

2025-06-19

今日工作内容：（部署模型，测试模型）

- SARIMIA + 随机森林

遇到的问题：

- 模型更新数据后重新训练，效果变差（原因：6月15日之后的数据趋势明显变低）

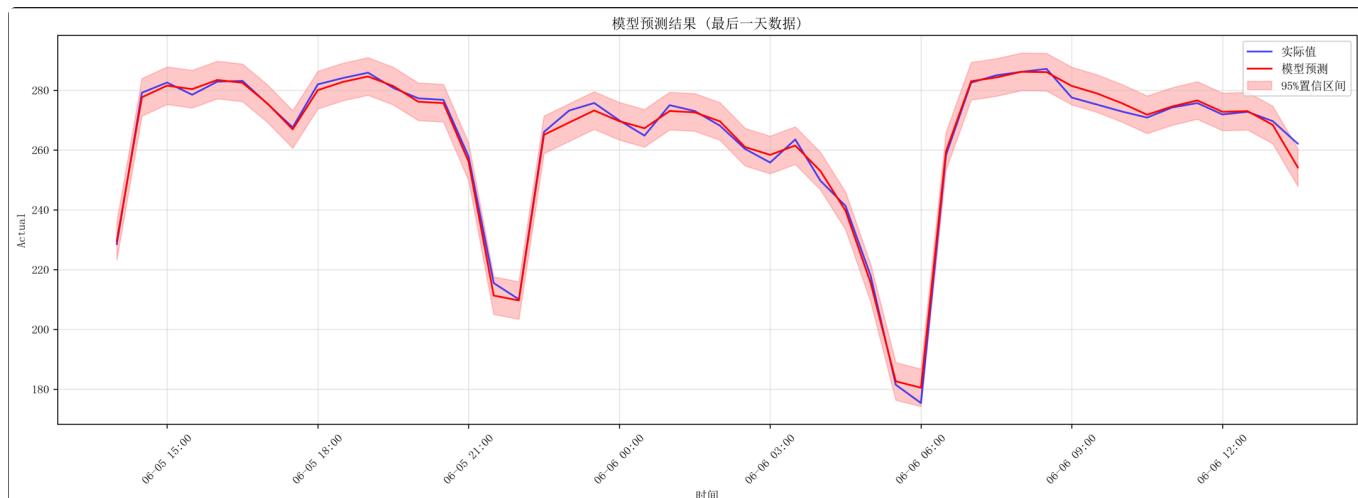
测试的过程：

- 测试模型SARIMIA在6月10日-6月12日的数据作为训练，预测13日效果较好 ($\text{mse}=40.9431$)
- 6月10日-6月13日数据作为训练集，预测14日效果开始变差 ($\text{mse}=1701.0832$)

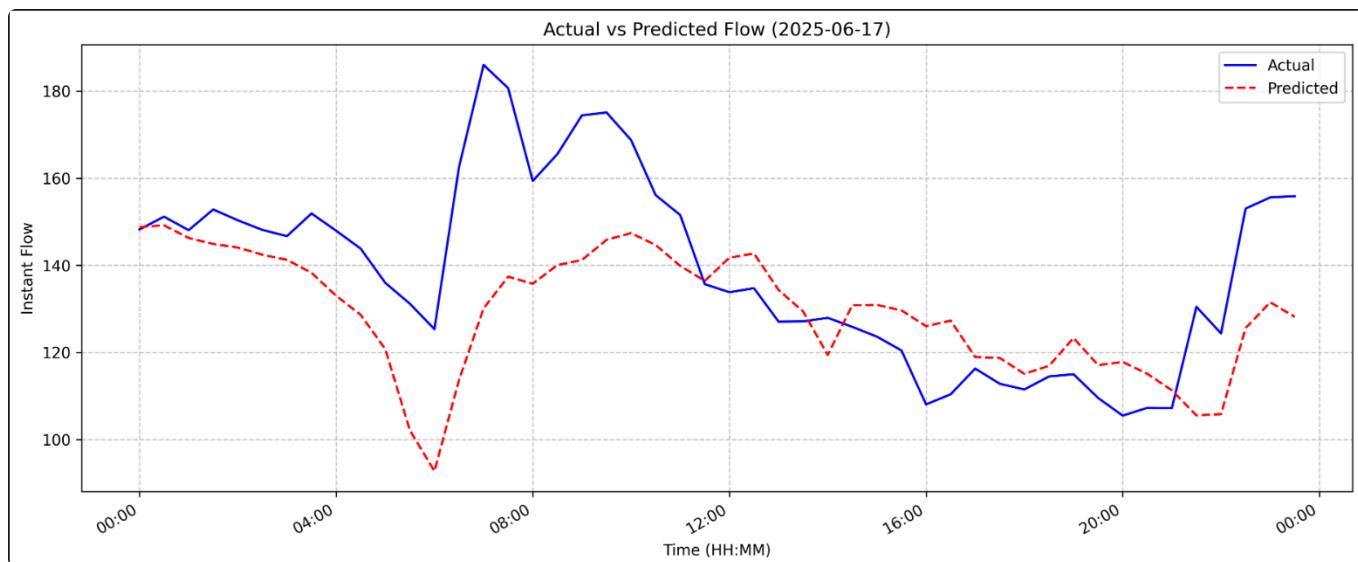
结论：

- SARIMIA只适合用于周期性明显的时段做预测
- SARIMIA不适合用全部数据喂模型，需要找有规律的一段区间来喂。需要结合时序图来选择训练集
- 针对6月15日产气量发生变化的情况，可能需要做个时序分类。并且想办法把实际生产计划编码到预测中。

SARIMIA + 随机森林



采用完整数据集训练，并预测最后一天

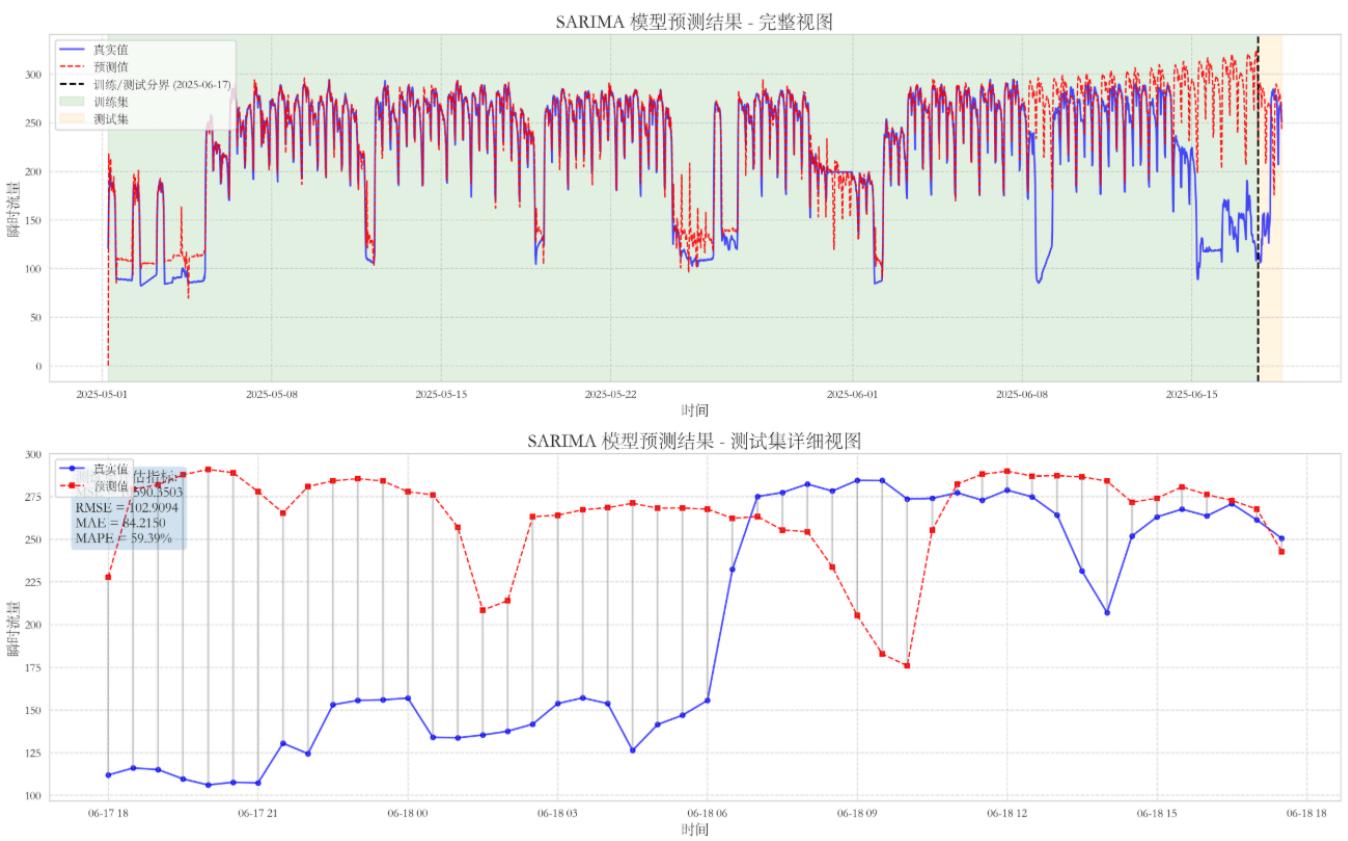


Test MSE: 389.99

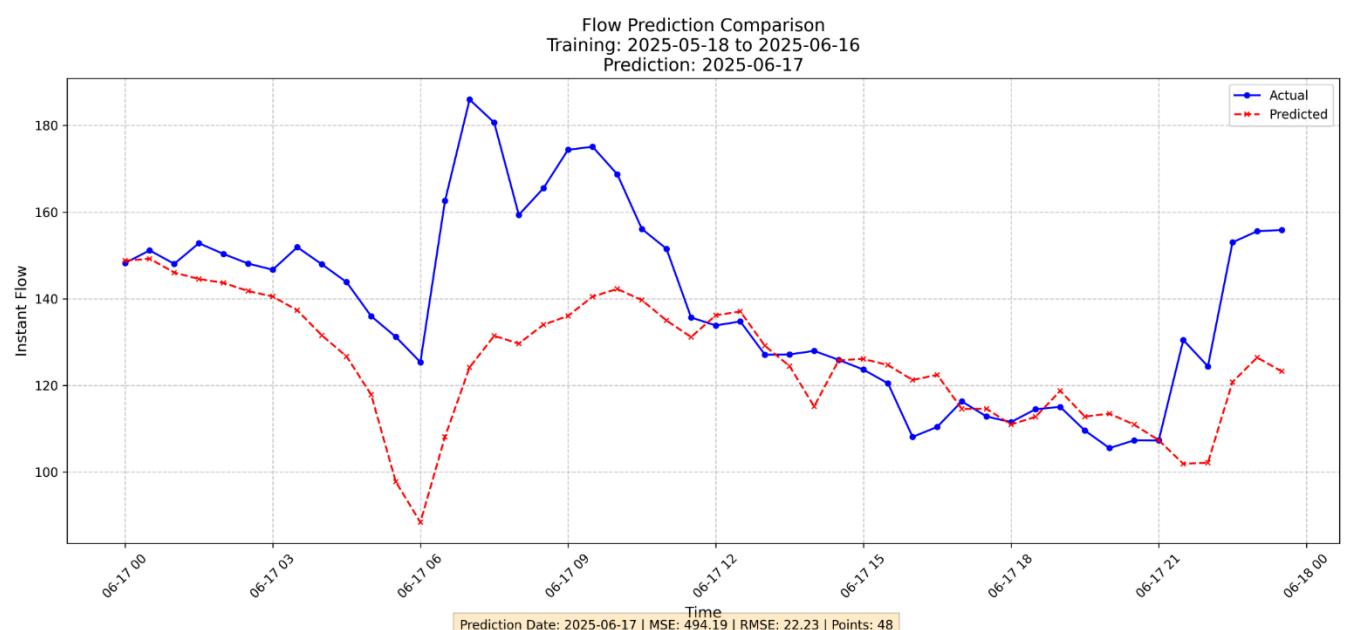
在完整数据集上的表现



对比6_13(数据到6-6)



只用最近一个月的数据



Using training data from 2025-05-18 00:00:00 to 2025-06-16 23:30:00

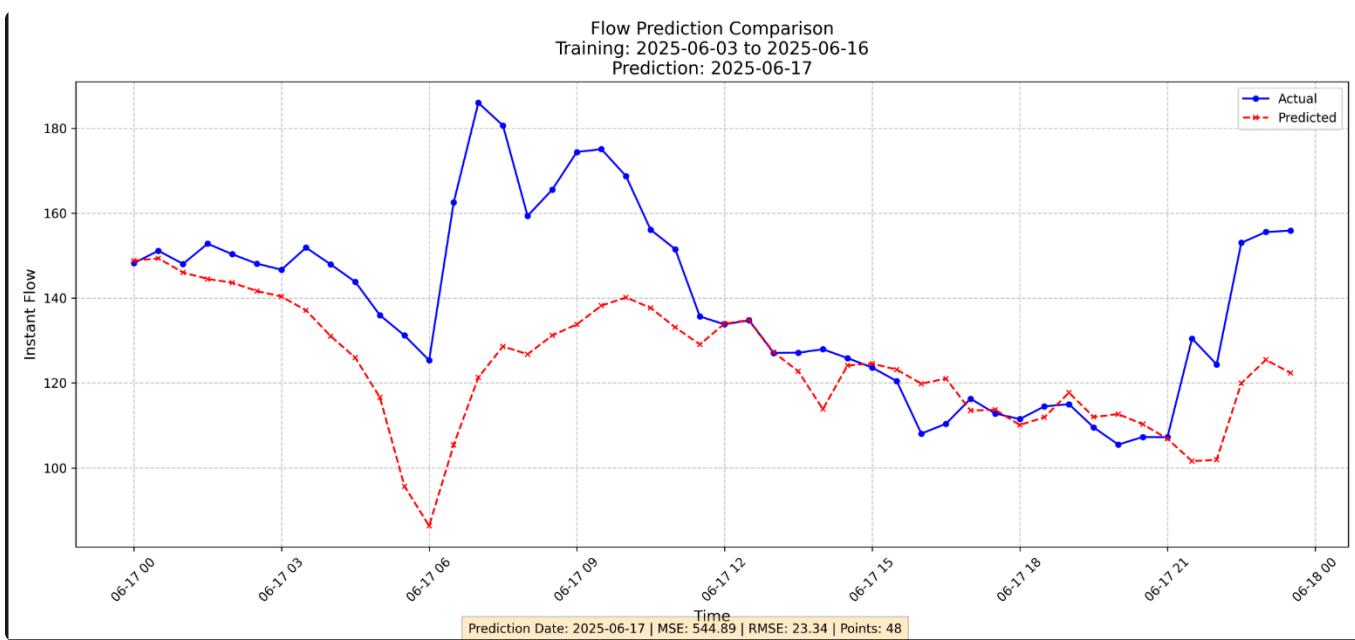
Using test data from 2025-06-17 00:00:00 to 2025-06-17 23:30:00

Training points: 1440, Test points: 48

Test MSE: 494.19

RMSE: 22.23

近两周的数据



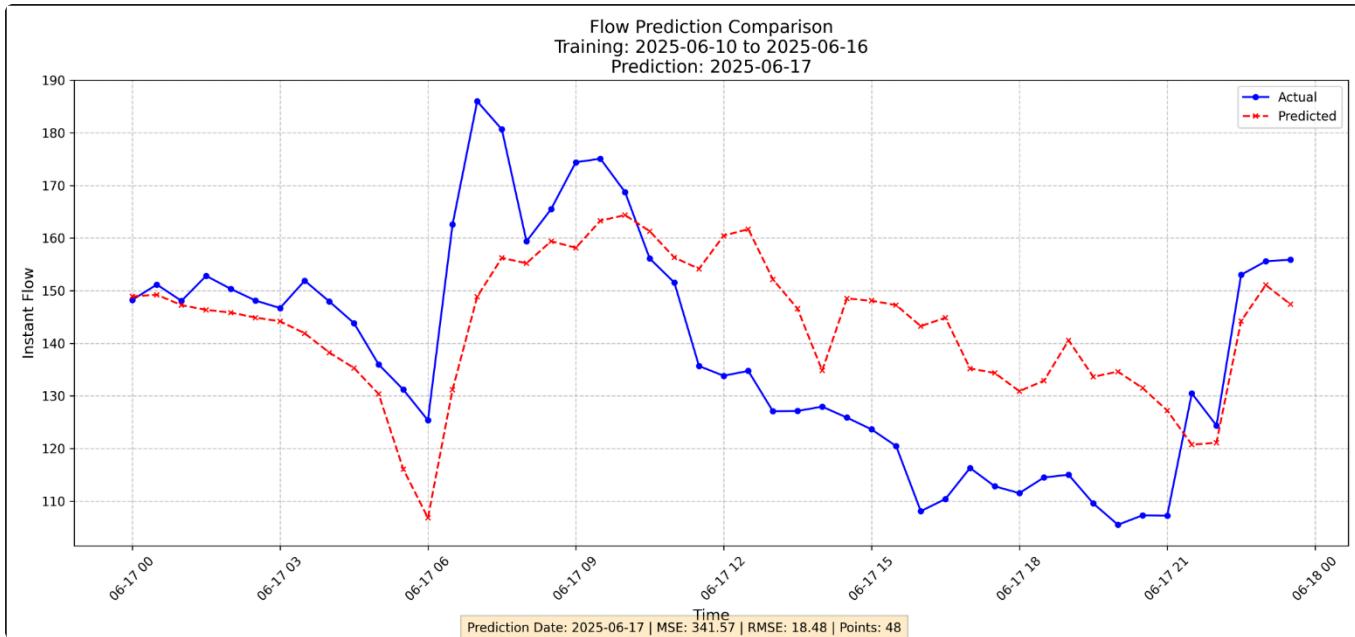
Using test data from 2025-06-17 00:00:00 to 2025-06-17 23:30:00

Training points: 672, Test points: 48

Test MSE: 544.89

RMSE: 23.34

近一周



Using training data from 2025-06-10 00:00:00 to 2025-06-16 23:30:00

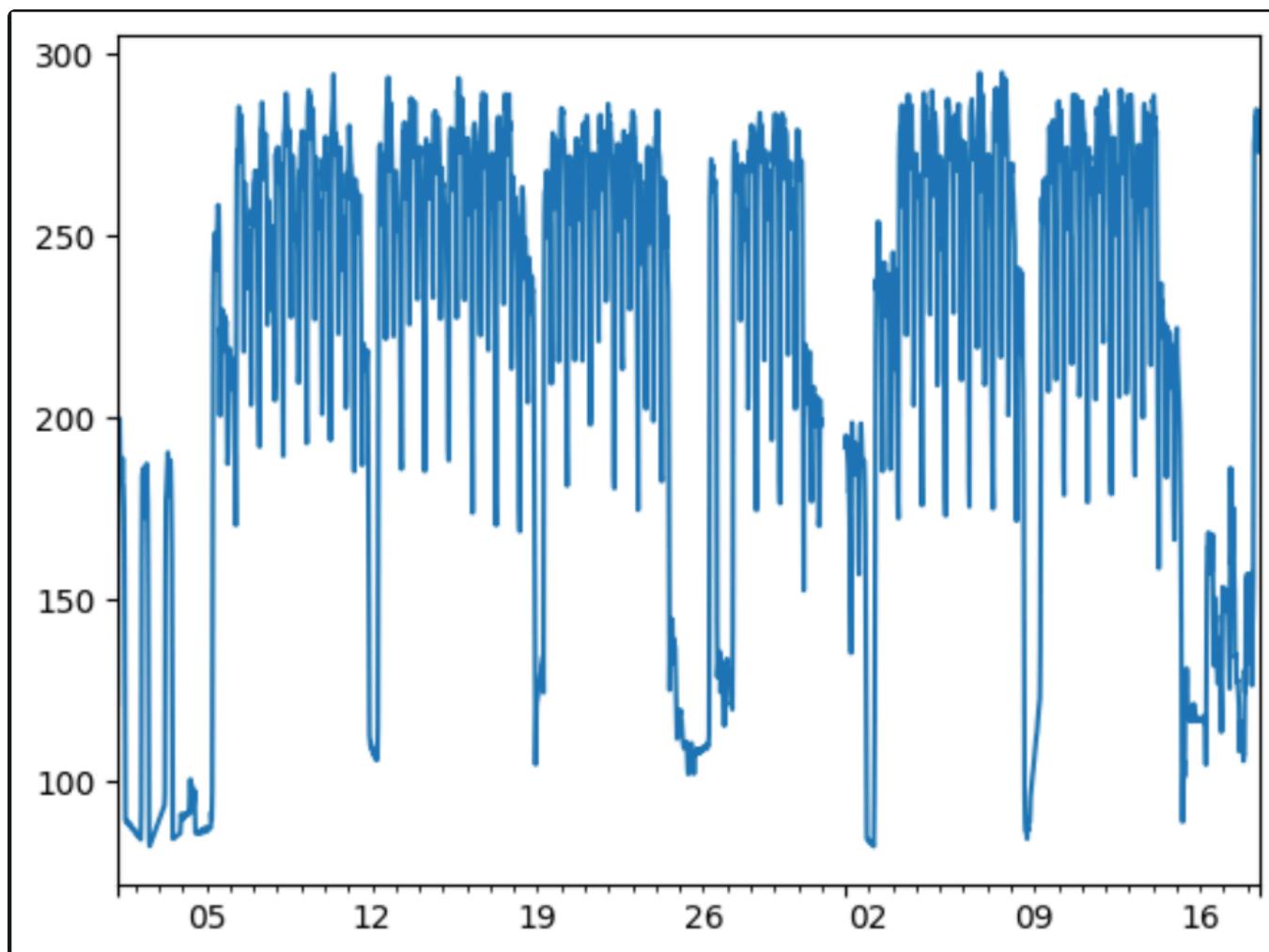
Using test data from 2025-06-17 00:00:00 to 2025-06-17 23:30:00

Training points: 336, Test points: 48

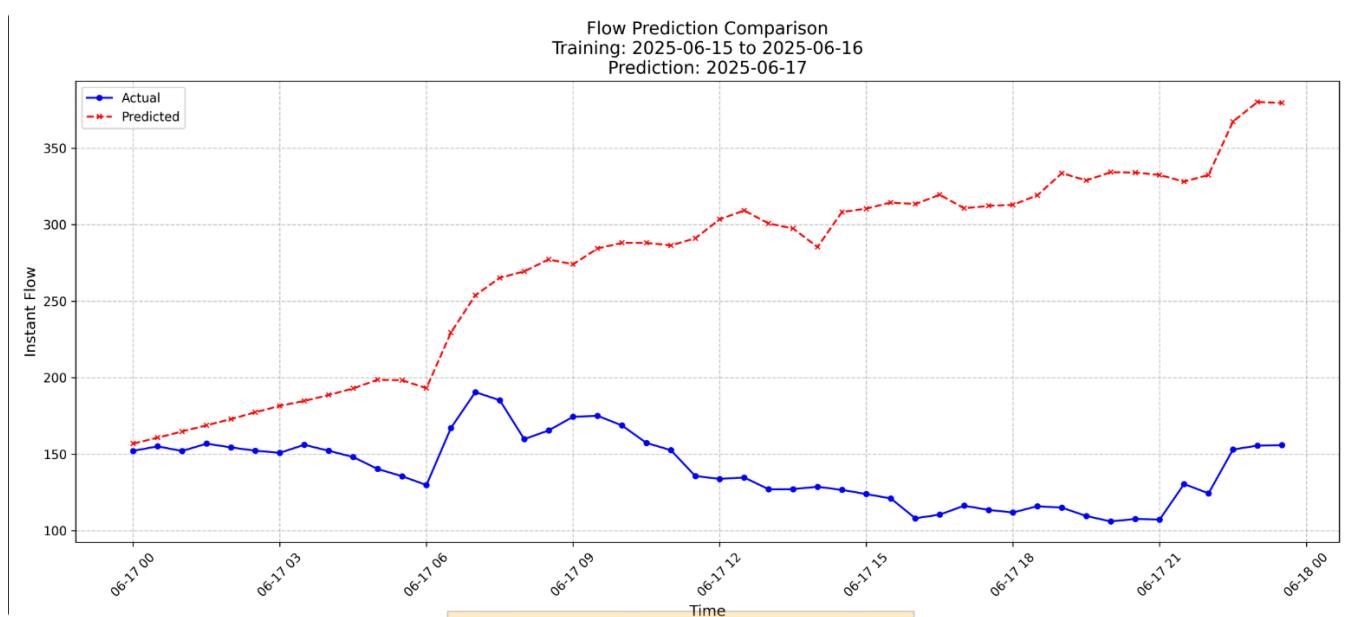
Test MSE: 341.57

RMSE: 18.48

整体的数据走势



近两天



压力预测

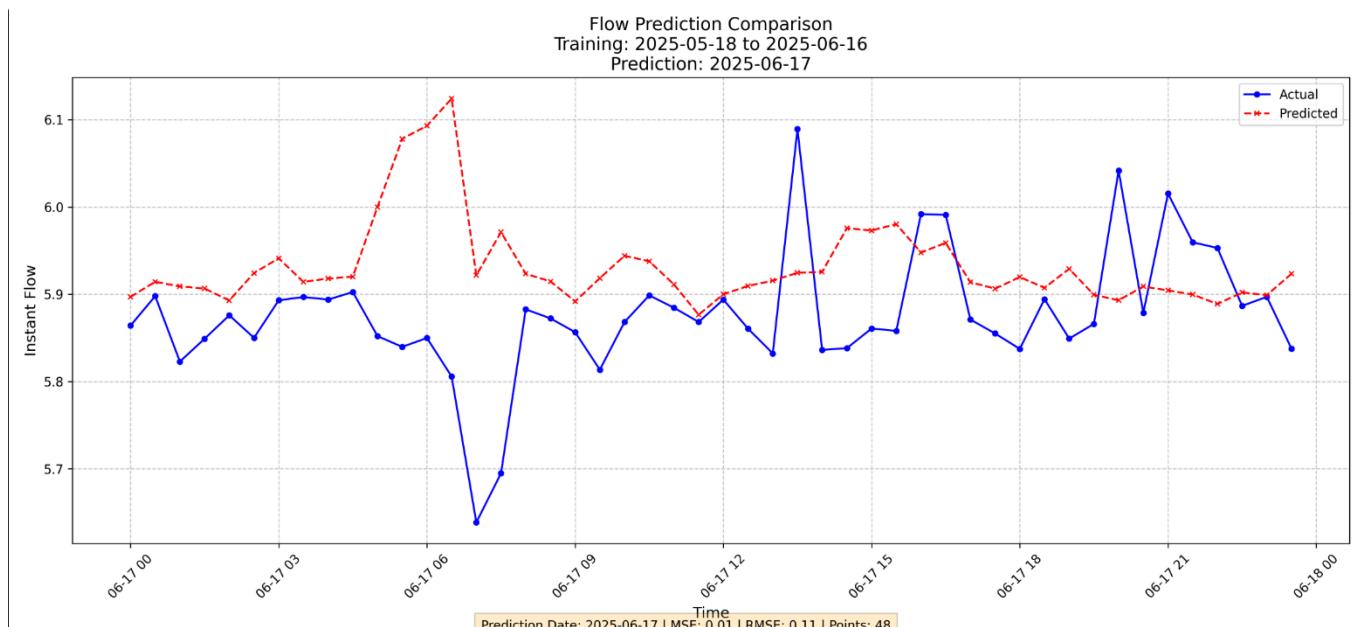
近一个月数据

Using training data from 2025-05-18 00:00:00 to 2025-06-16 23:30:00

Using test data from 2025-06-17 00:00:00 to 2025-06-17 23:30:00

Test MSE: 0.01

RMSE: 0.11



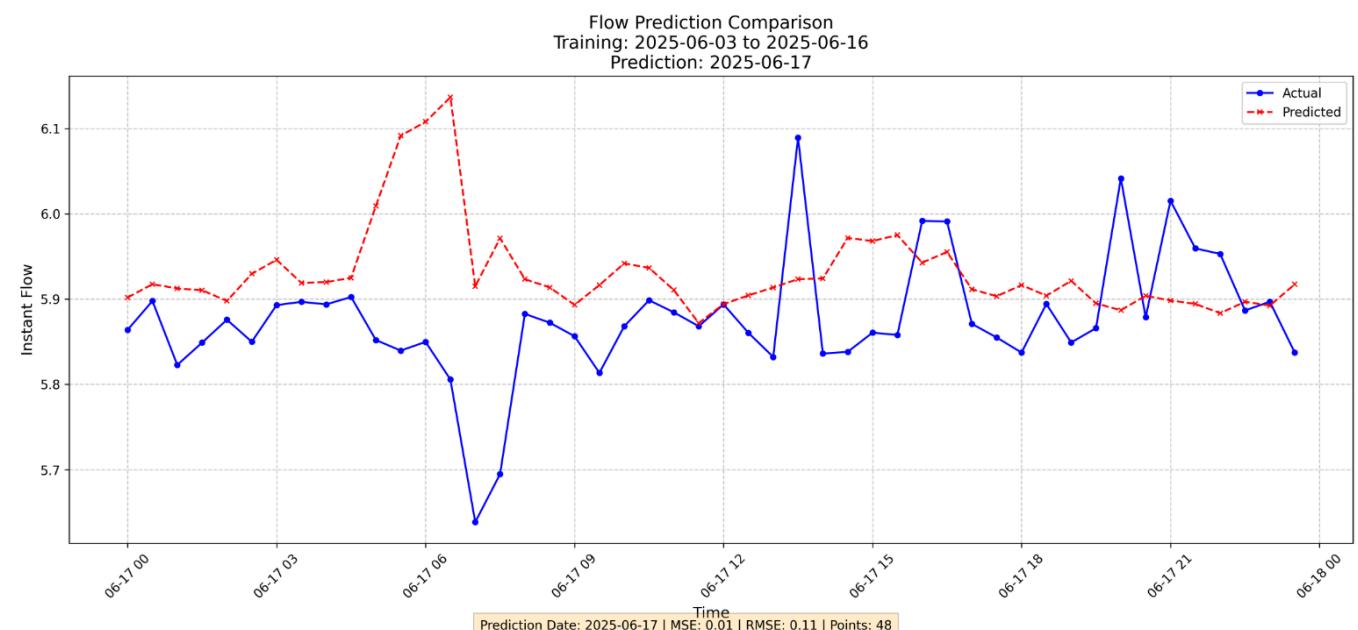
近两周数据

Using training data from 2025-06-03 00:00:00 to 2025-06-16 23:30:00

Using test data from 2025-06-17 00:00:00 to 2025-06-17 23:30:00

Test MSE: 0.01

RMSE: 0.11

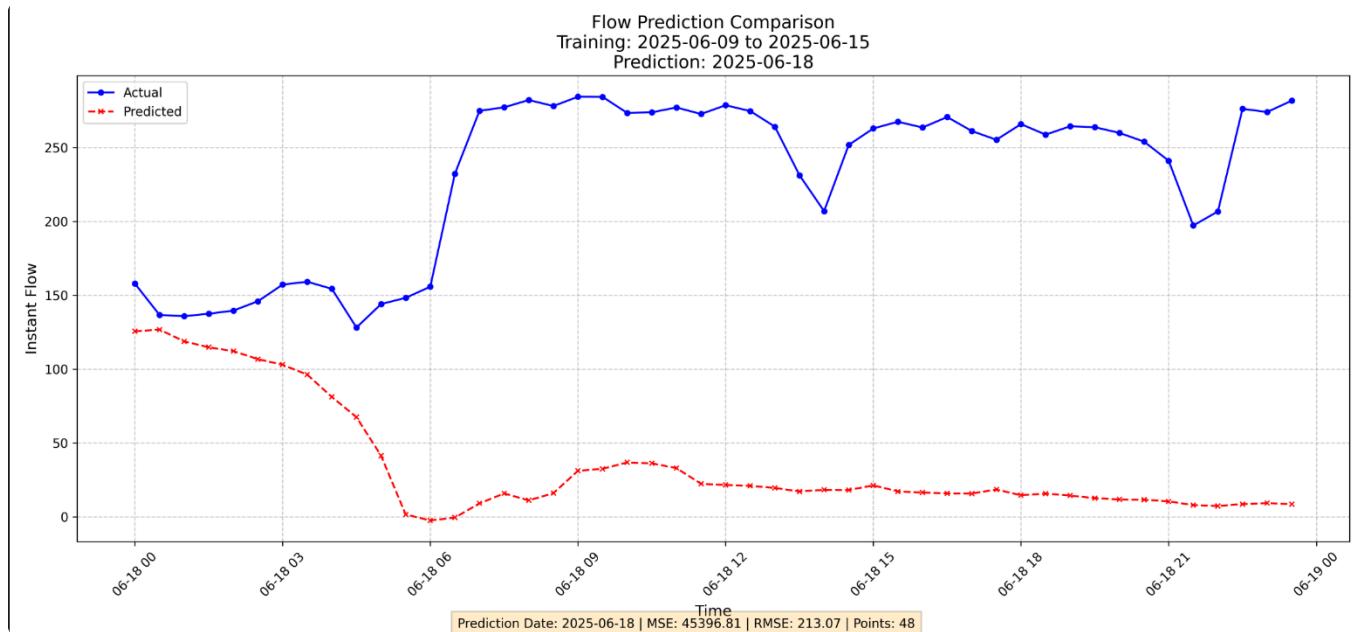


一周预测

Training date range: 2025-06-09 00:00:00 to 2025-06-15 23:30:00

Test date range: 2025-06-18 00:00:00 to 2025-06-18 23:30:00

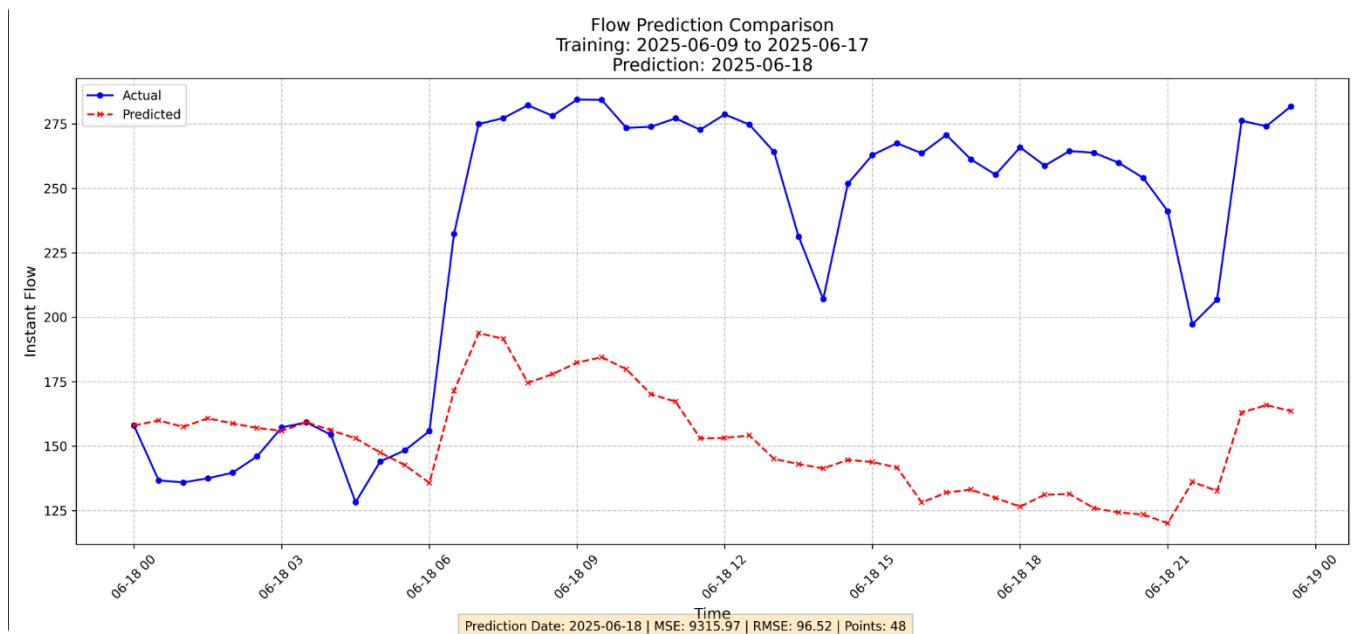
Training points: 336 (7.0 days), Test points: 48



Training date range: 2025-06-09 00:00:00 to 2025-06-17 23:30:00

Test date range: 2025-06-18 00:00:00 to 2025-06-18 23:30:00

Training points: 432 (9.0 days), Test points: 48



加入随机森林

Training date range: 2025-06-09 00:00:00 to 2025-06-17 23:30:00

Test date range: 2025-06-18 00:00:00 to 2025-06-18 23:30:00

Training points: 432 (9.0 days), Test points: 48

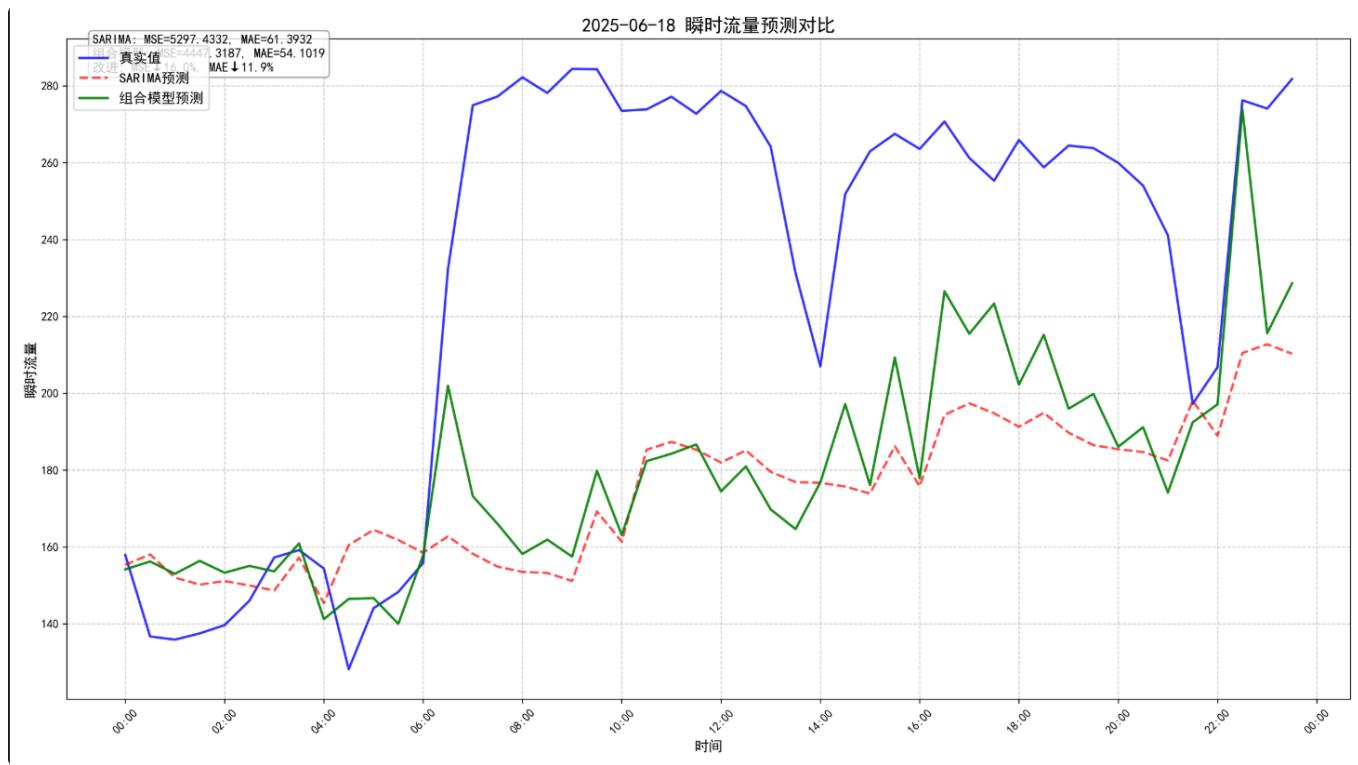
SARIMA Test MSE: 5297.4332

Combined Model Test MSE: 4447.3187

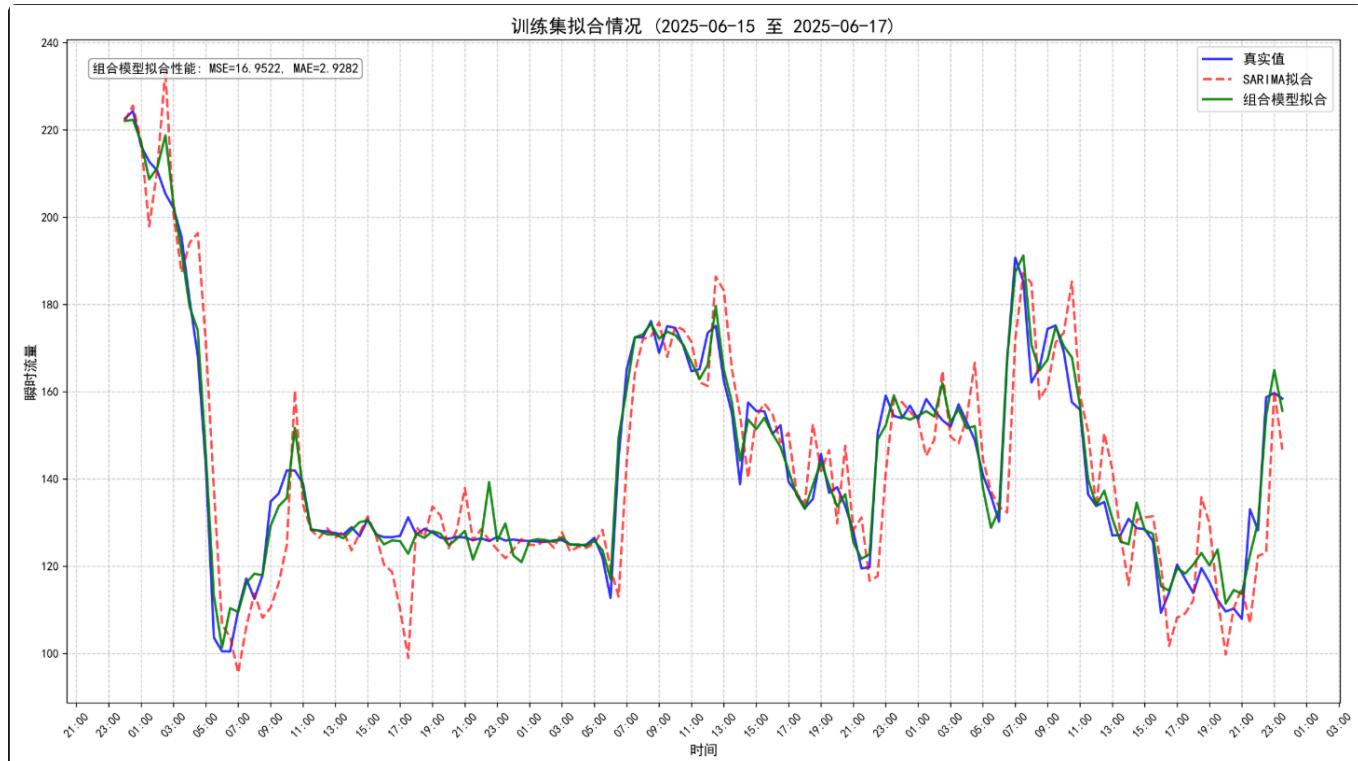
SARIMA Test MAE: 61.3932

Combined Model Test MAE: 54.1019

模型改进: MSE降低 16.05%, MAE降低 11.88%



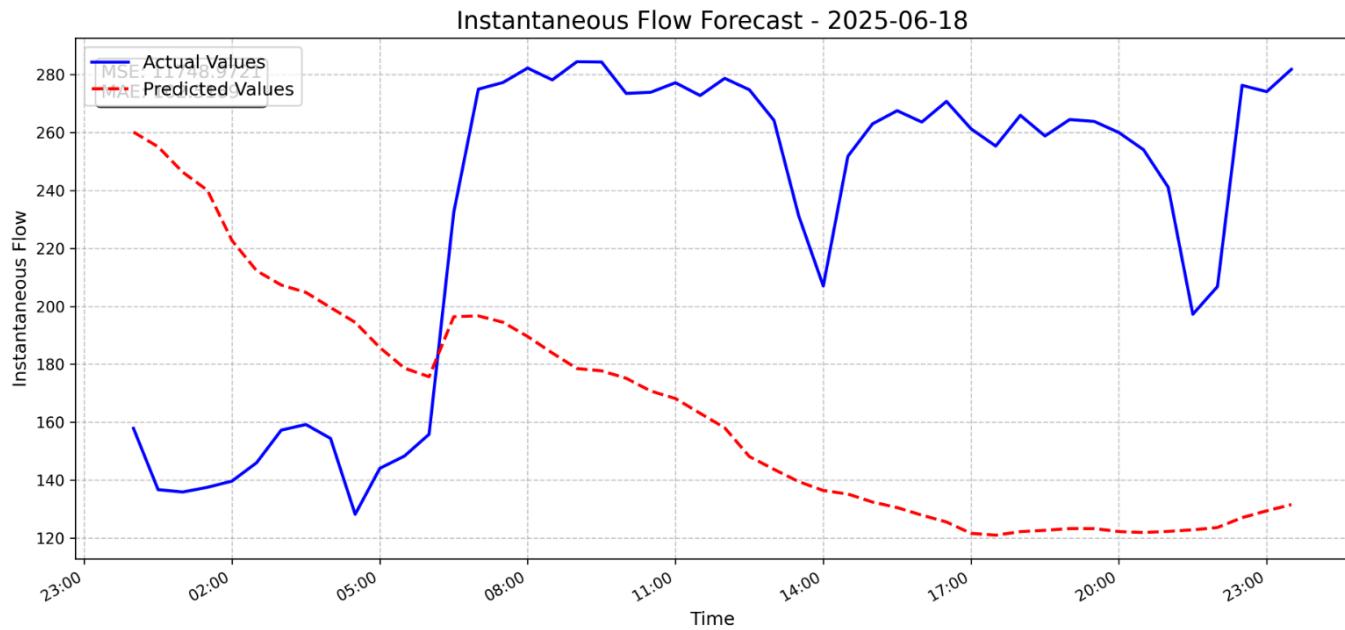
取最后3天的效果



直接用随机森林

测试集MSE: 11748.9721

测试集MAE: 102.5169

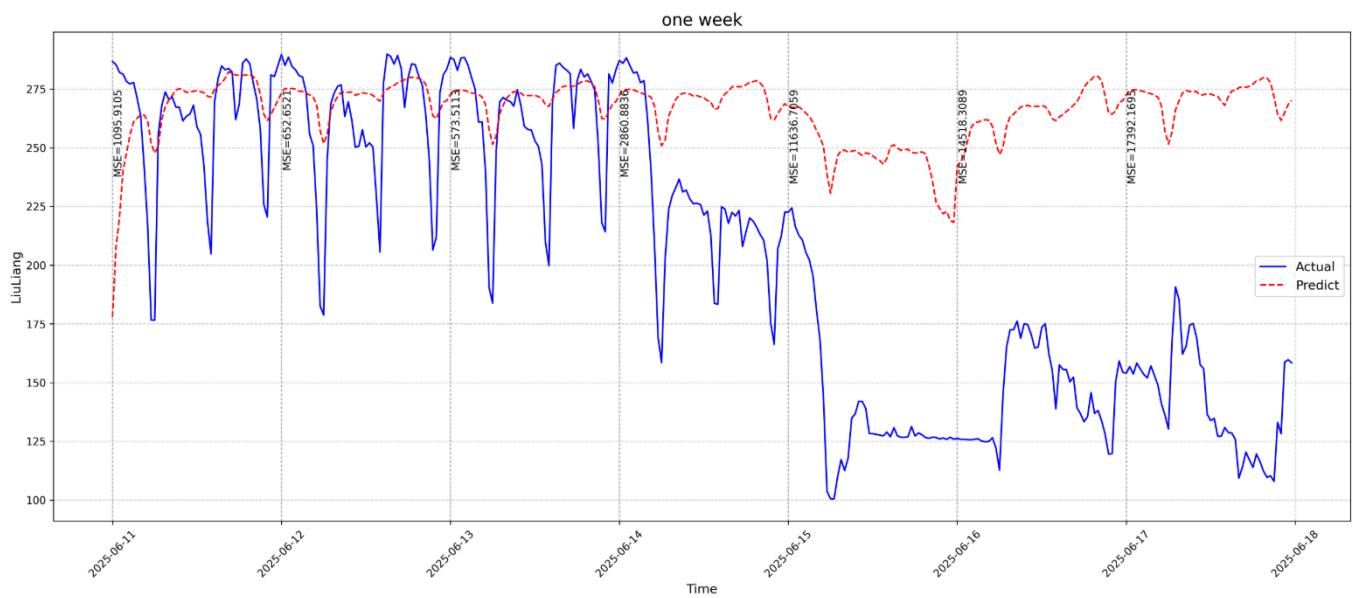


Random Forest Model | Prediction Date: 2025-06-18

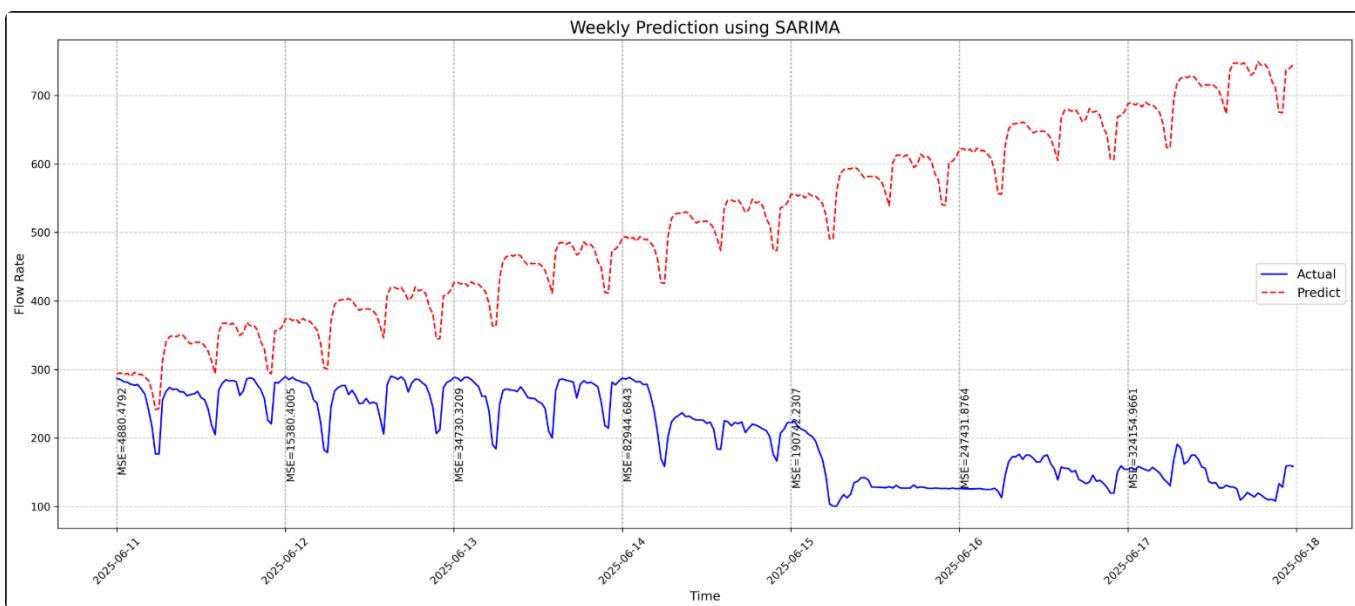
最佳参数: {'max_depth': 20, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 100}

测试集MSE: 6961.4488

测试集MAE: 65.3109



对比采用SARIMIA的预测效果



在6月15日之后，真实值就小于200了

测试SARIMIA

训练数据时间范围: 2025-06-10 00:00:00 至 2025-06-12 23:30:00

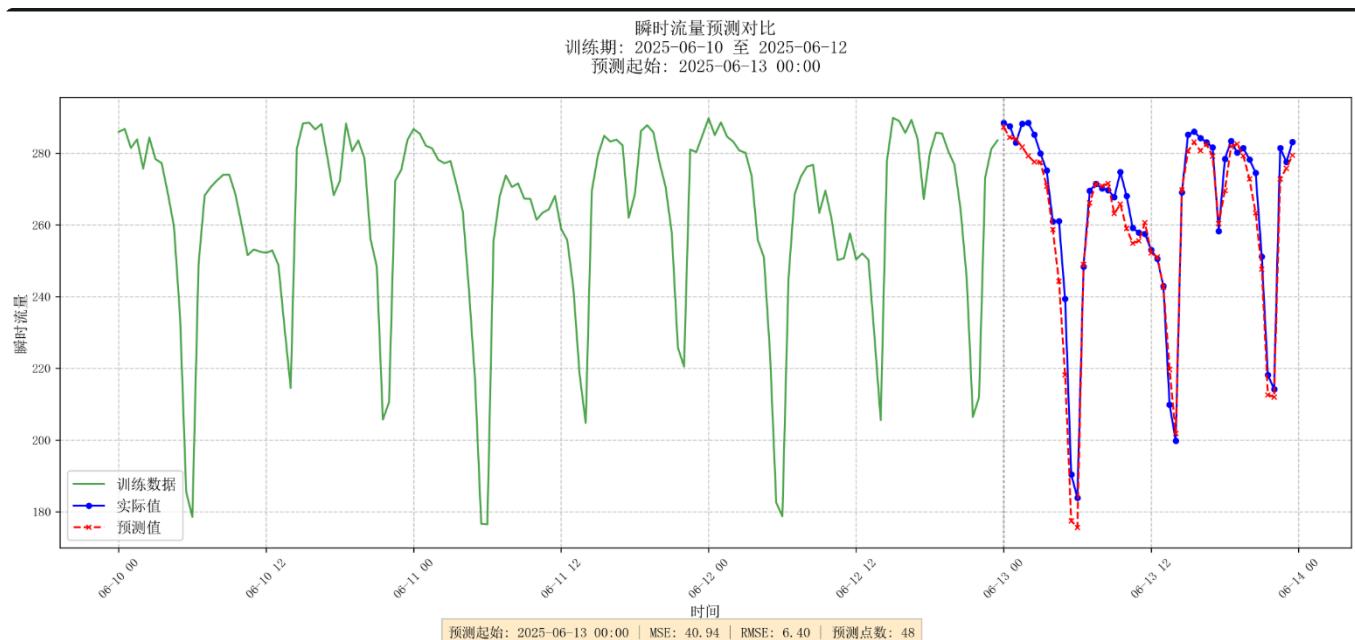
训练数据点数: 144

测试数据时间范围: 2025-06-13 00:00:00 至 2025-06-13 23:30:00

测试数据点数: 48

预测点数: 48

评估指标: MSE=40.9431, RMSE=6.3987



训练数据时间范围: 2025-06-10 00:00:00 至 2025-06-13 23:30:00

训练数据点数: 192

测试数据时间范围: 2025-06-14 00:00:00 至 2025-06-14 23:30:00

测试数据点数: 48

预测点数: 48

评估指标: MSE=1701.0832, RMSE=41.2442

