NO_2 Forecast with Lightweight Machine Learning Models Based on POD Dimensionality Reduction

Marco Stricker [1] Marlon Nuske [1] Dinesh Krishna Natarajan [1] Andreas Dengel [1,2]

[1]DFKI - German Research Center For Artificial Intelligence [2]Technical University Of Kaiserslautern

Motivation

We investigate the potential of dimensionality reduction for improving the forecasting capabilities of recurrent neural networks and perform a thorough investigation of its performance for forecasting NO_2 concentrations. As one major air pollutant NO_2 can cause respiratory symptoms or even premature death. With the ability to forecast distribution of gases, precautions can be taken, such as issuing health warnings in the case of NO_2. Traditionally, physical models are used to perform forecasts; they are, however, often computationally expensive. Complex deep learning models can be used as an alternative. They are, however, difficult to optimize and require a lot of training data, which is not always available.
To overcome these challenges, we reduce the dimensionality of the input data before training a neural network.

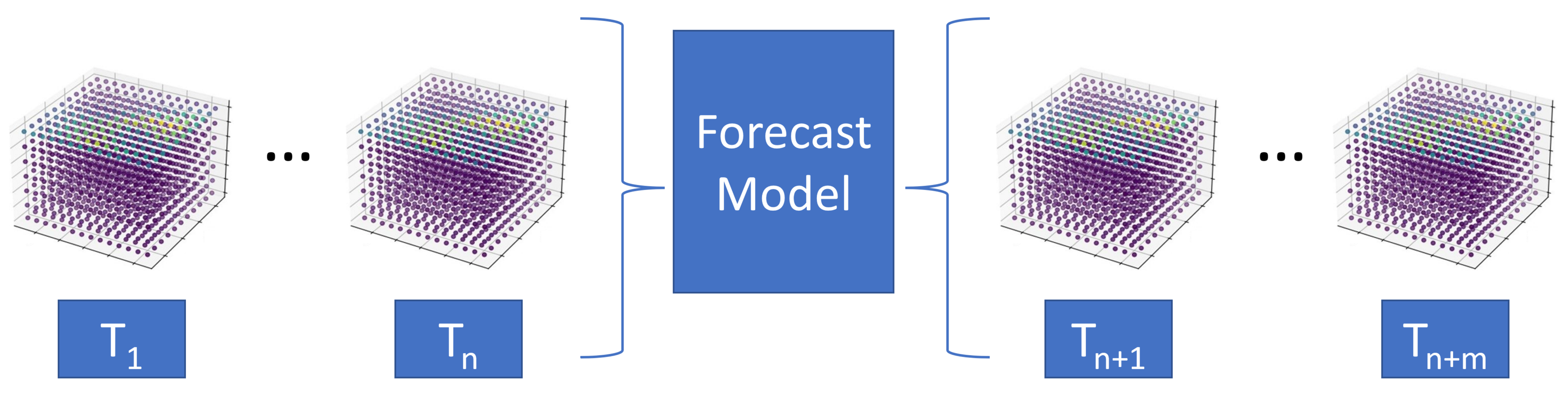


Figure 1. Problem: Predicting an output time series based on an input timeseries, where each timestep contains many 3D datapoints

Data

- NO_2 values from CAMS dataset
- One datapoint corresponds to one timestep
- Each datapoint has a Latitude, Longitude and Height
- We performed multiple experiments with different ways of scaling input data
- We flatten our input data into a space x time matrix before feeding it to our network, as shown in Figure 2

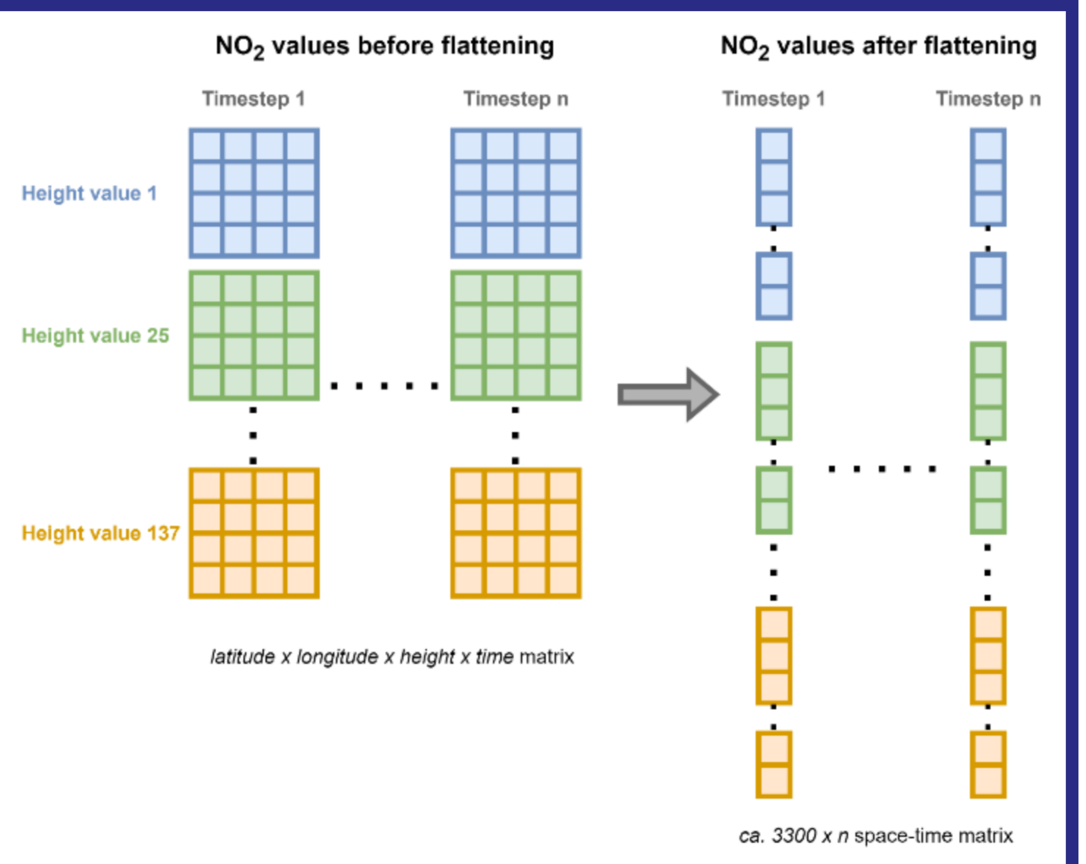


Figure 2. Flattening the high dimensional input data into a space x time matrix

Dimensionality Reduction

- Data = Linear Combination of modes and coefficients
- Method: Proper Orthogonal Decomposition (POD)
- Spatial Modes are the same for all timesteps
- Temporal Coefficients are specific to each timestep
- Make predictions based only on coefficients

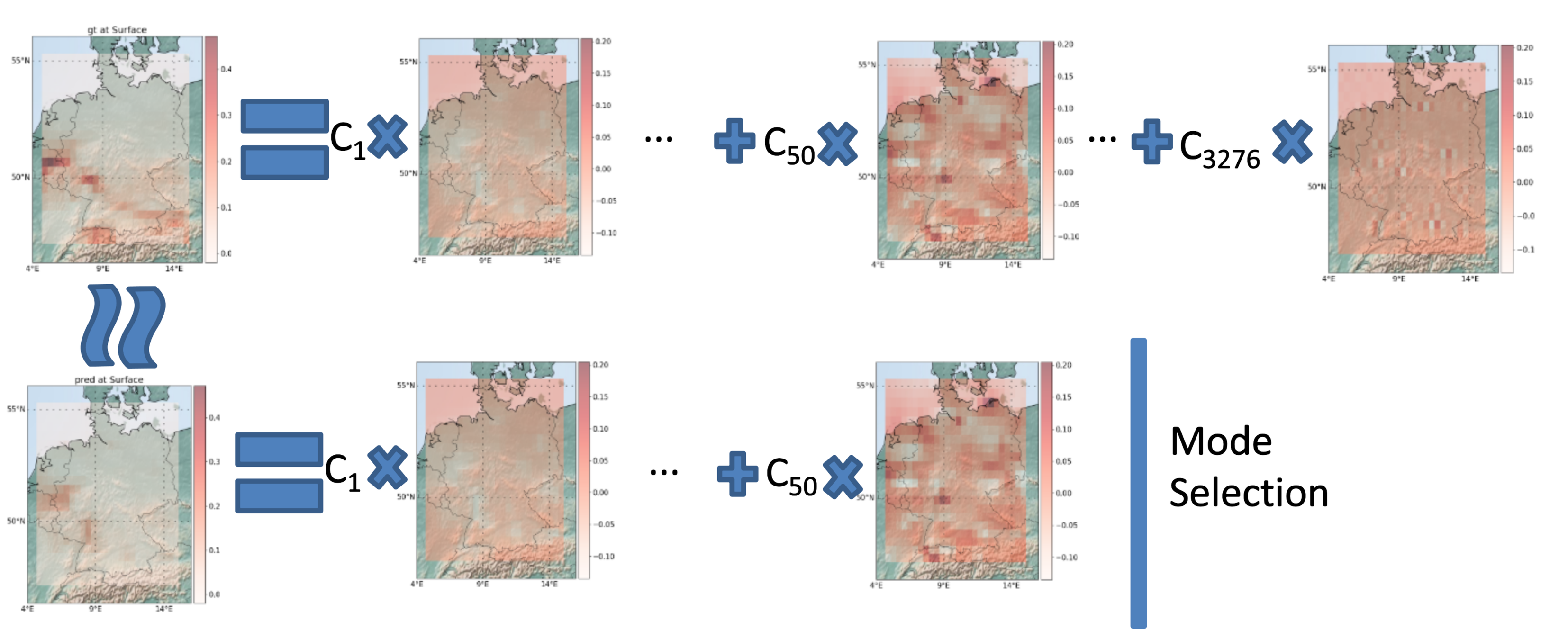


Figure 3. Transforming data into a representation of spatial modes and temporal coefficients. Selecting the most important modes achieves data compression

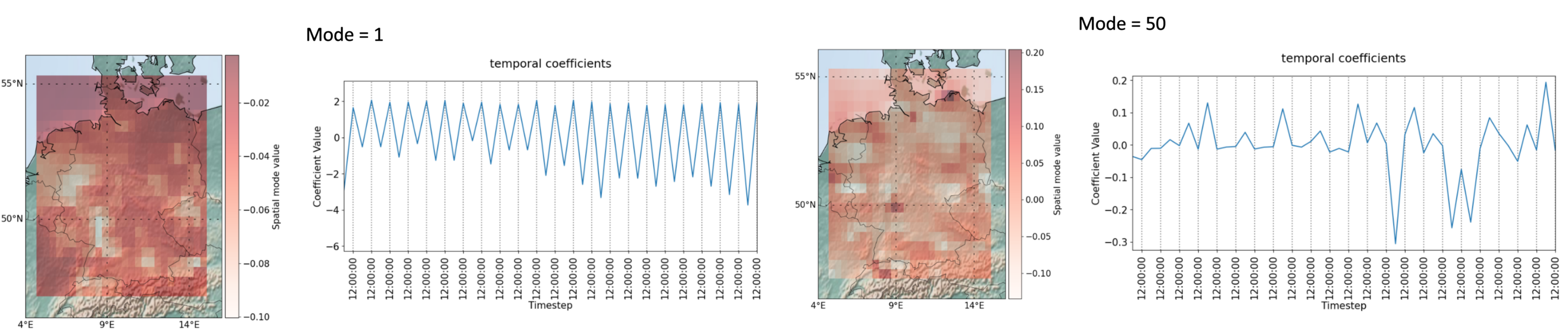


Figure 4. Example spatial modes with corresponding temporal coefficients.

Our Pipeline

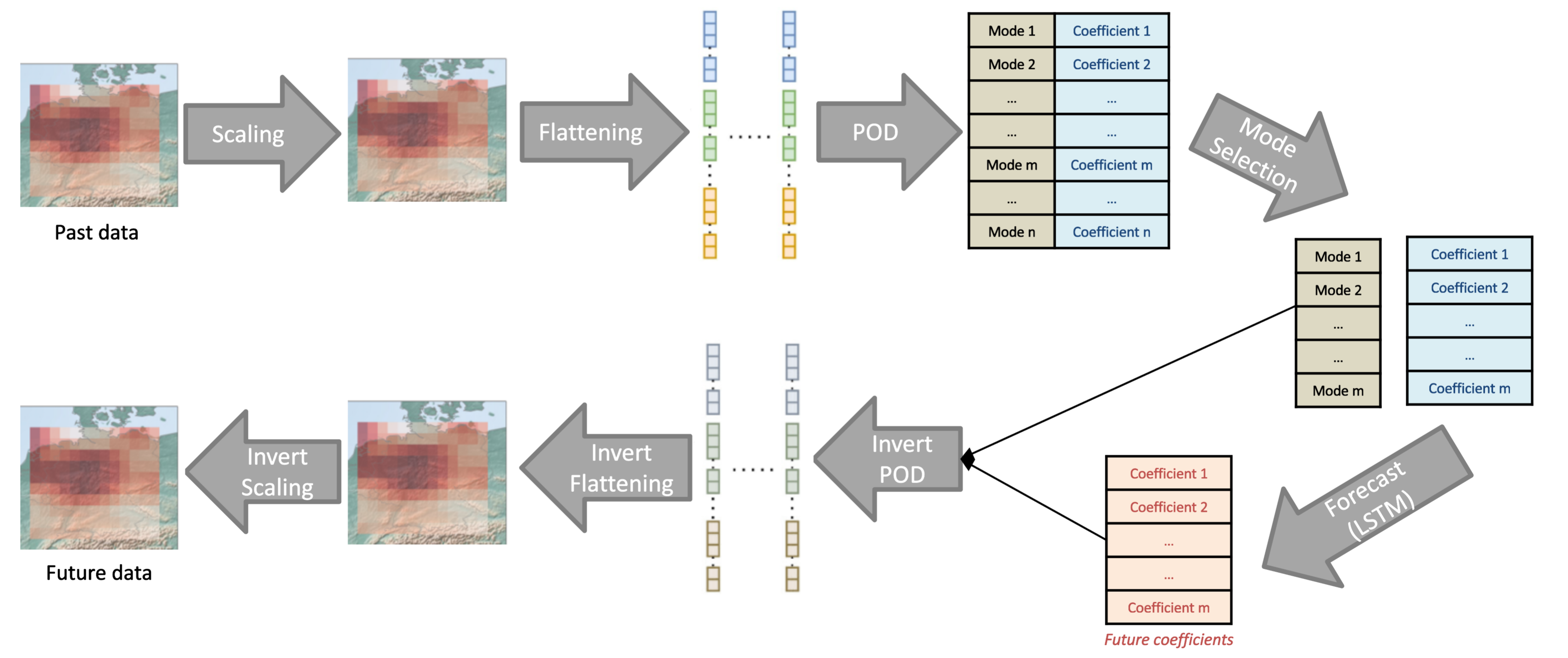
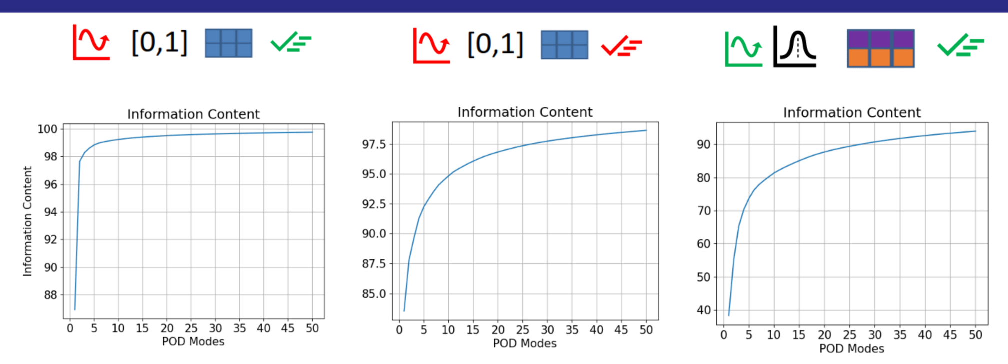


Figure 5. Our proposed pipeline.

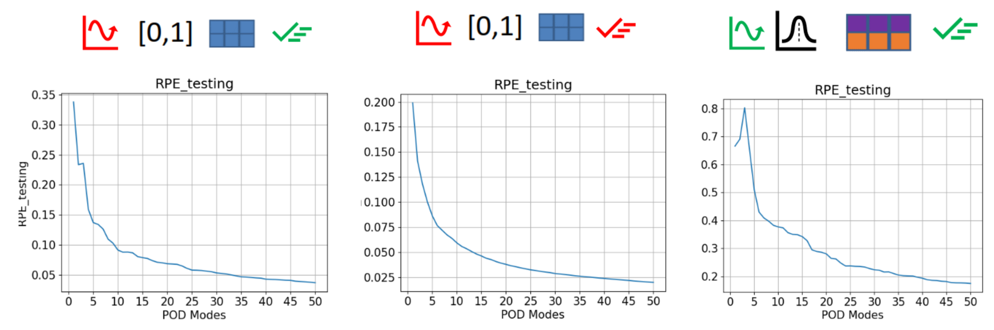
Forecast Model

- Recurrent Neural Network (LSTM/GRU)
- Split Data into Train/Validation/Testing
- Input: 10 days, Output: 5 days
- Benchmark Method: Same procedure without dimensionality reduction

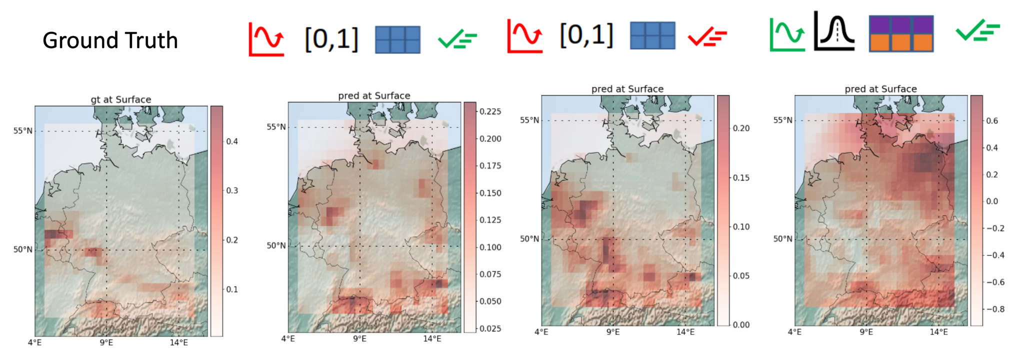
Compression Evaluation



(a) Information Content



(b) Relative Projection Error

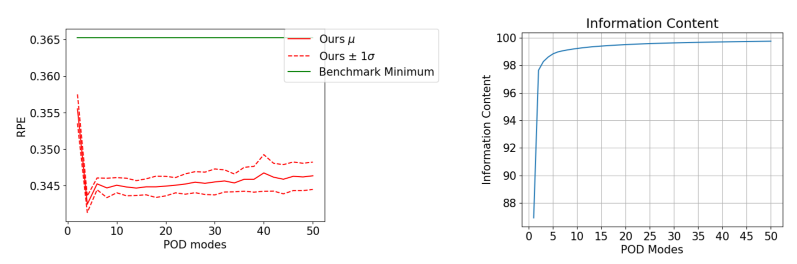


(c) Sample Visualizations

Figure 6. Evaluating the reconstruction of the input based on the reduced dimensionality. Shown are the results for different experimental settings.

- Higher IC implies that more information is retained
- Lower RPE means that our reconstruction contains smaller mistakes
- Settings for best IC are not resulting in best RPE
- The experimental configurations include: removal of fluctuations, no scaling/standardization/normalization, calculating the scaling over the whole input or per space axis, scale height levels independently

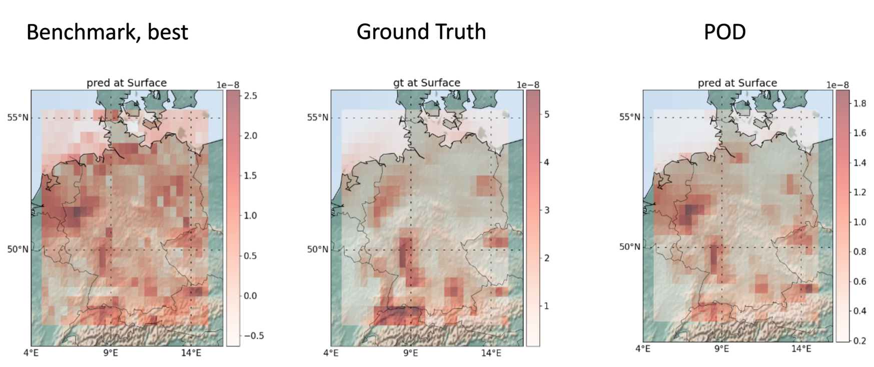
Forecast Evaluation



(a) Forecast RPE Evaluation against benchmark

	Ours (4 modes)	Benchmark (Input 3276)
Mean RPE	**0.3424 ± 0.0011**	0.4650 ± 0.0647
Min RPE	**0.3411**	0.3653
Time for POD	49.84 seconds	-
Time to Train	152.50 seconds	268.89 seconds
Total Time	**202.34 seconds**	268.89 seconds

(b) Tabular Evaluation



(c) Sample Visualizations

Figure 7. Evaluating our forecast model against the benchmark

Conclusion

Due to the data compression we were able to train less complex networks, which require less training data and are easier to optimize compared to deep state-of-the-art models.