Predicting the Weather using Explainable Machine Learning

By: Laurits L. Fauli, Jeppe P.G. Petersen and Niels A. Davidsen

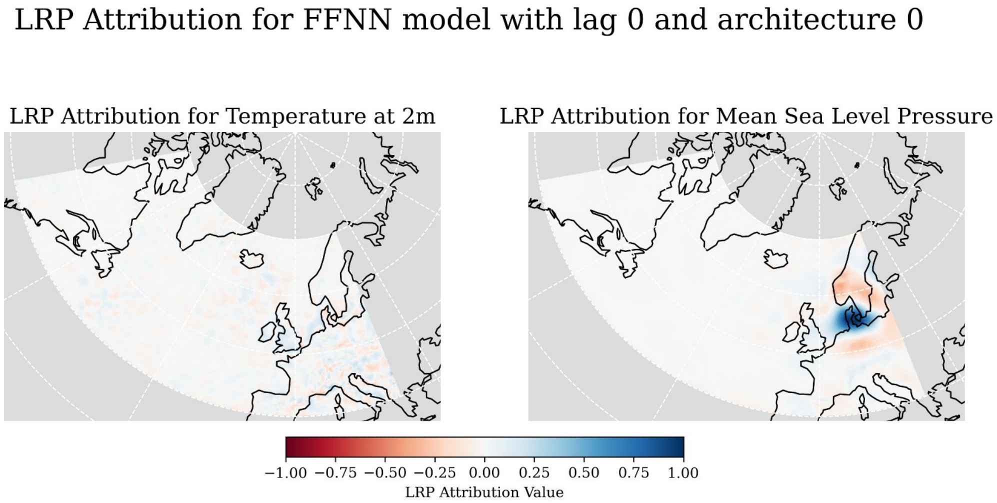
Introduction

- How do we surpass the black box problem?
- Can we train a FFNN to predict precipitation and use it to understand the temporal signal?
- Inspired and supervised by Jonathan Melcher

Disclaimer: AI tools have been used during development of the project

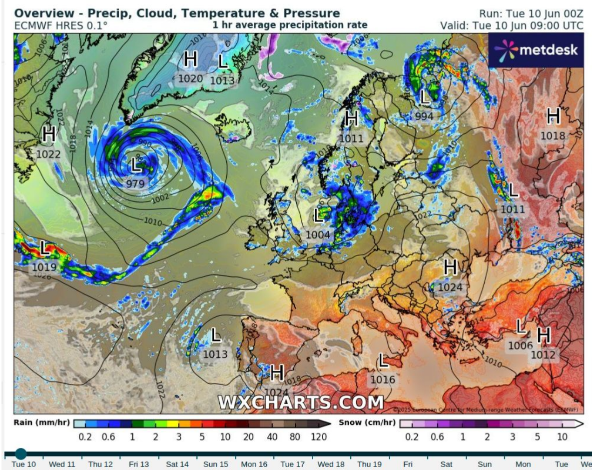
Summary of Results

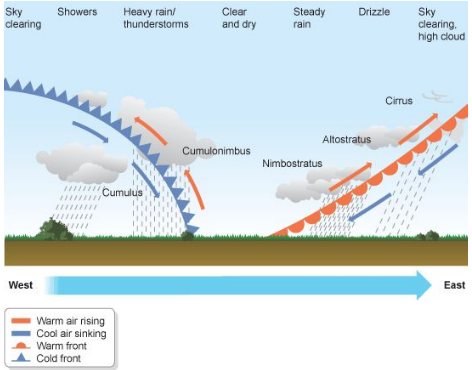
- Our FFNN used pressure in the region around Denmark, to predict the precipitation over Denmark!

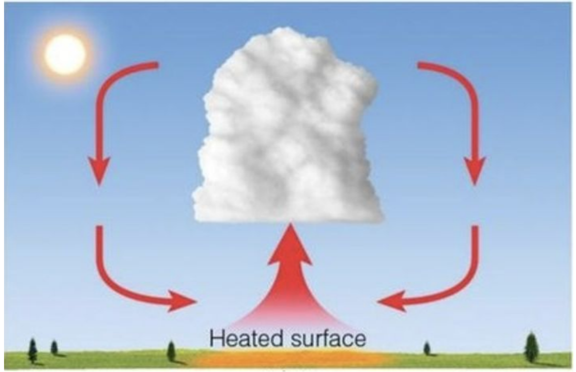


Precipitation in Denmark

- Governed by pressure
- Temperature dependent (warm fronts, cold fronts, and convection)





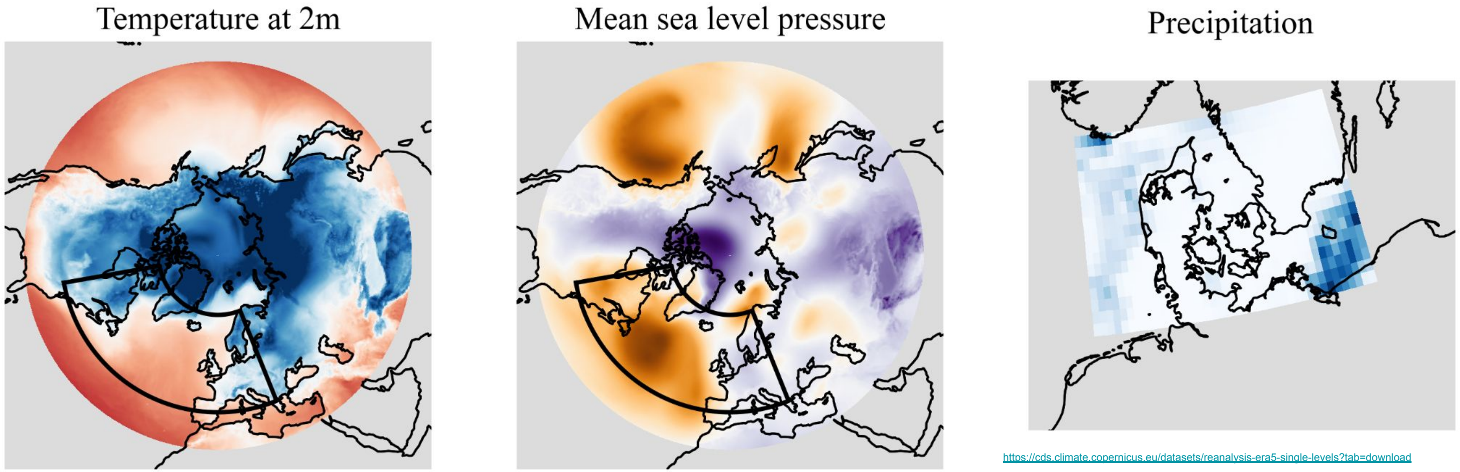


source: google image

Methods

The Data

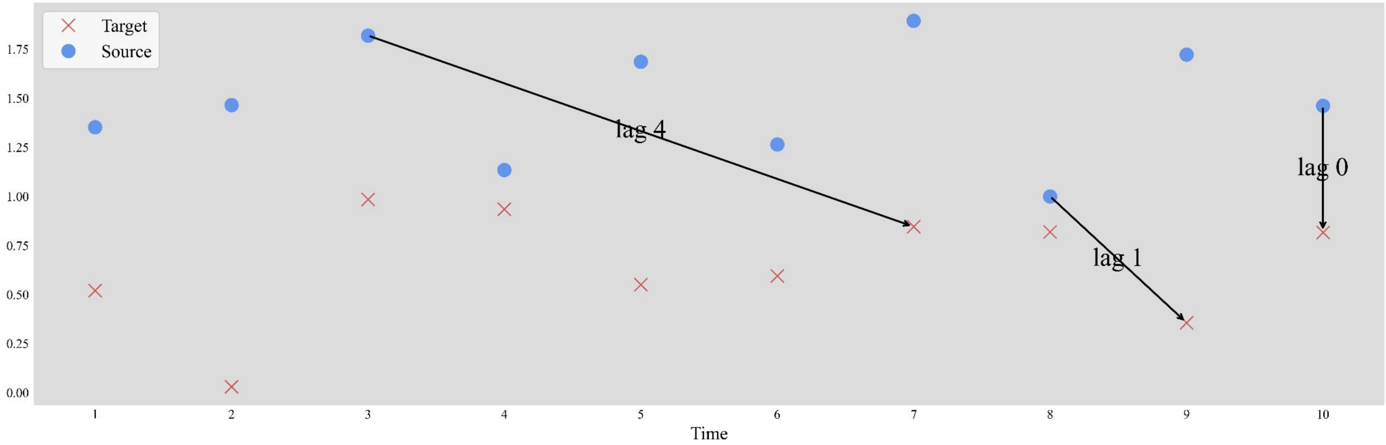
- ERA5 reanalysis, 1940 to 2023. Downloaded from Copernicus Climate data store. Observations assimilated with climate model
- Did temporal standard normalization across days, and chose sub-region
- Did spatial mean over precipitation, resulting in a regression problem for one value



https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download

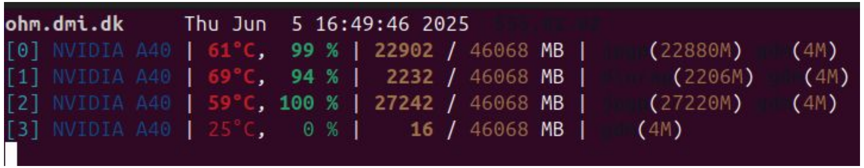
The Data

- Tested different temporal lags, i.e. can msl 5 days ago predict precip. today?



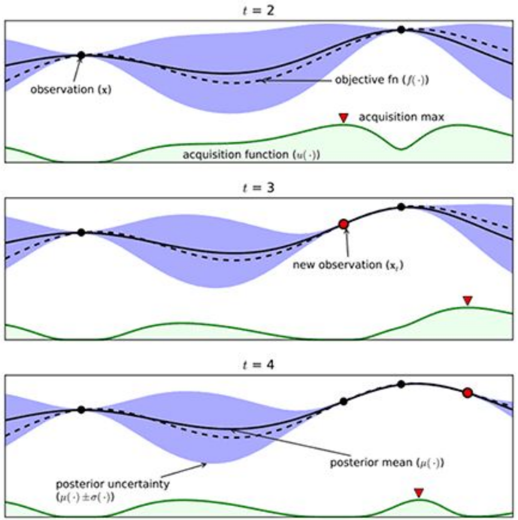
The Data and the GPU

- 91 GB. (30681,2,120,408) on form (days, fields, lat_index, lon_index). (30681,) for target.
- 5 % test, then 20 % validation
- Using DMI's supercomputer for storage and computation
- 4 x NVIDIA A40
- Model has trainable parameters of order 100e6
- Aprox. 12 seconds pr epoch.



Hyper parameter optimization

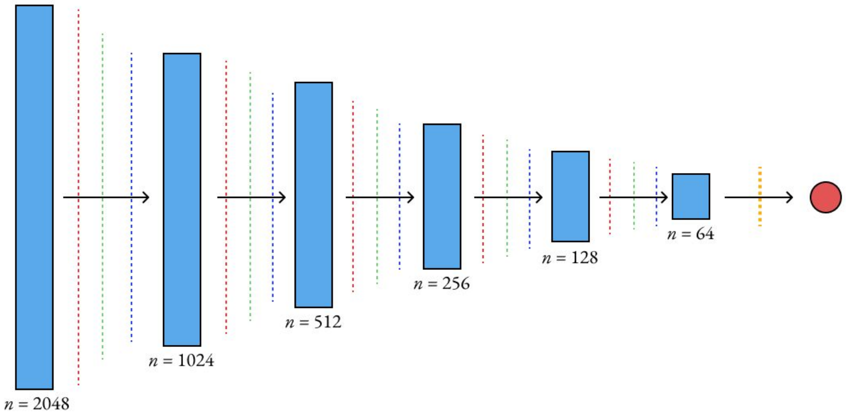
- Using Optuna to optimize using bayesian optimization
  - Number and size of layers, dropout-rate, learning-rate, weight-decay
- Optimized for lag 0 and lag 3, minimizing Mean-Sqaure-Error (mse)
  - lag 0: 56 trials
  - lag 3 109 trials
  - lag 0: 0.0065 best mse
  - lag 3: 0.0090 best mse



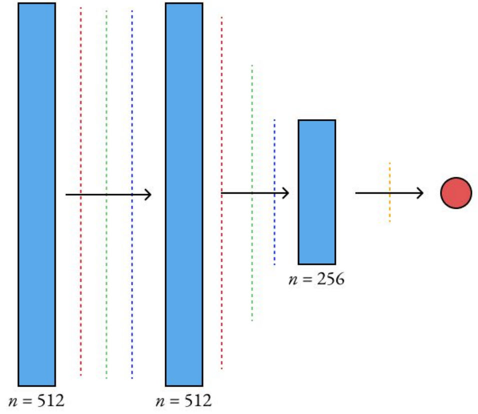
https://www.google.com/url?sa=i&url=https%3A%2F%2Fpolyaxon.com%2Fdocs%2Fautomation%2Foptimization-engine%2Fbayesian-optimization%2F&psig=AOvVaw3Jd2iJZGGnsgwgmSygu6G4&ust=1749563416975000&source=images&cd=vfe&opi=89978449&ved=0CBQQjRxqFwoTCIDNxeG95I0DFQAAAAAdAAAAABA8

Architectures

Architecture 0



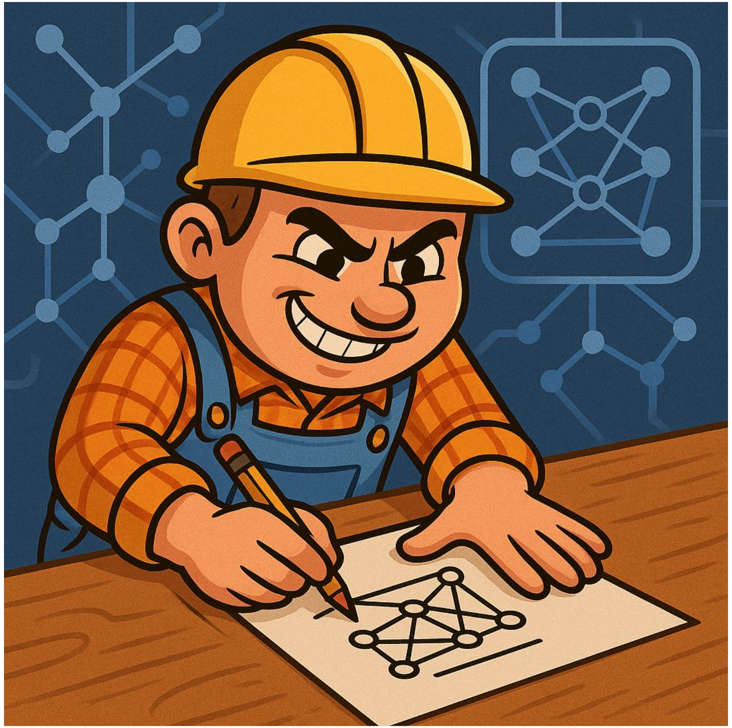
Architecture 3



Batch Normalization ELU activation Dropout Sigmoid activation

Architecture specifications for FFNN

- MAE Loss function
- ADAM optimizer
- ReduceLROnPlateau scheduler
- Early stopping
- Weight Initialization

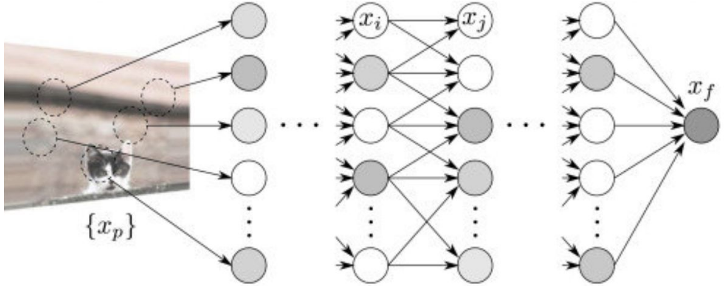


source: Chat gpt, with promt “bob the builder scheming with neural network stuff in the background

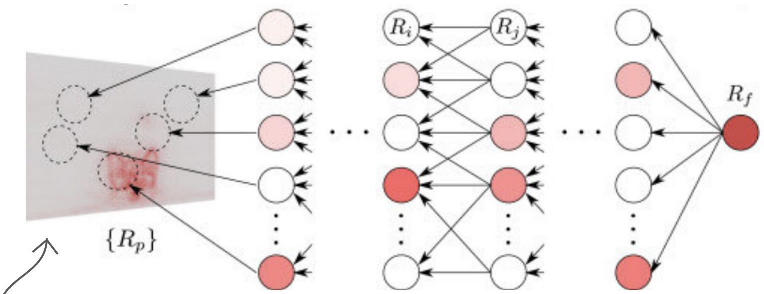
Layer-Wise Relevance Propagation

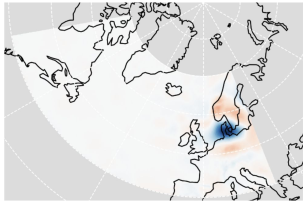
- A look into the machinery
- $$R_i = \sum_j \frac{a_i w_{ij}}{\epsilon + \sum_{0,j} a_i w_{ij}} R_j$$

Forward pass



Backward propagation

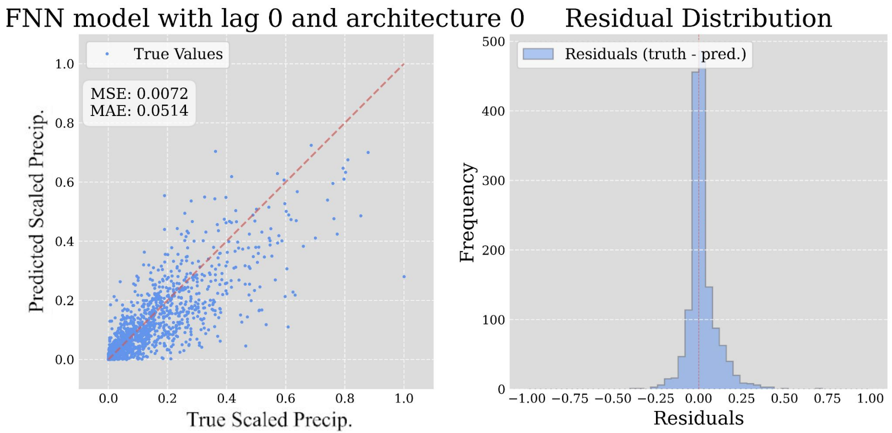


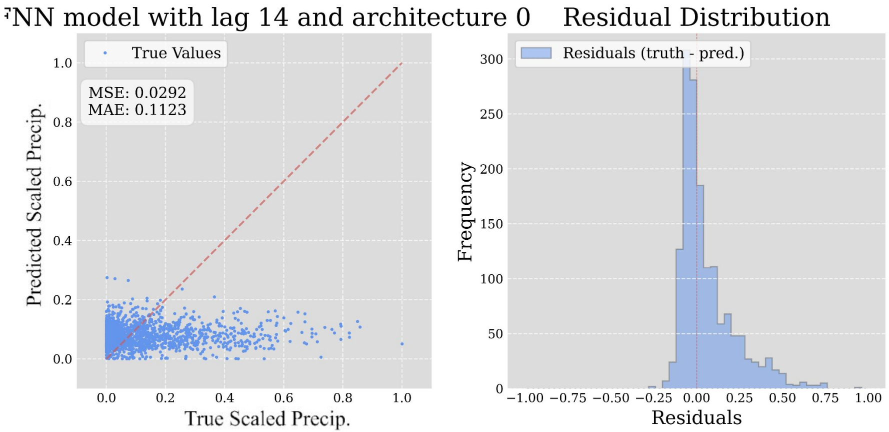


FFNN Results

Results for the FFNN

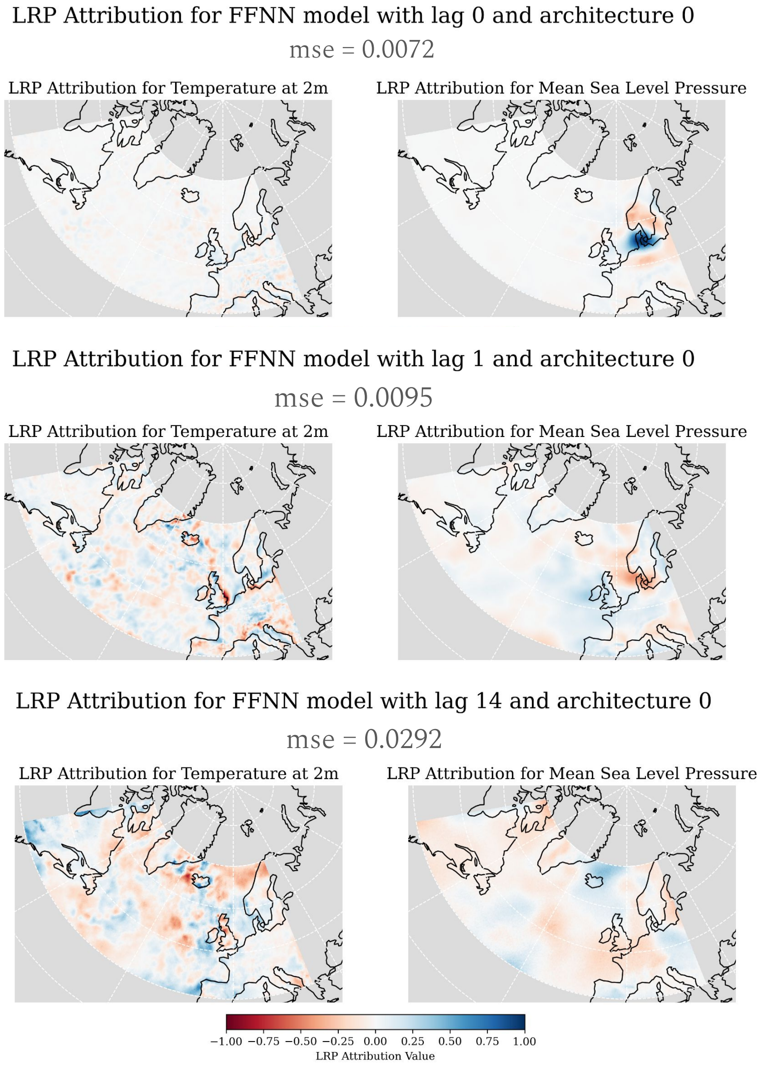
- Good predictability at 0 time lag
- Decreasing predictability for increasing time lag
- Similar results for different architectures at the same time lag
- Results dependent on the chosen test data





LRP of the FFNN

- For lag 0: Large attribution from pressure and to the predicted precipitation over DK
- LRP shows attribution patterns even for increasing time lag but mind the MSE
- NB! These plots show the SUM of the LRP across all test data.

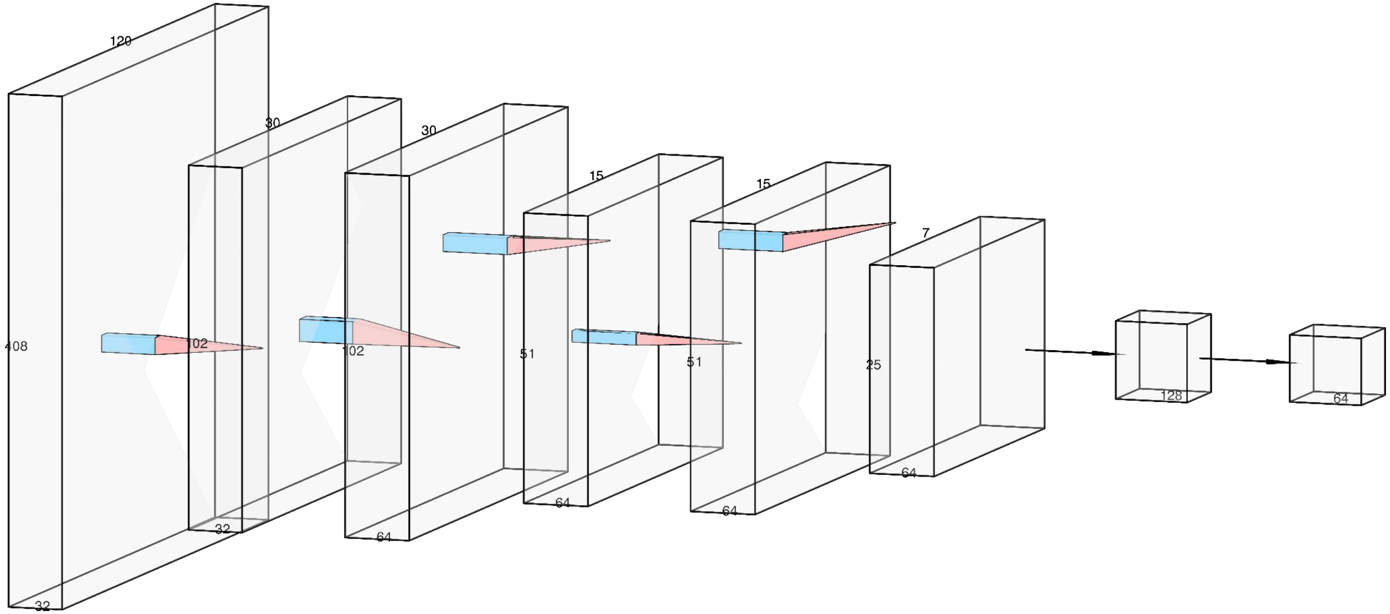


CNN

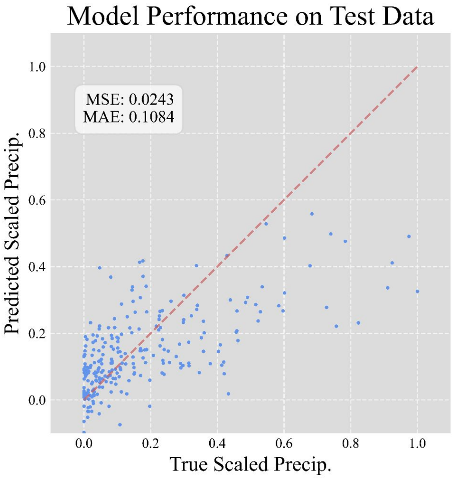
Convolutional Neural Network

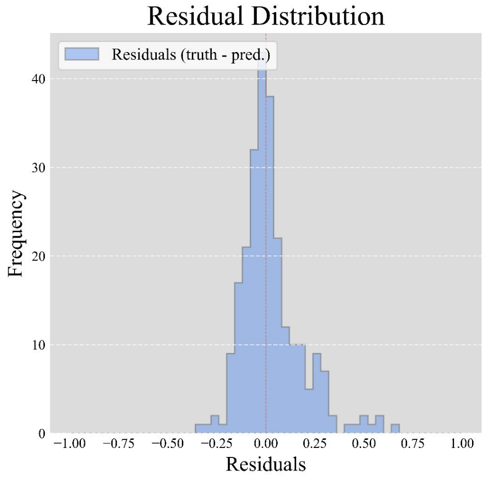
- More typical approach to image processing
- Only trained on five years data
- Tricky to implement LRP

Convolutional Neural Network

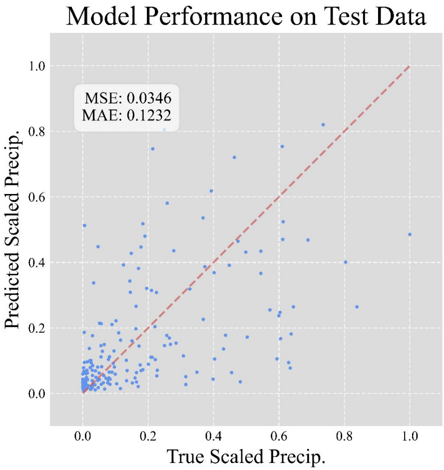


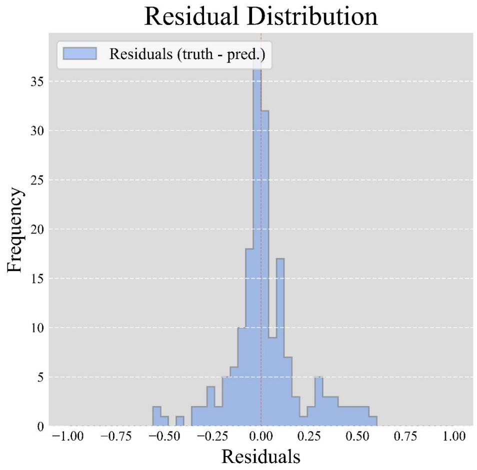
CNN results (Lag 0)





FFNN results (Lag 0)





Discussion and Future Work

What we would have done differently

- Tried different relevant variables, especially temperature at 850 hpa.
- Propper cross validation
- Use different data loading method
- Changing the method of splitting data to improve reproducibility between results
- Further explore subsampling

Future work

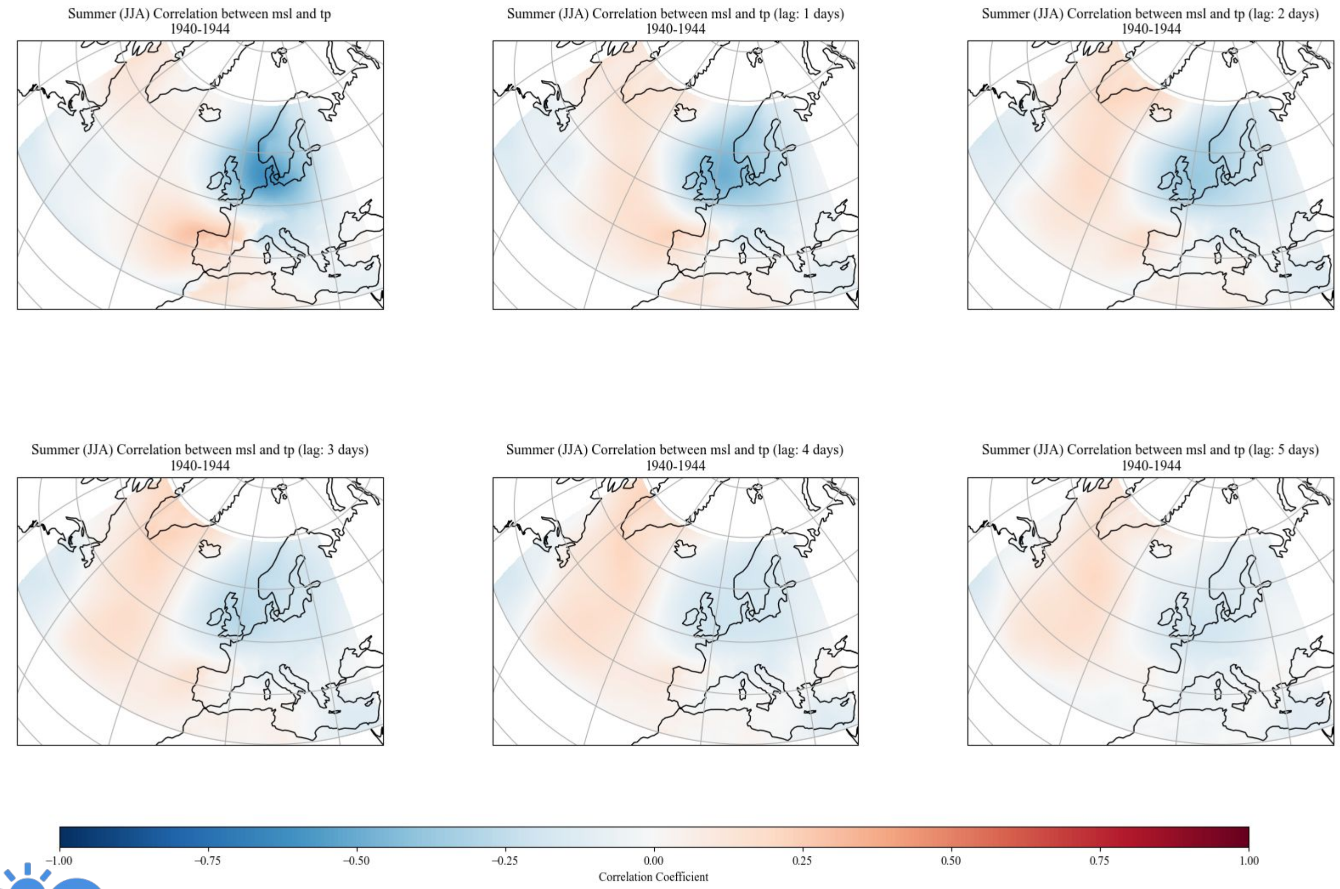
- LSTM to capture time dependencies
- Make CNN compatible with LRP and HPC, and hyper optimize
- Do clustering on LRP results, or only LRP on good predictions

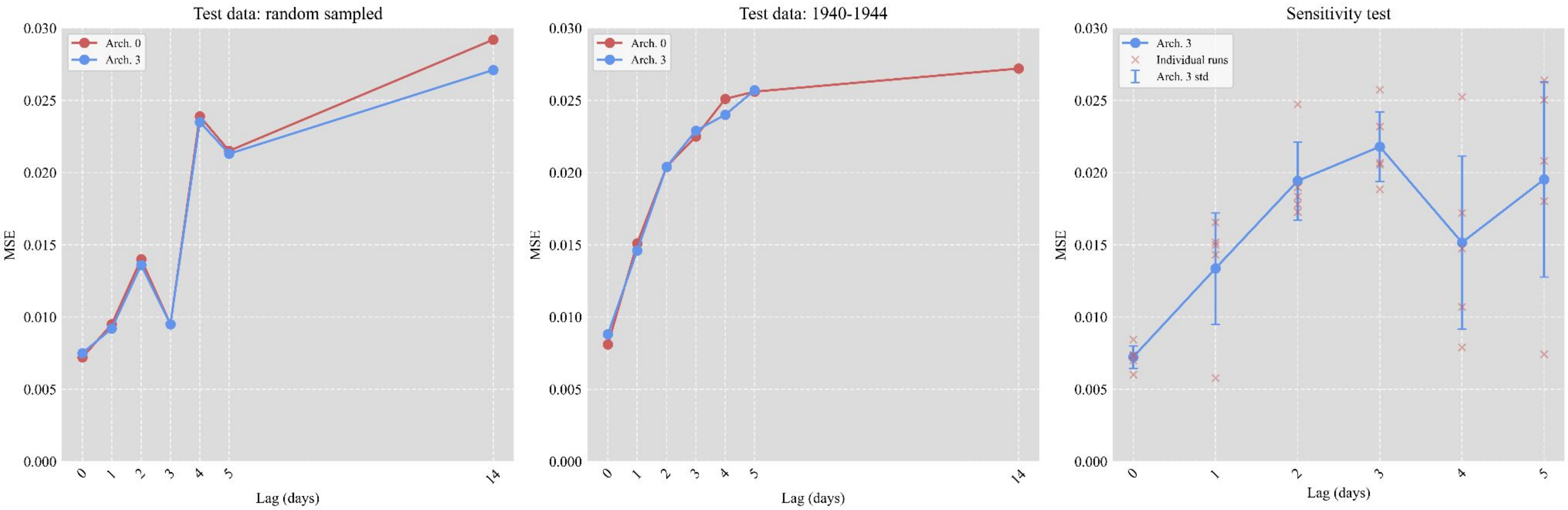
Thank you!

A special thanks to Jonathan Melcher
and DMI

Appendix

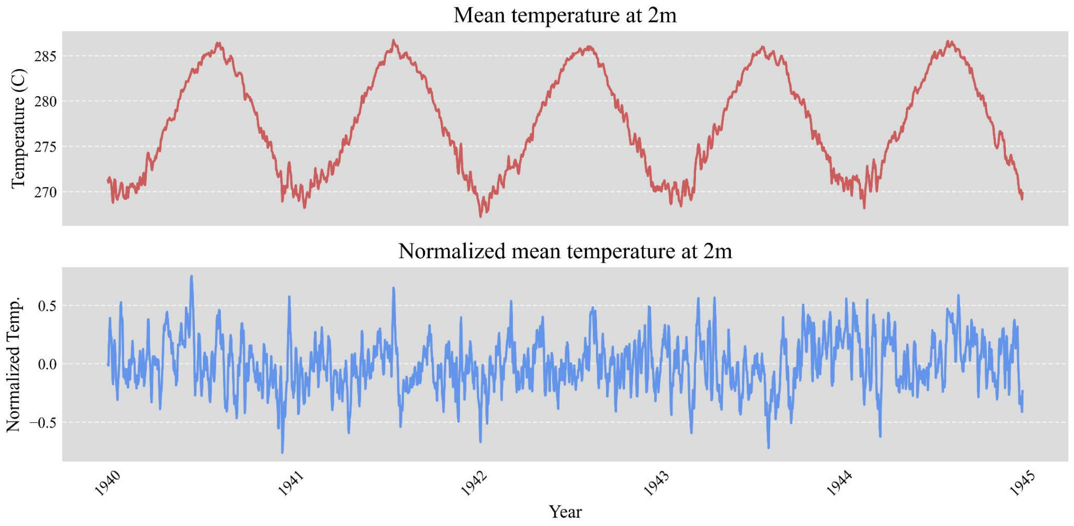
Linear Correlation between Mean sea level pressure and Total Precipitation




Test data: random sampled
Arch. 0
Arch. 3
MSE
Lag (days)
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
1
2
3
4
5
14
Test data: 1940-1944
Arch. 0
Arch. 3
MSE
Lag (days)
Sensitivity test
Arch. 3
Individual runs
Arch. 3 std
MSE
Lag (days)

Data preprocessing

- Did temporal standard normalization across days



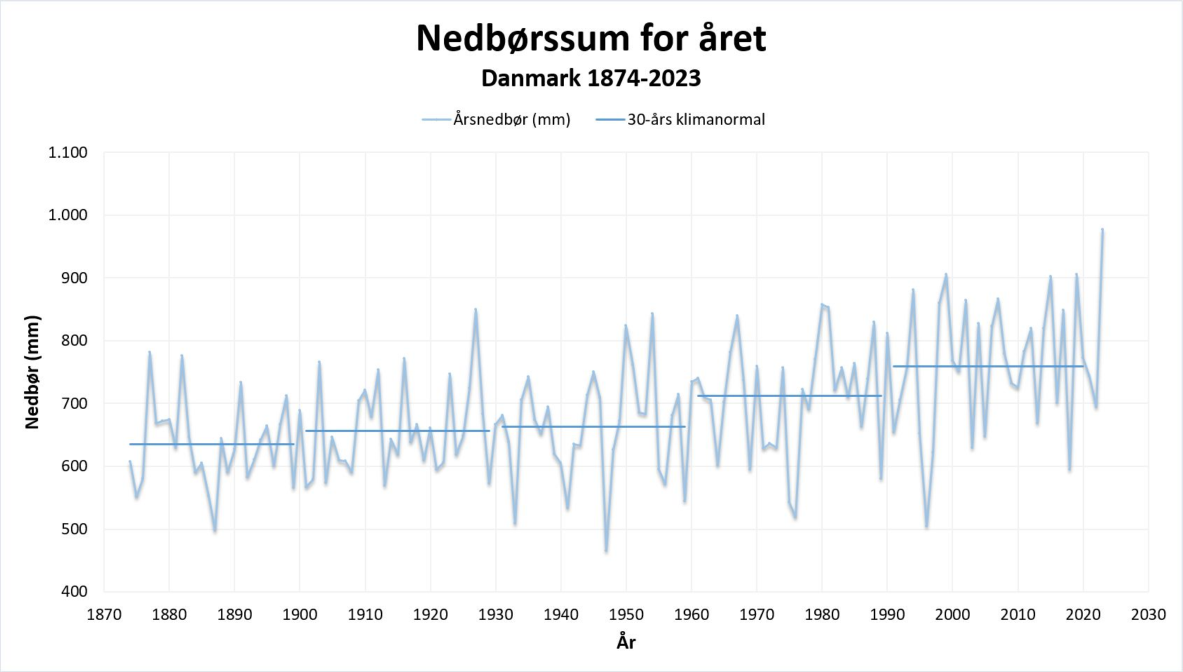
Architecture lag 1
{'hidden_dims': [2048, 1024, 512, 256, 128, 64],
'dropout_rate': 0.24174314484213788,
'learning_rate': 9.95041716902193e-05,
'weight_decay': 7.4804330990821515e-06}

Architecture lag 3
{'hidden_dims': [512, 512, 256],
'dropout_rate': 0.17315195126013944,
'learning_rate': 1.0140932207450912e-05,
'weight_decay': 9.733043688044047e-05}

CNN Architecture

```
regular_conv_layers = [
    (32, 3, 4),
    (64, 3, 2),
    (64, 3, 2)
]
regular_fc_layers = [128, 64]

epochs=100,
batch_size=256,
learning_rate=1e-4,
validation_split=0.2,
weight_decay=1e-5,
patience=5,
factor=0.5,
early_stopping_patience=10,
loss_function='MSE'
```



https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.thelocal.dk%2F20240722%2Fdenmarks-last-12-months-were-the-wettest-ever&psig=AOvVaw05B4ePhWYeNzzXq93-3ZRW&ust=1749649001862000&source=images&cd=vfe&opi=89978449&ved=0CBcQjhxqFwoTCNDL58v85o0DFQAAAAAdAAAAABAw