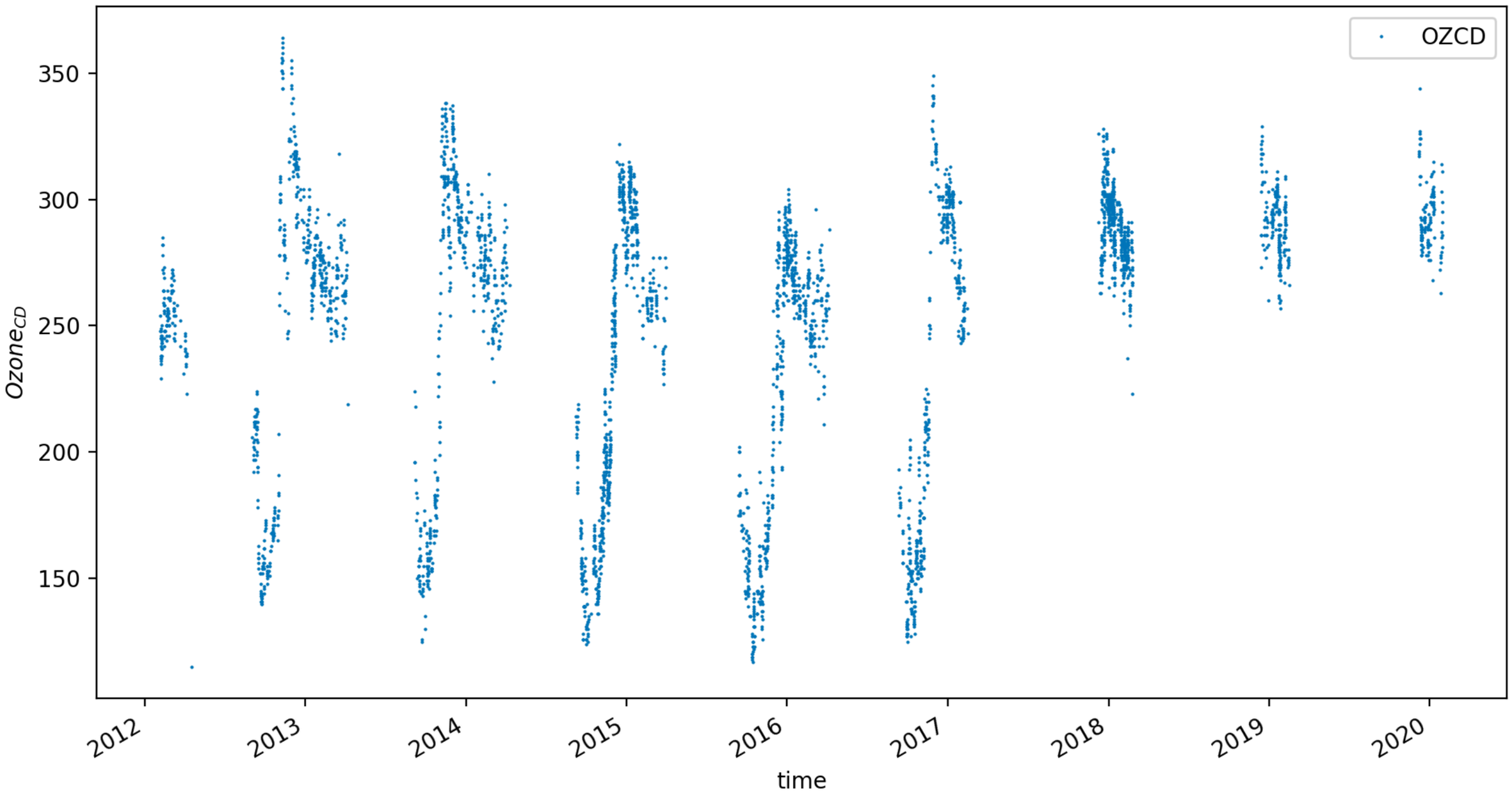
Update on progress

- Time interpolation model is set up
    - Cannot naively use seasonal trend decomposition due to missing data, noise and outliers
    - Identified features
    - Tried several models: tsdl , GAM (mgcv), GP (numpyro)

Next step:

- Regress on observables:
    - scattered light wave-length,
    - atmospheric pressure,
    - solar zenith angle
- Need some details:
    - How are measurements usually taken (location, timing)
    - Do we expect Ozone to change within a day (observations have minutely resolution by minute)
    - Confirm some of data fields and their dimensions

Peter Yatsyshin

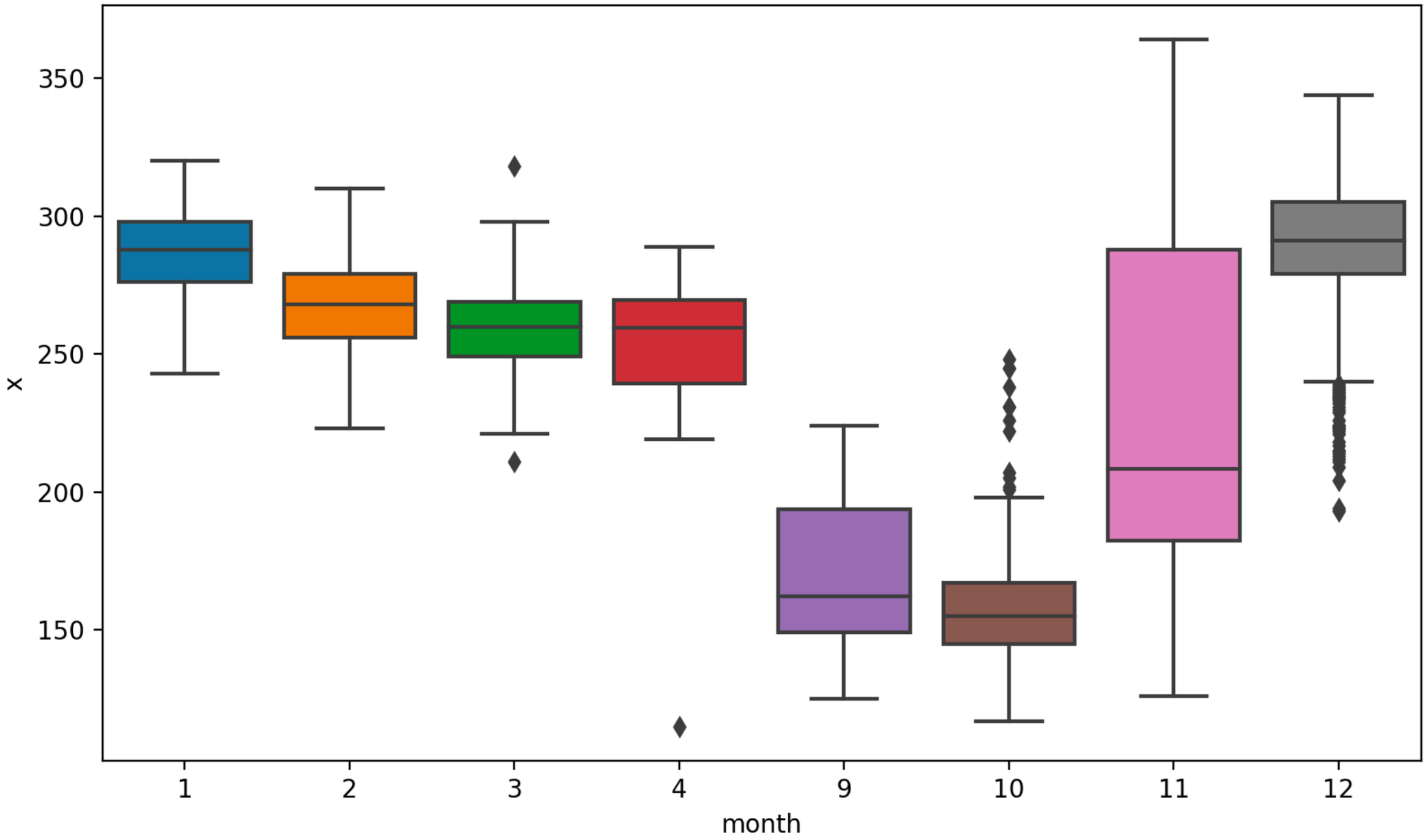


- Minute resolution
- Measured during Jan - April and Sep - Dec
- Missing data, noise, outliers

Peter Yatsyshin

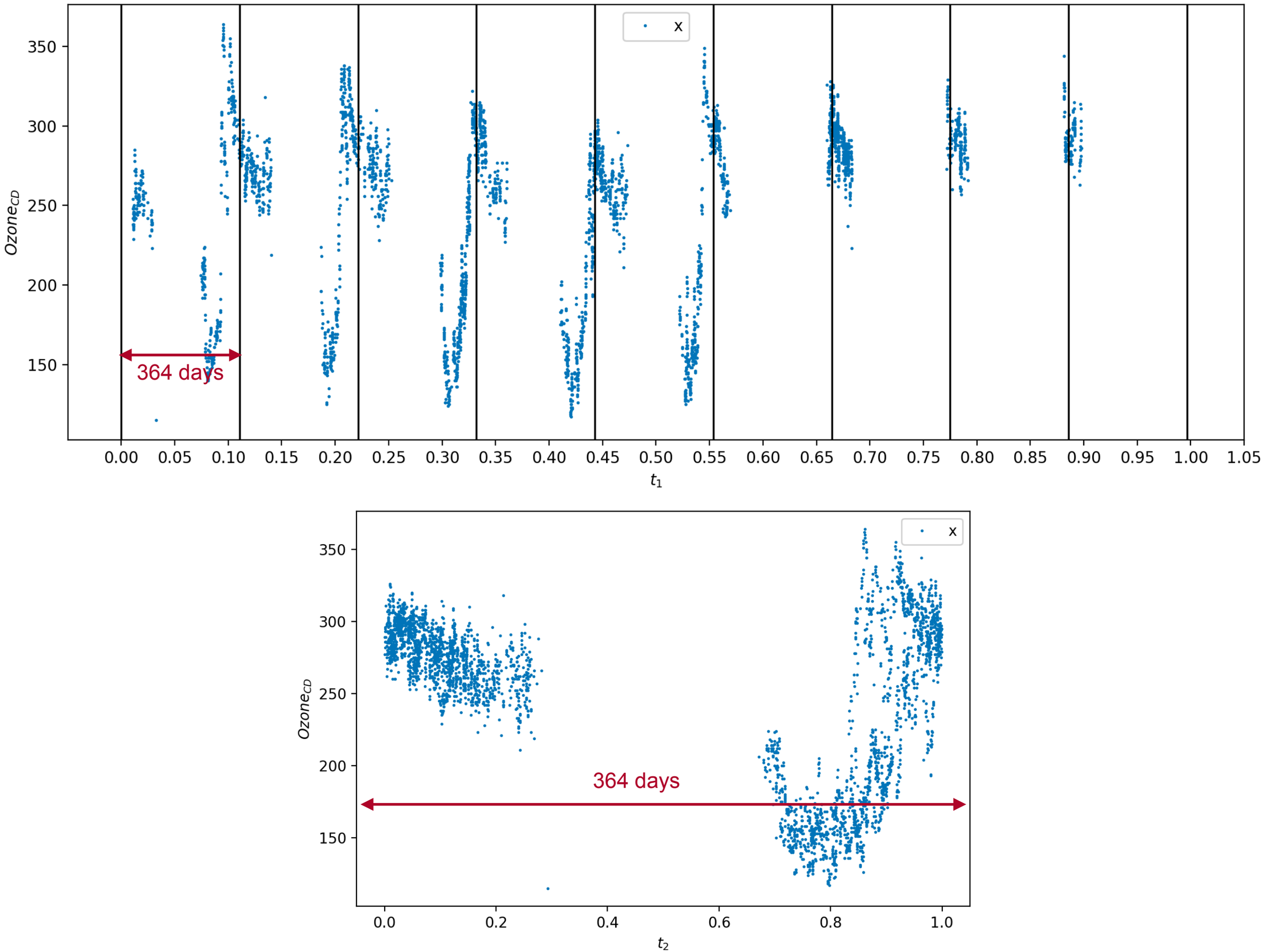
Seasonality? Yes

Let’s consider annual feature



Peter Yatsyshin

Set up features

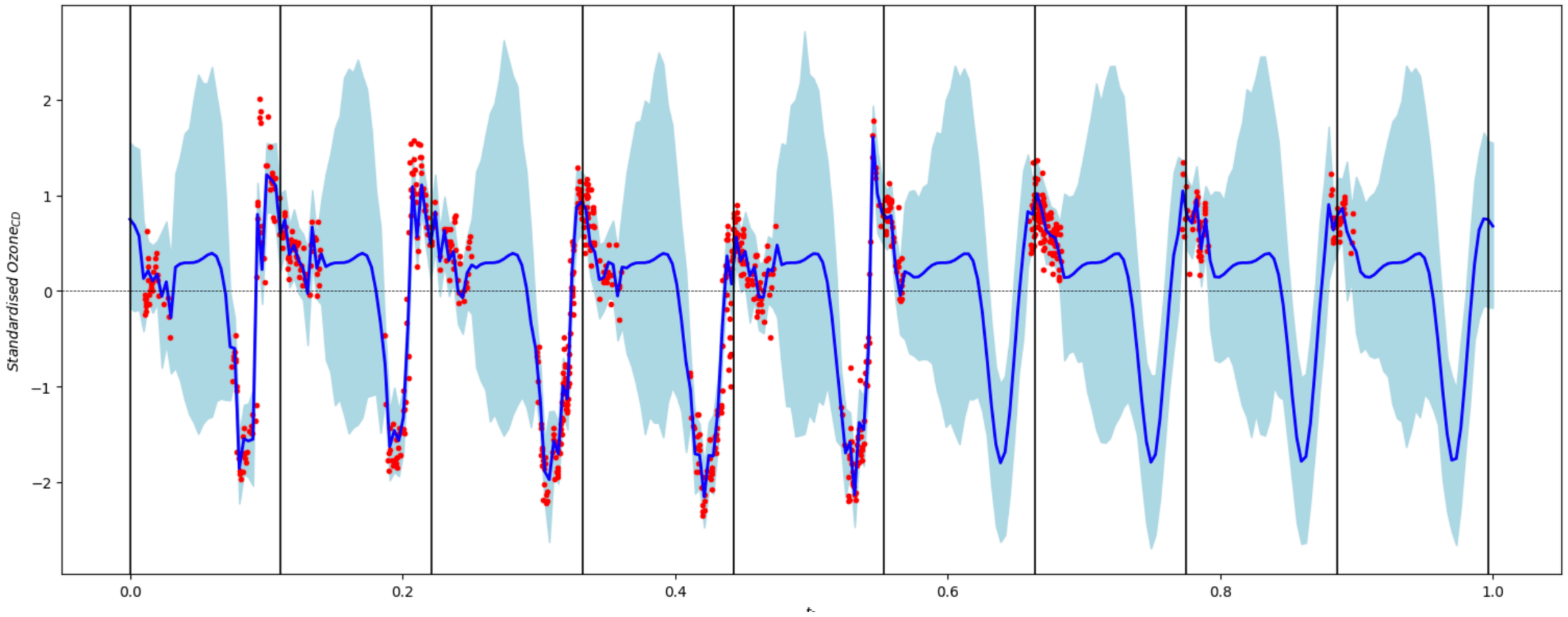


Simple model, which is easy to extend to other features

$$f = f_1(t_1) + f_2(t_2),$$
$$f_1 \sim GP(0, K1),$$
$$f_2 \sim GP(0, K2),$$

where $t_1 \in [0,1]$ and $t_2 \in [0,1]$

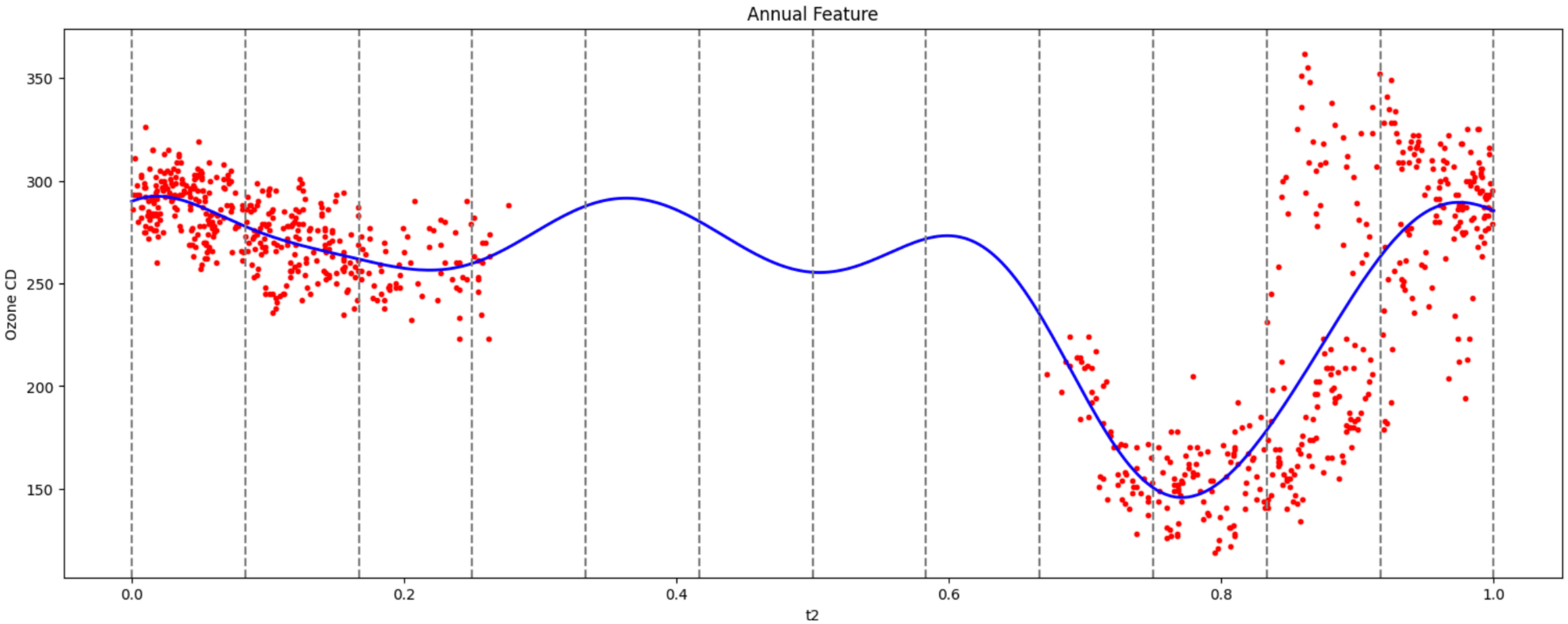
While t_1 corresponds to rescaled measurement time, t_2 changes within one year (52 weeks).



This model was set up in numpyro. See our GitHub repo.

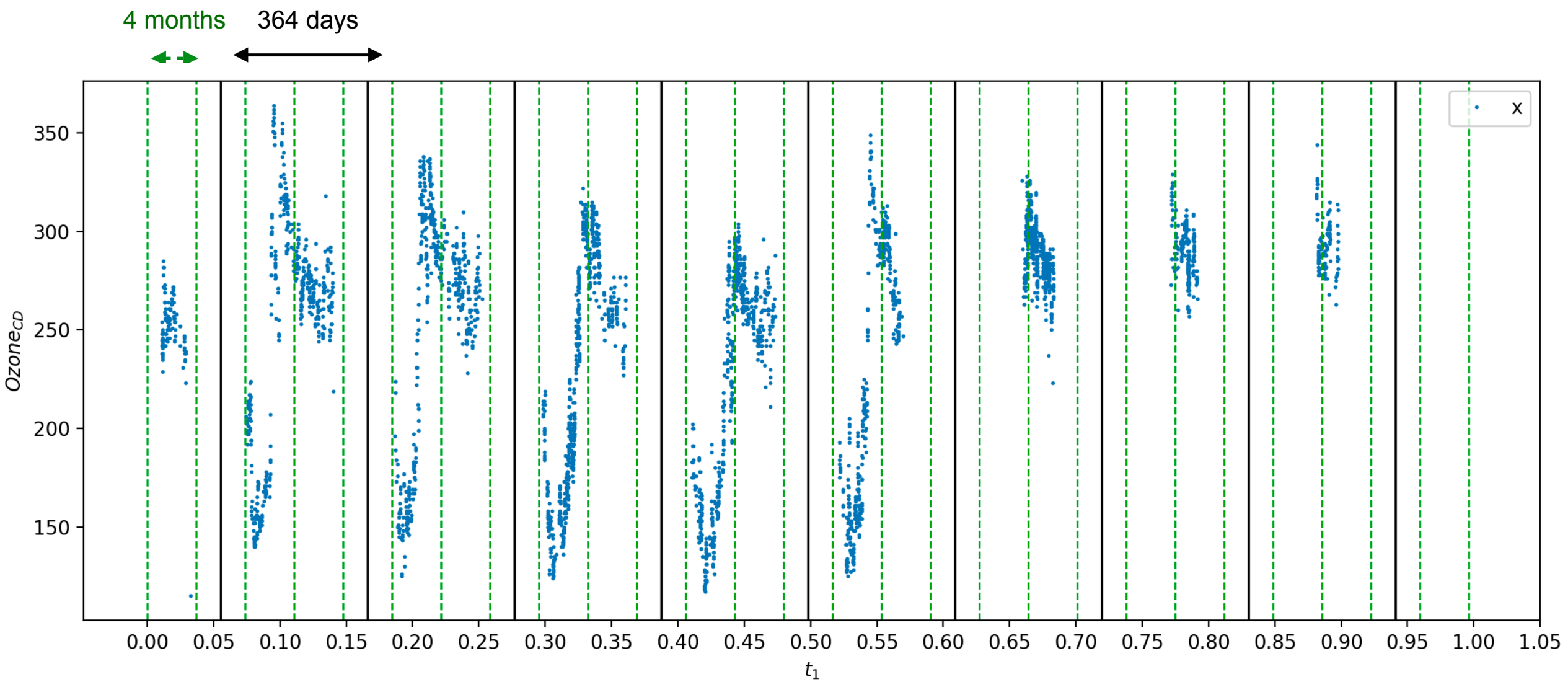
Peter Yatsyshin

Fit on the annual feature space



Peter Yatsyshin

Some refinement is in order



- Better to start the year on May 1
- Tighten priors
- Maybe, bin to daily
- Transform
  - log
  - diff
  - diff(log)

Peter Yatsyshin