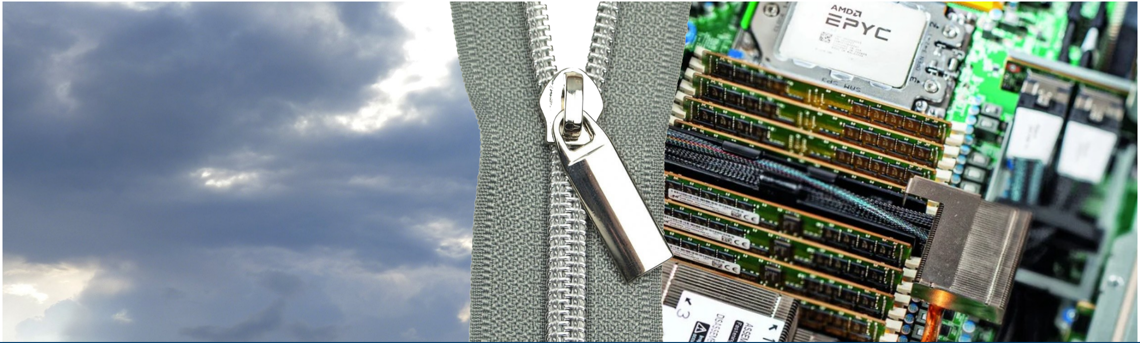


Representing chemical history of ozone

a method development study for deep learning models

30-04-2021 | Felix Kleinert[FZJ, BN], Lukas H. Leufen[FZJ, BN], Aurelia Lupascu[IASS], Tim Butler[IASS], and Martin G. Schultz[FZJ] | f.kleinert@fz-juelich.de | vEGU 2021 ITS4.4/AS4.1

UNIVERSITÄT BONN
IASS POTSDAM Institute for Advanced Sustainability Studies
JÜLICH Forschungszentrum

Outline

Overall goal

Improve quality of station based near-surface ozone predictions

Goal of this method development study

- Design different concepts of representing chemical history of air parcel (capturing transport)
- Benchmark concepts against baseline model which is not aware of chemical history
- Compare the change in prediction quality

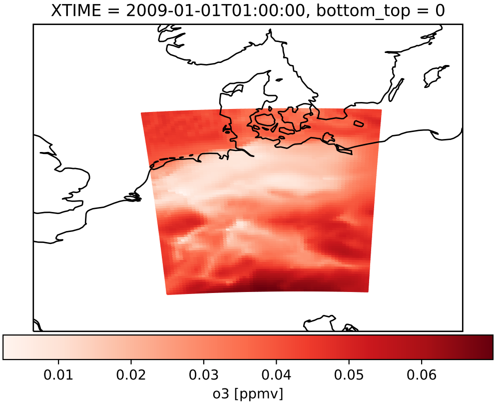
General prediction task

”Use chemical and meteorological model data of the past few hours to predict near-surface ozone concentrations for a lead time of up to two hours”

Data: WRF-Chem simulation

Model data

- We use WRF-Chem data as the gridded model data provides an internally consistent dataset with complete spatial coverage and no missing values
- Changes in forecasting performance can be attributed to concepts rather than a change in data availability (irregular distribution of measurement stations)
- Short period where advection is dominant (Jan to Mar 2009)



Info

In this study, we use model data (see also [3]) to train neural networks. Later on, we will apply the best working concept to measurement data of the Tropospheric Ozone Assessment Report (TOAR) database [6], e.g. to improve the IntelliO3-ts model [2].

Supervised machine learning

Supervised machine learning approach

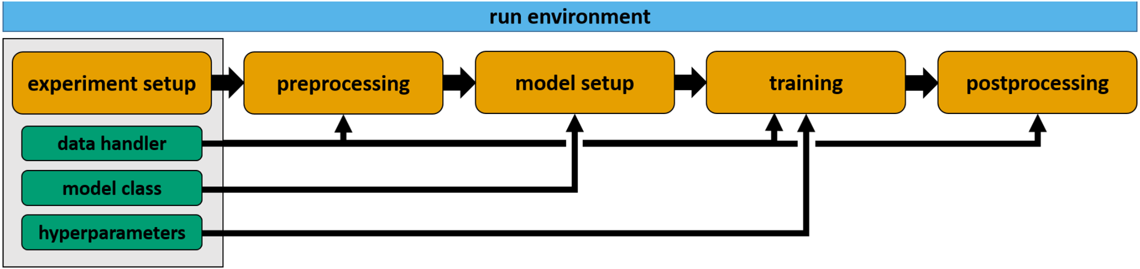
- Task: Find unknown function f which maps input pattern ($\mathbf{X}$) to corresponding labels/ ground truth ($\mathbf{y}$)
- Machine learning model is estimator ($\hat{f}$) which maps $\mathbf{X}$ to an estimate $\hat{\mathbf{y}}$ of the ground truth
- Goodness of estimate quantified by error/ loss function
- Generally ground truth and estimates differ $f(\mathbf{X}) = \mathbf{y} \neq \hat{\mathbf{y}} = \hat{f}(\mathbf{X})$

None neural network competitor models

- Ordinary least-squares (ols)
- Persistence (persi)

Learning Framework: MLAir

- We carry out our experiments with MLAir [5]
- We extended the code to represent chemical history (`DataHandlers`)

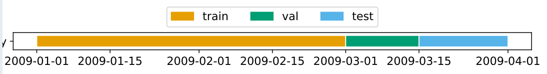


Visualization of the MLAir standard setup (from [5, Fig. 1])

Baseline model ($\hat{f}_b$)

Data

- Data from 8 surface grid boxes
- Input (**X**)
  - Variables: O_3, NO, NO_2, CO, T2, wdir10, wspd10, PSFC, PBLH, CLDFRA
  - Time: t_{-6} to t_0, hourly
  - Shape: (# samples, # prev time steps, 1, # vars)
- Target variable (**y**): O_3 for t_1 to t_2



Data split

Model

- 2× CNN layers [4] where vars are used as channels
- 3 × 1 kernel, 16 and 32 filters
- Symmetric padding, ELU activation [1]
- 2× FC layer, ELU and linear activation

Baseline prediction task

"Use chemical and meteorological model data of the past few hours **at a specific grid box** to predict near-surface ozone concentrations for a lead time of up to two hours **at that particular grid box.**"

Representing chemical history

- The ozone concentration at a particular point in space and time depends on the chemical history of an air parcel (where it has been before).
- The baseline model only captures the history of the fixed grid box rather than the air parcel.

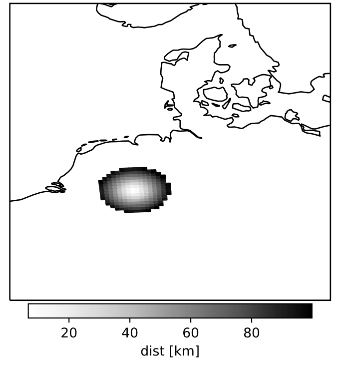
$\rightarrow$ Ideally we would use backward trajectories

- Backward trajectories require full field of model data (not directly available when using measurement data)

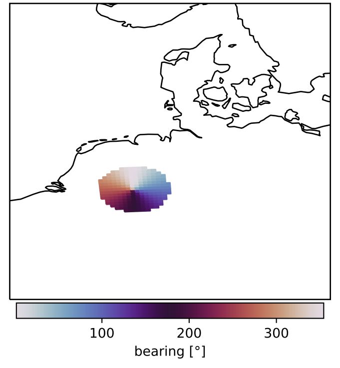
$\Rightarrow$ Approximation by wind sector

Sectorial preprocessing

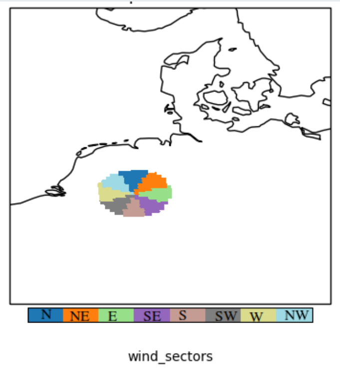
Exemplary point of interest (POI)



- Distance from POI to neighbouring grid boxes (here max. 100 km)



- Bearing from POI to neighbouring grid boxes



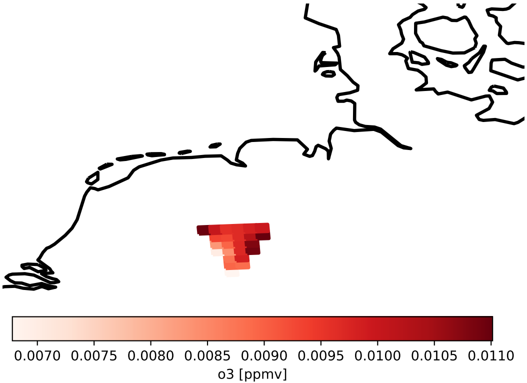
- Assign each grid box to a sector used for aggregation

Sector model ($\hat{f}_s$)

Data preprocessing

- Data of Baseline Model
- \+ Aggregated (mean) data of each var from upstream sector determined by wind direction at t_0 e.g. mean over N sector (see right figure as an example for O_3)
- Used as additional variables (channels in neural network) $\rightarrow$ 20 variables (10 (grid box) +10 (sector))

$\rightarrow$ Additional variables represent sector



Sector prediction task

"Use chemical and meteorological model data of the past few hours at a specific grid box **and spatially aggregated sector data of the current wind direction (upstream)** to predict near-surface ozone concentrations for a lead time of up to two hours at that particular grid box."

Extended sector model ($\hat{f}_{es}$)

Data preprocessing

- Data of Baseline model
- Data of Sector model

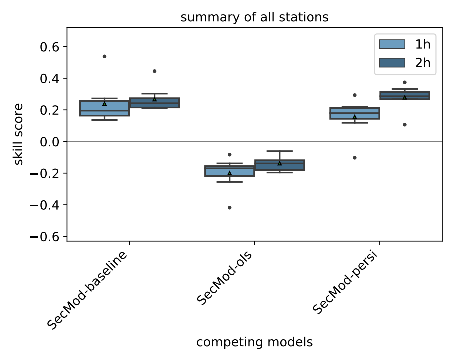
\+ Sector data of the two adjacent sectors
  - Aggregated (mean) data of each variable from adjacent sectors of wind direction at t_0 (left and right sectors, e.g. NW, and NE if main sector is N)
  - Used as additional variables (channels in neural network) $\rightarrow$ 40 variables (10 (grid box) +10 (main sector) +10 (left sector) +10 (right sector))

Extended sector prediction task

"Use chemical and meteorological model data of the past few hours at a specific grid box, spatially aggregated sector data of the current wind direction (upstream) **and spatially aggregated sector data of the two adjacent sectors** to predict near-surface ozone concentrations for a lead time of up to two hours at that particular grid box."

Results: Model comparison

Sector model

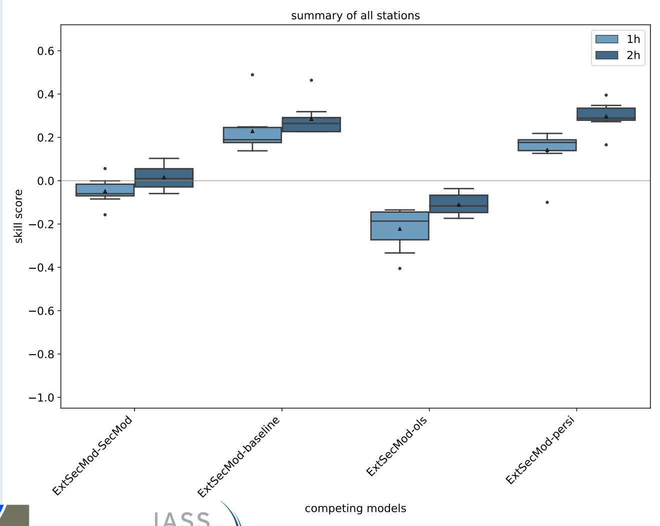


$\Rightarrow$ 20% improvement over baseline

Skill score

$$SS \begin{cases} >0, & \text{first model outperforms second,} \\ =0, & \text{first model as good as second,} \\ <0, & \text{second model outperforms first.} \end{cases}$$

Extended sector model



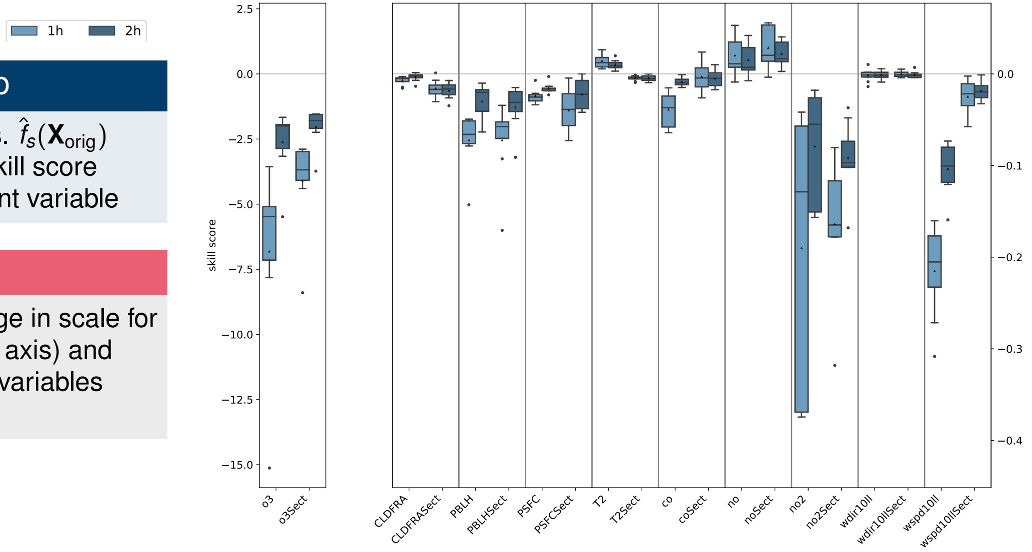
Sector model - importance of variables

Bootstrap

$\hat{f}_s(\mathbf{X}_{\mathrm{boot}})$ vs. $\hat{f}_s(\mathbf{X}_{\mathrm{orig}})$
negative skill score
$\Rightarrow$ important variable

Info

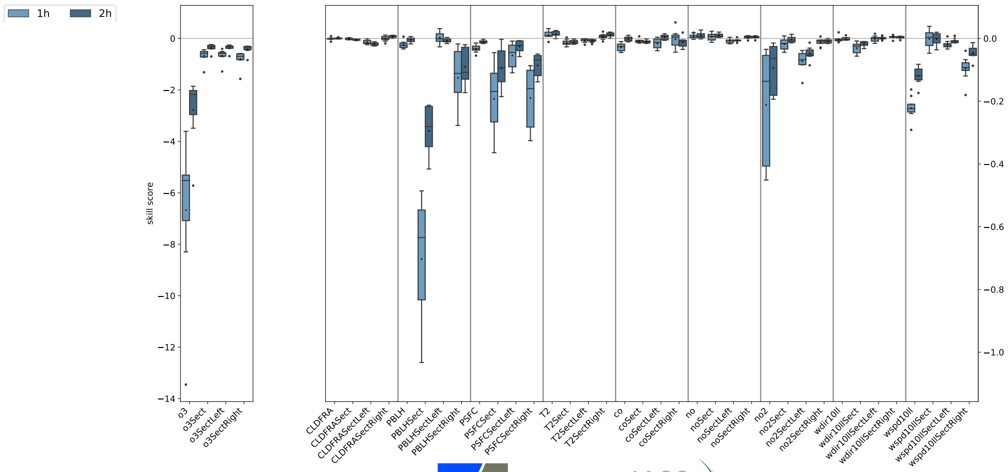
Note change in scale for ozone (left axis) and remaining variables (right axis)



Extended sector model - importance of variables

Bootstrap

$\hat{f}_{es}(\mathbf{X}_{\text{boot}})$ vs. $\hat{f}_{es}(\mathbf{X}_{\text{orig}})$: negative skill score $\Rightarrow$ important variable



Summary & Outlook

Summary

- Including aggregated information of one sector improves the forecast
- Including neighbouring sectors do not further improve the forecasts
- Influence of sector variables differ

⇒ Adding information of one sector seems to be promising also for larger networks

Outlook and further steps

- Increase of lead time
- Perform experiments on larger data set (expected to beat ols)
- Implement spatial interpolation as proposed by [7]
- Integration of best working method into IntelliO3-ts [2]

Acknowledgement

http://www.intelliaq.eu

European Research Council

Established by the European Commission

Funding is provided through
ERC Advanced grant
ERC-2017-ADG #787576
by Martin Schultz

UNIVERSITÄT BONN

IASS POTSDAM Institute for Advanced Sustainability Studies

JÜLICH Forschungszentrum

References I

[1] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. Fast and accurate deep network learning by exponential linear units (elus), 2016.

[2] F. Kleinert, L. H. Leufen, and M. G. Schultz. Intellio3-ts v1.0: a neural network approach to predict near-surface ozone concentrations in germany. Geoscientific Model Development, 14 (1):1–25, 2021. doi:10.5194/gmd-14-1-2021. URL `https://gmd.copernicus.org/articles/14/1/2021/`.

[3] F. Kuik, A. Kerschbaumer, A. Lauer, A. Lupascu, E. von Schneidemesser, and T. M. Butler. Top–down quantification of NO_x emissions from traffic in an urban area using a high-resolution regional atmospheric chemistry model. Atmospheric Chemistry and Physics, 18(11): 8203–8225, 2018. doi:10.5194/acp-18-8203-2018. URL `https://acp.copernicus.org/articles/18/8203/2018/`.

[4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, Nov 1998. ISSN 0018-9219. doi:10.1109/5.726791.

References II

[5] L. H. Leufen, F. Kleinert, and M. G. Schultz. Mlair (v1.0) – a tool to enable fast and flexible machine learning on air data time series. Geoscientific Model Development, 14(3):1553–1574, 2021. doi:10.5194/gmd-14-1553-2021. URL https://gmd.copernicus.org/articles/14/1553/2021/.

[6] M. G. Schultz, S. Schröder, O. Lyapina, O. Cooper, I. Galbally, I. Petropavlovskikh, E. Von Schneidemesser, H. Tanimoto, Y. Elshorbany, M. Naja, R. Seguel, U. Dauert, P. Eckhardt, S. Feigenspahn, M. Fiebig, A.-G. Hjellbrekke, Y.-D. Hong, P. Christian Kjeld, H. Koide, G. Lear, D. Tarasick, M. Ueno, M. Wallasch, D. Baumgardner, M.-T. Chuang, R. Gillett, M. Lee, S. Molloy, R. Moolla, T. Wang, K. Sharps, J. A. Adame, G. Ancellet, F. Apadula, P. Artaxo, M. Barlasina, M. Bogucka, P. Bonasoni, L. Chang, A. Colomb, E. Cuevas, M. Cupeiro, A. Degorska, A. Ding, M. Fröhlich, M. Frolova, H. Gadhavi, F. Gheusi, S. Gilge, M. Y. Gonzalez, V. Gros, S. H. Hamad, D. Helmig, D. Henriques, O. Hermansen, R. Holla, J. Huber, U. Im, D. A. Jaffe, N. Komala, D. Kubistin, K.-S. Lam, T. Laurila, H. Lee, I. Levy, C. Mazzoleni, L. Mazzoleni, A. McClure-Begley, M. Mohamad, M. Murovic, M. Navarro-Comas, F. Nicodim, D. Parrish, K. A. Read, N. Reid, L. Ries, P. Saxena, J. J. Schwab, Y. Scorgie, I. Senik, P. Simmonds, V. Sinha, A. Skorokhod, G. Spain, W. Spangl, R. Spoor, S. R.

UNIVERSITÄT BONN

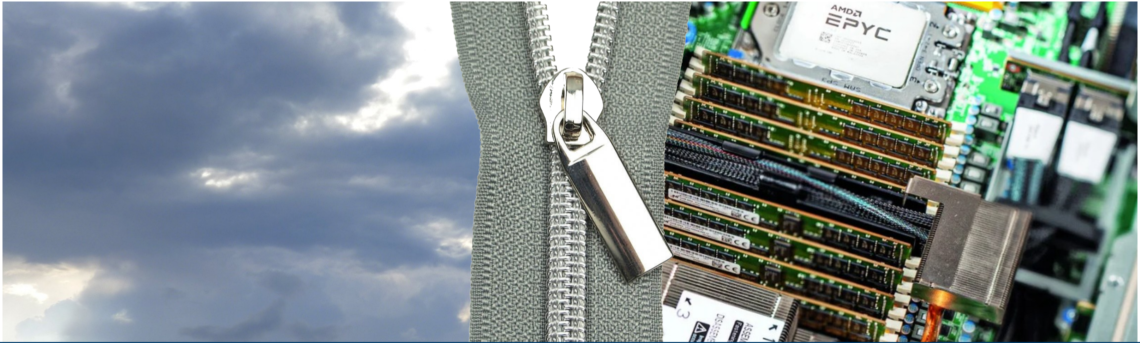
IASS POTSDAM Institute for Advanced Sustainability Studies

JÜLICH Forschungszentrum

References III

Springston, K. Steer, M. Steinbacher, E. Suharguniyawan, P. Torre, T. Trickl, L. Weili, R. Weller, X. Xu, L. Xue, and M. Zhiqiang. Tropospheric ozone assessment report: Database and metrics data of global surface ozone observations. Elementa: Science of the Anthropocene, 5(0):58, 2017. ISSN 2325-1026. doi:10.1525/elementa.244. URL http://www.elementascience.org/article/10.1525/elementa.244/.

[7] X. Yi, J. Zhang, Z. Wang, T. Li, and Y. Zheng. Deep Distributed Fusion Network for Air Quality Prediction. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 965–973, London United Kingdom, July 2018. ACM. ISBN 978-1-4503-5552-0. doi:10.1145/3219819.3219822. URL https://dl.acm.org/doi/10.1145/3219819.3219822.



Representing chemical history of ozone
a method development study for deep learning models

30-04-2021 | Felix Kleinert[FZJ, BN], Lukas H. Leufen[FZJ, BN], Aurelia Lupascu[IASS], Tim Butler[IASS], and Martin G. Schultz[FZJ] | f.kleinert@fz-juelich.de | vEGU 2021 ITS4.4/AS4.1

UNIVERSITÄT BONN

IASS POTSDAM Institute for Advanced Sustainability Studies

JÜLICH Forschungszentrum