

Вопросы и рабочие замечания по анализу летних данных

1. Данные и разметка

- В наличии только летние данные;
 - Разметка участков `bl` и `stat` — только авторазметка, ручная отсутствует.
 - Доля точек с `bl=True` очень мала.
 - Необходима проверка параметров и логики авторазметки: какие критерии и пороги? Почему так мало выделено плато?
-

2. Использование данных для моделей

- Для baseline и чувствительности брать только стационарные (`stat=True`) и, для baseline, только `bl=True`.
 - Использовать ли концентрации других газов в качестве признаков (feature engineering) — отдельный вопрос для анализа.
-

3. Кросс-чувствительность и корреляция газов

- Для анализа кросс-чувствительности обучить линейную модель сигнала датчика по всем концентрациям:

$$\text{Signal_NO2} = \beta_0 + \beta_1 * [\text{NO2}] + \beta_2 * [\text{CO}] + \beta_3 * [\text{SO2}] + \dots + \varepsilon$$

- Проверить, есть ли значимые коэффициенты при “чужих” газах.
 - Обязательно оценить корреляцию концентраций между газами (иначе возможна мультиколлинеарность и ложные β).
-

4. Модель сигнала и признаки

- Базовая модель: сумма baseline (функция T, RH) и чувствительность к газу.
 - В случае сильного дрефта — рассмотреть добавление временного тренда или других корректирующих факторов.
 - Производные по времени и rolling-признаки использовать только вне плато, для анализа динамики, не для baseline.
-

5. Дообучение baseline и drift

- Рассмотреть возможность дообучения baseline по выделенным плато на уличных данных.
 - Выделять дополнительные baseline-участки автоматически (по std сигнала, стабильности T/RH и синхронизации по датчикам).
 - Оценить drift baseline и влияние деградации сенсоров.
-

6. Вопросы для коллег и задачи

- Как работает авторазметка bl/stat, есть ли возможность скорректировать параметры?
- Запросить данные по датчикам G3 (для расширения анализа).
- Есть ли дополнительные уличные данные (с других сезонов или других постов)?
- Какие физические формулы могут дополнительно связывать параметры (например, зависимости сигналов от T, RH)?