

基于K-means聚类方法和相似时段的特高压并联电抗器油温预测

谭风雷¹, 陈昊¹, 何嘉弘²

(1. 国网江苏省电力有限公司检修分公司, 江苏 南京 211102; 2. 东南大学 电气工程学院, 江苏 南京 210096)

摘要:为及时评估特高压并联电抗器内部的绝缘状态,提出了一种特高压并联电抗器油温预测方法。首先基于特高压并联电抗器油温曲线的“单峰单谷”变化趋势,利用K-means聚类方法将油温曲线划分为低谷时段和高峰时段2个部分。然后分析了特高压并联电抗器油温的主要影响因素,并根据K-means聚类和时间“近大远小”原则,基于气象、油温和时间3个因素对油温历史样本进行3次分类后,选择得到了相似时段。最后详细给出了基于相似时段的油温预测方法的计算流程。将所提方法应用到江苏地区某特高压并联电抗器油温的预测工作中,结果表明所提方法的平均误差为1.03%,最大误差为4.84%,具有较高预测精度,从而验证了方法的有效性。

关键词:特高压并联电抗器;油温预测;单峰单谷;K-means聚类;相似时段

中图分类号:TM 472

文献标志码:A

DOI:10.16081/j.epae.202103001

0 引言

随着我国经济的快速增长,特高压交流输变电工程得以全面建设,变压器、电抗器、断路器等核心电力设备^[1-4]在特高压电网得到了大量运用,提高此类设备的可靠性对保证国内经济发展具有重要意义。目前,特高压并联电抗器^[5-8]广泛应用在1000 kV特高压输电线路中,其主要作用有:降低特高压输电线路空载或轻载时的容升效应,抑制过电压;优化特高压线路电压与无功分布;通过与中性点电抗器的配合,降低潜供电流,提高特高压线路瞬时故障的重合闸成功率。由于容量大、电压高,特高压并联电抗器内部发热严重,容易引起绝缘损坏,目前直接判断特高压并联电抗器内部绝缘状态较为困难,可借助绕组温度(后文简称为绕温)和顶层油温(后文简称为油温)进行间接判断^[9],若能提前预测绕温和油温,即可及时评估特高压并联电抗器的绝缘状态,提高其运行可靠性和安全性。

目前,已有针对特高压并联电抗器绕温和油温的研究,主要包括热点温升计算、绕温异常分析等。文献[10]计算了特高压干式平波电抗器的温度场,得到了温度分布,并通过温升试验验证了计算结果

的正确性。文献[11]以某特高压并联电抗器绕温异常为例,在充分研究异常原因的基础上,通过试验确定了异常原因并提出了解决措施。文献[12]采用场-路耦合的电磁学方法,建立了高压电抗器三维温升模型,结果可为电抗器温度的在线监测提供技术支持。但现有的研究仅限于高压电抗器内部温度的实时计算,国内外尚未有专家学者开展高压电抗器绕温和油温预测的研究,但鉴于预测工作的必要性,尤其是特高压并联电抗器一旦发生故障,电力系统将损失大量负荷,不利于电力系统的稳定性^[13],通过预测特高压并联电抗器的绕温和油温有助于及时评估其内部绝缘状态,对提高电力系统稳定性具有实际意义。

由于目前尚未有文献开展特高压并联电抗器油温的预测研究,本文没有可借鉴的思路。鉴于电力负荷预测技术^[14-15]已相对成熟,且两者同属于电力预测领域,本文考虑在充分研究特高压并联电抗器油温变化特性的基础上,借鉴电力负荷预测领域的相关思路进行特高压并联电抗器油温预测:首先基于油温变化曲线研究了油温特性及其主要影响因素;然后基于气象、油温和时间这3个因素对油温历史样本进行了3次分类,得到了待预测时段的相似时段;最后合理选择了预测方法,并通过预测江苏地区某特高压并联电抗器的油温验证了方法的有效性。

1 特高压并联电抗器油温特性

1.1 油温曲线时段划分

本文分析油温特性时,以江苏地区某特高压并联电抗器2019年夏季70 d(6月17日—8月25日)的油温数据作为研究对象,根据这70 d的油温数据计算每个时刻油温的平均值,绘制油温日平均变化

收稿日期:2020-04-18;修回日期:2021-01-04

基金项目:国家自然科学基金资助项目(51807028);江苏省自然科学基金资助项目(BK20170672);国网江苏省电力有限公司科技项目(J2018014)

Project supported by the National Natural Science Foundation of China(51807028), the Natural Science Foundation of Jiangsu Province(BK20170672) and the Science and Technology Project of State Grid Jiangsu Electric Power Co., Ltd.(J2018014)

曲线如图1所示。

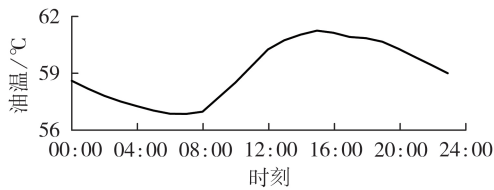


图1 油温日平均变化曲线

Fig.1 Daily average change curve of oil temperature

分析图1可知,油温随着时间的变化而变化,呈现明显的“单峰单谷”现象,即油温在清晨时达到谷值,在下午时达到峰值。根据油温“单峰单谷”的变化趋势,本文基于K-means聚类方法^[16-19]将油温曲线划分为低谷时段和高峰时段2个部分,具体步骤如下。

(1)对江苏地区某特高压并联电抗器2019年夏季的油温进行预处理,并设置时段数为2。

(2)根据图1,选择油温日平均曲线对应的谷值时刻和峰值时刻作为初始时段中心,并根据式(1)计算各时刻油温与初始时段中心的距离。

$$E_{ijl} = \sum_{k=1}^{k_{\max}} (M_{lk} - C_{ijk})^2 \quad (1)$$

其中, $k_{\max}$ 为样本数量; l 表示时刻, l 取值为0、1、2、…、23时分别对应00:00、01:00、02:00、…、23:00; C_{ijk} 为第 k 个样本本日经第 j 次迭代后第 i 个时段中心的油温值; M_{lk} 为第 k 个样本本日第 l 个时刻的油温值; E_{ijl} 为经第 j 次迭代后第 l 个时刻的油温与第 i 个时段中心的距离。

(3)将24个时刻按照最小距离原则分配给2个时段,即:

$$\begin{cases} \text{第 } l \text{ 个时刻属于第 1 个时段} & E_{1jl} \leq E_{2jl} \\ \text{第 } l \text{ 个时刻属于第 2 个时段} & E_{1jl} > E_{2jl} \end{cases} \quad (2)$$

(4)根据时段划分结果构建新的时段中心,如式(3)所示。

$$C_{i(j+1)k} = \frac{1}{N_{ij}} \sum_{l_0=1}^{N_{ij}} M_{f_{ij}(l_0)k} \quad (3)$$

其中, N_{ij} 为经第 j 次迭代后第 i 个时段的时刻数量; $f_{ij}(l_0)$ 为经第 j 次迭代后第 i 个时段第 l_0 个时刻对应的实际时刻值。

(5)根据式(4)建立目标评价函数 G_j 。

$$G_j = \sum_{i=1}^2 \sum_{k=1}^{k_{\max}} (M_{lk} - C_{ijk})^2 \quad (4)$$

若目标评价函数取得最小值,则迭代完成,此时的分类结果即为最优分类,油温峰值所在时段为高峰时段,油温谷值所在时段为低谷时段,否则返回步骤(2)继续迭代。

基于上述步骤,将江苏地区某特高压并联电抗器2019年夏季的油温曲线每日24个时刻划分为高

峰时段和低谷时段2个部分,其中,高峰时段为11:00—23:00,在15:00时油温达到最大值;低谷时段为00:00—10:00,在07:00时油温达到最小值。

1.2 油温影响因素分析

根据特高压并联电抗器的实际运行情况,其油温主要受到气象、设备状态和无功变化3个因素的影响。

(1)气象因素。本文结合气象的定义,重点研究温度、湿度、气压、风速、光照、降雨这6种气象因素。考虑到不同因素对油温的影响不同,本文采用相关系数衡量气象因素与油温之间的线性相关性,计算表达式为:

$$R_{xy} = \frac{\sum_{i=1}^{N_x} \left(x_i - \frac{1}{N_x} \sum_{i=1}^{N_x} x_i \right) \left(y_i - \frac{1}{N_y} \sum_{i=1}^{N_y} y_i \right)}{\sqrt{\sum_{i=1}^{N_x} \left(x_i - \frac{1}{N_x} \sum_{i=1}^{N_x} x_i \right)^2} \sqrt{\sum_{i=1}^{N_y} \left(y_i - \frac{1}{N_y} \sum_{i=1}^{N_y} y_i \right)^2}} \quad (5)$$

其中, R_{xy} 为自变量与因变量的相关系数; x_i 为自变量; N_x 为自变量样本数量; y_i 为因变量; N_y 为因变量样本数量。本文将6种气象因素作为自变量,油温作为因变量,由式(5)可得各气象因素与油温的相关系数见表1。

表1 气象因素与油温的相关系数

Table 1 Correlation coefficients between meteorological factors and oil temperature

气象因素	相关系数	气象因素	相关系数
温度	0.7918	风速	-0.3493
湿度	-0.0075	光照	0.1128
气压	-0.0017	降雨	-0.1735

根据相关系数的定义^[20],当相关系数的绝对值小于0.1时,表示自变量、因变量之间几乎没有相关性;否则表示自变量、因变量之间存在相关性。分析表1可知,油温与温度、光照呈正相关关系,与湿度、气压、风速和降雨呈负相关关系,同时油温与温度的相关性较高,与湿度、气压的相关性较低。故本文只研究温度、风速、光照和降雨这4种气象因素。

(2)设备状态因素。不同工况条件(如带故障运行与健康运行、夏季运行与冬季运行)下,油温变化较大。设备状态难以量化,考虑到短周期内可以认为设备状态保持不变,即设备状态可以通过时间来衡量,故本文将对设备状态的研究转化为对时间因素的研究。

(3)无功变化因素。考虑到特高压并联电抗器的无功主要受电压和频率影响。其中,特高压电网的频率波动一般控制在 ± 0.2 Hz范围内,即波动范围为 $\pm 4\%$;电压控制在1020~1070 kV范围内,正常波动一般控制在 ± 10 kV,即波动范围为 $\pm 1\%$;无功波动

较小,一般控制在 $\pm 2\%$ 以内,故无功变化不是油温的主要影响因素。为了进行进一步的分析,在特高压并联电抗器额定容量为240 Mvar、额定电压为1 000 kV,且处于额定运行状态的条件下,绘制油温与无功的散点图如图2所示。分析图2可知,油温变化与无功分布没有显著的相关性。另外,计算可得油温与无功的相关系数为0.085,根据相关系数定义,可认为无功变化与油温无相关性,故本文将忽略无功变化这一影响因素。

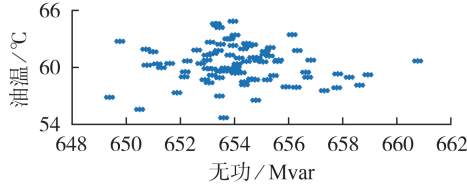


图2 油温与无功的散点图

Fig.2 Scatter plot of oil temperature and reactive power

2 特高压并联电抗器油温预测

2.1 油温归一化处理

根据前文分析可知,特高压并联电抗器的油温曲线分为低谷时段和高峰时段2个部分,考虑到2个部分的变化趋势不同,为了提高预测精度,本文对2个部分分别进行讨论。另外,本文在进行油温数据预测前,对油温进行归一化处理,具体的处理步骤如下。

(1)按照式(6)计算低谷和高峰时段的油温平均值作为各时段的基准油温,基准油温曲线如图3所示。

$$\begin{cases} B_1 = \frac{1}{11k_{\max}} \sum_{l=0}^{10} \sum_{k=1}^{k_{\max}} M_{lk} \\ B_2 = \frac{1}{13k_{\max}} \sum_{l=11}^{23} \sum_{k=1}^{k_{\max}} M_{lk} \end{cases} \quad (6)$$

其中, B_1 为低谷时段的基准油温; B_2 为高峰时段的基准油温。

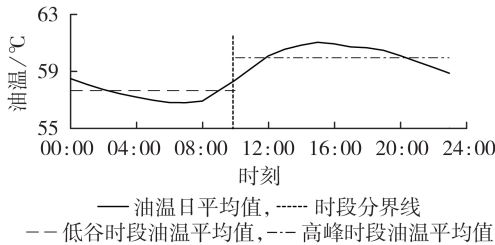


图3 基准油温曲线

Fig.3 Base oil temperature curve

(2)利用历史样本在各时刻的油温减去基准油温,得到新的油温 M'_{lk} 如式(7)所示。

$$M'_{lk} = \begin{cases} M_{lk} - B_1 & 0 \leq l \leq 10 \\ M_{lk} - B_2 & 11 \leq l \leq 23 \end{cases} \quad (7)$$

(3)按照式(8)对 M'_{lk} 进行归一化处理。

$$M''_{lk} = \frac{M'_{lk} - [\min(M'_{lk}) - C_0 [\max(M'_{lk}) - \min(M'_{lk})]]}{(1 + 2C_0) [\max(M'_{lk}) - \min(M'_{lk})]} \quad (8)$$

其中, M''_{lk} 为归一化油温; C_0 为归一化系数。

为简化描述,下文以高峰时段为例进行说明。

2.2 历史样本时段的3次分类

(1)基于气象因素的初分类。

对历史油温数据进行归一化处理后,本文基于气象因素,采用K-means聚类方法对归一化数据进行二分类,其计算过程与1.1节中对油温曲线的分段方法类似,区别在于步骤(2),即对油温曲线进行分段时,只考虑油温这1个因素,而基于气象因素进行二分类时,需考虑温度、风速、光照和降雨4种因素且各因素的权重不同,故本文提出了基于相关性系数优化的欧氏距离,则基于气象因素的K-means聚类方法的步骤(2)中,各样本日油温与初始中心的距离可表示为:

$$E_{ijk} = x_w \sum_{l=1}^L (W_{lk} - W_{ijl})^2 + x_F \sum_{l=1}^L (F_{lk} - F_{ijl})^2 + x_G \sum_{l=1}^L (G_{lk} - G_{ijl})^2 + x_J \sum_{l=1}^L (J_{lk} - J_{ijl})^2 \quad (9)$$

其中, L 为时段长度,可取值为11或13; x_w 为温度相关系数; W_{ijl} 为经第 j 次迭代后的第 i 个分类中心在第 l 时刻的温度值; W_{lk} 为第 k 个样本日第 l 时刻的温度值; x_F 为风速相关系数; F_{ijl} 为经第 j 次迭代后的第 i 个分类中心在第 l 时刻的风速值; F_{lk} 为第 k 个样本日第 l 时刻的风速值; x_G 为光照相关系数; G_{ijl} 为经第 j 次迭代后的第 i 个分类中心在第 l 时刻的光照值; G_{lk} 为第 k 个样本日第 l 时刻的光照值; J_{ijl} 为经第 j 次迭代后的第 i 个分类中心在第 l 时刻的降雨值; J_{lk} 为第 k 个样本日第 l 时刻的降雨值; x_J 为降雨相关系数。

进行基于气象因素的初分类后,历史样本被分为2类,其中待预测时段所在类即为相似类Con1,而另一类则为非相似类Con2。

对某待预测日高峰时段的相似类进行分析,结果如图4所示。显然相似类中大部分油温归一化曲线变化趋势相近,但也有少数油温曲线相差较大,若将该部分油温纳入预测算法中进行考虑,则势必会影响预测效果,因此本文考虑对基于气象因素进行

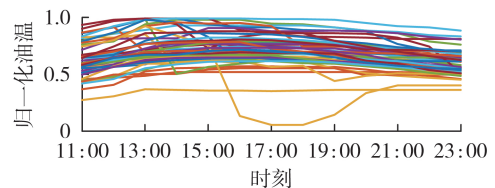


图4 高峰时段归一化油温曲线

Fig.4 Normalized oil temperature curves of peak period

初分类后得到的相似类 Con1 再进行二分类,将分类中对应样本时段数量较大的类称为相似时段类,而另一类则为非相似时段类。

(2) 基于油温特性的再分类。

为进一步选择相似时段,基于油温特性对相似类 Con1 进行再分类。考虑到油温特性主要包括油温大小和油温变化率2个部分,本文将从这2个方面对相似类 Con1 分别进行再分类。油温大小是指每个时刻对应的油温数据;油温变化率是指当前时刻油温相对于前一时刻油温的变化率,如式(10)所示。

$$R_{lk} = \frac{M_{lk} - M_{(l-1)k}}{M_{(l-1)k}} \quad (10)$$

其中, R_{lk} 为第 k 个样本在第 l 个时刻的油温变化率。

再分类过程仍然采用 K -means 聚类方法进行基于油温大小和油温变化率的二分类,具体过程与 1.1 节类似,区别仅在于步骤(2)中计算距离的对象不同,本文不再赘述。

基于油温大小对 Con1 进行再分类后,形成2个分类 Con1a1 和 Con1a2,两者对应的样本时段数量分别为 N_{a1} 、 N_{a2} ,具体说明如下:

a. 若 $N_{a1} > N_{a2}$,则认为 Con1a1 为基于油温大小的相似时段类 Con1a, N_{a1} 为基于油温大小的分类样本时段数 N_a , Con1a2 为基于油温大小的非相似时段类;

b. 若 $N_{a1} < N_{a2}$,则认为 Con1a2 为基于油温大小的相似时段类 Con1a, N_{a2} 为基于油温大小的分类样本时段数 N_a , Con1a1 为基于油温大小的非相似时段类;

c. 若 $N_{a1} = N_{a2}$,则认为 Con1a1+Con1a2 为基于油温大小的相似时段类 Con1a, $N_{a1}+N_{a2}$ 为基于油温大小的分类样本时段数 N_a ,没有基于油温大小的非相似时段类。

同理,基于油温变化率对 Con1 进行再分类后,形成2个分类 Con1b1 和 Con1b2,两者对应的样本时段数量分别为 N_{b1} 、 N_{b2} ,具体分析与基于油温大小的分类类似,本文不再赘述。

(3) 基于时间因素的末分类。

基于气象因素进行初分类和基于油温特性进行再分类后,得到了待预测时段的相似时段类 Con1a 和 Con1b,下面先对相似时段类 Con1a 和 Con1b 进行比较,得到基于时间因素末分类的研究对象后,基于时间因素进一步选择待预测时段的相似时段。根据时间因素“近大远小”原则可知,历史样本时段距离待预测时段越远(近),其对应的时间相关度越低(高),则基于时间因素对历史样本时段进行末分类的具体过程如下。

a. 当 $N_a > N_b$ 时,表示基于油温大小的再分类更有效,所以选择 Con1a 作为末分类的研究对象。当 $N_a \geq 20$ 时,选择与待预测时段时间距离最近的 20 个

样本时段作为基于油温大小的历史样本相似时段预测集 Conya; 当 $N_a < 20$ 时,选择全部样本时段作为基于油温大小的历史样本相似时段预测集 Conya。

b. 当 $N_a < N_b$ 时,表示基于油温变化率的再分类更有效,故选择 Con1b 作为末分类的研究对象。当 $N_b \geq 10$ 时,选择与待预测时段时间距离最近的 10 个样本时段作为基于油温变化率的历史样本相似时段预测集 Conyb; 当 $N_b < 10$ 时,选择全部样本时段作为基于油温变化率的历史样本相似时段预测集 Conyb。

c. 当 $N_a = N_b$ 时,表示基于油温大小与基于油温变化率的再分类效果相近,故选择 Con1a 和 Con1b 作为研究对象。针对 Con1a: 当 $N_a \geq 20$ 时,选择与待预测时段时间距离最近的 20 个样本时段作为基于油温大小的历史样本相似时段预测集 Conya; 当 $N_a < 20$ 时,选择全部样本时段作为基于油温大小的历史样本相似时段预测集 Conya; 针对 Con1b: 当 $N_b \geq 10$ 时,选择与待预测时段时间距离最近的 10 个样本时段作为基于油温变化率的历史样本相似时段预测集 Conyb; 当 $N_b < 10$ 时,选择全部样本时段作为基于油温变化率的历史样本相似时段预测集 Conyb。

根据历史样本时段的分类步骤,给出分类流程图如附录 A 图 A1 所示。

2.3 油温预测方法

本文油温预测方法的总体思路如下:

(1) 将待预测日划分为高峰和低谷 2 个时段分别进行研究;

(2) 进行基于气象因素的初分类,从气象因素方面选择待预测时段的相似时段;

(3) 在初分类的基础上,进行基于油温特性的再分类,再从油温特性方面选择待预测时段的相似时段;

(4) 在初分类和再分类的基础上,进行基于时间因素的末分类,从时间因素方面选择待预测时段的相似时段;

(5) 经过 3 次分类后,可得到待预测时段的最优相似时段,根据最优相似时段的分类结果和样本数量,结合相似日的思想,合理选择油温预测方法,即可实现油温预测。

考虑到本文在进行基于油温特性的初分类时,分别从油温大小和油温变化率 2 个方面进行了研究,故进行油温预测时,也应结合分类特点分别采用不同的方法,具体说明如下。

(1) 当 $N_a > N_b$ 时,采用 BP 神经网络。首先将 Conya 内历史样本时段对应的温度、风速、光照和降雨量作为神经网络的输入变量集,将 Conya 内历史样本时段对应的归一化油温作为神经网络的输出变量集;然后合理选择神经网络的结构并设置主要参数,训练 BP 神经网络;最后将待预测时段对应的温

度、风速、光照和降雨量作为训练神经网络的输入变量,此时神经网络的输出值 M'_{a_l} 的反归一化结果即为油温预测值 M_{a_l} ,反归一化表达式为:

$$M_{a_l} = M'_{a_l} \left(1 + 2C_0 \right) \left(\max(M'_{lk}) - \min(M'_{lk}) \right) + \left[\min(M'_{lk}) - C_0 \left(\max(M'_{lk}) - \min(M'_{lk}) \right) \right] + B_2$$

$$11 \leq l \leq 23, 1 \leq k \leq k_{\max} \quad (11)$$

(2)当 $N_a < N_b$ 时,采用油温趋势变化法。首先定义综合气象因素如式(12)所示。

$$Q_x(k, l) = \frac{1}{4} \left(\frac{W_{lk}}{W_{l0}} x_w + \frac{F_{lk}}{F_{l0}} x_F + \frac{G_{lk}}{G_{l0}} x_G + \frac{J_{lk}}{J_{l0}} x_J \right) \quad (12)$$

其中, $Q_x(k, l)$ 为历史样本第 k 个样本日第 l 个时刻相对于待预测时段第 l 个时刻的综合气象因素。

然后利用 Conyb 内历史样本时段计算油温变化率 S_{c_l} :

$$S_{c_l} = \frac{1}{N_b} \sum_{k=1}^{N_b} \frac{M_{lk} - M_{(l-1)k}}{M_{(l-1)k}} Q_x(k, l) \quad (13)$$

最后利用 Conyb 内历史样本时段计算归一化油温 M'_{b_l} :

$$M'_{b_l} = (1 + S_{c_l}) \frac{\sum_{k=1}^{N_b} M_{(l-1)k} Q_x(k, l)}{N_b} \quad (14)$$

对 M'_{b_l} 进行反归一化处理后即可得到油温预测值 M_{b_l} 。

(3)当 $N_a = N_b$ 时,采用BP神经网络和油温趋势变化法,该情况下的油温预测值即为 M_{a_l} 和 M_{b_l} 的平均值。

3 算例分析

本文以江苏地区某特高压并联电抗器(保定天威保变公司生产,型号为BKDF-240000/1000)2019年夏季6月17日—8月15日60d的油温作为历史样本,以2019年夏季8月16日—8月25日10d的油温(见附录B表B1)作为预测样本。特高压并联电抗器的后台监控系统一般采集2种油温,分别简称为油温1和油温2,由于篇幅限制,表B1仅给出油温1的数据。基于油温1、2验证本文方法的有效性。

3.1 本文方法的预测精度分析

利用本文方法对预测样本的油温进行预测,误差如图5所示。分析图5可知,对于油温1、2,本文方法的预测误差都能很好地控制在5%以内,具有较高的预测精度。

为进一步分析本文方法的精度,绘制油温预测曲线,如附录B图B1所示。可见,对于油温1、2,本文方法的预测效果最好,长短期记忆(LSTM)神经网络、差动整合移动平均自回归(ARIMA)模型和BP神

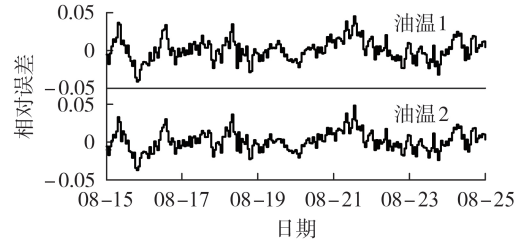


图5 油温预测误差

Fig.5 Forecasting error of oil temperature

经网络的预测效果次之,趋势变化法的预测效果相对较差。不同预测方法的预测结果对比如表2所示。根据表2可知,本文方法的平均预测误差为1.030%,小于LSTM神经网络的平均预测误差1.240%、ARIMA模型的预测误差1.630%、BP神经网络的平均预测误差1.680%和趋势变化法的平均预测误差1.930%;同时本文方法的标准差也相对较小,即预测误差波动较小,预测效果好,从而验证了本文方法的有效性与可行性。

表2 不同方法的预测结果对比

Table 2 Comparison of forecasting results among different methods

方法	时段	预测误差 / %			标准差
		最大值	最小值	平均值	
本文方法	低谷时段	3.810	0.009	1.000	0.0126
	高峰时段	4.840	0.010	1.040	0.0130
	所有时段	4.840	0.009	1.030	0.0131
趋势变化法	低谷时段	5.810	0.020	2.200	0.0181
	高峰时段	6.020	0.006	1.700	0.0199
	所有时段	6.020	0.006	1.930	0.0210
ARIMA 模型	低谷时段	5.620	0.027	1.520	0.0121
	高峰时段	5.350	0.019	1.720	0.0114
	所有时段	5.620	0.019	1.630	0.0135
BP 神经网络	低谷时段	4.660	0.003	1.530	0.0193
	高峰时段	4.470	0.003	1.810	0.0165
	所有时段	4.660	0.003	1.680	0.0197
LSTM 神经网络	低谷时段	3.770	0.027	1.040	0.0095
	高峰时段	5.160	0.011	1.400	0.0116
	所有时段	5.160	0.011	1.240	0.0125

为进一步验证本文所提出方法的有效性,本文采用均方误差MSE(Mean Squared Error)、均方根误差RMSE(Root Mean Squared Error)、平均绝对误差MAE(Mean Absolute Error)和决定系数R2(R-Square)4个指标来评价预测效果。将这4个指标值分别表示为 λ_{MSE} 、 λ_{RMSE} 、 λ_{MAE} 、 λ_{R2} ,指标定义及计算公式见附录C。

根据附录C式(C1)计算各类预测方法的指标值,具体结果如表3所示。显然本文方法的MSE、RMSE和MAE最小,R2最大,表明本文方法具有较好的预测效果。

3.2 本文方法的鲁棒性验证

为验证本文方法的鲁棒性,对预测模型连续计

表3 模型评价指标

Table 3 Model evaluation indexes

方法	λ_{MSE}	λ_{RMSE}	λ_{MAE}	λ_{R2}
本文方法	0.548 2	0.740 4	0.583 3	0.846 4
趋势变化法	2.911 5	1.706 3	1.220 1	0.184 3
ARIMA模型	1.632 9	1.277 9	0.983 0	0.542 5
BP神经网络	1.633 3	1.278 0	0.984 1	0.542 1
LSTM神经网络	0.808 1	0.898 9	0.677 4	0.773 6

算了10次,误差如表4所示。由表可知,平均预测误差的平均值、最小值、最大值分别为1.05%、1.03%、1.10%,最大预测误差控制在4.82%~4.98%范围内,最小误差控制在0.009%~0.011%范围内,上述误差随着运算次数增加而产生的波动都较小,从而表明本文方法具有较强的鲁棒性。

表4 不同运算次数下本文方法的预测误差

Table 4 Forecasting error of proposed method under different calculation times

运算次数	平均误差 / %	最大误差 / %	最小误差 / %
第1次	1.03	4.84	0.009
第2次	1.03	4.82	0.010
第3次	1.10	4.95	0.010
第4次	1.03	4.90	0.010
第5次	1.03	4.85	0.009
第6次	1.10	4.98	0.011
第7次	1.03	4.85	0.009
第8次	1.03	4.91	0.009
第9次	1.05	4.91	0.011
第10次	1.07	4.92	0.011

4 结论

(1)考虑到特高压并联电抗器油温曲线的“单峰单谷”变化趋势,基于K-means聚类方法将油温曲线分为低谷、高峰2个时段,其中低谷时段为00:00—10:00,高峰时段为11:00—23:00。

(2)对特高压并联电抗器油温的影响因素进行分析,得出气象和设备状态是主要影响因素,其中气象因素主要包括温度、风速、光照和降雨,设备状态因素通过时间因素进行衡量。

(3)对历史油温数据进行归一化处理后,利用K-means聚类方法和时间“近大远小”原则,基于气象因素、油温特性和时间因素对历史油温数据进行了3次分类,并根据分类结果对特高压并联电抗器的油温进行预测。

(4)本文以江苏地区某特高压并联电抗器的油温历史数据为例,分析比较了各类预测方法的MSE、RMSE、MAE和R2,结果表明本文方法具有较高的预测精度和有效性;此外针对预测模型连续计算了10次,结果表明随着运算次数的增加,本文方法的预测效果波动较小,从而验证了本文方法的鲁棒性。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 黄新波,张龙,朱永灿,等. 基于功角特性的干式空心电抗器匝间绝缘在线监测技术[J]. 电力自动化设备,2019,39(2):143-148.
HUANG Xinbo, ZHANG Long, ZHU Yongcan, et al. Online inter-turn insulation monitoring technology based on power angle characteristics for dry-type air-core reactor[J]. Electric Power Automation Equipment, 2019, 39(2): 143-148.
- [2] 亓孝武,李可军,于小晏,等. 基于核极限学习机和Bootstrap方法的变压器顶层油温区间预测[J]. 中国电机工程学报,2017,37(19):5821-5828.
QI Xiaowu, LI Kejun, YU Xiaoyan, et al. Transformer top oil temperature interval prediction based on kernel extreme learning machine and Bootstrap method[J]. Proceedings of the CSEE, 2017, 37(19): 5821-5828.
- [3] 李琳,谢裕清,刘刚,等. 油浸式电力变压器饼式绕组温升的影响因素分析[J]. 电力自动化设备,2016,36(12):83-88.
LI Lin, XIE Yuqing, LIU Gang, et al. Influencing factor analysis for disc-type winding temperature rise of oil-immersed power transformer[J]. Electric Power Automation Equipment, 2016, 36(12): 83-88.
- [4] 李可军,亓孝武,魏本刚,等. 基于核极限学习机误差预测修正的变压器顶层油温预测[J]. 高电压技术,2017,43(12):4045-4053.
LI Kejun, QI Xiaowu, WEI Bengang, et al. Prediction of transformer top oil temperature based on kernel extreme learning machine error prediction and correction[J]. High Voltage Engineering, 2017, 43(12): 4045-4053.
- [5] 张琦雪,黄海晏,陈祥,等. 静态阻塞滤波器的电抗器匝间故障及其保护的分析和[J]. 电力自动化设备,2017,37(5):74-79.
ZHANG Qixue, HUANG Haiyan, CHEN Xiang, et al. Analysis of SBF reactor inter-turn fault and corresponding protection[J]. Electric Power Automation Equipment, 2017, 37(5): 74-79.
- [6] 郑涛,赵彦杰. 超/特高压可控并联电抗器关键技术综述[J]. 电力系统自动化,2014,28(7):127-135.
ZHENG Tao, ZHAO Yanjie. Overview of key techniques of EHV/UHV controllable shunt reactor[J]. Automation of Electric Power Systems, 2014, 28(7): 127-135.
- [7] 徐征宇,李金忠,葛栋,等. 特高压并联电抗器高性能隔声装置研制[J]. 高电压技术,2018,44(7):2276-2283.
XU Zhengyu, LI Jinzhong, GE Dong, et al. Development of high performance sound insulation device for UHV shunt reactor[J]. High Voltage Engineering, 2018, 44(7): 2276-2283.
- [8] 杨志超,吴奕,王坚,等. 一种主变压器热点温度实时计算解析模型[J]. 电力自动化设备,2016,36(11):147-151.
YANG Zhichao, WU Yi, WANG Jian, et al. Analytical model for real-time calculating hot-spot temperature of main transformer[J]. Electric Power Automation Equipment, 2016, 36(11): 147-151.
- [9] 李可军,徐延顺,魏本刚. 基于PSO-HKELM的变压器顶层油温预测模型[J]. 高电压技术,2018,44(8):2501-2508.
LI Kejun, XU Yanshun, WEI Bengang, et al. Prediction model for top oil temperature of transformer based on hybrid kernel extreme learning machine trained and optimized by particle swarm optimization[J]. High Voltage Engineering, 2018, 44(8): 2501-2508.
- [10] 肖彤霞,李琳,纪锋,等. 特高压干式平波电抗器损耗与热点温升计算及试验研究[J]. 高压电器,2018,54(9):177-182,190.
XIAO Caixia, LI Lin, JI Feng, et al. Calculation and test of the loss and hot-point temperature rise of the UHV dry-type smoothing reactor[J]. High Voltage Apparatus, 2018, 54(9): 177-182, 190.

- [11] 刘杰,费韬,雷振洲,等.一起特高压交流变电站高压并联电抗绕组温度异常分析与处理[J].变压器,2017,54(7):65-68.
LIU Jie, FEI Tao, LEI Zhenzhou, et al. Analysis and treatment on abnormal winding temperature for UHV substation HV shunt reactor[J]. Transformer, 2017, 54(7): 65-68.
- [12] 吴书煜,马宏忠,魏旭,等.高压电抗器匝间短路三维模型计算与分析[J].电力自动化设备,2019,39(4):148-154.
WU Shuyu, MA Hongzhong, WEI Xu, et al. Calculation and analysis of three-dimensional model for turn-to-turn short circuit of high voltage reactor[J]. Electric Power Automation Equipment, 2019, 39(4): 148-154.
- [13] 李金忠,张丹丹,徐征宇,等.有限元法分析隔声装置对特高压并联电抗器散热性能的影响[J].高电压技术,2017,43(3):822-827.
LI Jinzhong, ZHANG Dandan, XU Zhengyu, et al. Heat dissipation performance of ultra-high voltage shunt reactor with sound insulation based on finite element method[J]. High Voltage Engineering, 2017, 43(3): 822-827.
- [14] 陈昊,万秋兰,王玉荣.基于厚尾均值广义自回归条件异方差族模型的短期风电功率预测[J].电工技术学报,2016,31(5):91-98.
CHEN Hao, WAN Qiulan, WANG Yurong. Short-term wind power forecast based on fat-tailed generalized autoregressive conditional heteroscedasticity-in-mean type models[J]. Transactions of China Electrotechnical Society, 2016, 31(5): 91-98.
- [15] 陈昊,高山,王玉荣,等.基于广义自回归条件异方差偏度峰度模型的风电功率预测方法[J].中国电机工程学报,2017,37(12):3456-3461.
CHEN Hao, GAO Shan, WANG Yurong, et al. Wind power forecasting method based on generalized autoregressive conditional heteroskedasticity with skewness and kurtosis model[J]. Proceeding of the CSEE, 2017, 37(12): 3456-3461.
- [16] 谭风雷,张军,马宏忠.基于趋势变化分段的电力负荷组合预测方法[J].华北电力大学学报(自然科学版),2020,47(2):17-24.
TAN Fenglei, ZHANG Jun, MA Hongzhong. Combined forecasting method of power load based on trend change division[J]. Journal of North China Electric Power University (Natural Science Edition), 2020, 47(2): 17-24.
- [17] 赵文清,龚亚强.基于Kernel K-means的负荷曲线聚类[J].电力自动化设备,2016,36(6):203-207.
ZHAO Wenqing, GONG Yaqiang. Load curve clustering based on Kernel K-means[J]. Electric Power Automation Equipment, 2016, 36(6): 203-207.
- [18] 徐青山,吕亚娟,孙虹,等.大用户多维度可中断特性精细化分析[J].电工技术学报,2020,35(S1):284-293.
XU Qingshan, LÜ Yajuan, SUN Hong, et al. Refined analysis of large-consumers' interruptible features from multi-dimension[J]. Transactions of China Electrotechnical Society, 2020, 35(S1): 284-293.
- [19] 程泽,刘冲,刘力.基于相似时刻的光伏出力概率分布估计方法[J].电网技术,2017,41(2):448-454.
CHENG Ze, LIU Chong, LIU Li. A method of probabilistic distribution estimation of PV generation based on similar time of day[J]. Power System Technology, 2017, 41(2): 448-454.
- [20] 胡琛,张竹,焦洋,等.基于随机矩阵理论的电子式互感器误差状态相关性分析方法[J].电力自动化设备,2018,38(9):45-53.
HU Chen, ZHANG Zhu, JIAO Yang, et al. Error state correlation analysis based on random matrix theory for electronic transformer[J]. Electric Power Automation Equipment, 2018, 38(9): 45-53.

作者简介:



谭风雷

谭风雷(1989—),男,重庆人,工程师,硕士,主要研究方向为电力电子技术在电力系统中的应用(E-mail: 220122094@seu.edu.cn);

陈昊(1980—),男,江苏南京人,研究员级高级工程师,博士,主要研究方向为电力系统保护与控制(E-mail: pingfengma@126.com);

何嘉弘(1988—),男,江苏南京人,讲师,博士,研究方向为高电压与绝缘技术(E-mail: hejiahong@seu.edu.cn)。

(编辑 任思思)

Oil temperature forecasting of UHV shunt reactor based on K-means clustering method and similar period

TAN Fenglei¹, CHEN Hao¹, HE Jiahong²

(1. Maintenance Branch Company of State Grid Jiangsu Electric Power Co., Ltd., Nanjing 211102, China;

2. School of Electrical Engineering, Southeast University, Nanjing 210096, China)

Abstract: A method of oil temperature forecasting of UHV (Ultra High Voltage) shunt reactor is proposed to evaluate the insulation state of UHV shunt reactor in time. Firstly, based on the "single peak single valley" changing trend of UHV shunt reactor's oil temperature curve, the oil temperature curve is divided into two parts of valley period and peak period by using K-means clustering method. Then, the main influencing factors of the oil temperature of UHV shunt reactor are analyzed. According to K-means clustering and the time principle of "near big and far small", the similar period is obtained by three classifications based on weather, oil temperature and time factors. Finally, the calculation process of the oil temperature forecasting method based on similar period is given in detail. The proposed method is applied to the oil temperature forecasting work of a UHV shunt reactor in Jiangsu area, the results show that the average error of the proposed method is 1.03% and the max error is 4.84%, which shows the high forecasting accuracy and effectiveness of the proposed method.

Key words: UHV shunt reactor; oil temperature forecasting; single peak single valley; K-means clustering; similar period

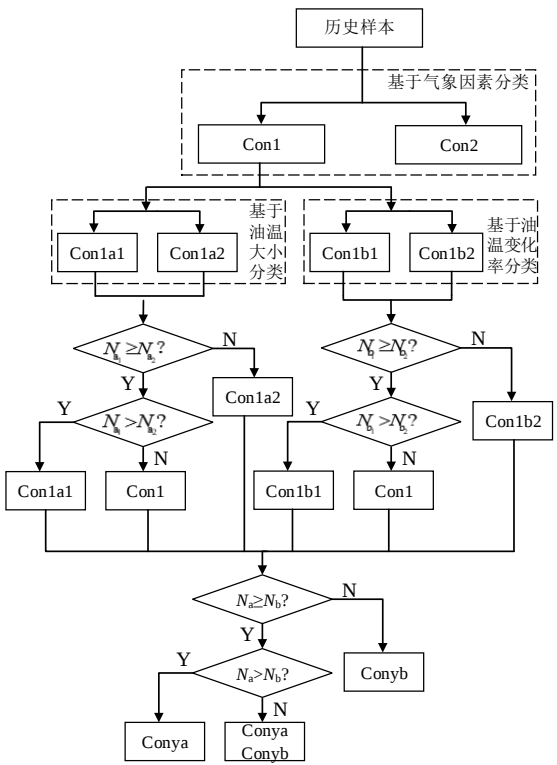


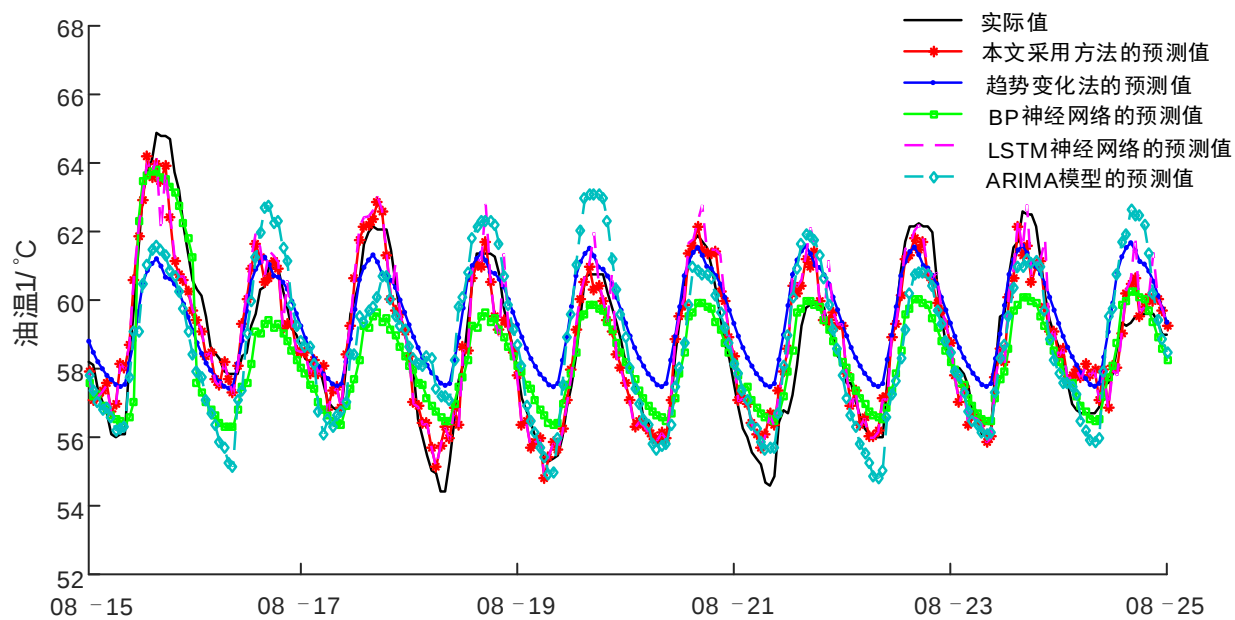
图 A1 历史样本时段分类流程

Fig.A1 Classification process of historical sample period

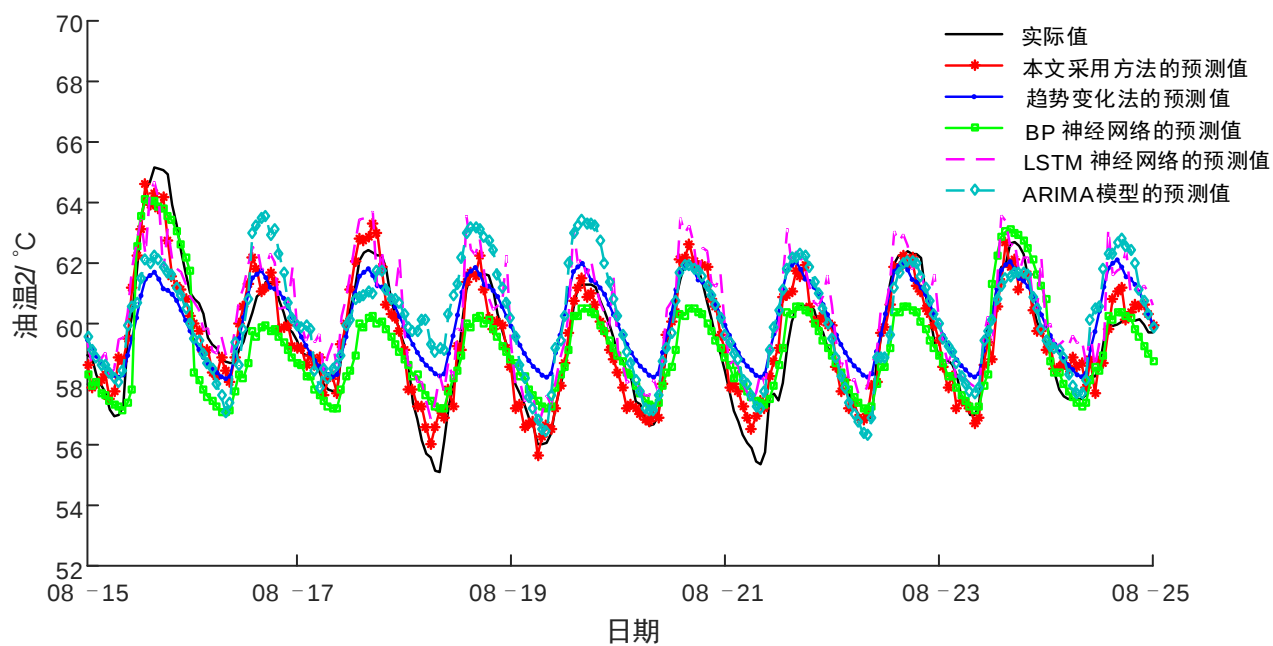
附录 B

表 B1 实际油温数据（8 月 16 日—8 月 25 日）
Table B1 Actual oil temperature (August 16—August 25)

时刻	油温/℃									
	8 月 16 日	8 月 17 日	8 月 18 日	8 月 19 日	8 月 20 日	8 月 21 日	8 月 22 日	8 月 23 日	8 月 24 日	8 月 25 日
00:00	58.2	60.3	58.3	57.9	57.1	58.0	56.9	57.6	58.0	58.3
01:00	58.1	60.1	58.2	56.5	56.8	57.6	56.3	57.6	58.0	57.6
02:00	57.4	59.5	58.2	56.0	56.4	57.3	56.0	57.1	57.9	57.0
03:00	57.1	58.6	58.1	55.5	56.0	56.5	55.6	56.6	57.5	56.8
04:00	56.7	58.5	57.6	55.0	55.9	56.4	55.4	56.3	56.7	56.7
05:00	56.1	58.3	57.1	55.0	55.3	56.0	55.2	56.1	56.4	56.7
06:00	56.0	57.9	56.9	54.4	55.3	55.8	54.7	56.0	56.1	56.7
07:00	56.1	57.9	56.8	54.4	55.4	55.9	54.6	56.0	56.1	56.7
08:00	56.1	57.9	56.8	55.3	55.5	55.9	54.9	56.0	56.2	56.9
09:00	58.2	57.9	58.0	56.4	56.0	56.9	56.4	57.6	58.0	57.6
10:00	59.8	58.4	59.6	57.7	56.9	58.4	56.8	58.9	59.5	57.6
11:00	61.5	58.5	60.0	58.1	57.5	59.4	56.7	59.7	59.6	57.9
12:00	63.1	59.3	60.6	59.3	58.4	60.4	57.2	60.8	60.8	58.6
13:00	63.7	59.6	61.6	60.5	59.5	61.1	58.1	61.7	61.5	59.3
14:00	64.1	60.3	62.0	61.4	60.7	61.7	59.3	62.2	61.7	59.3
15:00	64.9	60.4	62.2	61.4	60.8	61.9	59.8	62.2	62.6	59.3
16:00	64.8	61.0	62.1	61.4	60.8	61.7	59.9	62.2	62.5	59.5
17:00	64.8	60.9	62.1	61.4	60.8	61.5	59.9	62.2	62.5	59.6
18:00	64.7	60.9	62.1	61.3	60.8	61.0	59.9	62.2	62.2	59.6
19:00	63.7	60.3	61.5	60.8	60.2	60.3	59.8	62.0	61.2	59.6
20:00	63.2	59.8	60.5	59.8	59.8	59.5	59.1	60.6	60.1	59.6
21:00	62.2	59.2	59.6	59.2	59.3	58.7	58.9	59.5	59.4	59.3
22:00	61.4	59.0	59.0	58.6	58.8	58.3	58.4	58.7	59.0	59.0
23:00	60.6	58.6	58.6	58.0	58.2	57.6	58.2	58.2	58.6	59.0



(a) 油温 1



(b) 油温 2

图 B1 油温预测结果

Fig.B1 Forecasting result of oil temperature

附录 C

MSE 表示实际值与估算值之差平方的期望值，该参数可以评价数据的变化程度；**RMSE** 表示 **MSE** 的算术平方根；**MAE** 表示实际值与估算值绝对误差的平均值，该参数能够反映预测误差的实际情况；**R2** 可以衡量预测模型的拟合效果，取值范围在 0 到 1 之间。其中 **MSE**、**RMSE** 和 **MAE** 对应的值越小，表明预测效果越好，而 **R2** 对应的值越大，表明预测效果越好。**MSE**、**RMSE**、**MAE** 和 **R2** 的具体表达式如下：

$$\left\{ \begin{array}{l} \lambda_{\text{MSE}} = \frac{1}{N_n} \sum_{i=1}^{N_n} (z_i - \hat{z}_i)^2 \\ \lambda_{\text{RMSE}} = \sqrt{\frac{1}{N_n} \sum_{i=1}^{N_n} (z_i - \hat{z}_i)^2} \\ \lambda_{\text{MAE}} = \frac{1}{N_n} \sum_{i=1}^{N_n} |z_i - \hat{z}_i| \\ \lambda_{\text{R2}} = 1 - \frac{\sum_{i=1}^{N_n} (z_i - \hat{z}_i)^2}{\sum_{i=1}^{N_n} (z_i - \frac{1}{N_n} \sum_{i=1}^{N_n} z_i)^2} \end{array} \right. \quad (\text{C1})$$

其中， z_i 为第 i 个样本的实际值； $\hat{z}_i$ 为第 i 个样本的预测值； N_n 为样本数量。