Developing Deep Learning Methods for Surface NO_2 Estimation from GEMS Satellite Data (Session AS3.21, X5.82 | EGU23-9309)

Janek Gödeke[1], Peter Maaß[1], Andreas Richter[2], Kezia Lange[2], Hyunkee Hong[3], Hanlim Lee[4], Junsung Park[4]

[1]University of Bremen, Center for Industrial Mathematics, Germany
[2]Institute of Environmental Physics (IUP), University of Bremen, Germany
[3]National Institute of Environmental Research, Environmental Satellite Center, South Korea
[4]Pukyong National University, South Korea

1. Goal

Train Deep Neural Networks (DNN) for deriving estimates of NO_2 concentration at the earth's surface from:

- NO_2 tropospheric vertical column densities (VCDs),
- meteorological data,
- additional information, e.g. geographical coordinates.

2. Data in South Korea

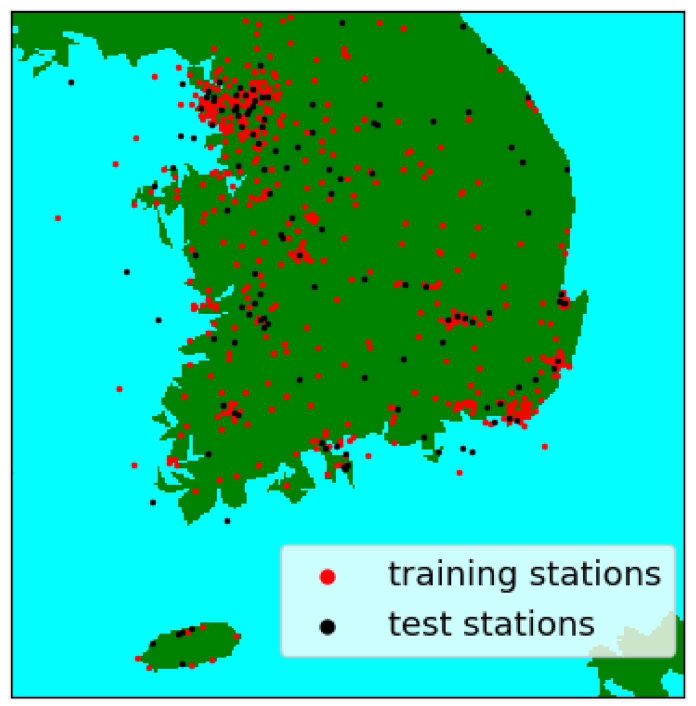
- **NO_2 VCDs:** Radiances and irradiances, observed by the Korean Geostationary Environmental Monitoring Spectrometer (GEMS)[(1)], are fed into the IUP NO_2 retrieval algorithm[(2)] to obtain vertical, tropospheric NO_2 columns. **Geostationarity enables hourly measurements!**

- **Meteorological data:** Copernicus ERA5 hourly data[(3)]: Evaporation, temperature at 2 m, boundary layer height, downward UV radiation at the surface, UV visible albedo for direct radiation, total O_3 column, total H_2O column, skin temperature, soil type.

- **NO_2 at Earth's surface:** In-situ observations from the air quality network of South Korea.

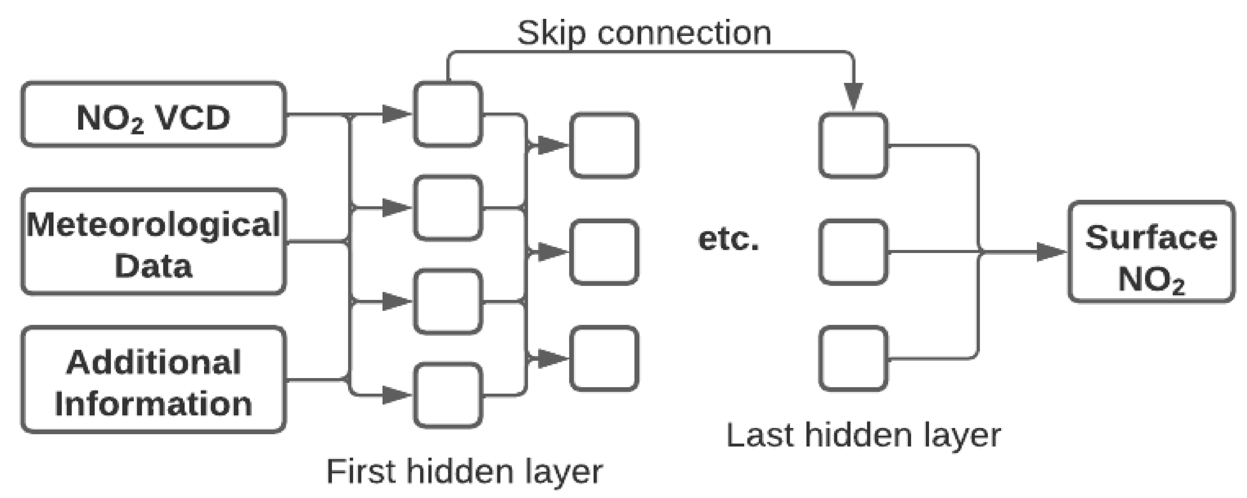
3. Strategy

- Collect different data at every in-situ station.
- Split in-situ stations into training and test stations.
- Train DNN only on data corresponding to the training stations.
- Validate the DNN on the test station data.



4. Neural Network

- The neural network is a mapping $\mathbb{R}^n \to \mathbb{R}$, where n is the number of input features.
- In order to avoid vanishing gradients in deep networks, skip connections between hidden layers are useful.



6. Conclusion and Next Steps

- Increase the performance of the DNN by **selecting more**, or more relevant, **input features**, e.g. population density, distance to nearest city, measurement time...
- Model performs best in regions where lots of training data points are located. **Increasing the size of the dataset** may lead to better results.
- **Optimize hyperparameters**, like number and width of hidden layers, learning rate, batch size, etc.
- Use multiple, **time-contiguous** measurements as an **input** of the neural network, not only measurements at a single measurement time.

References

(1) https://nesc.nier.go.kr/

(2) A. Richter et al. (2005), DOI:10.1038/nature04092.

(3) H. Hersbach et al. (2023), DOI:10.24381/cds.adbb2d47.

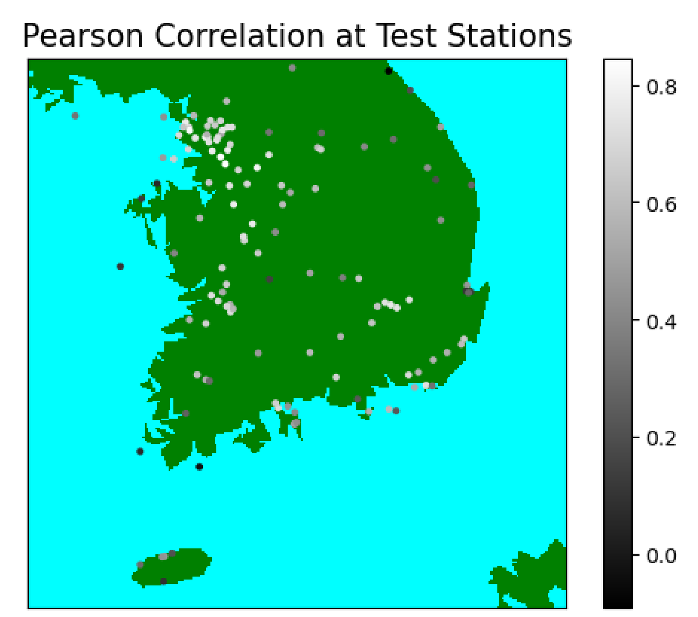
5. First Results

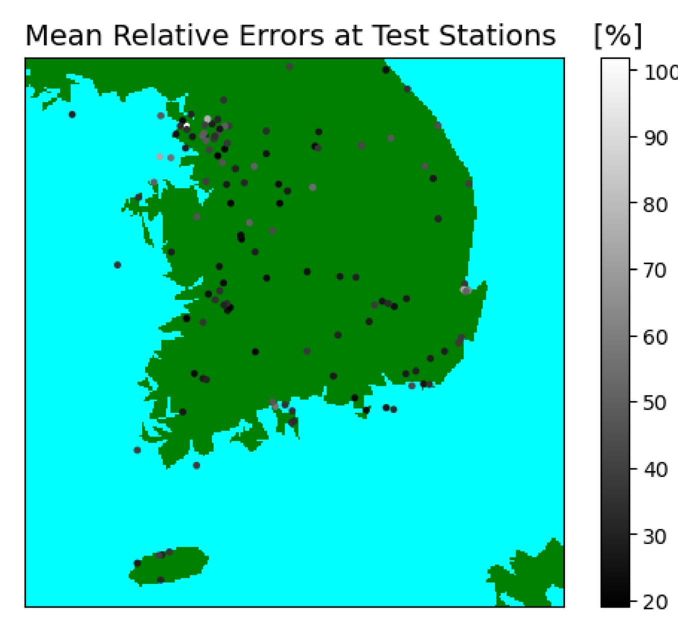
- Both training and test data points from June, July and August 2022.

- So far, only nine relevant meteorological input features were identified and used, see box 2.

- After filtering by qa-value>0.8:
  90.000 training data points,
  20.000 test data points.
  (Enabled by the geostationarity of GEMS)

- Comparison of in-situ NO_2 surface measurements and prediction of the DNN:





- Pearson correlation for all test data points: 0.65
- Mean relative error over all test data points: 0.35