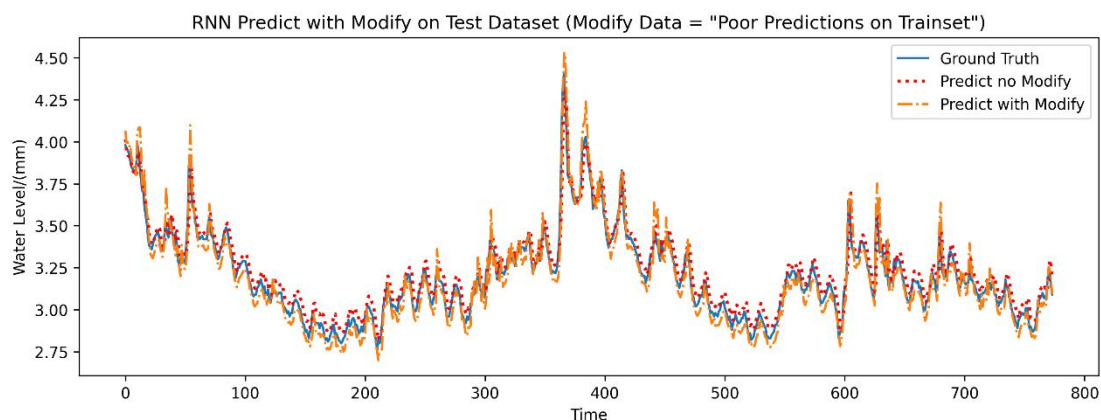


图 3.9 使用预测不佳数据确定训练数据的修正模型修正结果

但是这种做法的实践意义并不在于选取异常值，而是可以方便的在测试集上对我们想要修正的值进行“定位”，当我们认为测试集上预测不佳的点即为异常点时，此时不需要构造任何的统计量便可以确定异常值，并且本身这种做法也符合我们的目的：修正那些预测不佳的点。

进一步，我们可以使用这些点作为我们评价各种准则训练的修正模型的数据依据。

3.5 实验四：不同数据选择方式的训练的修正模型性能评估

我们对修正模型的性能从以下几个维度进行评估：

1. 训练集上确定的异常值个数 N ：

该指标代表了选取参与修正模型训练的数据个数，当 $N = n$ 时说明使用了全部的训练集参与训练，当 N 越大时越能反映训练集总体的分布情况，而越小越能反映局部的分布情况，如果我们异常值准则选取的数据能尽量反映“异常值”的分布情况，那么修正模型也能对异常值进行良好的修正。

2. 平均绝对误差 MAE，平均相对误差 MRE：

二者定义如下：

$$\text{MAE} = \frac{\sum_i |\text{pred}_i - \text{label}_i|}{N}, \text{MRE} = \frac{\sum_i |\text{pred}_i - \text{label}_i|}{\sum_i |\text{label}_i|}$$

其中 pred_i 为对某时刻下的修正值， label_i 为某时刻下的真实值， N 为确定的异常值个数。这两个指标能够反映我们修正后的值与真实值之间的偏差，但是由于我们目的也并不是对训练集上所有数据进行修正，所以这两个值并不代表我们的修正优劣，而是在某种意义上代表我们修正模型的修正程度。

3. 相对平均相对误差 RMRE：

由于MAE,MRE 并不能反映我们在想要进行修正数据上修正效果的优劣，于是我们引入了 RMRE，具体而言我们计算修正前后的预测值对应的平均相对误差 MRE_1, MRE_2 ，则 RMRE 定义如下：

$$RMRE = \frac{\sum_i |MRE_1^i - MRE_2^i|}{\sum_i |MRE_1^i|}$$

我们这里的MRE 选取的是 $MRE_j^i > 0.05$ 的点对应的MRE，它代表我们想要在训练集上修正的数据。

注：

- 1.以下的结果均为预测模型和修正模型都采用 SVR 的情况；
- 2.其中训练模型上我们想要修正的数据点序列为：[53 301 363 364 365 366 367 368 602 603 626 635 636 679]，我们的 RMRE 就是在这些点上的 MRE 基础上计算得到的，而 MRE 和 MAE 则是由整个测试集上的点决定的；
- 3.由于准则 3：使用训练集上预测不佳的数据作为异常值，作为我们实践时使用的实用准则，视具体情况而定而不具有普遍性，所以我们这里不予进行比较。

表 3.2 模型性能评估

Criterion	No Criterion	训练集	3σ准则 p=0.05,Zp=1.96	基于 Bootstrap 的准则	
				Zp p=0.05,Zp=1.96	上下分位点 low=2.5%,up=97.5%
N	3115	3115	56	2583	2627
MAE	0.0343	0.0217	0.0839	0.0230	0.0229
MRE	0.0107	0.0068	0.0262	0.0072	0.0072
RMRE	0	0.2306	0.3781	0.2243	0.2293

从上述表中我们可以看出：

- 1.3σ 准则确定的是数值上的离群点，因此 N 值较小，而基于 Bootstrap 的准则通过构造置信区间提供了更细粒度的筛选方法，于是筛选出了更多的“异常值”，因此 N 较大。
- 2.相对于我们的对照试验计算的 MAE 和 MRE，使用不同准则计算出的值更大，这反映了我们使用不同准则训练出的修正模型均具有一定的修正效果。
- 3.使用 Bootstrap 的准则时，无论使用哪种方式 RMSE 相对于对照试验都有下降，说明修正效果更好，而使用3σ 原则不降反升，说明对并不服从正态分布的数据使用该原则是武断的，应使用更加符合实际数据分布的准则来训练修正模型。