



DOI:10.12404/j.issn.1671-1815.2308734

引用格式:赵勇,尹爱军,成浩,等.天然气区域生产中的关联工艺参数协同预测与预警[J].科学技术与工程,2025,25(2):560-566.

Zhao Yong, Yin Aijun, Cheng Hao, et al. Collaborative prediction and warning of associated process parameters in natural gas regional production[J]. Science Technology and Engineering, 2025, 25(2): 560-566.

天然气区域生产中的关联工艺参数协同预测与预警

赵勇¹, 尹爱军^{2*}, 成浩³, 李骞¹, 纪林呈¹, 吴倩颖¹

(1. 中国石油天然气股份有限公司西南油气田分公司重庆气矿, 重庆 400021; 2. 重庆大学机械与运载工程学院, 重庆 400044;
3. 广西大学土木建筑工程学院, 南宁 530004)

摘要 为提前发现天然气区域生产过程中管线、装置可能发生的超压、泄露等异常工况,现有工业控制报警系统不能准确反映设备真实状态,而采用单参数的预警误判率较高,实用性不足。提出了一种天然气区域生产各上、下游场站关联工艺参数的协同预测与预警方法。针对天然气区域内场站多、生产过程复杂、监测数据多样的特点,首先对每个场站的参数进行降维,提取各场站的关键工艺参数;再对关键参数进行关联度评估及分组,以同组的高关联参数为自变量,建立多元非线性套索回归预测模型;同时,建立关键参数的长短期记忆预测模型;结合两种模型预测结果的对比分析确定动态阈值,对区域生产进行协同预警。通过重庆气矿某井区开展的有效性验证,结果表明该方法不仅能有效减少单值异常出现的预警误判,又可提前定位异常场站和异常点位,具有较高的实用价值。

关键词 天然气区域生产;报警;工艺参数;关联度;协同预警

中图分类号 TE863 TP183; 文献标志码 A

Collaborative Prediction and Warning of Associated Process Parameters in Natural Gas Regional Production

ZHAO Yong¹, YIN Ai-jun^{2*}, CHENG Hao³, LI Qian¹, JI Lin-cheng¹, WU Qian-ying¹

(1. Chongqing Gas Mine of Southwest Oil and Gas Field Branch of Petrochina, Chongqing 400021, China;
2. College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China;
3. School of Civil Engineering and Architecture, Guangxi University, Nanning 530004, China)

[Abstract] In order to detect the abnormal working conditions such as overpressure and leakage, that may occur in pipelines and installations in the process of natural gas regional production, the current industrial control and alarm systems cannot accurately reflect the real state of the equipment, and the single-parameter early warning has a higher rate of error judgement, which is insufficient in practicality. A collaborative prediction and warning method for process parameters related to upstream and downstream stations in a natural gas production area was tested. Aiming at the characteristics of natural gas region with many stations, complex production process and diverse monitoring data, firstly, the parameters of each station were downgraded to extract the key process parameters of each station. Then, the key parameters are evaluated and grouped by correlation, and a multivariate nonlinear lasso regression prediction model was established with the highly correlated parameters in the same group as the independent variables. At the same time, a long and short-term memory prediction model was established for the key parameters, and a comparison analysis of the prediction results was performed to determine the dynamic prediction and early warning of natural gas production. Comparative analysis of the prediction results of the two models was used to determine the dynamic thresholds for coordinated early warning of regional production. The results show that the method can not only effectively reduce the misjudgment of single-value anomalies, but also locate the anomalous stations and points, which is of high practical value.

[Keywords] natural gas regional production; alarm; process parameters; correlation degree; collaborative early warning

天然气是一种清洁、高效、低碳的能源,是中国能源结构调整和优化的重要组成部分^[1]。天然气区域生产要经过采、输、增压、脱水等工艺流程,生

产场站分布较远,管线、装置受内外部因素的影响可能发生超压、泄露等异常情况^[2],需在短时间内进行应急联动处置,调整各关联生产单元,最大程

收稿日期:2023-11-08; 修订日期:2024-10-18

基金项目:国家自然科学基金(52275518)

第一作者:赵勇(1972—),男,汉族,四川绵阳人,软件设计师。研究方向:自动化及深度学习。E-mail:562427095@qq.com。

*通信作者:尹爱军(1978—),男,汉族,湖南邵阳人,博士,教授,博士研究生导师。研究方向:大数据分析,智能诊断与预测。E-mail:aijun.yin@cqu.edu.cn。

投稿网址:www.stae.com.cn

度地降低异常情况对整个区域生产系统的影响。

报警是通过组态软件远程读取现场可编程逻辑控制器(programmable logic controller, PLC)或远程终端单元(remote terminal unit, RTU)的压力、流量、液位、频率等实时数据,当数据超过人工设置的报警上、下限值时,触发报警。目前,天然气生产场站以单值报警为主,忽略了生产装备间的连续性和强关联性,易受设备自身断电、故障、校验、检修的影响。使得测量值大幅偏离正常范围,触发大量无效报警,淹没了有效报警^[3]。由于报警限值采用人工设置,对现场工况的动态变化存在观察不及时、修改滞后的现象。

针对这些问题,中外也提出了一些预测措施,如文献[4]基于GRU改进RNN神经网络的飞机燃油流量预测,门控循环单元(gated recurrent unit, GRU)、循环神经网络(recurrent neural network, RNN)虽对参数的短期预测具有一定的准确度,然而采用了多次连乘运算,不能有效记忆较长的历史信息,容易出现梯度消失等问题。文献[5]基于二次模态分解的LSTM短期电力负荷预测,长短期记忆模型(long short-term memory, LSTM)能有效地记忆较长的历史信息,预测准确度有了显著改善,但在短期预测中准确度仍存在不足。文献[6]综合岭回归和SARIMA方法在桥梁健康监测数据分析中的应用,季节性差分自回归移动平均模型(seasonal autoregressive integrated moving average, SARIMA)多元线性回归对短期预测具有较高的准确度,但对长期预测,会随着时间的增加,误差逐步增大。

综上所述,对时序非稳态数据的预测,仅采用LSTM的单参数预测,将存在单点故障风险、参数间的协同不够充分;而仅采用RNN或多元回归,存在梯度消失或梯度爆炸风险,对较长时间的预测准确度不足。

如何充分挖掘区域内的海量数据,找出影响区域运行的关键参数,并精确预测,提前预警,可结合多元回归以及LSTM预测的优点,开展关联数据的协同预测和深度区域协同预警(deep regional synergy for early warning, DRSEW)研究,对提前发现并定位异常,提升整区域的安全生产管控能力,具有重要现实意义。

1 协同预警基础理论

1.1 深度区域协同预警框架

天然气生产工艺流程复杂,在同区域的传输系统中各场站监测参数存在互相耦合,关联性强。深度区域协同预警流程如图1所示,其实现流程如下。

(1)采集生产区域内各场站的监测参数数据,对数据预处理,按时间对齐、量岗进行归一化操作。

(2)对每个场站采集的各类参数进行主成分分析(principal component analysis, PCA)降维,降低数据处理量、剔除高冗余数据,确定关键工艺参数。

(3)对各场站关键参数开展关联度评估,按关联度强弱进行分组,建立关联参数集。

(4)分别以各场站关键数据为自变量同时建立多元非线性套索回归预测模型和LSTM预测模型。

(5)对比两个预测模型预测结果,确定动态阈值,协同关联参数的总体变化趋势,进行预警判定。

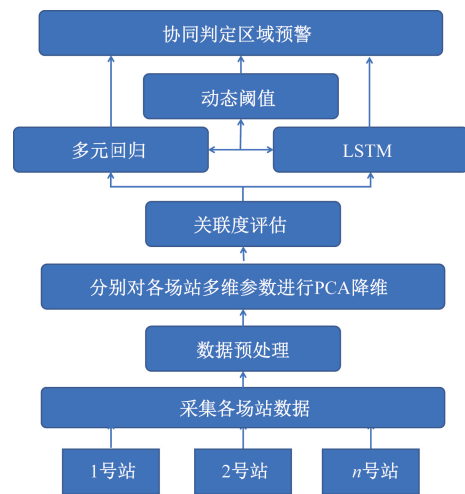


图1 深度区域协同预警流程

Fig. 1 The tree of deep regional synergy for early warning

1.2 参数降维与关联分析

通过PCA分析对采集的油压、套压、一级节流压力、差压、流量等多个参数矩阵降维。PCA通过线性变换将原始数据变换为一组各维度线性无关的表示,用于提取原始数据的主要特征。原始数据经过矩阵转换后,采用零均值化,其协方差矩阵及其特征值和特征向量,再将原始数据映射至特征向量空间,实现原始数据的降维。如有 n 行 m 列的原始数据矩阵 X ,经零均值化后的协方差矩阵可表示为

$$C = \frac{1}{m}XX^T \quad (1)$$

式(1)中: X 为原始矩阵; m 为矩阵的列数; C 为协方差矩阵,其特征值为 λ_i ,其中 $i = 1, 2, \dots, n$,特征向量为 C_i ,将 C_i 构成的矩阵与原始矩阵 X 相乘,就得到了降维后的主要数据,如图2所示。

降维后,选择对方差贡献最大的参数进行关联分析,灰色关联分析(grey relation analysis, GRA)是灰色系统理论的一个应用,关联度是系统之间的两个因素随着不同对象或时间而变化的关联性大小的度量。GRA分析是将因素之间发展趋势的相异

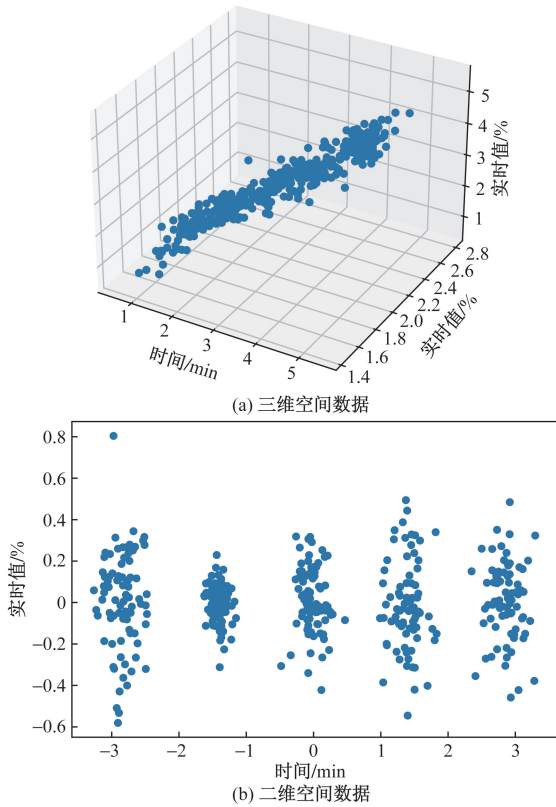


图2 PCA 3维降至2维
Fig. 2 PCA reduce 3D to 2D

或相似程度作为评价因素之间关联程度的一种方法^[9]。GRA的基本思想是将因素的数据序列作为依据,通过数学的方法来比较因素之间的几何对应关系。如果系统有 n 个变量,经归一化后的数据集表示为 $\mathbf{X} \in \mathbf{R}^{m \times n}$,母序列 X_0 与各子序列 $X_i (i = 1, 2, \dots, n-1)$ 之间第 k 时刻的关联系数 $\zeta_i(k)$ 可表示为

$$\zeta_i(k) = \frac{a + pb}{|X_0(k) - X_i(k)| + pb} \quad (2)$$

式(2)中: $a = \min_i \min_k |X_0(k) - X_i(k)|$,为两级最小差; $b = \max_i \max_k |X_0(k) - X_i(k)|$,为两级最大差; p 为分辨系数(一般取 0.5); $X_0(k)$ 为母序列集合中的第 k 个单元; $X_i(k)$ 为子序列集合中的第 k 个单元。

X_0 与 X_i 之间的灰色关联度 r_i 为

$$r_i = \frac{1}{m} \sum_{k=1}^m \zeta_i(k) \quad (3)$$

1.3 多元回归与 LSTM

多元线性回归采用了最小二乘法的思想,用线性拟合非线性,通过梯度下降逐步逼近最优解。但对波动较大的数据进行预测时,如自变量较少,导致维度不足,影响预测精度。而多元非线性回归增加了自变量的维度,不仅可提升自变量的变化幅

度,也考虑了自变量之间的相互制约关系,公式为

$$\hat{y} = \beta_0 + \sum_{i=1}^n \beta_i X_i^j X_{i+1}^{j+1}, \quad i = 1, 2, \dots, n; \quad j = 0, 1, 2, \dots, n \quad (4)$$

式(4)中: \hat{y} 为预测值; β_0, β_i 为回归系数; X_i^j 为自变量的 j 次方。

多元非线性回归易出现过拟合,变量组合较多,因此采用套索回归(lasso regression, LASSO)改善优化逻辑^[7],作为一种收缩估计方法,其基本思想是在回归系数的绝对值之和小于一个常数的约束条件下,使残差平方和最小化,从而能够产生某些严格等于 0 的回归系数,虽然牺牲了部分偏差值,但一定程度上避免了过拟合的问题,增加了预测的准确性^[8-9]。其参数估计被定义为

$$\beta_{\text{Lasso}} = \operatorname{argmin}^2 \left\| \mathbf{Y} - \sum_{j=1}^p \mathbf{X}_j \beta_j \right\|^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (5)$$

式(5)中: λ 为正则化非负参数; β_j 为回归系数; \mathbf{X}_j 为因变量矩阵, $\mathbf{X}_j = (X_{1j}, X_{2j}, \dots, X_{nj})^T$, 其中 $j = 1, 2, \dots, p$; $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$ 为自变量。

由于 L_1 正则化项采用的是绝对值之和,导致损失函数有不可导的点,采用坐标轴下降法得以解决^[10]。将多元非线性回归不同组合的自变量代入套索回归进行参数筛选,降低运算复杂度,可提升预测精度,避免过拟合。

LSTM 在 RNN 的基础上增加了 3 个门控,即遗忘门、输入门和输出门,控制细胞对当前值进行保留或删除操作,如图 3 所示。主要原理如下。

遗忘门为

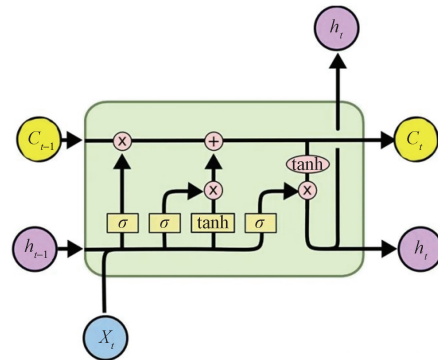
$$f_t = \sigma(W_f[h_t - 1, x_t] + b_f) \quad (6)$$

输入门为

$$i_t = \sigma(W_i[h_t - 1, x_t] + b_i) \quad (7)$$

输出门为

$$o_t = \sigma(W_o[h_t - 1, x_t] + b_o) \quad (8)$$



X_t 为输入数据; σ 为 Sigmoid 函数; Tanh 为双曲正切函数; C_{t-1} 为上一时段细胞值; C_t 为更新后的细胞值

图3 LSTM 结构

Fig. 3 LSTM structure

更新细胞为

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (9)$$

$$C_t = f_i C_{t-1} + i_t \tilde{C}_t \quad (10)$$

更新隐藏层为

$$h_t = o_t \tanh C_t \quad (11)$$

式中: σ 为 Sigmod 函数; W_f, W_i, W_o, W_c 为权重; x_t 为输入的数据; b_f, b_i, b_o, b_c 为截距。

LSTM 预测对长期的历史信息有记忆作用, 能实现较长时间的预测, 但短期预测相对较差。

1.4 动态阈值与协同预警

通过对比 LSTM 和多元回归两种模型的预测结果, 当两者预测值接近时, 可由预测均值确定阈值。由于, 时序函数满足正态分布, 其概率密度公式为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-u)^2}{2\sigma^2}\right] \quad (12)$$

式(12)中: u 为均值; σ 为标准差; σ^2 为方差。

根据概率分布, 接近 95.44% ~ 99.73% 的数据将分布在预测均值与 2 ~ 3 倍预测标准差的和与差之间, 可据此确定预警限值。

结合现场实际数据的历史变化趋势, 当 $f(t)$ 是关于时间 t 的函数, 同时满足狄利克雷条件, 可对该时序数列进行傅里叶变换, 公式为

$$F(\omega) = F[f(t)] = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \quad (13)$$

式(13)中: ω 为角频率; t 为时间。

可将多元非线性函数转换到复频域, 分析并计算出函数变化周期, 按周期分隔计算预警阈值; 也可通过 SARIMA 进行变换, 分析大致周期, 得到分段预警限值。

协同预警是基于 50% 关联参数的预测值均达到阈值触发区域预警; 当关联参数中只有一个参数达到阈值, 不触发区域预警; 当某一个参数的两种预测变化率出现不同步变动, 或某一个参数出现不同于其余关联参数的预测趋势, 可判断该场站或该参数出现异常。

区域协同预警可有效减少单个参数异常引发的预警误判, 可快速定位异常场站。

2 分析与验证

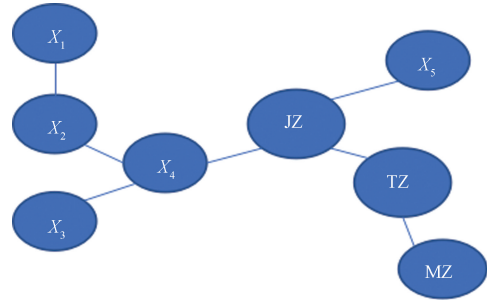
2.1 数据来源及主成分分析

为了验证 DRSEW 方法在真实作业数据中的有效性, 使用重庆气矿 2023 年 9 月 14 日—10 月 14 日黄 202 井区生产数据进行实验, 该井区共包含 8 个场站, 分布如图 4 所示。

原始数据共 120 维数据, 通过对各站的 PCA 分析, 取方差贡献较大的为关键参数, 总数据维度降

为 8 维, 参数名称及方差贡献如表 1 所示。

每个参数原始采集频率为 30 次/min, 考虑到各个场站采集轮询时间不尽相同, 而后续计算均需数据在时间上严格对齐, 为确保预测的准确性, 将各场站原始数据以 1 条/min 进行收缩对齐, 最终得到每个参数 24 d 的数据, 即 35 699 条, 各站关键压力数据如图 5 所示。



$X_1 \sim X_5$ 为采气站; JZ 为汇集站; TZ 为处理站; MZ 为终端站

图 4 黄 202 井区场站分布图

Fig. 4 Distribution map of H202 well area

表 1 特征参数

Table 1 Characteristic parameter

编号	关键参数	方差贡献	编号	关键参数	方差贡献
P1	X_4 出站压力	0.37	P5	JZ 进站压力 X_4	0.23
P2	X_1 出站压力	0.34	P6	JZ 进站压力 X_5	0.19
P3	X_3 出站压力	0.36	P7	TZ 出站压力	0.35
P4	JZ 出站压力	0.21	P8	TZ 进站压力	0.32

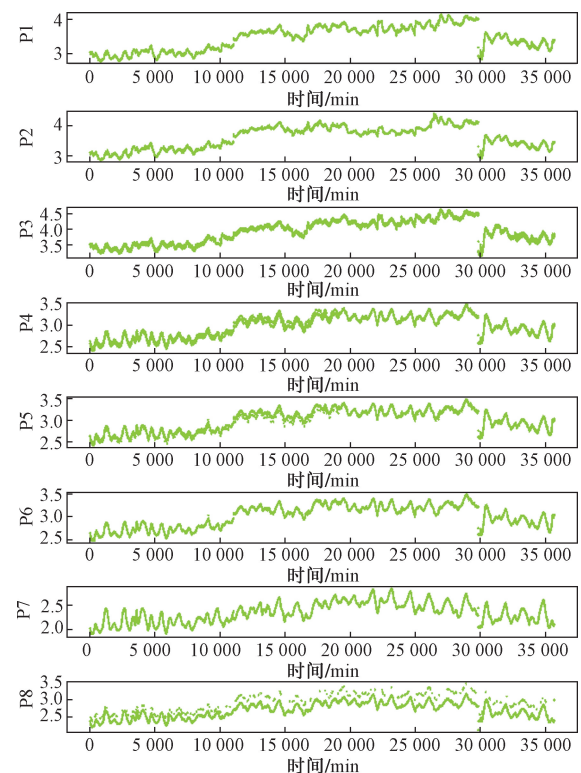


图 5 原始数据

Fig. 5 Raw data

2.2 关联分析

使用灰色关联分析计算各参数之间的关联度,结果如图6所示。可见,不同场站参数间关联度差异明显,P1、P2、P3 关联度超过 0.95,P4、P5、P6、P7、P8 之间关联度较大,其值均超过了 0.89,可以将参数分为两组。

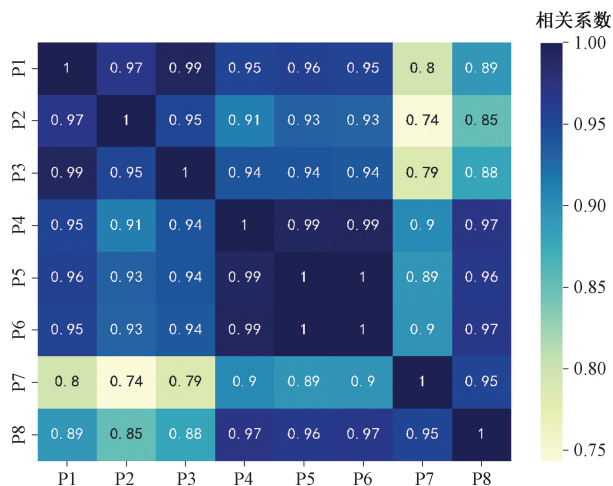


图6 灰色关联度

Fig. 6 Grey correlation degree

2.3 趋势预测

使用 Z-Score 的归一化方式对原始监测数据进行预处理,将数据集以 8:2 的比例划分为训练集、验证集,通过训练集完成模型超参数的训练,验证集用于统计实验指标。选择了目前实际应用较为广泛的预测方法,包括历史均值法 (historic average, HA)、核岭回归法 (kernel ridge, KR)^[11]、随机森林算法 (random forest, RF)^[12-13]、极端梯度提升算法 (gradient boosting, XGBoost)^[14-15],与模型进行对比验证。采用平均绝对误差 (mean absolute error, MAE) 与均方根误差 (root mean squared error, RMSE) 评价预测模型的准确性。

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{x}_i - x_i| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2} \quad (15)$$

式中: n 为测试样本总量; \hat{x}_i 为测试样本预测值; x_i 为观测值。

MAE 衡量了预测值与观测值之间的绝对误差, RMSE 表示预测值与观测值之间的偏差,这两个指标越小,模型拟合效果越好、精确度越高。

利用各参数过去的监测数据预测其此后 10 min 的情况,即时间窗口设为过去,预测窗口设为 10。LASSO 的正则化系数设为 $\lambda = 5 \times 10^{-5}$ 。为综合评价各预测方法的预测性能,统计在不同方法下各监

测参数预测评价指标的加权和,得到加权均方根误差 nMAE 以及加权平均绝对误差 nRMSE。表 2 展示了各预测方法在验证集上的评价指标。

表 2 不同方法预测性能评价

Table 2 Predictive performance evaluation of different approaches

方法	nMAE/%	nRMSE/%
HA	0.114	0.140
KernelRidge	0.092	0.114
RF	0.301	0.382
XGBoost	0.297	0.381
LASSO	0.079	0.102
DRSEW	0.066	0.087

可以看出,DRSEW 在 nMAE、nRMSE 两种加权评价指标中均取得了出色的表现。相较于对比方法中的最优结果,该模型的 nMAE 和 nRMSE 相比于 LASSO 分别降低了 16.45%、14.71%。由于作为天然气生产系统的监测参数表现出时序非平稳性,基于统计学的预测方法 HA、KernelRidge、LASSO 表现出不俗的效果;而基于集成学习树模型的 RF、XGBoost 预测效果较差,原因在于树模型更多适用于逻辑回归预测上,在时序自回归预测方面表现效果不佳^[16]。

为观测模型与真实数据的匹配情况,按照预测值与真实值的相对误差小于 0.05 计算符合率,超过 0.05 的按不符合计算,预测结果如图 7 所示。

可见,在经过 3 次迭代后,所有预测符合率均超过 99%,证明预测模型具有较高的拟合精度,与生产实时工况的变化吻合。

2.4 区域协同预警

通过对比,多元回归与 LSTM 预测值误差小于 0.05,经傅里叶与自回归分析,确定周期为 4 d,即 5 760 min 为一个周期,按周期计算动态阈值,如图 8 所示。

由图 8 可知,在第一个周期内第 2 520 点,即经历 2 520 min 后,第二组 P4~P8 压力均超出预警上限,在第 4 900 点,第一组 P1~P3 均超过上限;在第二个周期内第 11 150 点 P1~P8 所有压力点同时超出上限,由于现场工况发生变化,动态阈值触发有效预警;在第 4 个周期的第 23000 点,仅 P7 压力超出阈值,但其余压力均未达到阈值,未触发区域预警;同时发现 P8 压力点在多个周期内多次出现数据跳变,基于 DRSEW 的区域协同逻辑,判断 P8 压力数据为异常,应对该点数据监测模块及回路进行校验,以排除故障。

由于预测模型设置的预测窗口期为 10,即可提前 10 min 预警,修改模型预测窗口,可将预测时间提前到 30 min 以上。该方法对单个离群点噪声不敏感,可以有效减少区域预警误判。

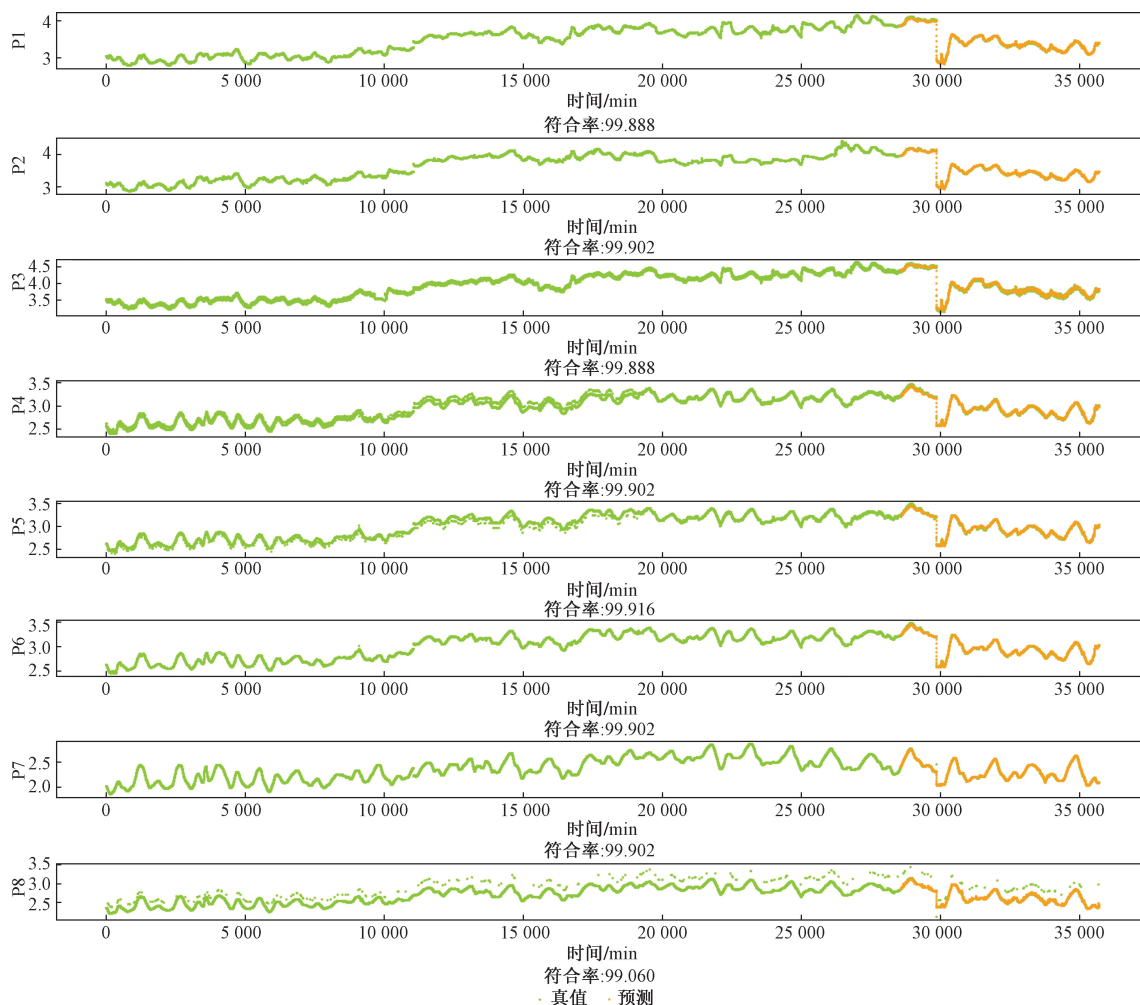


图7 预测效果

Fig. 7 Prediction effect

3 结论

在区域化生产中进行准确的预测与预警,得出如下结论。

(1)应同时建立单参数预测与多参数预测模型,对比预测结果,及时纠偏。如采用 LSTM 单参数和多元回归多参数两种预测模型,对比预测结果,当一个预测结果发现异常,可结合另一模型快速判定。

(2)对时序非稳态序列数据进行预警,应按周期进行动态阈值限定,由于工况变化的不确定性,固定阈值难以实现有效的预警。

(3)多参数协同预警能实现区域异常情况的快速发现,可有效降低异常响应时间,对区域化生产企业有着非常重要的理论意义和工程应用价值。

参 考 文 献

[1] 石云,潘继平,王恺,等.“双碳”背景下天然气与新能源融合

发展路径及策略[J]. 油气与新能源, 2023, 35(4): 1-6.

Shi Yun, Pan Jiping, Wang Kai, et al. The path and strategy for the integrated development of natural gas and new energy under the background of “dual carbon” [J]. Oil and Gas and New Energy, 2023, 35(4): 1-6.

[2] 温玉臣. 油田天然气集气与处理加工工艺的探讨[J]. 中国石油和化工标准与质量, 2023, 43(13): 157-159.

Wen Yuchen. Exploration of gas gathering and processing technology for oilfield natural gas [J]. Chinese Petroleum and Chemical Standards and Quality, 2023, 43(13): 157-159.

[3] 赵长江. 天然气输送联动联锁安全控制研究应用[J]. 石化技术, 2022, 29(8): 200-202.

Zhao Changjiang. Research and application of interlocking safety control in natural gas transportation [J]. Petrochemical Technology, 2022, 29(8): 200-202.

[4] 陈聪, 候磊, 李乐乐, 等. 基于 GRU 改进 RNN 神经网络的飞机燃油流量预测 [J]. 科学技术与工程, 2021, 21(27): 11663-11674.

Chen Cong, Hou Lei, Li Lele, et al. Research on aircraft fuel flow based on recurrent neural network [J]. Science Technology and Engineering, 2021, 21(27): 11663-11674.

[5] 张淑娴, 江文韬, 陈玉花, 等. 基于二次模态分解的 LSTM 短期电力负荷预测 [J]. 科学技术与工程, 2024, 24(7):

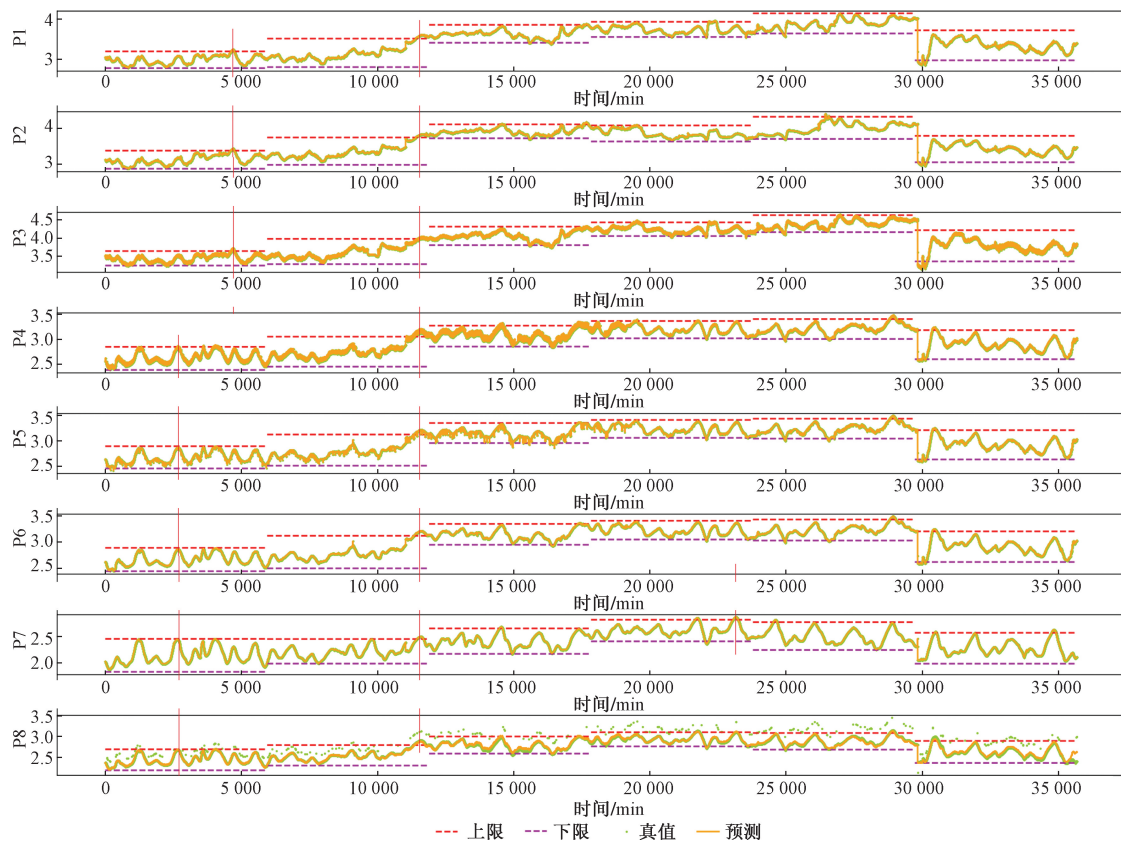


图8 协同预测和预警

Fig. 8 Collaborative prediction and early warning

2759-2767.

Zhang Shuxian, Jiang Wentao, Chen Yuhua, et al. Short-term power load forecasting based on quadratic mode decomposition through LSTM[J]. Science Technology and Engineering, 2024, 24(7): 2759-2767.

- [6] 谌楨文, 常军. 综合岭回归和 SARIMA 方法在桥梁健康监测数据分析中的应用[J]. 科学技术与工程, 2023, 23(20): 8846-8854.

Chen Zhenwen, Chang Jun. Application of integrated ridge regression and SARIMA methods in the analysis of bridge health monitoring data[J]. Science Technology and Engineering, 2023, 23(20): 8846-8854.

- [7] Lee J H, Shi Z, Gao Z. On LASSO for predictive regression[J]. Journal of Econometrics, 2022, 229(2): 322-349.

- [8] 胡记磊, 吴文良, 王璟. 基于自适应 LASSO 的逻辑回归砂土液化判别模型[J]. 三峡大学学报(自然科学版), 2023, 45(2): 67-72.

Hu Jilei, Wu Wenliang, Wang Jing. Logistic regression sand liquefaction discriminative model based on adaptive LASSO[J]. Journal of China Three Gorges University (Natural Science Edition), 2023, 45(2): 67-72.

- [9] 常明明. 三类群 Lasso 正则化回归模型及其在高维数据分析中的应用[D]. 新乡: 河南师范大学, 2020.

Chang Mingming. Lasso regularized regression model with three groups and its application in high-dimensional data analysis[D]. Xinxiang: Henan Normal University, 2020.

- [10] 邢耀. 胖(高)大数据的 LASSO Logistic 模型构建与应用[D]. 济南: 山东大学, 2022.

Xing Yao. Building and application of Lasso Logistic model of fat

(high) Big data[D]. Jinan: Shandong University, 2022.

- [11] Exterkate P, Groenen P J F, Heij C, et al. Nonlinear forecasting with many predictors using kernel ridge regression[J]. International Journal of Forecasting, 2016, 32(3): 736-753.

- [12] Massaoudi M, Chihi I, Sidhom L, et al. Enhanced random forest model for robust short-term photovoltaic power forecasting using weather measurements[J]. Energies, 2021, 14(13): 1-8.

- [13] 王健豪, 方庆, 罗霄, 等. 一个基于随机森林的微博转发预测算法[J]. 计算机科学, 2014, 34(4): 62-64.

Wang Jianhao, Fang Qing, Luo Xiao, et al. A Weibo forwarding prediction algorithm based on random forest[J]. Computer Science, 2014, 34(4): 62-64.

- [14] Mateo J, Rius-Peris J M, Marana-Perez A I, et al. Extreme gradient boosting machine learning method for predicting medical treatment in patients with acute bronchiolitis[J]. Biocybernetics and Biomedical Engineering, 2021, 41(2): 792-801.

- [15] 陈向俊, 傅军平, 陈栋栋, 等. 基于极端梯度提升回归模型的电梯钢丝绳磨损预测方法[J]. 机电工程, 2022, 39(4): 554-560.

Chen Xiangjun, Fu Junping, Chen Dongdong, et al. A prediction method for elevator steel wire rope wear based on extreme gradient lifting regression model[J]. Mechanical and Electrical Engineering, 2022, 39(4): 554-560.

- [16] 罗常伟, 王双双, 尹峻松, 等. 集成学习研究现状及展望[J]. 指挥与控制学报, 2023, 9(1): 1-8.

Luo Changwei, Wang Shuangshuang, Yin Junsong, et al. Current situation and prospect of Ensemble learning research[J]. Journal of Command and Control, 2023, 9(1): 1-8.