2025-06-13

工作内容：模型训练和测试
模型：SARIMIA+LGB
方法：SARIMIA预测结果作为LGB模型的输入
特征工程：
- 趋势特征
- 频域特征
结果：

- SARIMIA mse=35.66698
- SARIMIA + LGB (无特征) mse = 33.948
- SARIMIA + LGB（频谱能力 + 宽展） mse = 30.2246（目前最优👇）

其他模型测试：

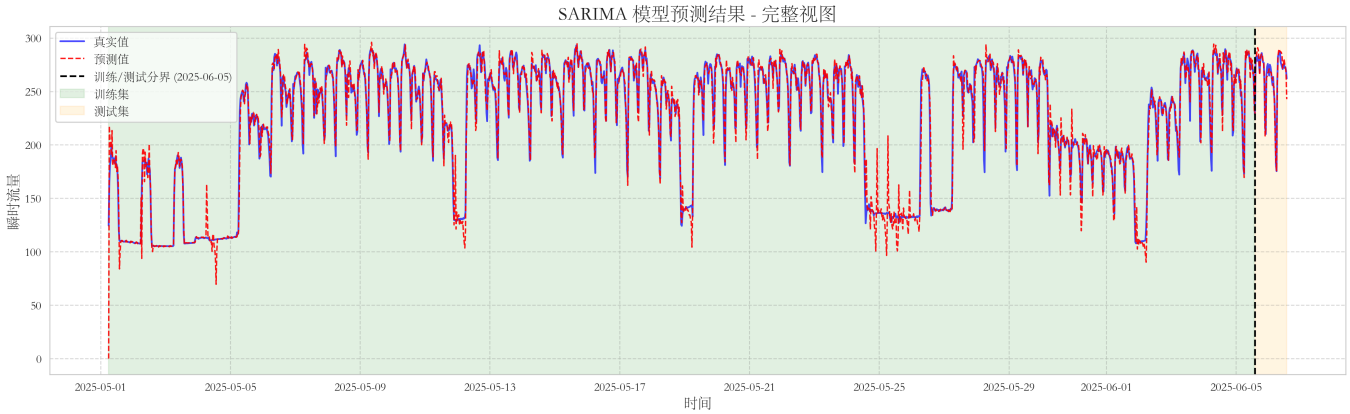
- 单LGB预测： MSE=-
- 单XGBoost预测：MSE=69679.1421
- Prohet预测：MSE=39.4408

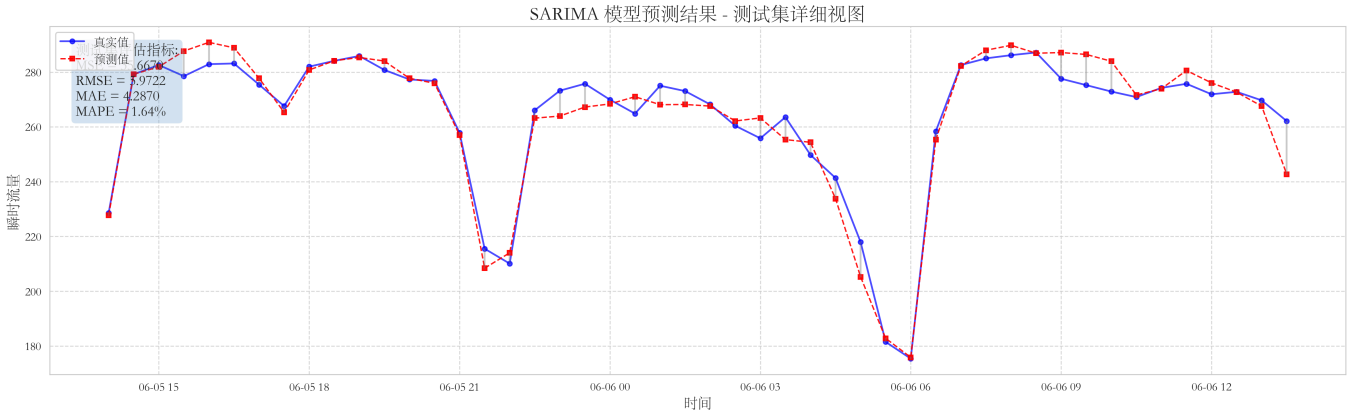
一阶段

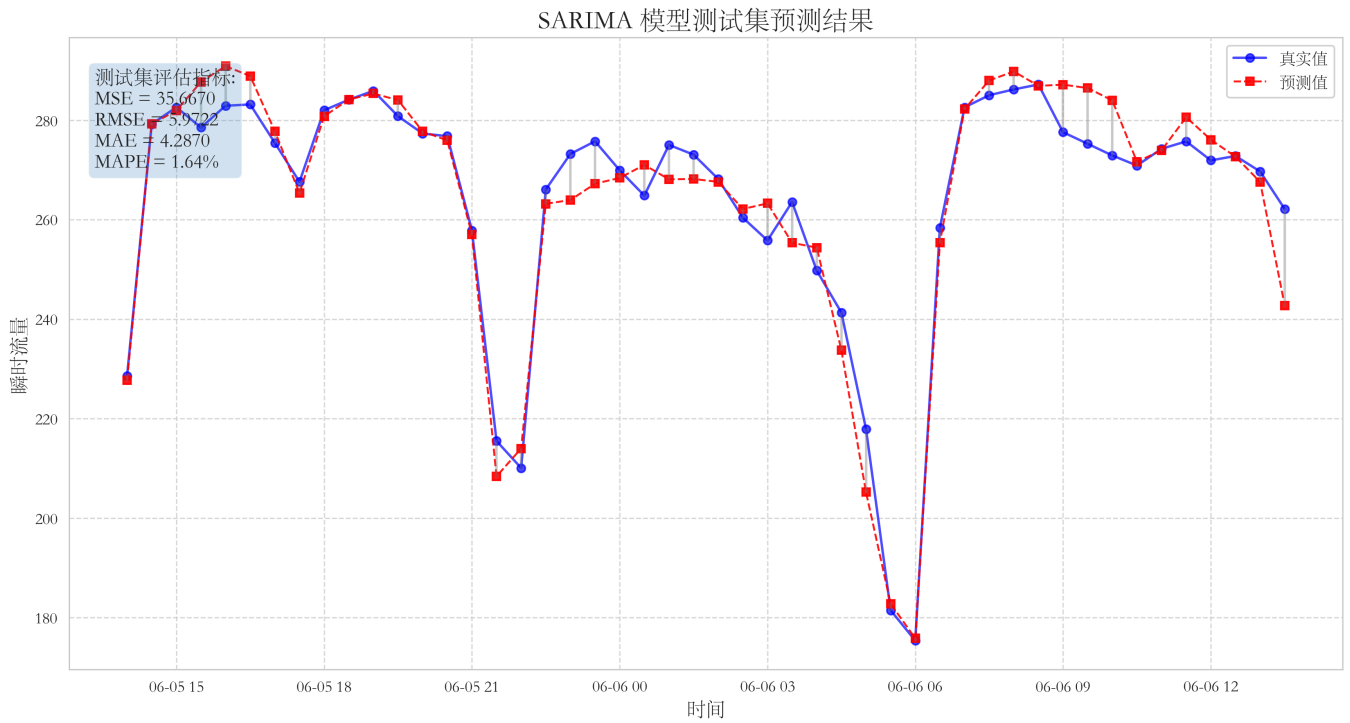
采样频率：30分钟
数据范围: 2025-05-01 06:00:00 到 2025-06-06 13:30:00
总计 1744 个30分钟间隔的数据点
每天有 48 个数据点
训练集大小: 1696 个点 (约 35.3 天)
测试集大小: 48 个点 (1天)

完整数据集上的表现

数据集	MSE	RMSE	MAE	MAPE(%)	数据点数量
训练集	99.74449	9.987216	5.650278	2.832987	1696
测试集	35.66698	5.972183	4.287041	1.64227	48







二阶段

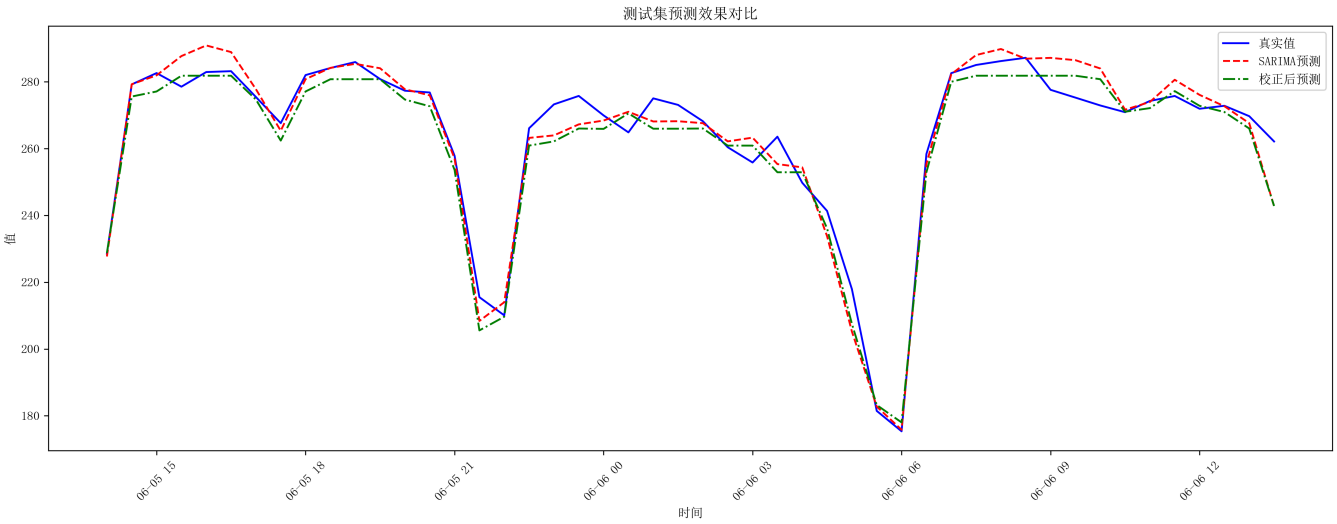
无额外特征的LGB

模型：LGB

特征：SARIMIA预测值

目标：真实值
策略：采用时序的三折交叉验证
tscv = TimeSeriesSplit(n_splits=3)

效果：



#训练集评估结果：
最终预测 RMSE: 9.1482 | MSE: 83.6897 | MAE: 5.8464 | MAPE: 3.08%
相比原始SARIMA提升: RMSE提升: 10.98% | MSE提升: 20.75%

#验证集评估结果：
最终预测 RMSE: 9.1056 | MSE: 82.9119 | MAE: 6.3890 | MAPE: 3.16%
相比原始SARIMA提升: RMSE提升: -4.42% | MSE提升: -9.05%

#测试集评估结果：
最终预测 RMSE: 5.8264 | MSE: 33.9468 | MAE: 4.5531 | MAPE: 1.74%
相比原始SARIMA提升: RMSE提升: 2.44% | MSE提升: 4.82%

训练集: RMSE=9.1482, MSE=83.6897, MAE=5.8464, MAPE=3.08%
验证集: RMSE=9.1056, MSE=82.9119, MAE=6.3890, MAPE=3.16%
测试集: RMSE=5.8264, MSE=33.9468, MAE=4.5531, MAPE=1.74%

加入趋势分解的LGB

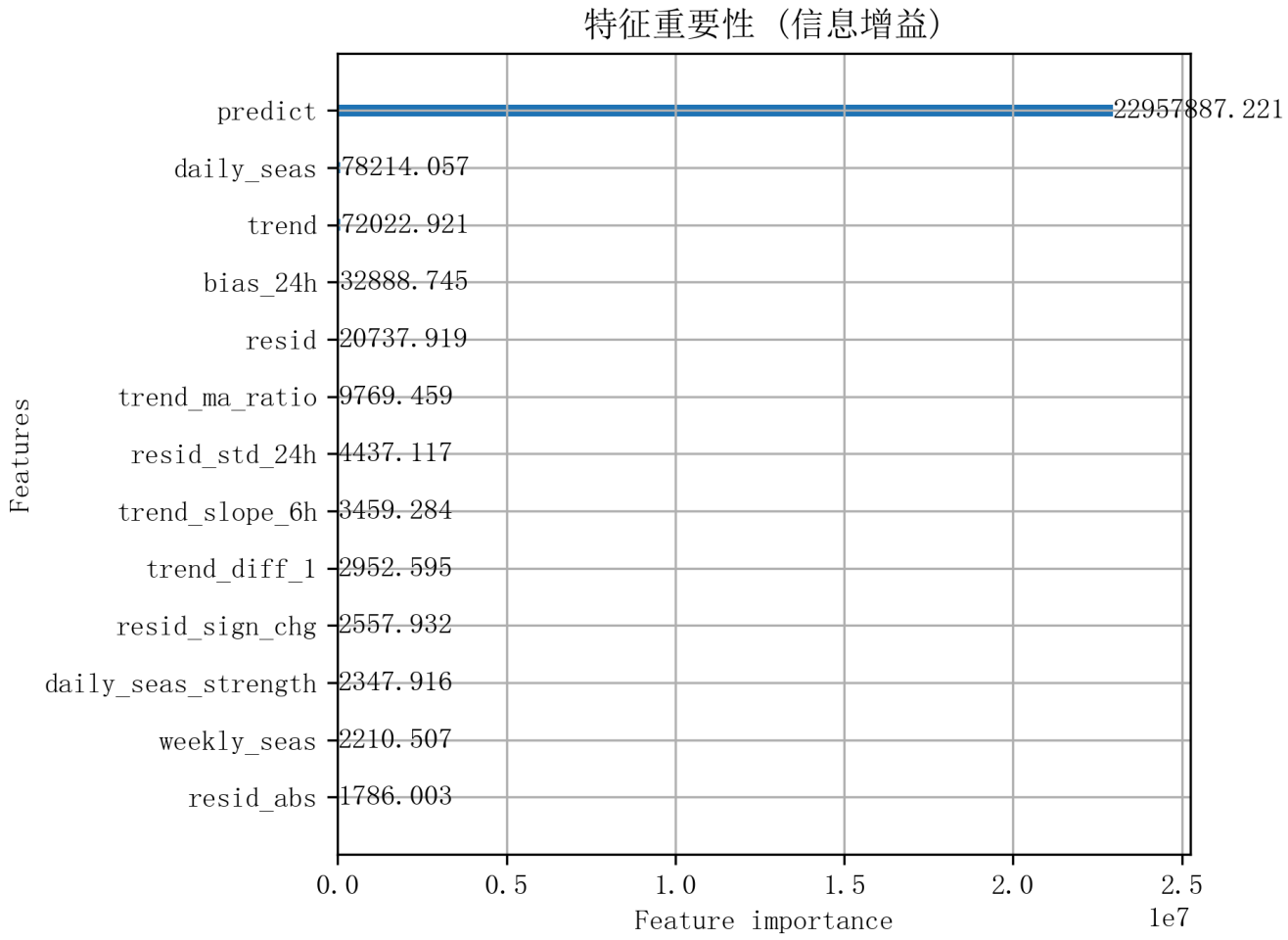
采用STL将时序分解为：趋势，结节性和残差

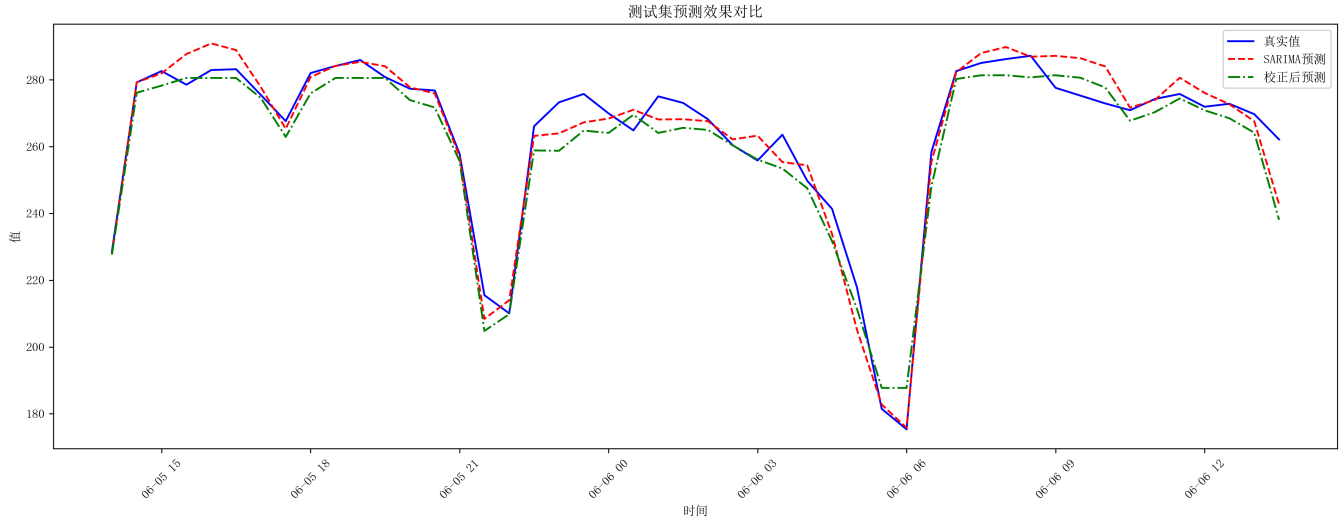
STL分解将时间序列Y_t分解为以下三个主要部分：

$$Y_t = T_t + S_t + R_t$$

其中：

- T_t是**趋势（Trend）**成分，表示数据的长期变化趋势。
- S_t是**季节性（Seasonal）**成分，表示周期性波动，具有固定的周期长度。
- R_t是**残差（Residual）**成分，表示随机波动或噪声部分。

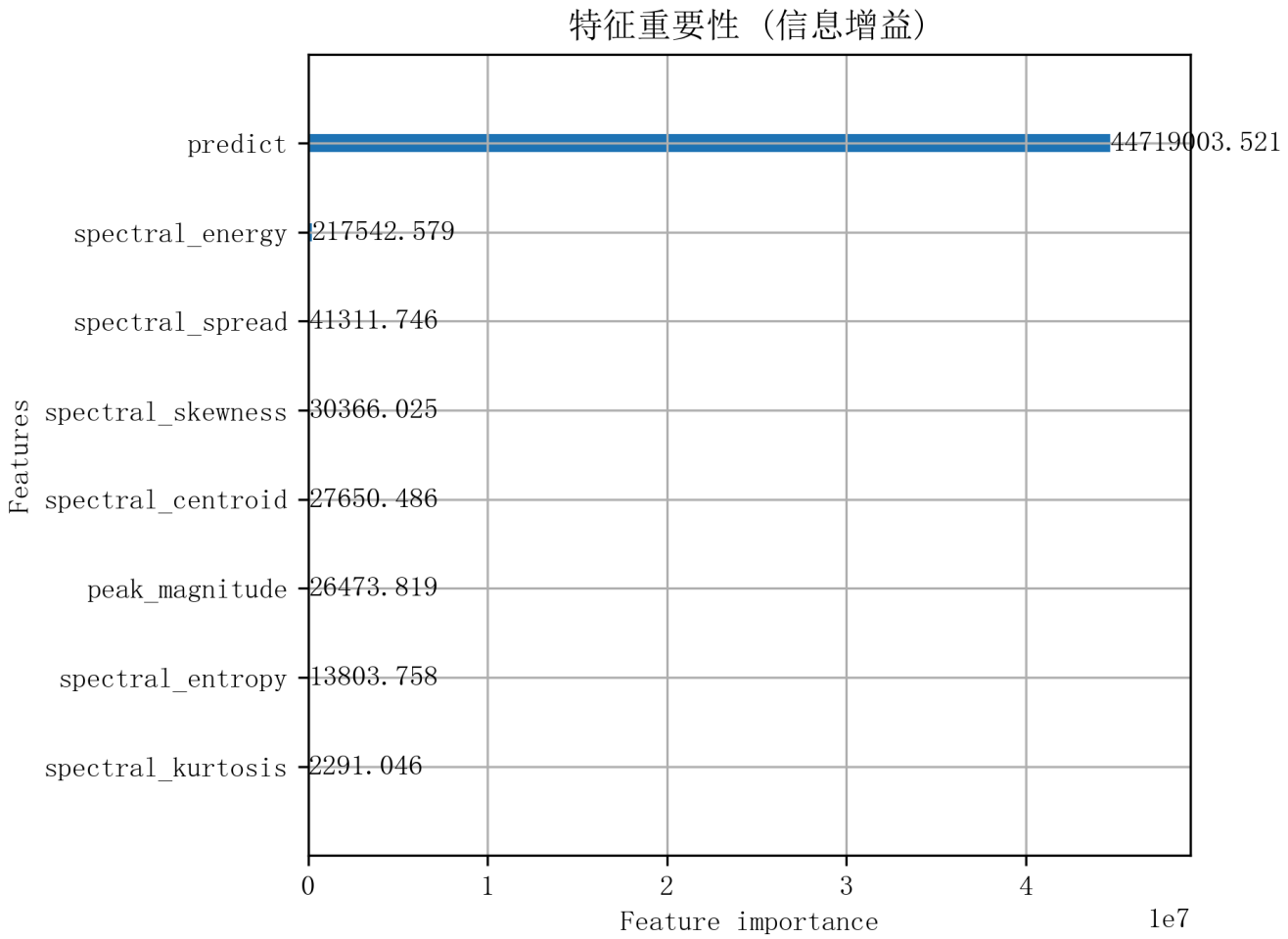


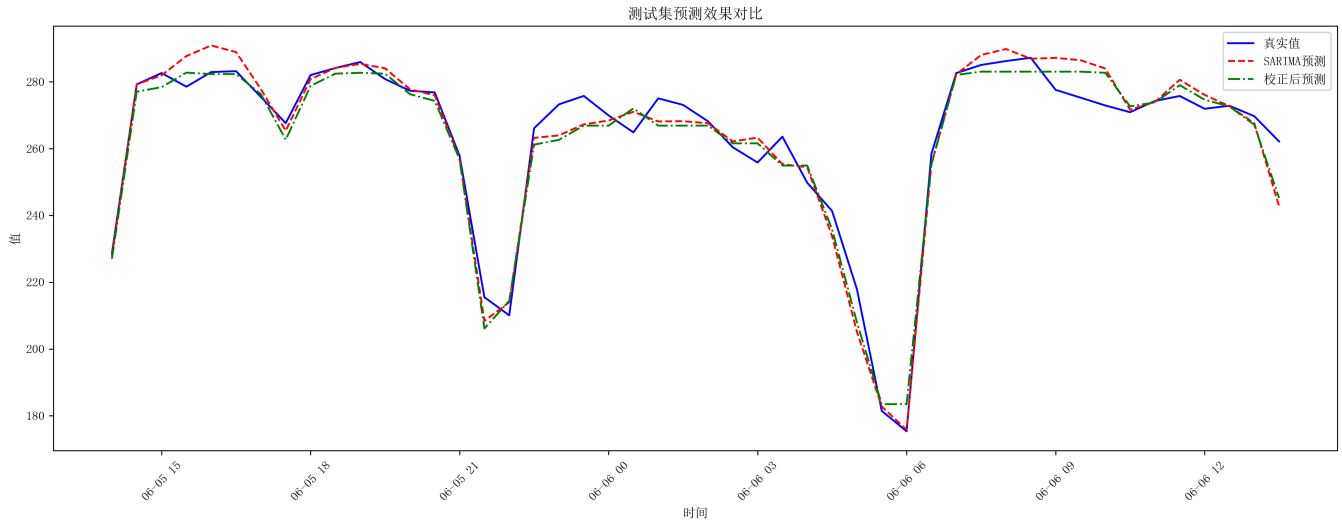


训练集: RMSE=8.1428, MSE=66.3050, MAE=5.4143, MAPE=2.87%
验证集: RMSE=8.9872, MSE=80.7701, MAE=6.6489, MAPE=3.44%
测试集: RMSE=6.8877, MSE=47.4404, MAE=5.3152, MAPE=2.07%
加入该特征后，预测性能有下降，该方法不可取

只选择额外加入日季节性和趋势性特征的预测结果，依然是下降的
训练集: RMSE=8.9240, MSE=79.6381, MAE=6.0642, MAPE=3.28%
验证集: RMSE=9.3862, MSE=88.1003, MAE=6.9643, MAPE=3.65%
测试集: RMSE=6.8933, MSE=47.5176, MAE=5.4078, MAPE=2.08%

时频转换



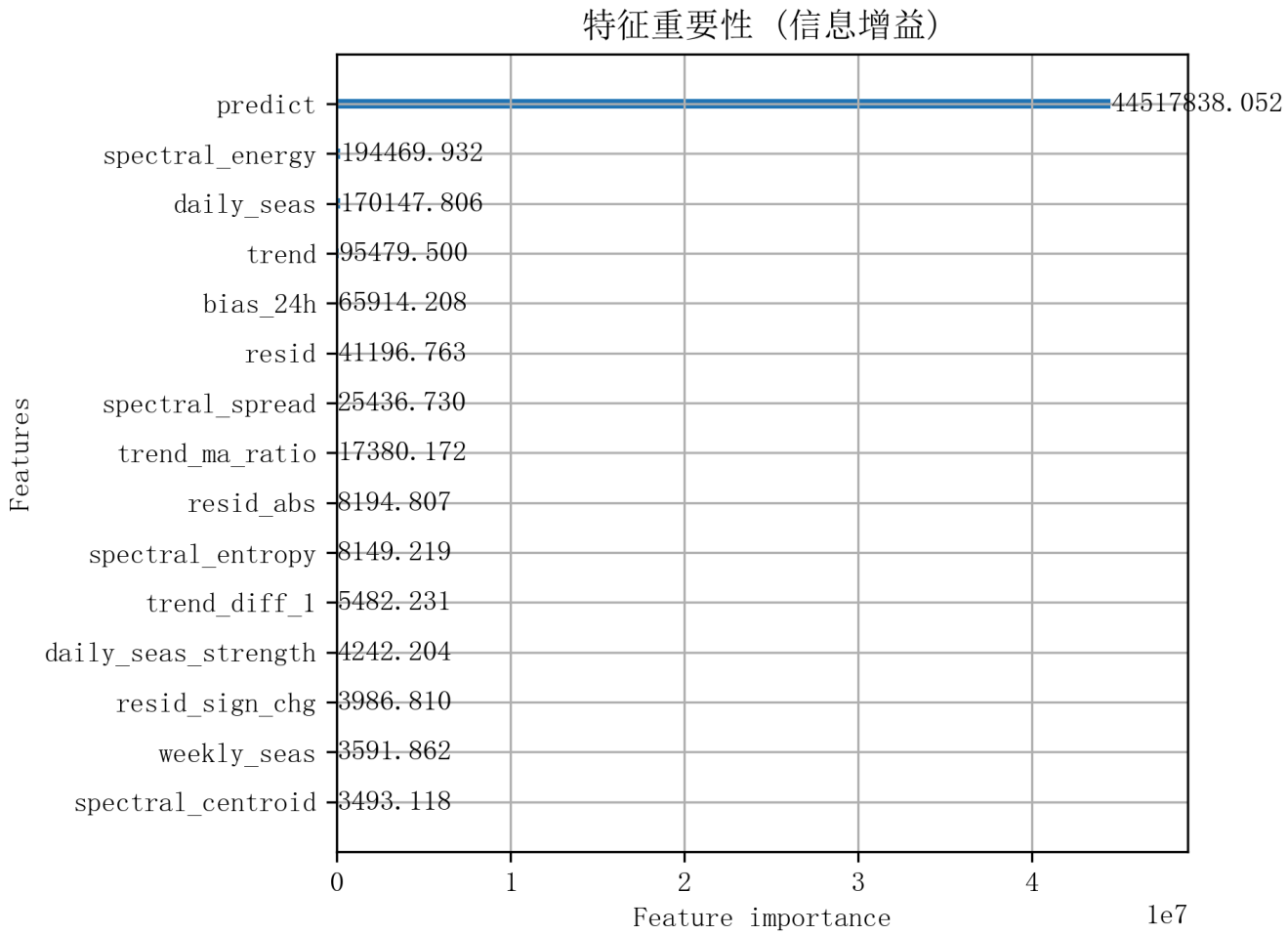


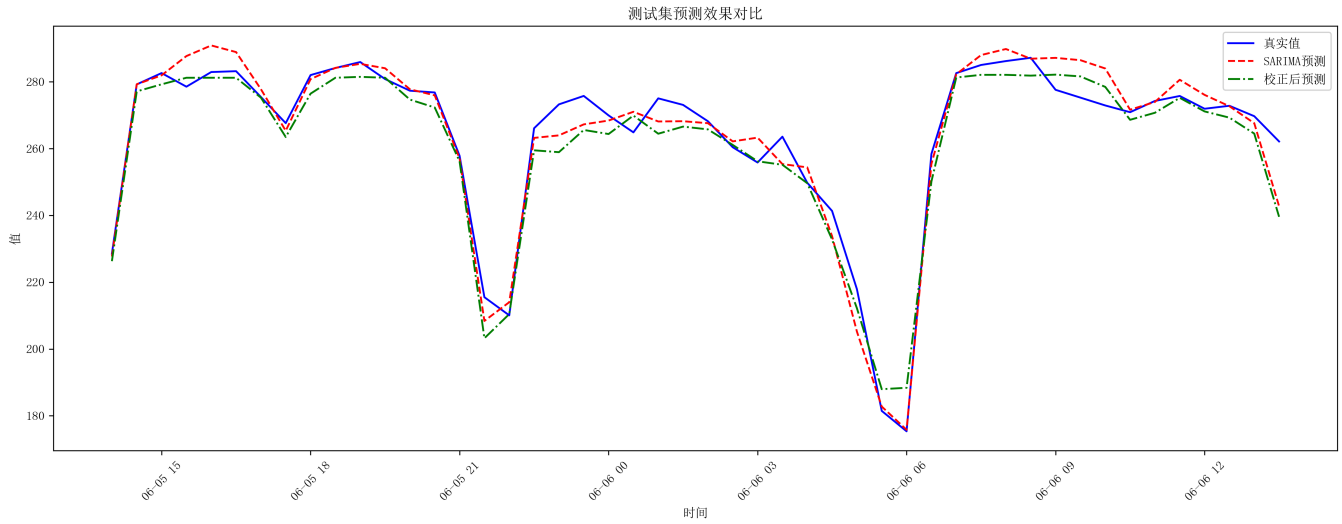
训练集: RMSE=7.9944, MSE=63.9105, MAE=4.6843, MAPE=2.26%
验证集: RMSE=9.3979, MSE=88.3203, MAE=6.8016, MAPE=3.50%
测试集: RMSE=5.5350, MSE=30.6366, MAE=4.3326, MAPE=1.69%

仅使用features = ['predict','spectral_energy','spectral_spread']

- spectral_energy：频谱能量
- spectral_spread：频谱展宽
  训练集: RMSE=8.2122, MSE=67.4400, MAE=4.7882, MAPE=2.31%
  验证集: RMSE=9.5016, MSE=90.2805, MAE=6.7170, MAPE=3.54%
  测试集: RMSE=5.4977, MSE=30.2246, MAE=4.1642, MAPE=1.62%

同时加入趋势和频域信息



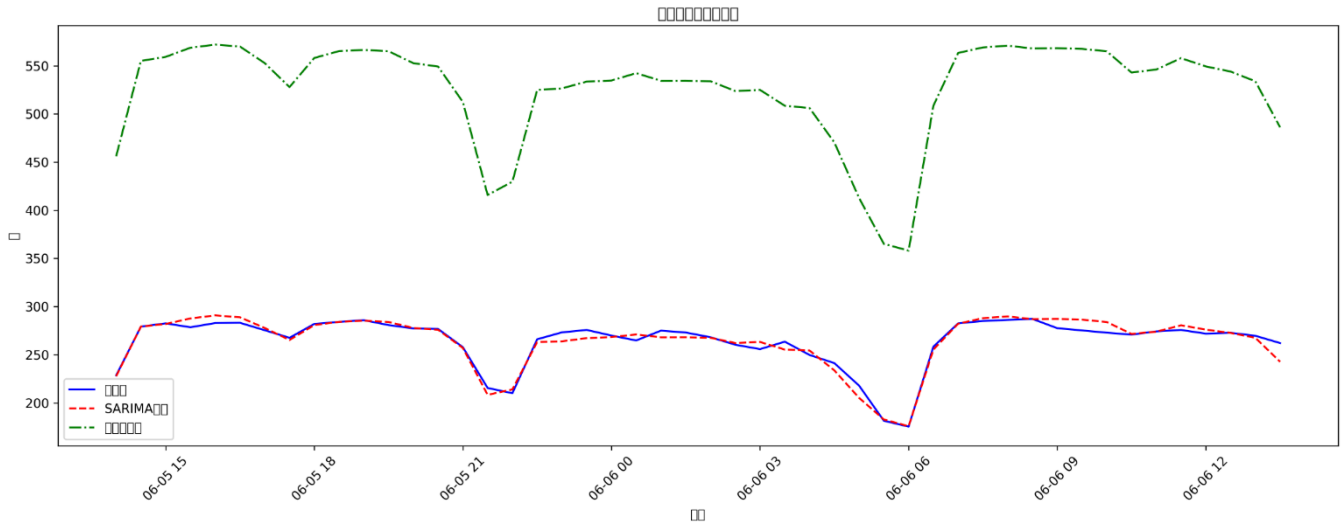


训练集: RMSE=7.5432, MSE=56.8996, MAE=4.8061, MAPE=2.48%
验证集: RMSE=9.5000, MSE=90.2506, MAE=6.9545, MAPE=3.71%
测试集: RMSE=6.5090, MSE=42.3668, MAE=4.8909, MAPE=1.92%

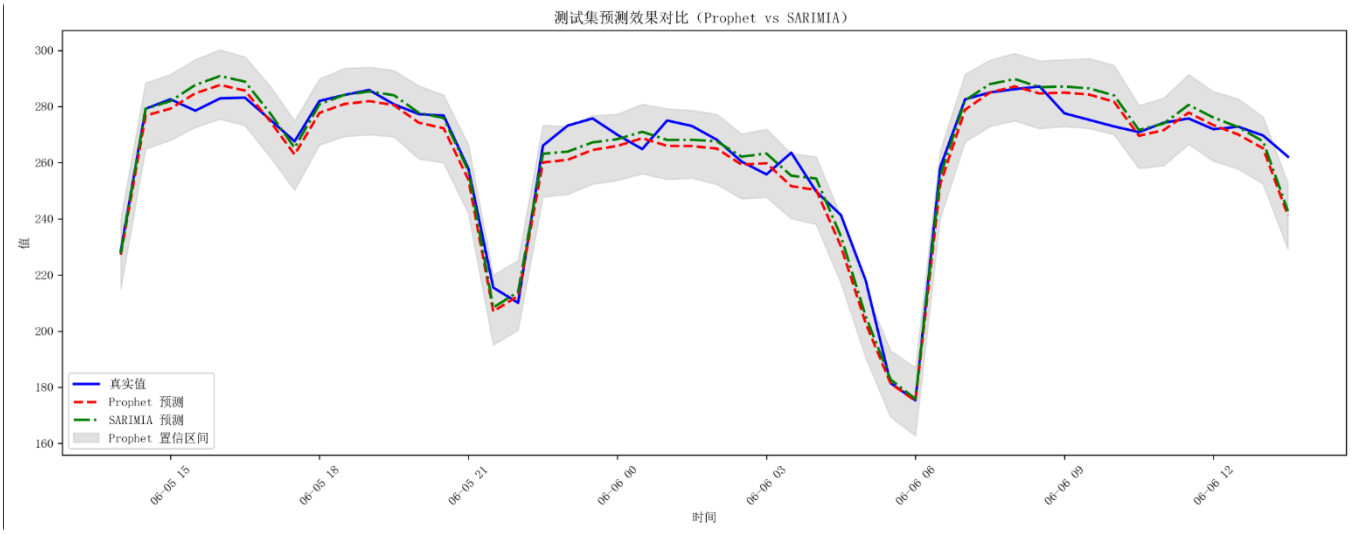
features = ['predict','spectral_energy','daily_seas']
训练集: RMSE=7.8932, MSE=62.3025, MAE=4.5751, MAPE=2.25%
验证集: RMSE=9.5662, MSE=91.5122, MAE=6.9238, MAPE=3.66%
测试集: RMSE=6.1242, MSE=37.5064, MAE=4.2706, MAPE=1.65%

XGBoost

偏移太厉害【无特征】
训练集: RMSE=235.0170, MSE=55232.9679, MAE=227.7065, MAPE=100.48%
验证集: RMSE=226.4754, MSE=51291.1026, MAE=221.6640, MAPE=100.47%
测试集: RMSE=263.9681, MSE=69679.1421, MAE=262.4687, MAPE=99.31%

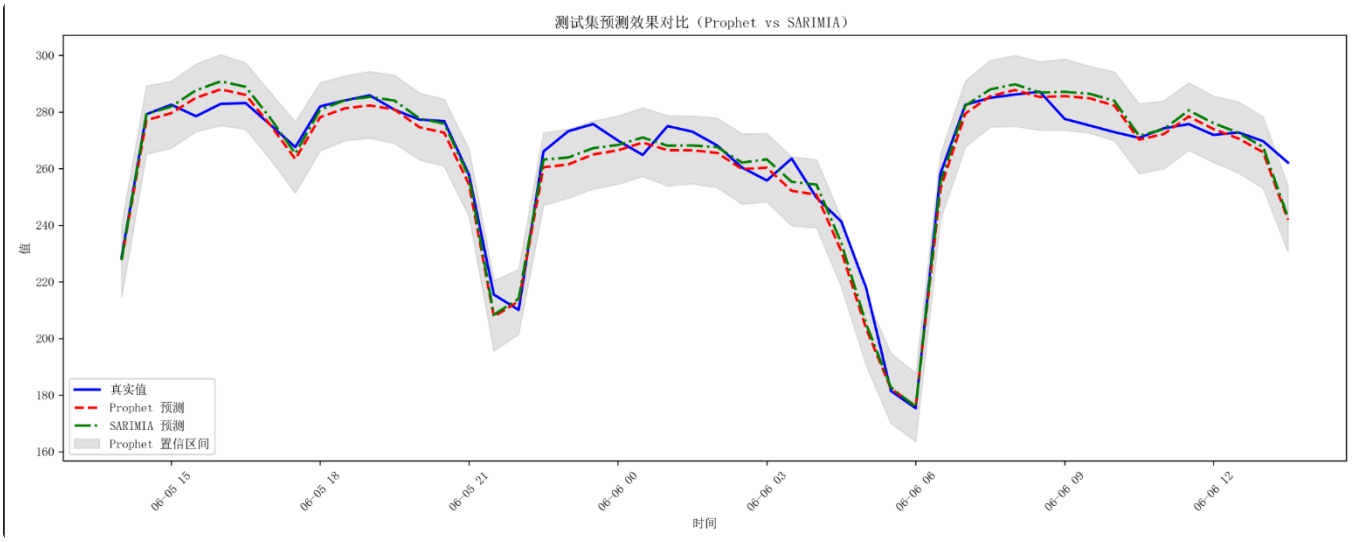


prohet



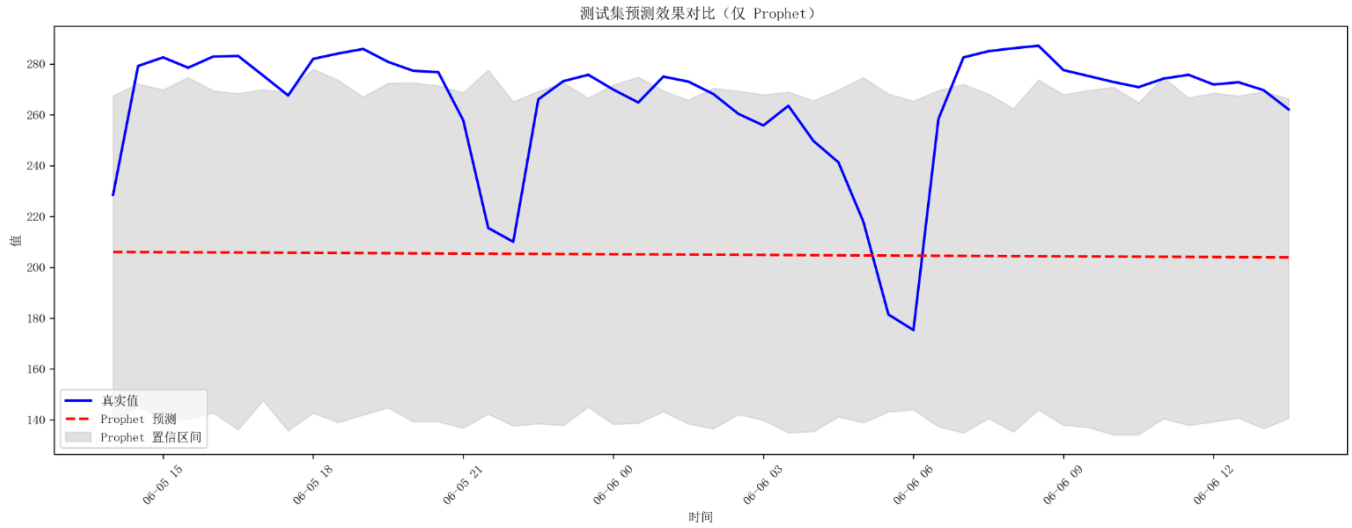
训练集: MSE=0.0000 | RMSE=0.0000 | MAE=0.0000 | MAPE=0.00%
验证集: MSE=72.8673 | RMSE=8.5362 | MAE=5.9708 | MAPE=2.93%
测试集: MSE=42.2817 | RMSE=6.5024 | MAE=4.8980 | MAPE=1.87%

无特征



训练集: MSE=0.0000 | RMSE=0.0000 | MAE=0.0000 | MAPE=0.00%
验证集: MSE=73.3028 | RMSE=8.5617 | MAE=5.9952 | MAPE=2.96%
测试集: MSE=39.4408 | RMSE=6.2802 | MAE=4.7290 | MAPE=1.81%

仅用历史值



训练集: MSE=0.0000 | RMSE=0.0000 | MAE=0.0000 | MAPE=0.00%
验证集: MSE=2406.6104 | RMSE=49.0572 | MAE=41.5670 | MAPE=20.50%
测试集: MSE=4147.0147 | RMSE=64.3973 | MAE=61.4042 | MAPE=22.73%