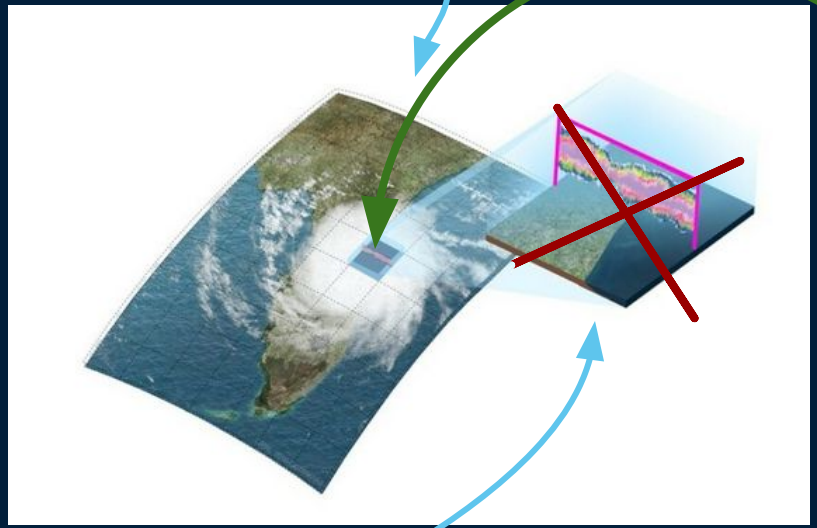
LEAP:
Atmospheric Physics using AI

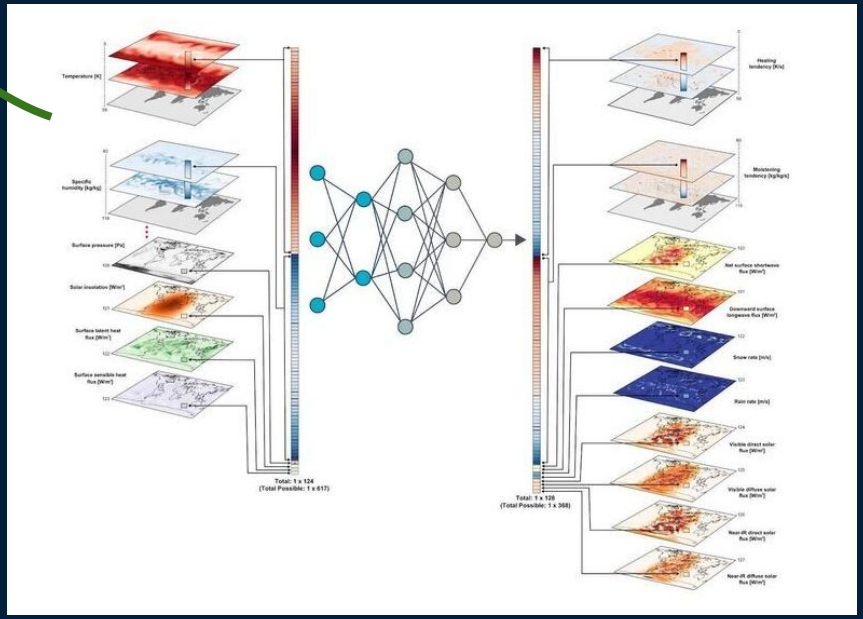
Final Presentation

Objective: a data-driven parametrization of small scale processes

numerical model



fine-scale model = expensive!!

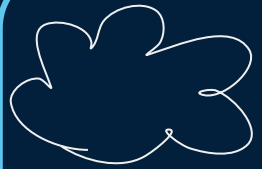


-> parametrizations ... not always very accurate

The Approach: Divide in 5 Categories

Temperature:

- heating tendency profile



Clouds:

- moistening tendency profile, liquid clouds & ice clouds

Wind:

- zonal acceleration
- meridional acceleration

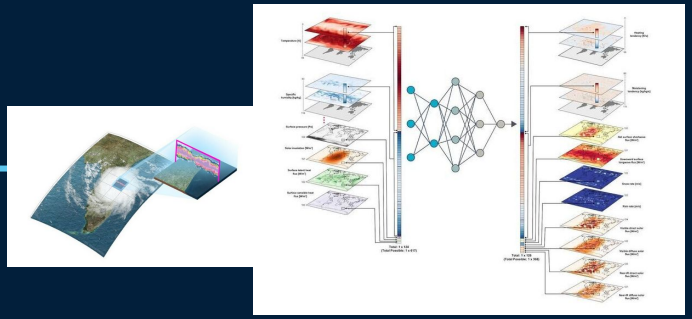
Radiation:

- shortwave radiation (5 individual parameters)
- longwave radiation

Precipitation:

- rain rate & snow rate

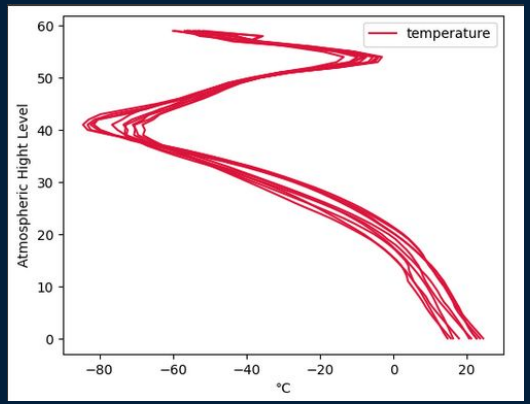
Data Pipeline: The Dataset



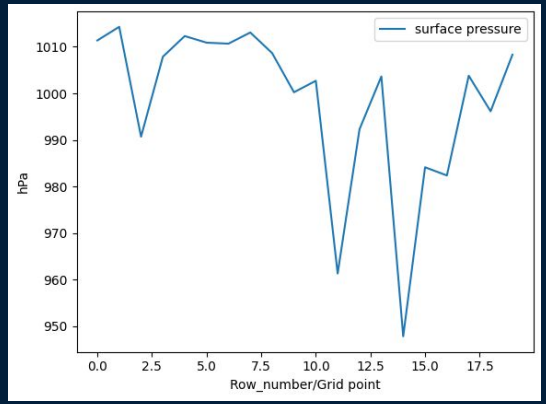
CSV file (~300GB) containing the training data:

- each row represents one point in space and time (~10 MIO rows!)
- 556 columns corresponding to **25 input** variables (from coarse model)
- 368 columns corresponding to **14 target** variables (from fine model)

some inputs/targets are vertical profiles:



others are just “scalars”:



Data Pipeline: Preprocessing

Feature Selection

-> based on physical knowledge and linear correlations

-> a dictionary that maps the relevant input feature to the targets

Normalization

-> column wise to not mix up different units

-> scaling with min and max to [-1,1]

Formatting & Splitting

-> 1D feature tensors and 1D target tensors of varying length

-> 100000 row train set, 80% train, 20% validation

-> 1000 row test set

A Family of Models:

train an individual model for each (sub)category:

- Temperature
- Clouds
- Wind (2 models)
- Radiation (2 models)
- Precipitation

baseline: simple MLP (two hidden layers,

ReLU,

number of neurons depend on input size)

A Family of Models:

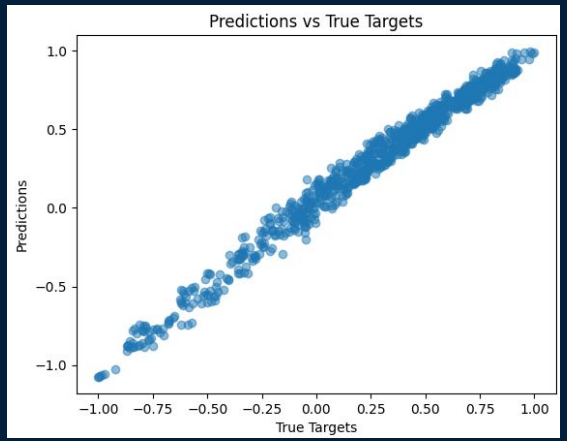
-> some targets have simple relationships to their input features:

-> a simple MLP is sufficient to get good predictions

example: longwave radiation

R^2 score = 0.98

$$R^2 = 1 - \frac{\sum (y_{true} - y_{pred})^2}{\sum (y_{true} - \overline{y_{true}})^2}$$



The figure shows (normalized) predictions of the MLP vs (normalized) true targets of a test set (1000 rows).

A Family of Models:

-> other targets have more complicated relationships:

-> need more sophisticated architecture than the simple MLP (e.g. transformer)

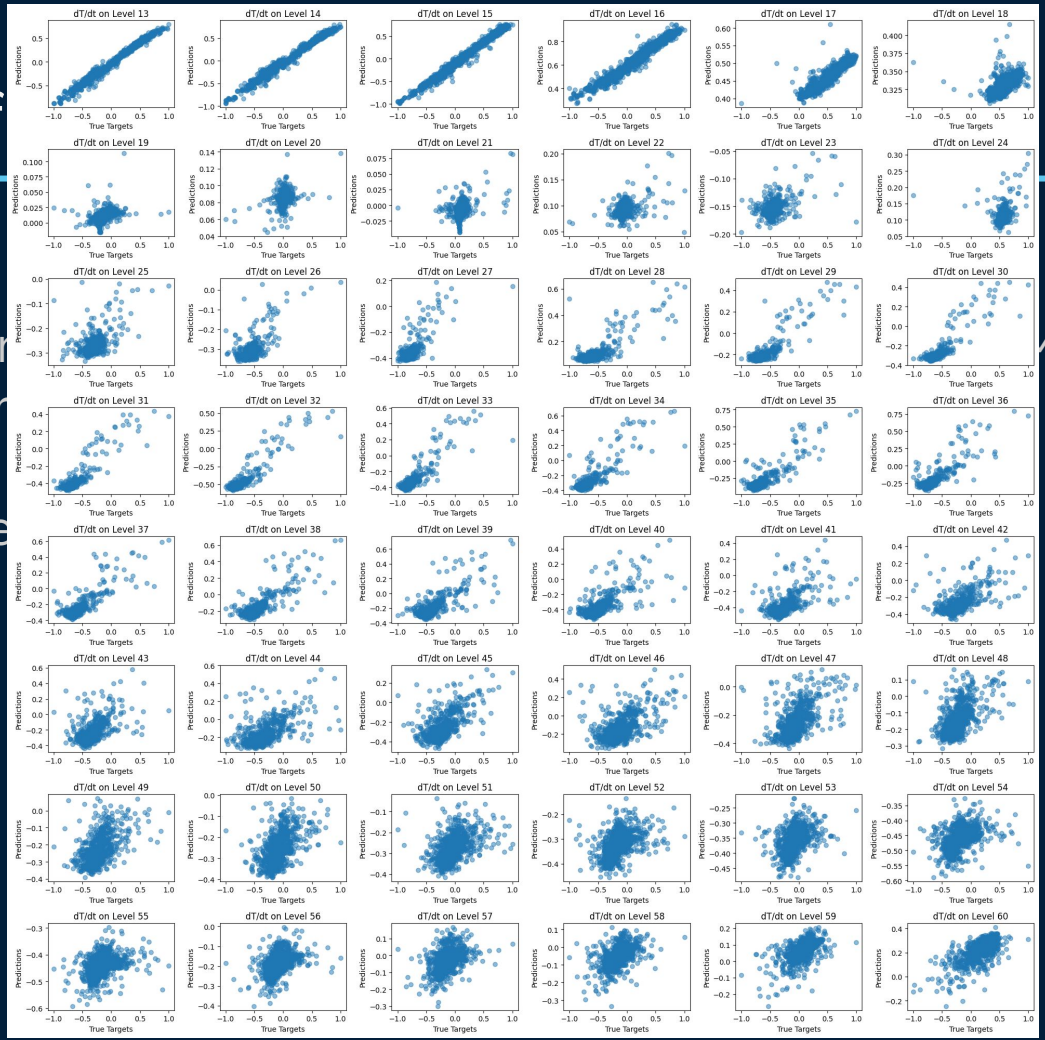
example: heating tendency profiles

A Family of

-> other targets

-> need mor MLP (e.g. transformer

example: he



Outlook for Remaining Time:

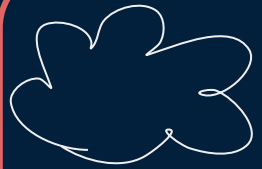
-> train a transformer model for the more complicated targets

-> explore if already well predictable targets can help predicting others

-> train as many models as possible on the full dataset

Summary in 5 Categories

Temperature:
- heating tendency profile



Clouds:
- moistening tendency profile, liquid clouds & ice clouds

Wind:
- zonal acceleration
- meridional acceleration

Radiation:
- shortwave radiation (5 individual parameters)
- longwave radiation

Precipitation:
- rain rate & snow rate