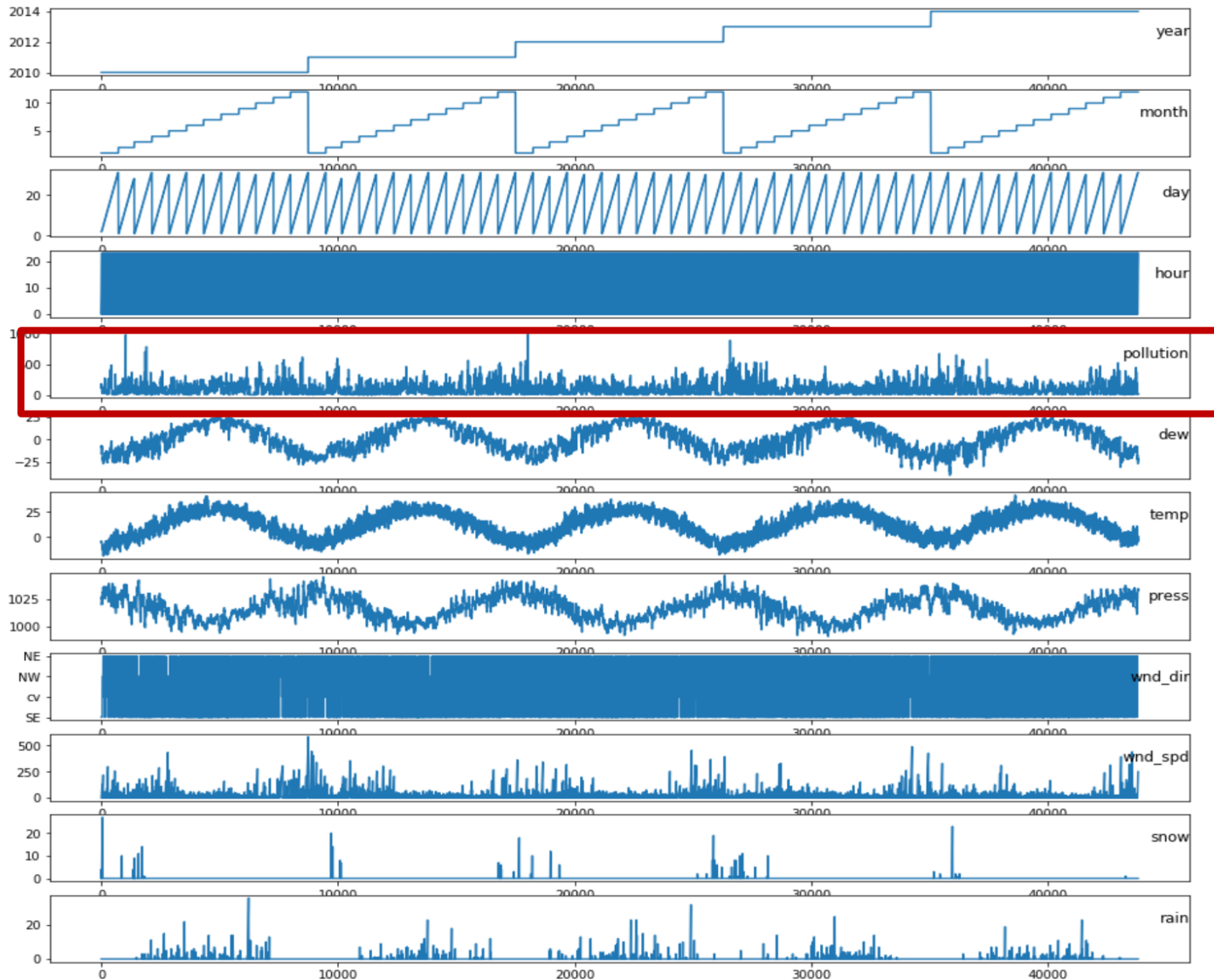


THIRD GRADED ASSIGNMENT

Multi-modal weather data (Beijing)

Dataset has 43800 samples and 12 features

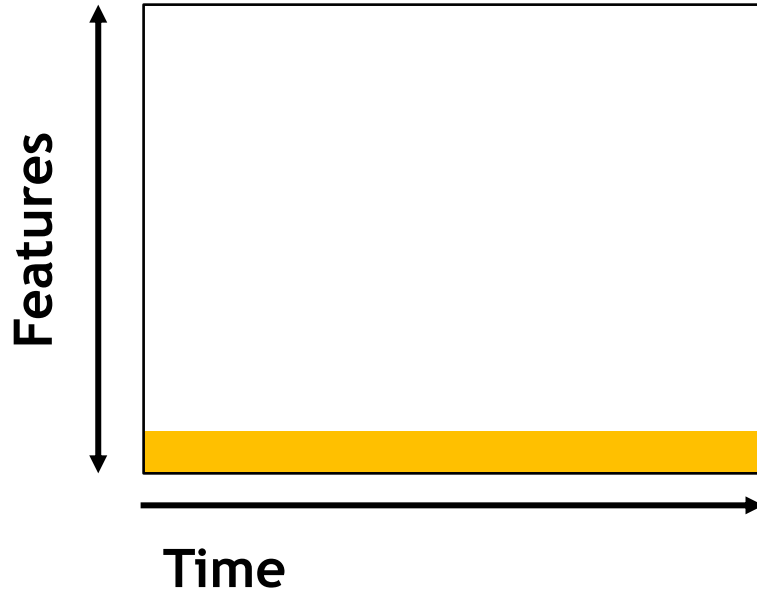


Hourly data
5 years

Task:
predict
pollution
a few hours
ahead

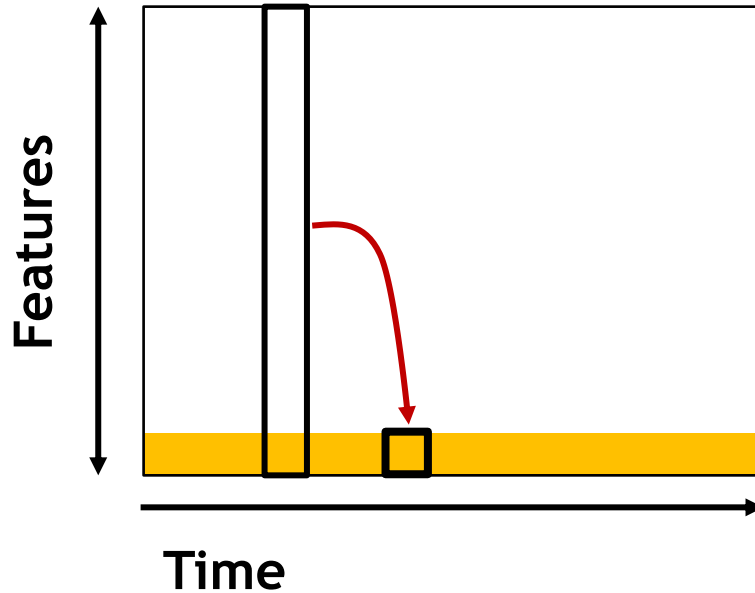
=regression!

Multi-step assignment



- **Step 1:**
multiple feature baseline
without history (Dense)
- **Step 2:**
sequential baselines using
current pollution +
pollution history
(1Dconv + RNN)
- **Step 3:**
multiple features + history
- **Step 4:** ensemble

Multi-step assignment: step 1

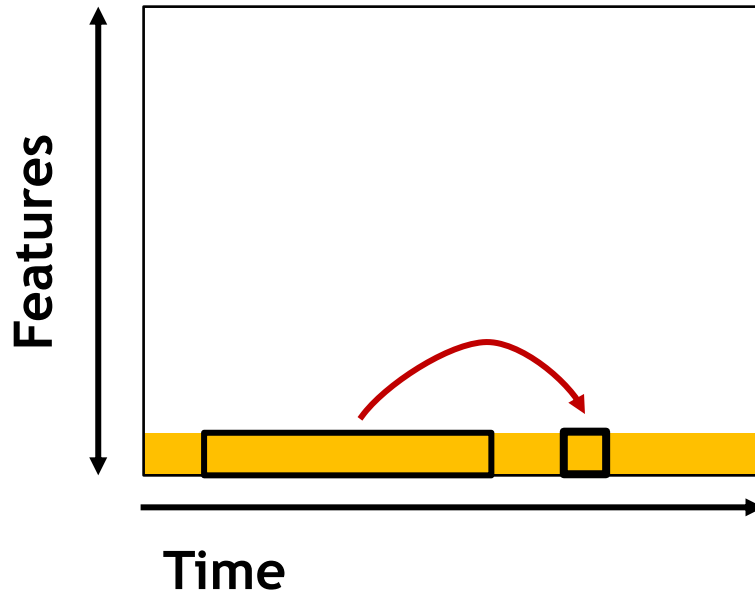


Step 1:

multiple feature baseline
without history

- Use all current features (including current pollution) to predict future pollution
- Use a dense network
- Explore feature importance and feature selection

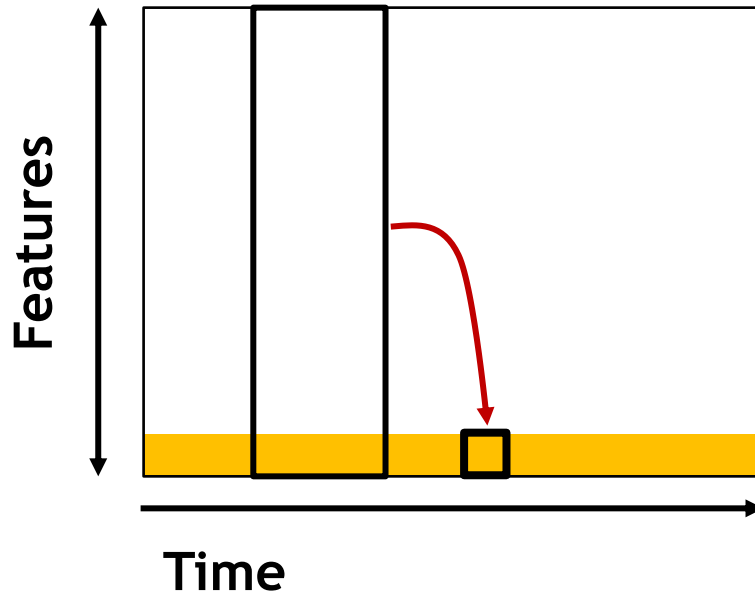
Multi-step assignment



Step 2:
single feature sequential
model

- Use pollution history to predict future pollution
- 2 types:
 - ◆ 1D conv (find good window size!)
 - ◆ LSTM or GRU (find good backprop depth)
- Compare and discuss

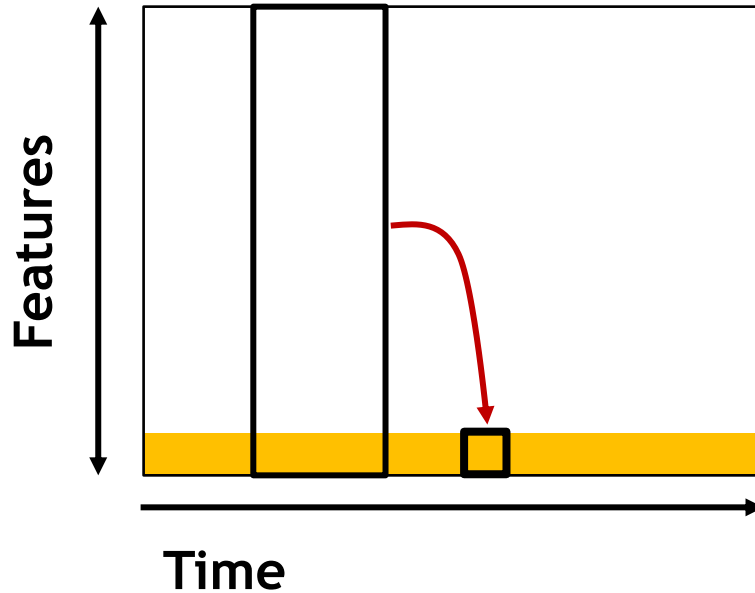
Multi-step assignment



Step 3:
multiple features + history

- To get started:
 - ◆ use experience from step 1 to decide which features to use
 - ◆ use experience from step 2 to decide which type of network to start with
- Can try more complex structures if you have time (bonus points)
 - ◆ stack of 1D conv + LSTM/GRU
 - ◆ stateful model

Multi-step assignment



Step 4:
ensemble

- Minimum: try to select promising models to combine (which? check theory) - can be models from different steps!
- Other routes (bonus points):
 - ◆ try to purposefully train models that make different mistakes (e.g., using different subsets of features)
 - ◆ combine models in ensemble and fine-tune or train combination weights
 - ◆ ...

Third graded assignment

- A lot to do: don't go too far on minor improvements
 - ◆ try to be efficient, based on what you already learned from previous assignments
 - ◆ Everyone needs to be able to do the actual coding - no grades for non-coding tourists!
- Reporting:
 - ◆ Methodology and results + plots for keypoint decision steps!
 - ◆ Learning convergence first!!
 - ◆ Here: 'powerful enough' is ill-defined, but since you have very little training data, don't go far!
 - ◆ Score mostly based on understanding and methodology!!
- If you prefer documenting a notebook:
 - ◆ make sure it is clearly documented (I will not search for snippets)
 - ◆ make sure all code is properly executed
 - ◆ export as html and (ideally) print to pdf
- Submit notebooks that reproduce final models and analyses plots
- Hand in a single zip-file through Ufora, before the deadline:

Friday, March 20th, 23:30