

Поиск аномалий в данных с дорожной  
инфраструктуры  
с помощью экспоненциальной скользящей  
средней

Анна Маненкова, МИВТ-23-8  
Елизавета Еськова, МИВТ-23-3-2

**Дорожно-транспортная инфраструктура требует разработки**  
**алгоритма машинного обучения** для автоматического  
**обнаружения аномалий в данных от датчиков, обеспечивая**  
**оперативное выявление неисправностей и своевременное**  
**техническое обслуживание оборудования**

Затрудняют работу по обслуживанию и поддержке инфраструктуры:

наличие некорректных данных

отсутствие данных (null)

# Подход

## Подготовка и анализ данных

Feature correlation for 30928 station:

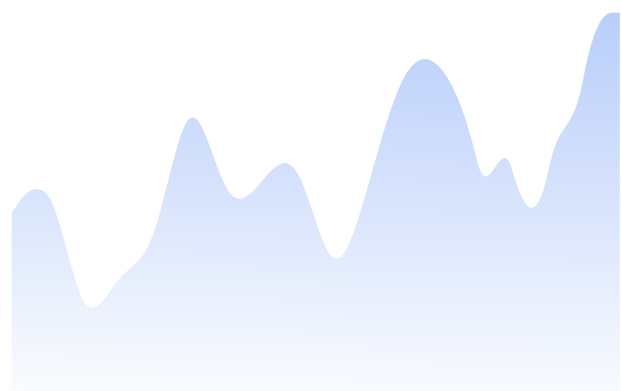
	Feature1	Feature2	Correlation
41	meteo_t_underroad	meteo_t_road	0.979978
27	meteo_wind_velocity	meteo_wind_gusts	0.954668
43	meteo_t_underroad	meteo_t_air	0.926252
61	meteo_t_road	meteo_t_air	0.919594
58	meteo_dew_point	meteo_t_air	0.648808
0	meteo_layer_type	meteo_cloudiness	0.567111
56	meteo_dew_point	meteo_t_road	0.505863
12	meteo_cloudiness	meteo_humidity	0.502819
34	meteo_humidity	meteo_t_road	-0.533152
30	meteo_humidity	meteo_t_underroad	-0.602907
59	meteo_dew_point	meteo_air_pressure	-0.697195

- сортировка данных по времени
- отдельная обработка данных по станциям

## Расчет ЕМА

Экспоненциальное  
скользящее среднее

+



=

$$EMA(t) = \alpha \cdot data(t) + (1 - \alpha) \cdot EMA(t - 1)$$

обеспечивает сглаженное представление  
данных,  
подчеркивая последние тенденции и  
затушевывая старые наблюдения.

## Обнаружение аномалий

**Аномалия, если:**

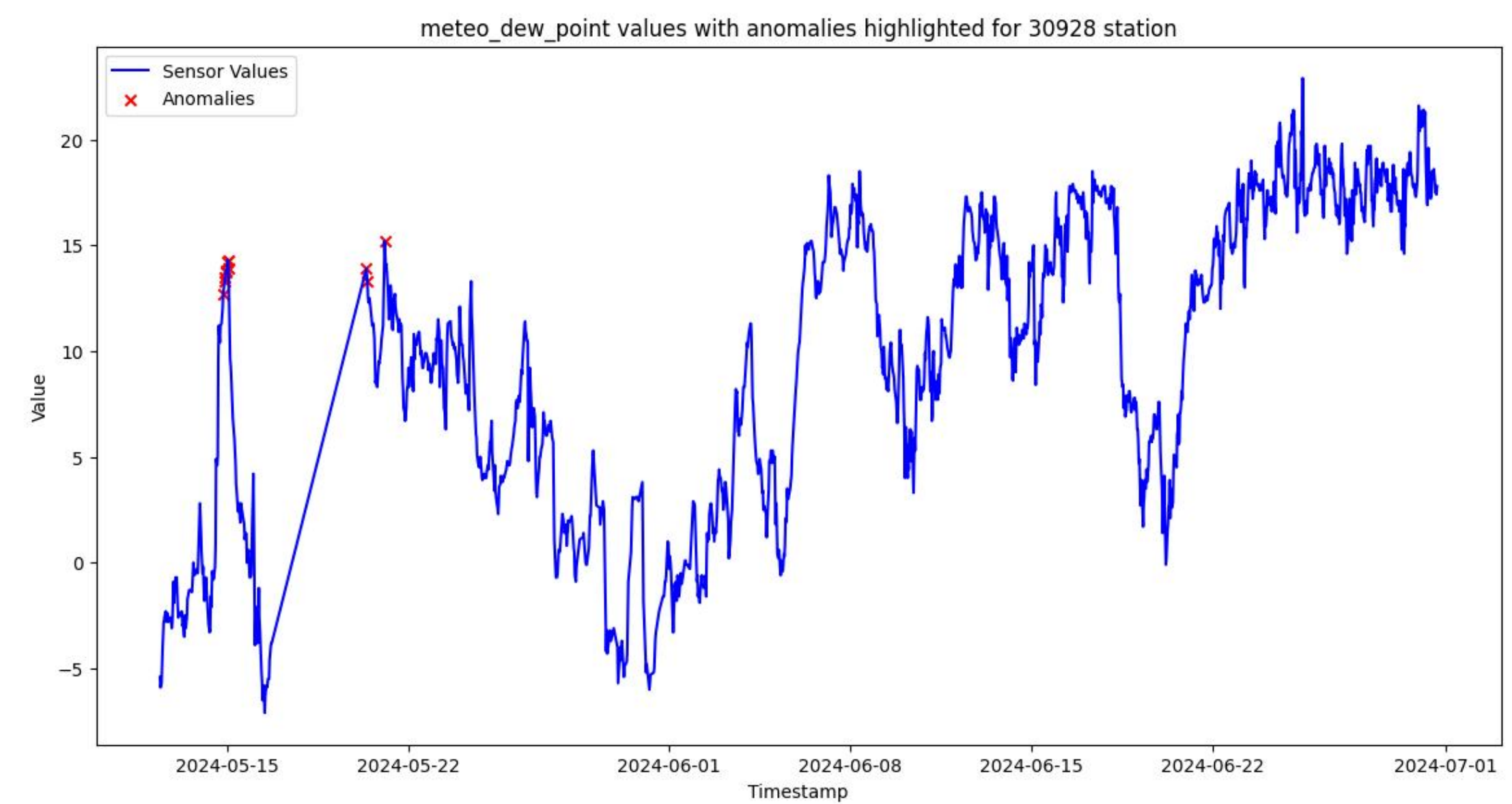
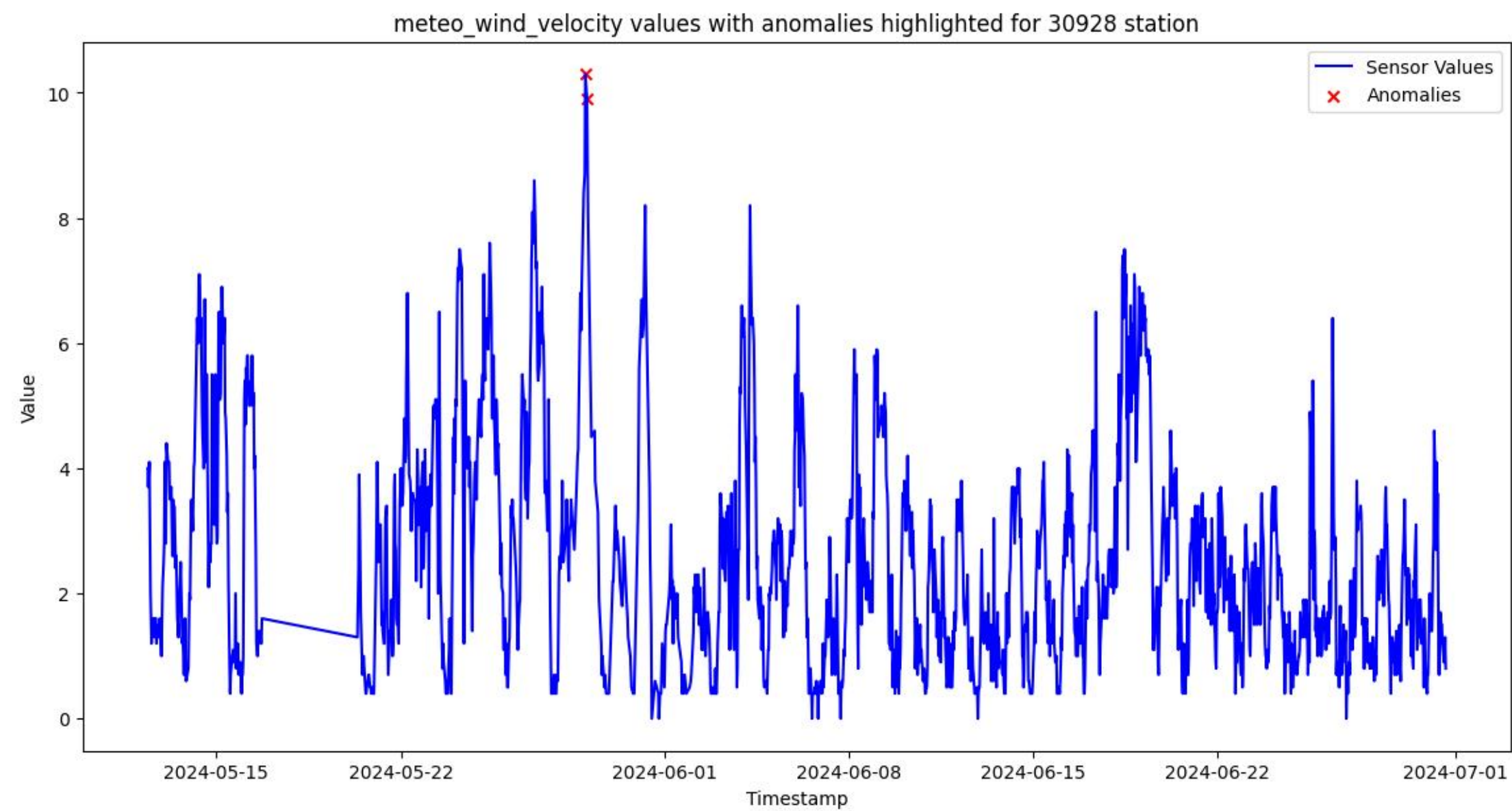
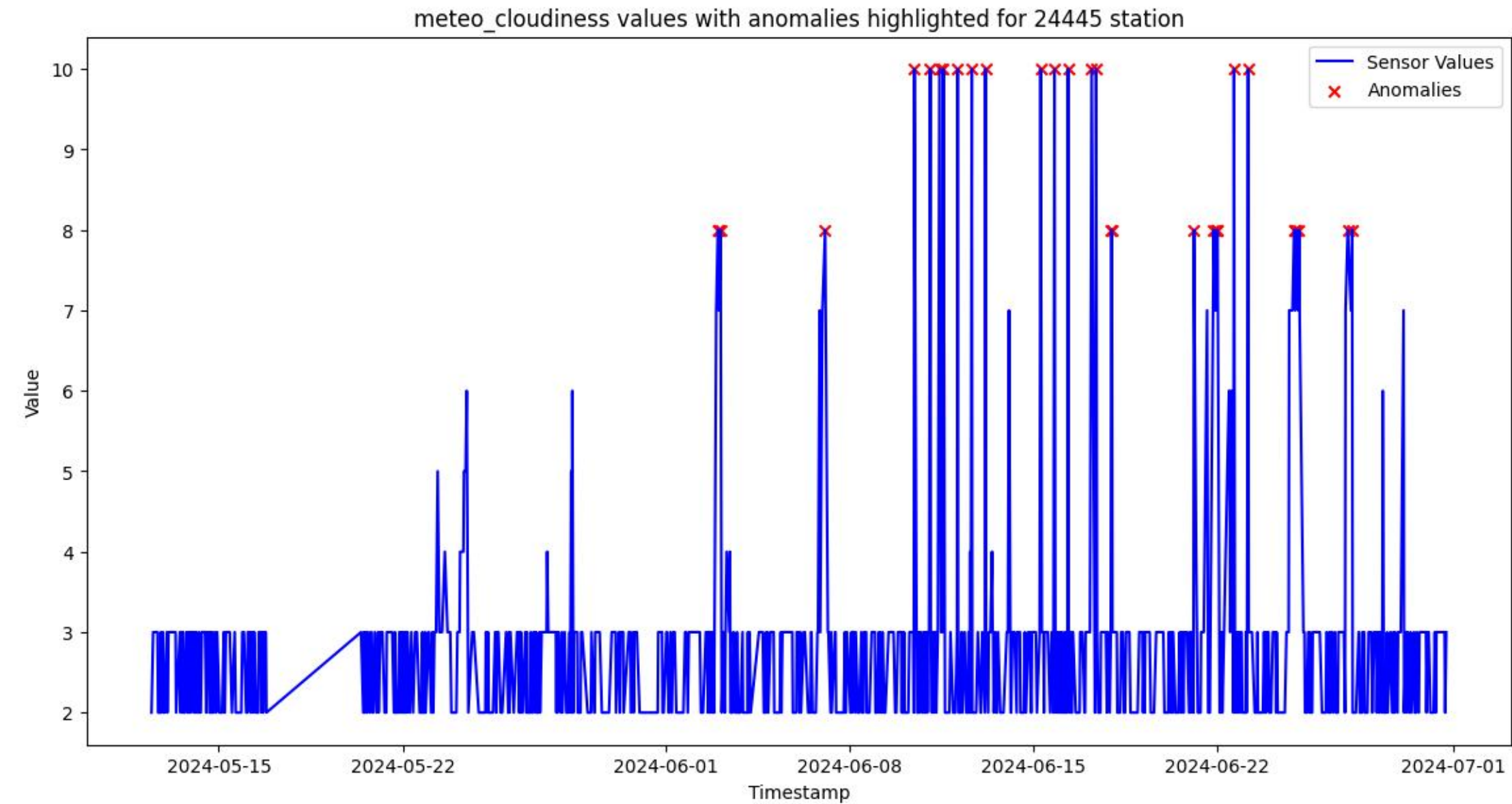
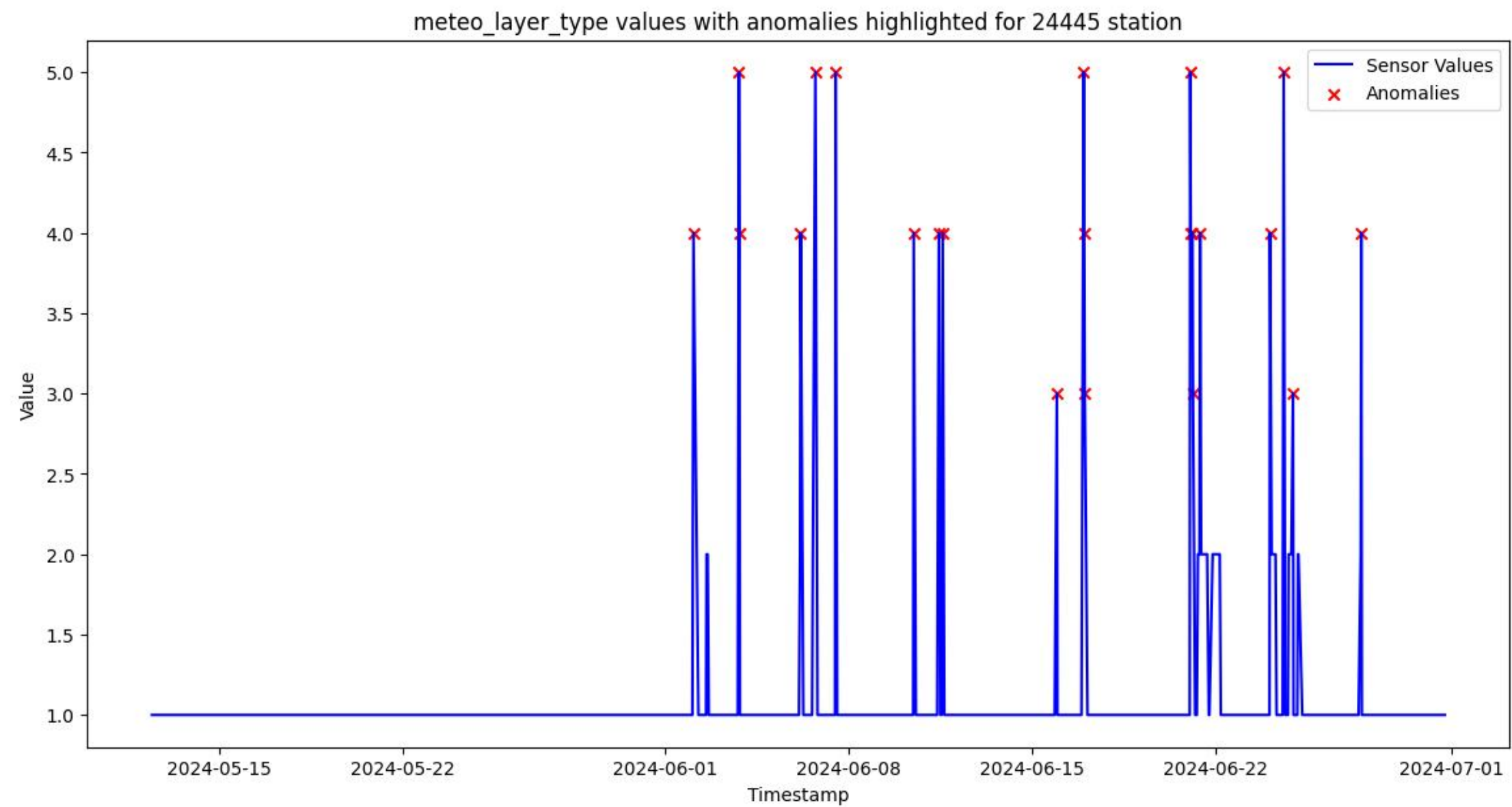
**абсолютное значение остатка** (разница  
между фактическими показаниями  
датчиков и соответствующими  
значениями ЕМА) **превышает заданный  
порог.**

**Порог = n \* стандартное отклонение  
остатков.**

тренировочные и тестовые данные = временной ряд → объединяем данные

# Визуализация

F1 macro = 0.18



**Спасибо**  
**за внимание!**