Photography encouraged

NEAR SURFACE OZONE PREDICTIONS BASED ON MULTIPLE ANN ARCHITECTURES

F. Kleinert[BN, FZJ], B. Gong[FZJ], M. Götz[FZJ, KIT*] and M. Schultz[FZJ]

April 8, 2019 | Felix Kleinert | f.kleinert@fz-juelich.de | Jülich Supercomputing Centre (JSC) | *currently at

UNIVERSITÄT BONN

JÜLICH Forschungszentrum

TROPOSPHERIC OZONE IS...

Photography encouraged

- ... harmful to people and crops as it affects the respiratory system and decreases crop yield (Cooper et al., 2014; Monks et al., 2015).
- ... a green house gas
- ... a secondary air pollutant, formed by precursors (highly non linear coupled cycles)

Research questions

- How well do artificial neural networks predict continuous ground level ozone concentrations in southern Germany (**regression**)?

  For threshold exceedance predictions (classification) refer to Bing Gong (EGU2019-10543) at 5 pm in Room L3

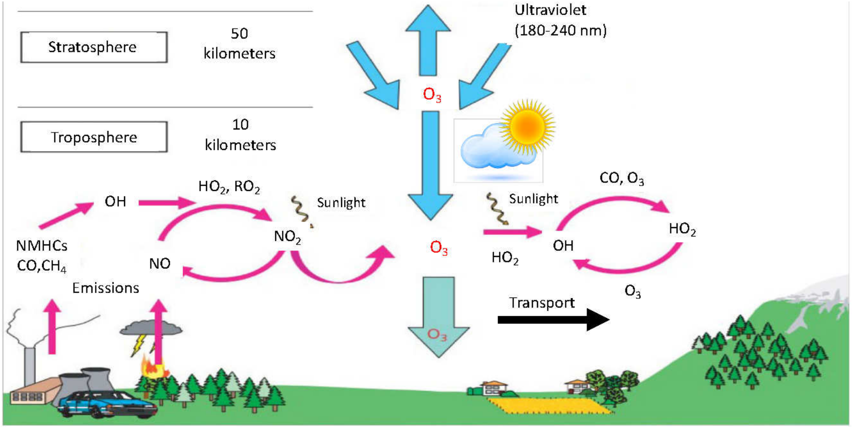
- What are the neural network's strengths and weaknesses for ground level ozone predictions?

Prediction task

"Use chemical and meteorological data of the last (few) days to predict daily maximum 8-hour average ozone concentration for a lead time of up to four days."

WHAT CONTROLS OZONE LOCALLY?

Chemistry and meteorology



Simplified schematic of tropospheric ozone production. Adopted from The Royal Society, 2008

DATABASE AND VARIABLE SELECTION

Station wise data (southern Germany)

- Chemical data are taken from Tropospheric Ozone Assessment Report (TOAR)
  Statistics according to the EU definition of ozone concentrations
- Meteorological data are taken from COSMO-REA6

→ Only stations with at least 3500 days of available data are selected (50 stations)

Variables and statistics

Variable	Description	Type	Unit
O_3	Daily maximum 8h average ozone	measur.	ppb
NO	Daily maximum 8h average nitrogen oxide	measur.	ppb
NO_2	Daily maximum 8h average nitrogen dioxide	measur.	ppb
u,v	Daily mean of wind's components	model rea	m/s
T-2M	Daily maximum of 2m temperature	model rea	$^\circ$C
relhum	Daily mean of relative humidity	model rea	%
pblheight	Daily maximum height of planetary boundary layer	model rea	m
cloudcover	Daily mean cloud cover	model rea	%

EXPERIMENT I:

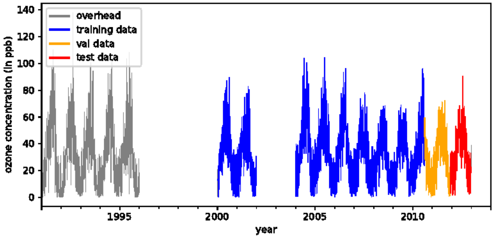
One network per station

Data preparation

- Input:
  - Create 2D data for each time step (daily data of 6 previous days $\times$ 9 variables)
  - Split time series into (**no** overlap)
    - training data (first 80 %)
    - validation data (following 10 %)
    - test data (remaining 10 %)
  - Standardise
- Output:
  - Create target vector of daily maximum 8-hour average ozone concentrations (lead time of up to 4 days)

Model setup

- 2 1D-convolutional layers
- 2 Fully-connected layers



Example time series used for training (blue), validation (yellow), and testing (red)

EXPERIMENT II:

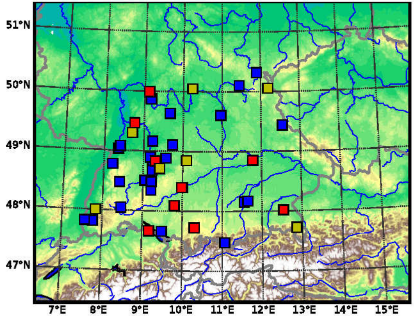
One network for all stations in southern Germany

Data preparation

- Input:
  - Create 2D data for each time step (daily data of 9 previous days $\times$ 9 variables)
  - Split stations into
    - training data (65 %)
    - validation data (15 %)
    - test data (20 %)
  - Standardise
- Output:
  - Create target vector of daily maximum 8-hour average ozone concentrations (lead time up to 4 days)

Model setup

- 4 1D-convolutional layers
- 3 Fully-connected layers



Stations in Southern Germany used for training (blue), validation (yellow), and testing (red).

VERIFICATION METHODS

Photography encouraged

Skill-Score

$$S = \frac{A - A_{\text{ref}}}{A_{\text{perf}} - A_{\text{ref}}},$$

where A is the score of interest (e.g. MSE)

Murphy and Winkler framework (1987)

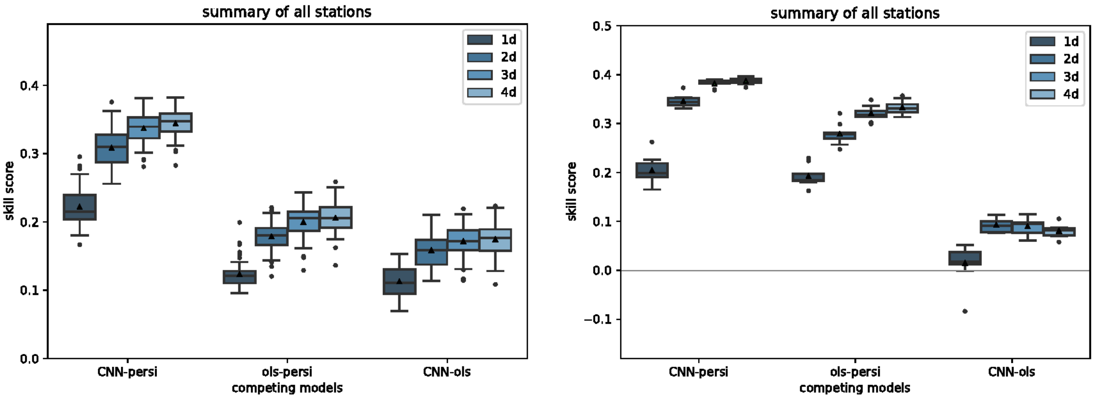
- Joint probability of model forecast (m) and observation (o)
- Calibration-Refinement Factorisation: $p(m, o) = p(o|m)\,p(m)$

Reference forecasts

Models

- Persistence forecast (persi)
- Ordinary least square (ols)
- Climatological forecasts (Murphy, 1988)
  - Single value prediction (station's mean)
  - Multi value prediction (station's monthly mean)

RESULTS: COMPETING MODELS

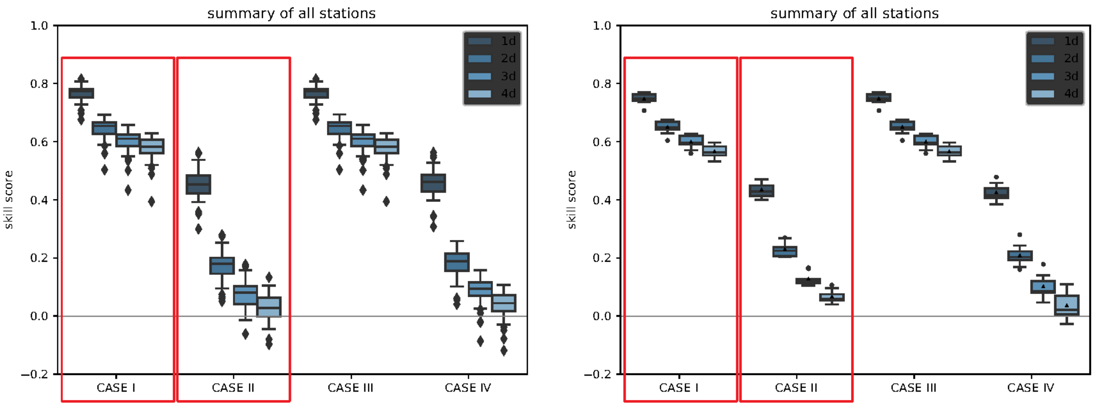


(a) Exp I, one network per station (b) Exp II, one network for all stations

[<model of interest> - < reference model>]

RESULTS: CLIMATOLOGY

Photography encouraged



(a) Exp I, one network per station

(b) Exp II, one network for all stations

CASE I: mean
CASE II: monthly mean } of station's test dataset

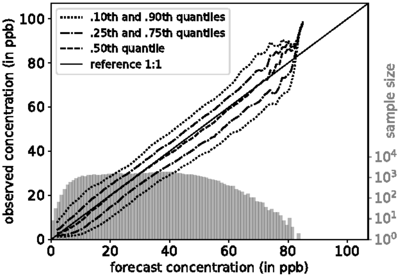
CASE III: mean
CASE IV: monthly mean } of station's full dataset

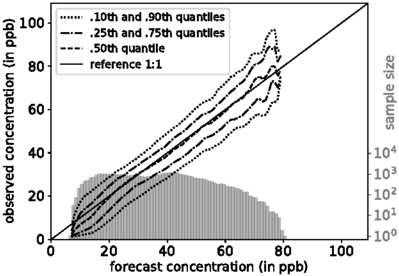
- CNNs capture the seasonal cycle of ozone

RESULTS: CONDITIONAL QUANTILES

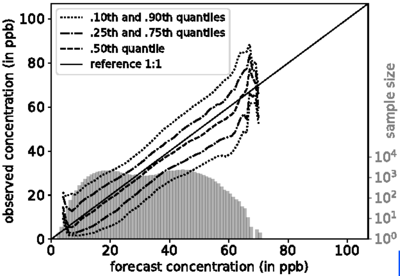
Photography encouraged

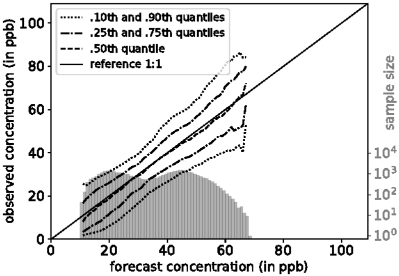
+1d





+4d





Exp I

Exp II

UNIVERSITÄT BONN

CONCLUSION

How well do artificial neural networks predict ground level ozone?

- CNNs outperform persistence, ols forecasts, and climatological forecasts (on the first two days)
- CNNs capture the seasonal cycle of ozone concentrations

What are the ANN's strengths and weaknesses of ground level ozone predictions?

+ Skilful predictions near the center of ozone distribution
+ At least climatological behaviour for longer lead times
+ Acceptable generalisation to unseen stations
- No skilful prediction at the distribution tails with increasing lead times

Outlook

- Make use of metadata (e.g. population density, night time light pollution, ...)
- Include regional information (e.g. neighbouring stations, weather maps, ...)

ACKNOWLEDGEMENT

Photography encouraged

IntelliAQ

http://www.intelliaq.eu

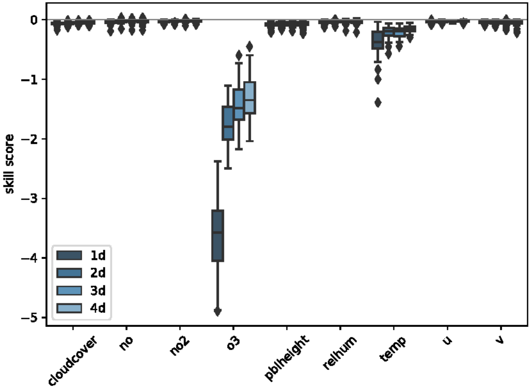
European Research Council

Established by the European Commission

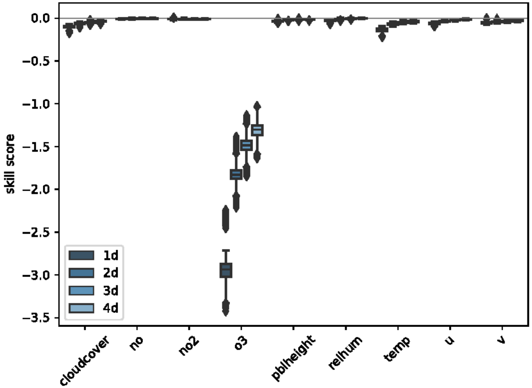
Funding is provided through
ERC Advanced grant
ERC-2017-ADG #787576
by Martin Schultz

Contact: f.kleinert@fz-juelich.de

RESULTS: IMPORTANCE OF INPUT VARIABLES



(a) Exp I, skill score of bootstraps with respect to original forecast



(b) Exp II, skill score of bootstraps with respect to original forecast