



AI CHALLENGER
全球AI挑战赛 2018

天气预报

参赛队：硬刚队

戴 玮 中科院自动化所

刘明桓 西南交通大学

邓金红 西南交通大学

问题描述

- 数据：每日的逐时温压湿风数据，包括观测值和睿图值（数值模式）
- 预测：未来37个时次的温度、湿度、风速
- 评价标准：**均方根误差RMSE和偏差BIAS**

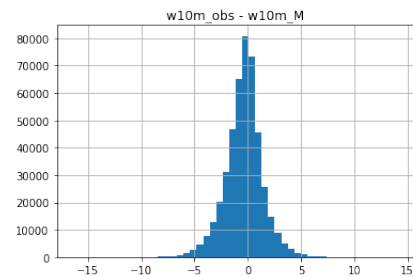
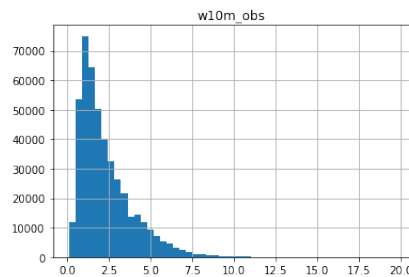
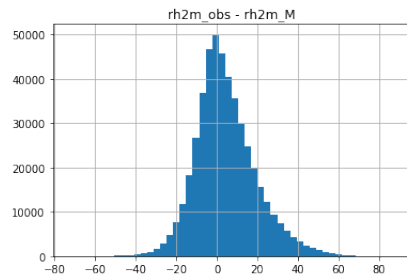
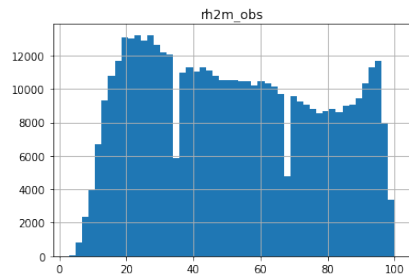
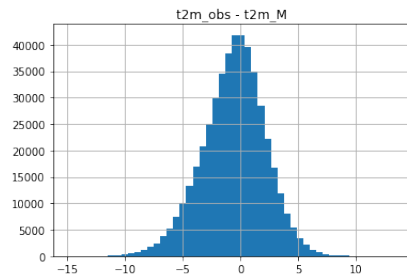
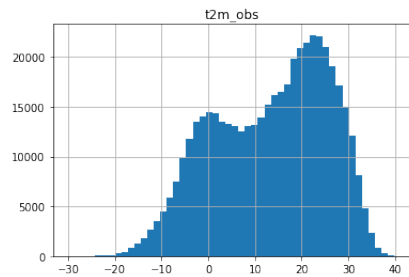
$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i^{obj} - X_i^{model})^2}{n}}$$

$$BIAS = \frac{\sum_{i=1}^n (X_i^{obj} - X_i^{model})}{n}$$

$$\text{Score} = (\text{RMSE}(\text{超算}) - \text{RMSE}(\text{选手})) / \text{RMSE}(\text{超算})$$

思路

- 评价标准是建立在已有睿图基础之上的
- 去预测真实值和睿图值之间的差，而不是直接预测真实值
- 更有利于拟合残差分布

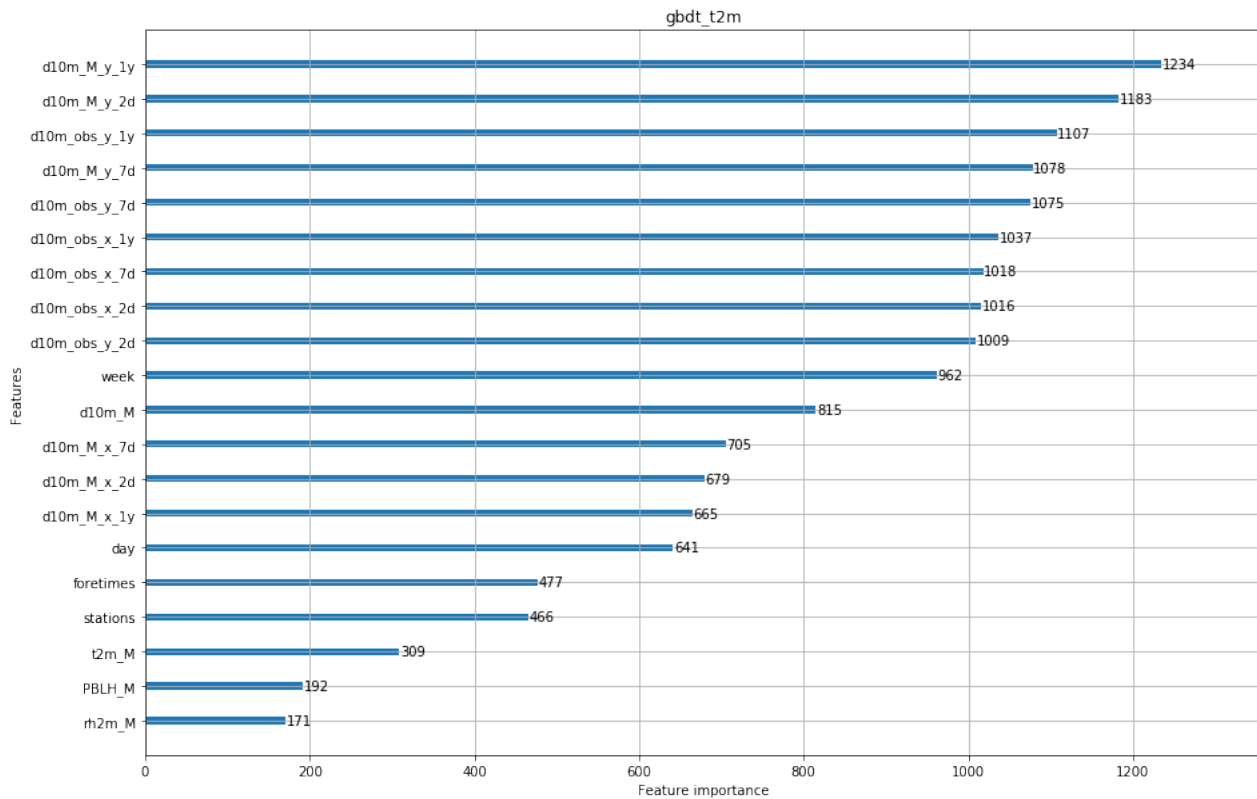


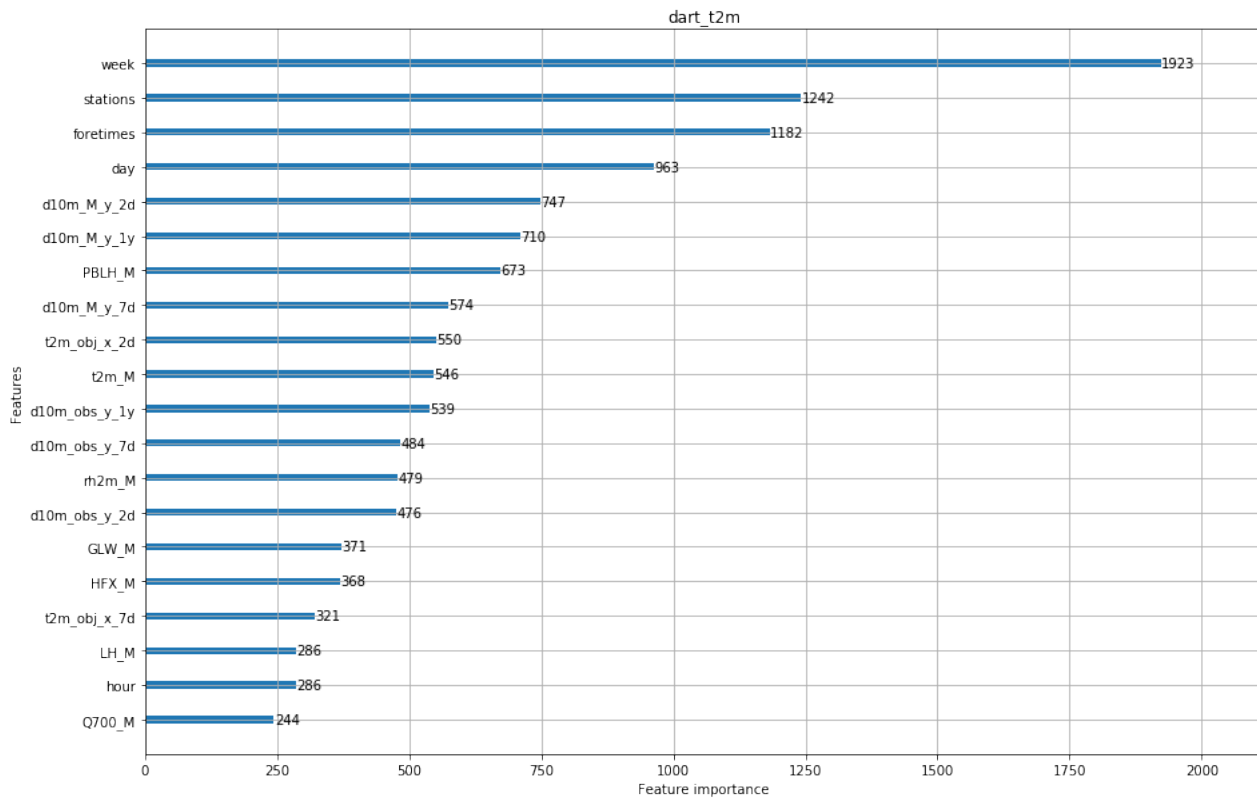
思路

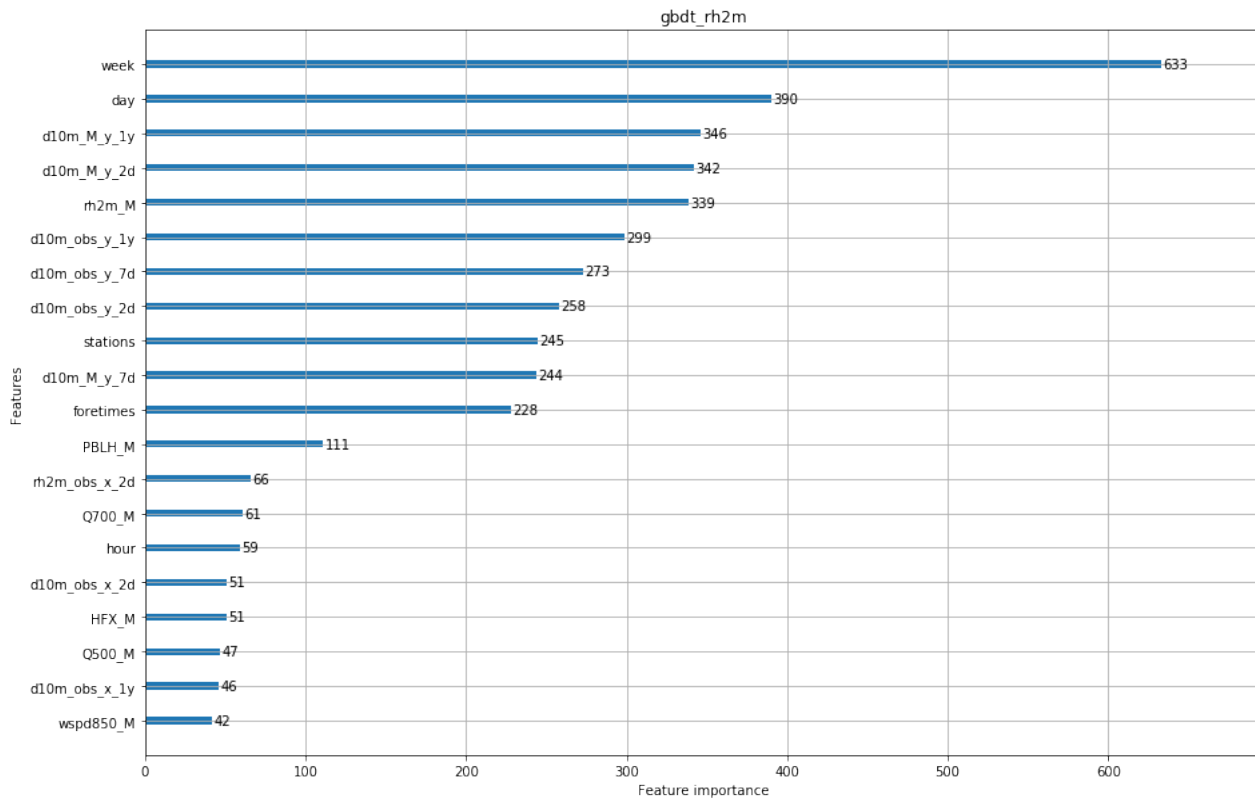
- 同一站点的天气在同一点钟具有相似性
- 利用了过去三个时间的特征：
 - 48小时前的同一点钟的天气
 - 一周一年前的同一点钟的天气
 - 一年前的同一点钟的天气
- 用所有站点的均值来平滑单一站点的差异

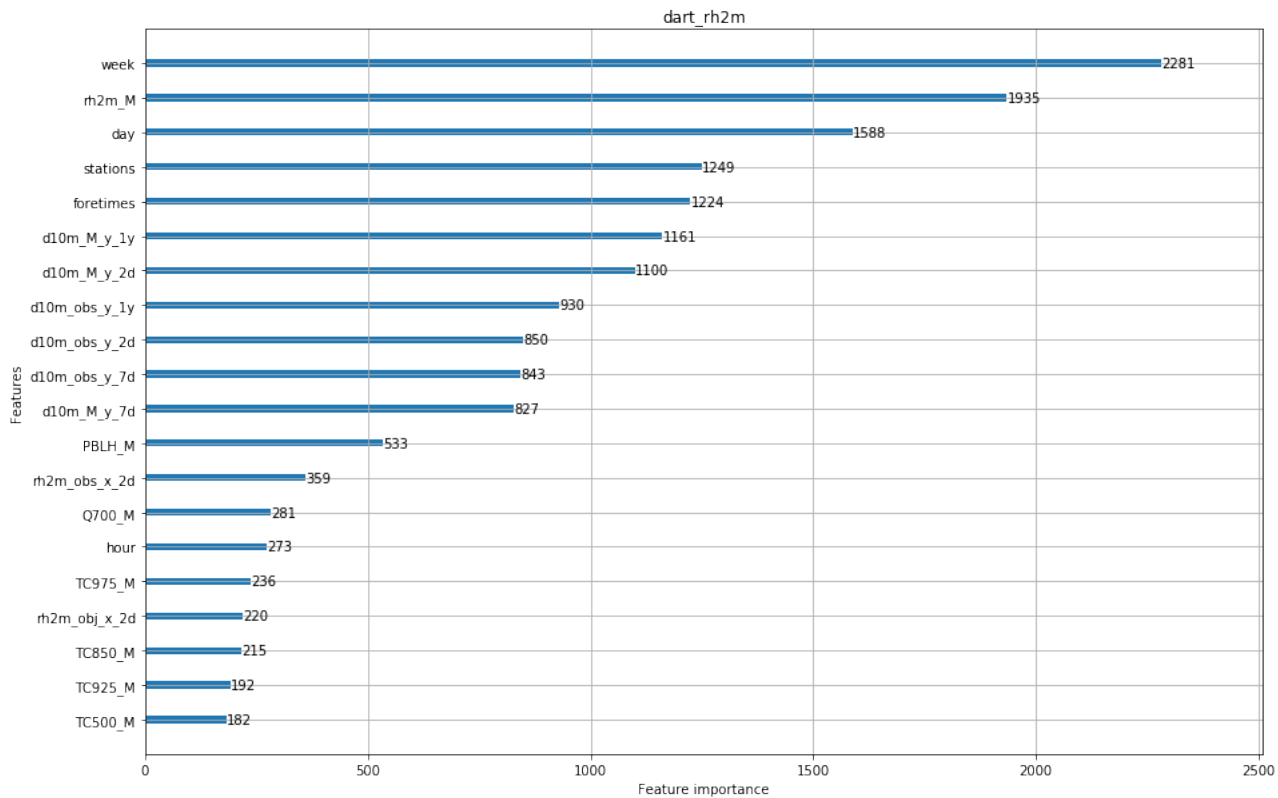
模型与算法

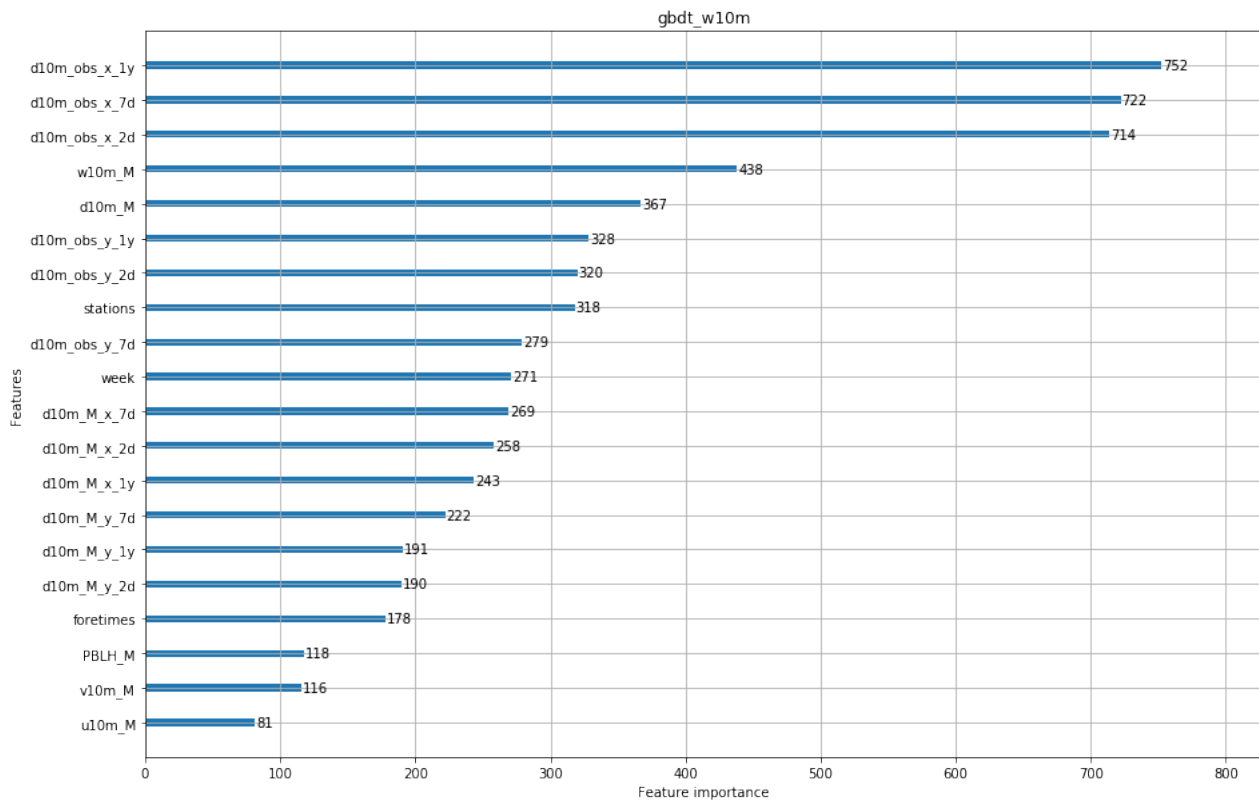
- Lightgbm
 - GBDT
 - DART
- 特征重要性

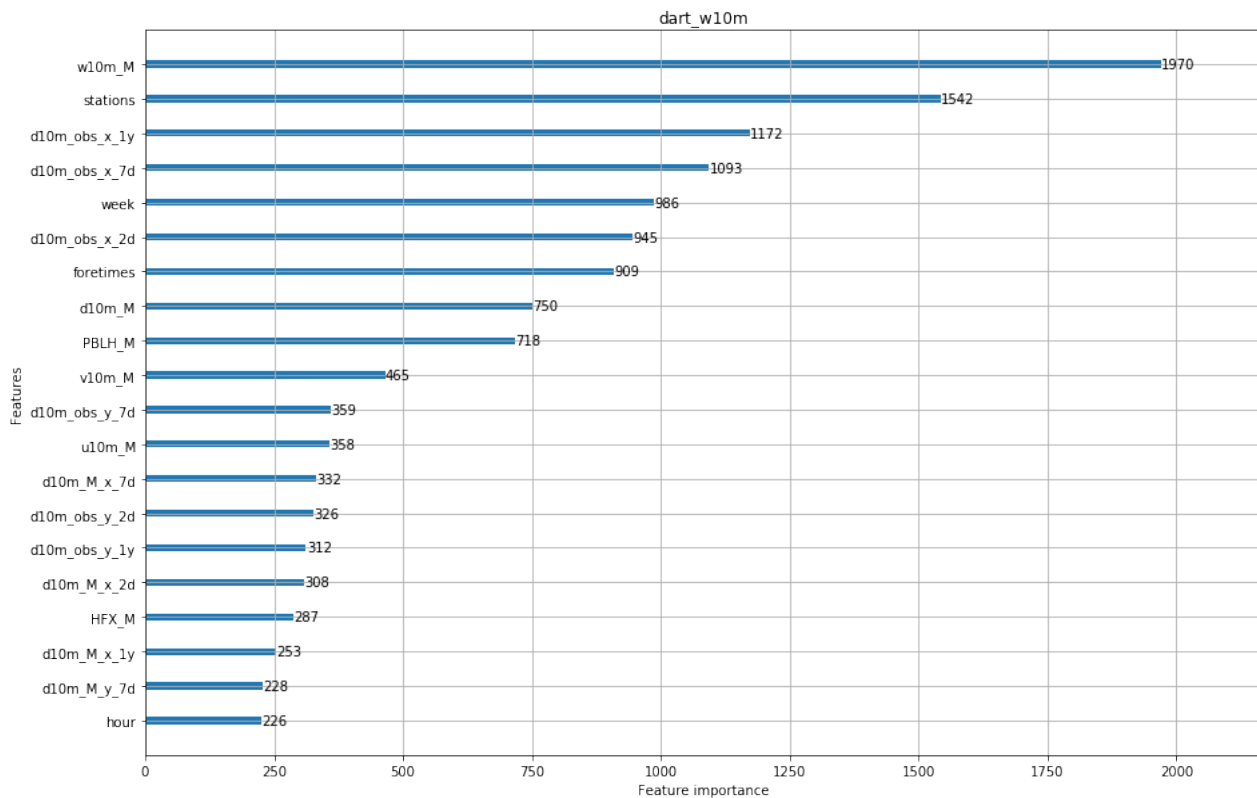




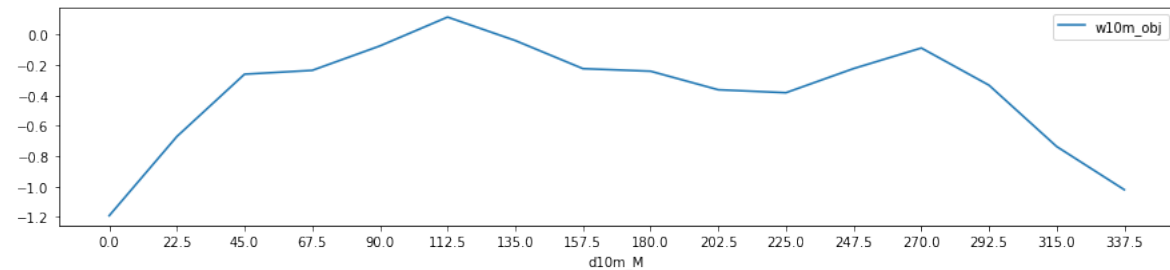
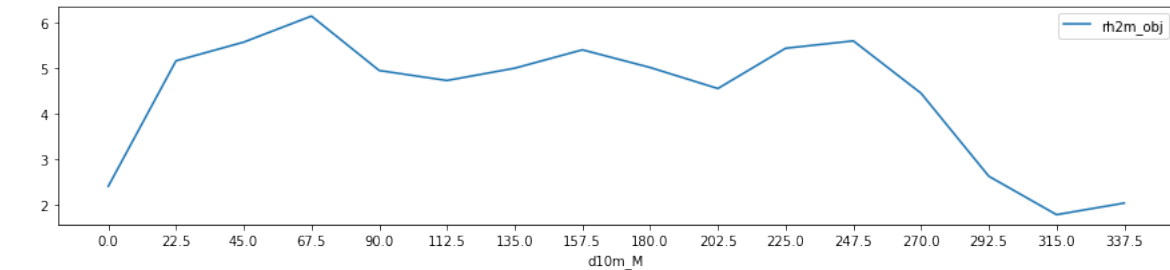
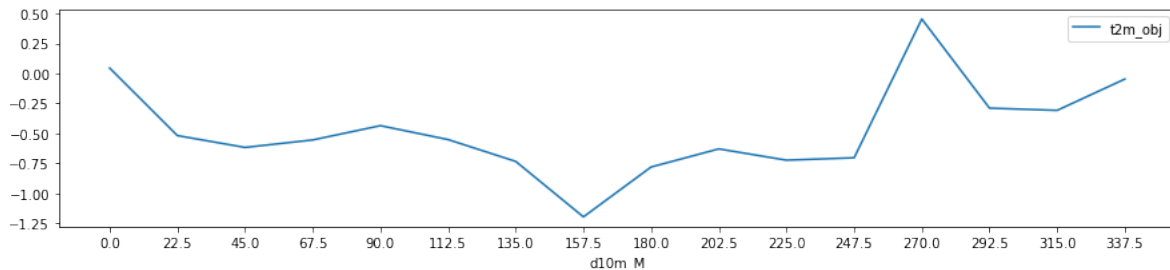




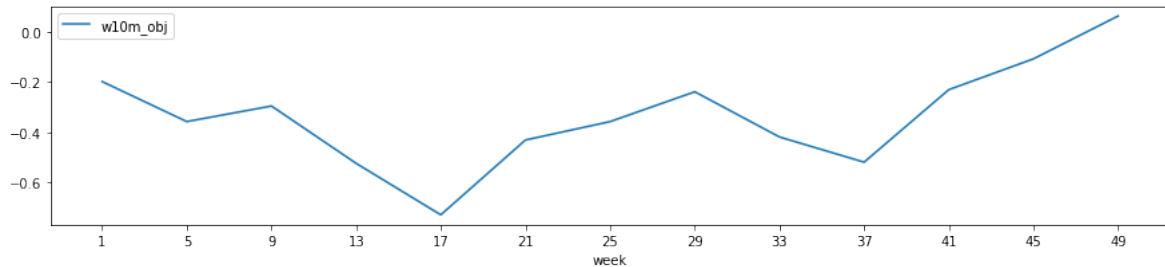
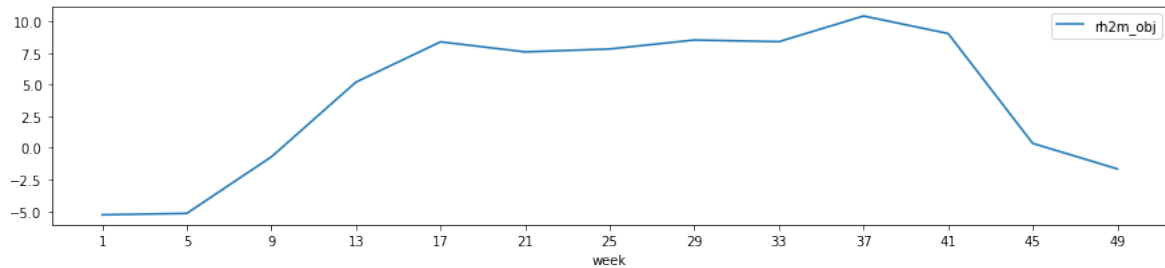
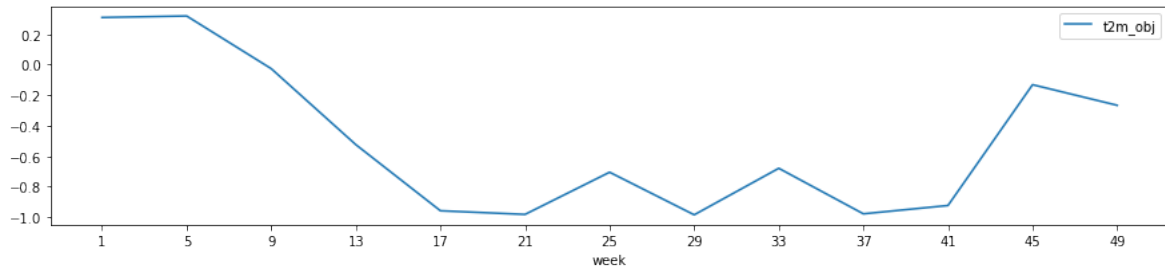




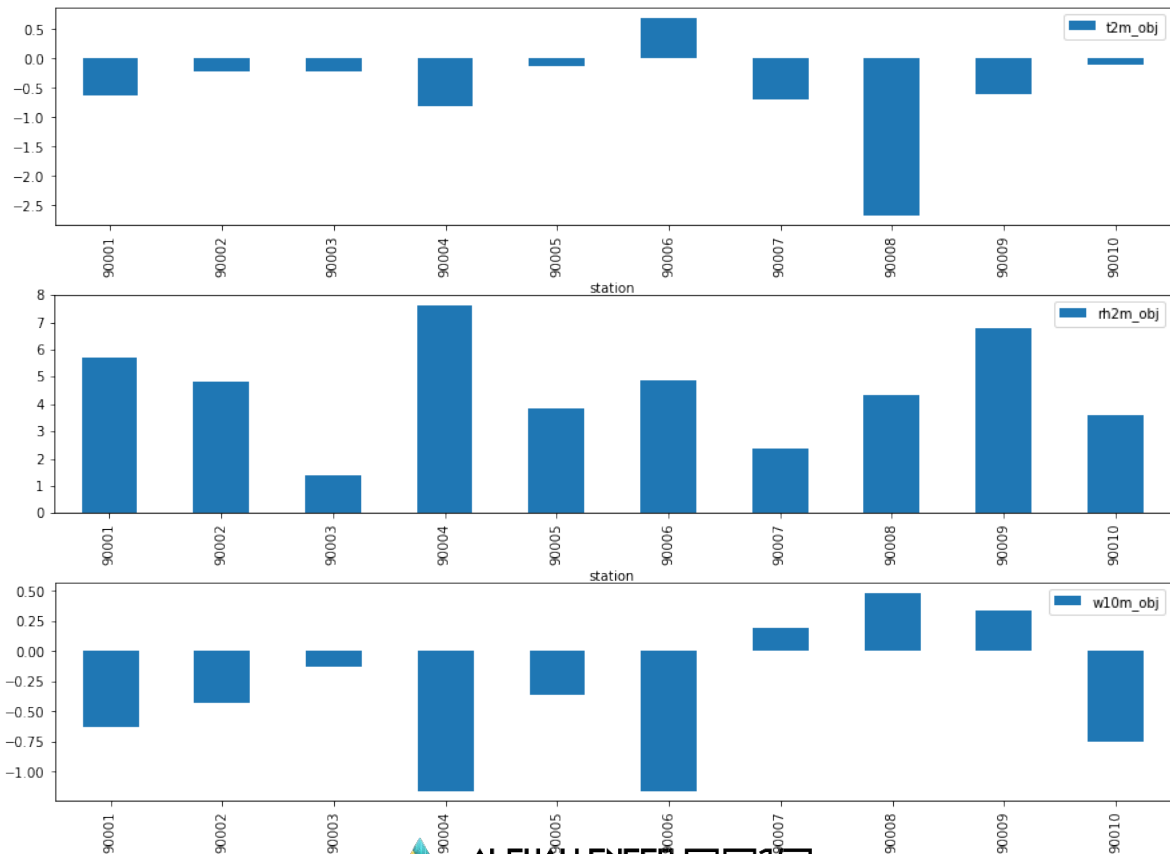
风向与预测目标的关系



周数与预测目标的关系



站点与预测目标的关系



难点

- 反常天气，比如冬天气温的异常升高
- 强对流天气，比如短时的狂风暴雨
- 对于数值模式和机器学习来说都很困难
- 也许只能结合雷达图等短临预报手段

疑问

- 气象站点的坐标以及空间信息
- 时间局限在10月末11月初
- 相对湿度的观测值似乎整体偏高

改进与展望

- 睿图值的修正
- 引入时序特征
- 数值模式和机器学习的优劣以及互补性
- 短临预报的特征和评价标准