PREDICTION OF DAILY MAXIMUM OZONE THRESHOLD EXCEEDANCES BY ARTIFICIAL INTELLIGENCE TECHNIQUES IN GERMANY

April 8, 2019 | Bing Gong, Felix Kleinert, Martin Schultz | Jülich Supercomputing Center

JÜLICH Forschungszentrum

Overview

1. Background
    - 1.1 Motivation
    - 1.2 Research Questions and Objectives
    - 1.3 Ground-level ozone

2. Methodology
    - 2.1 Data selection and preparation
    - 2.2 Machine learning mechanism
    - 2.3 Imbalanced data solution
    - 2.4 Evaluation metrics

3. Results
    - 3.1 Orig VS. SMOTE
    - 3.2 ML VS. CNN
    - 3.3 Prediction accuracy on 4 leading days
    - 3.4 Deep learning improvement

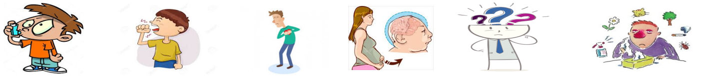
4. Conclusions and Future studies
    - 4.1 Conclusions and Future studies

Motivation
- Social Impact

Six common pollutants

PM | ***O_3*** | CO | SO_2 | NO_X | Lead

Health Effects

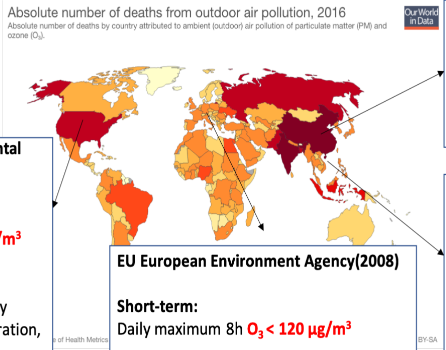


Asthma | ***Respiratory disease*** | ***Heart disease*** | ***Birth defects*** | ***Intellectual disorders*** | ***Immune system damage***

3 million deaths/year
(WHO, 2016)

US $225 billion/year
(World Bank, 2016)

Motivation
- Regulation and Standard



United States Environmental Protection Agency (2008)

Short-term:
Average of 8h **O_3 <140 µg/m^3**

Long-term:
Annual foruth-highest daily maximum 8-hour concentration, averaged over 3 years

China(MEPPRC, 2012)

Short-term:
Daily maximum 8h **O_3 < 160 µg/m^3**

EU European Environment Agency(2008)

Short-term:
Daily maximum 8h **O_3 < 120 µg/m^3**

Long-term:
Not to be exceeded on more than 25 days per year, averaged over three years

Hong Kong (APCO-Cap.311, 2012)

Short-term:
Daily maximum 8h **O_3 < 160 µg/m^3**

Long-term:
9 times of exceedance allowed per year.

Normativity of O_3 in United State, European Union, Mainland China, and Hong Kong

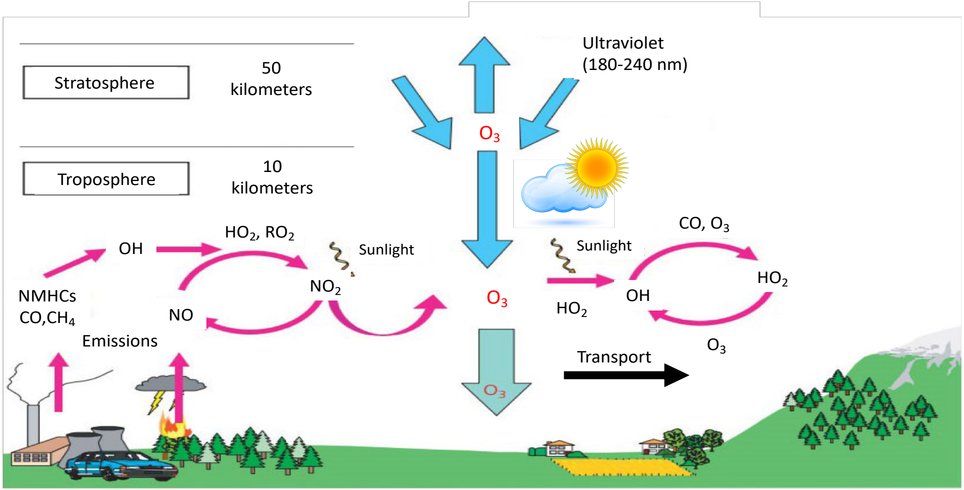
Research Questions and Objectives
- Predict the exceedances of the maximum 8h ozone concentrations in Germany

Research Questions:

- Can *__machine learning__* enhance the prediction accuracy for the air quality warning system ?
- How do different machine learning methods perform in the *__classification task__* of predicting threshold exceedances?
- *__How far in the future__* can make ozone prediction by machine learning/deep learning?

Map of the study region Baden-Württemberg in Germany (Source:onTheWorldMap.com)

Ground-level ozone
- Factors controlling ozone level



Schematic diagram for ozone formation (Fowler et al., 2008)

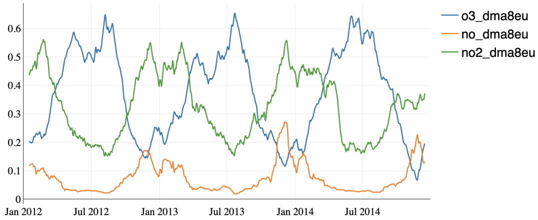
Ground-level ozone
- Key variables

Table: Pollution variables as inputs for modelling

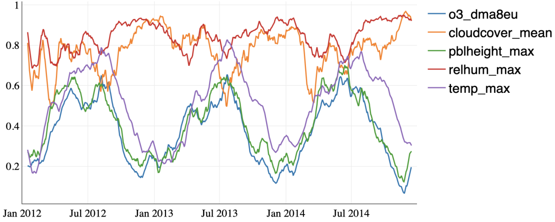
Variable	Description	Unit
o3_dma8eu	Daily maximum 8-hour average ozone	ppb
no_dma8eu	Daily maximum 8-hour nitrogen monoxide	ppb
no2_dma8eu	Daily maximum 8-hour nitrogen dioxide	ppb

Table: Meteorological variables as inputs for modelling

Variable	Description	Unit
pblheight_max	Maximum height of planetary boundary layer	m
relhum_max	Daily maximum relative humidity	%
temp_max	Daily maximum temperature	° C
u_mean	Daily mean u-component (zonal) of wind	m/s
v_mean	Daily mean v-component (meridional) of wind	m/s
cloudcover_mean	Daily mean total cloud cover	%



Moving average values of O_3, no_dma8eu and no2_dma8eu from 2012 to 2014 at Station DEBW010



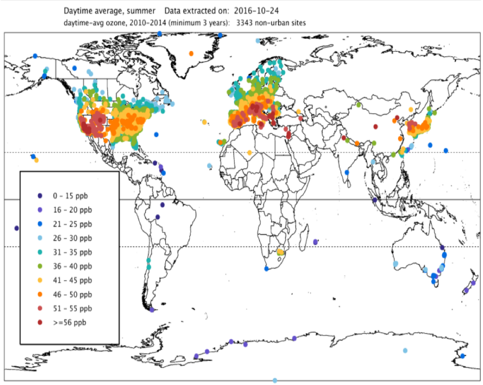
Moving average values of O_3, cloudcover_mean, pblheight_max, relhum_max, and temp_max from 2012 to 2014 at Station DEBW010

JÜLICH Forschungszentrum

Data selection and preparation

- Tropospheric Ozone Assessment Report (TOAR) Database for variable selection

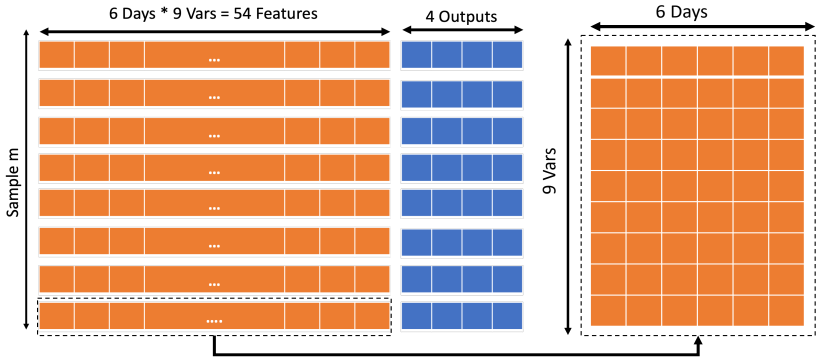
25 variables, more than ***10,000 measurement sites*** from over ***30 different sources*** around the world



TOAR
tropospheric ozone assessment report

Data selection and preparation

- Dataset structure for machine learning and Convolutional Neural Network

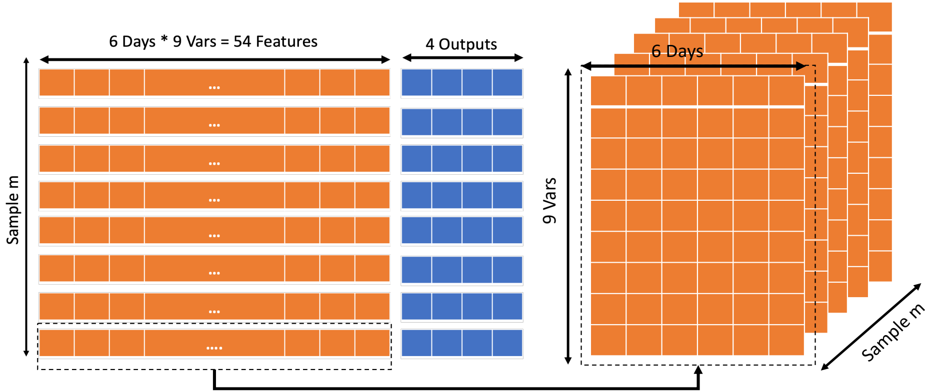


Data structure for machine learning

Data structure for 1D CNN

Data selection and preparation

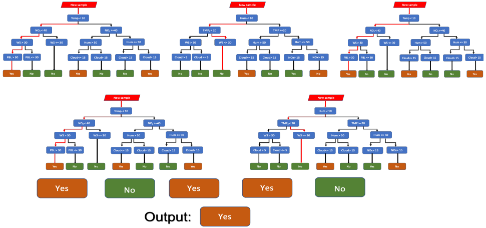
- Dataset structure for machine learning and Convolutional Neural Network



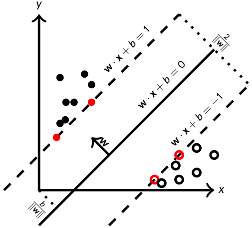
Data structure for machine learning

Data structure for 1D CNN

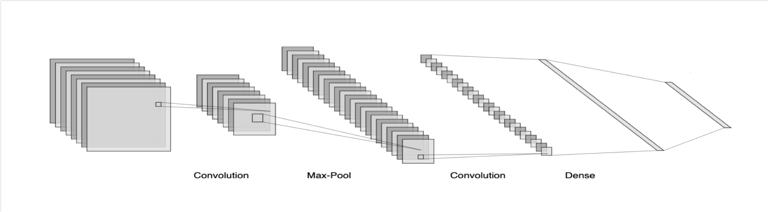
Machine learning mechanism
- Random Forest, Support Vector Machine, and Convolutional Neural Network



Random Forest



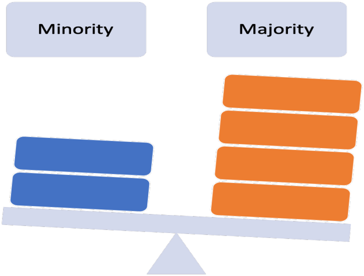
Support Vector Machine



1D CNN

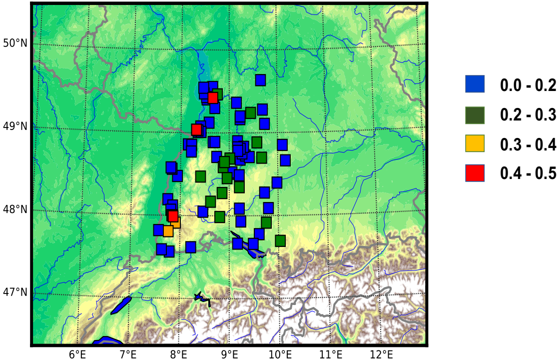
JÜLICH Forschungszentrum

Imbalanced data solution
- Imbalanced data issue



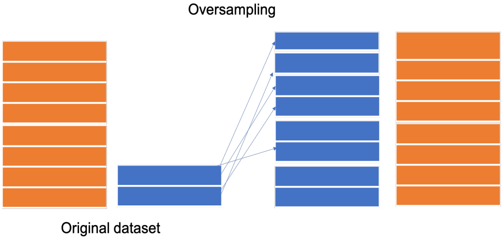
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

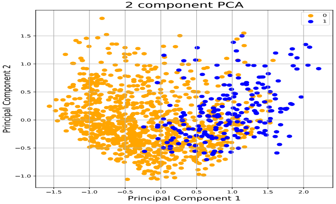
		Observed	
		Minority	Majority
Predicted	Minority	TP	FP
	Majority	FN	TN



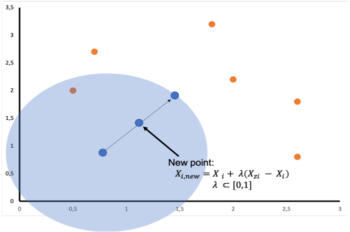
Exceedance frequency ($\frac{No.of\ Exceedances}{No.of\ samples}$) for all the monitoring stations in BW state

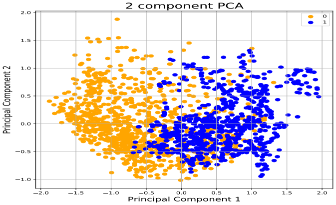
Imbalanced data solution
- Imbalanced data solution: Modifying class distribution





Scatter plot of raw data in 2 Dimensions





Scatter plot of SMOTE data in 2 Dimensions

Synthetic Minority Over-sampling Technique(SMOTE)

Evaluation metrics
- Classification evaluation metrics

Generic evaluation metric

		Observed		
		Positive	Negative	Total
Predicted	Positive	a	b	$a+b$
	Negative	c	d	$c+d$
	Total	$a+c$	$b+d$	N

- Accuracy = $\dfrac{a+d}{a+b+c+d}$
- Precision = $\dfrac{a}{a+b}$
- Recall = $\dfrac{a}{a+c}$

Evaluation metric for imbalanced data

- F1 - Score = $\dfrac{2 \cdot Recall \cdot Precison}{Recall + Precison}$
- G-means = $\dfrac{a \cdot d}{(d+c) \cdot (a+c)}$

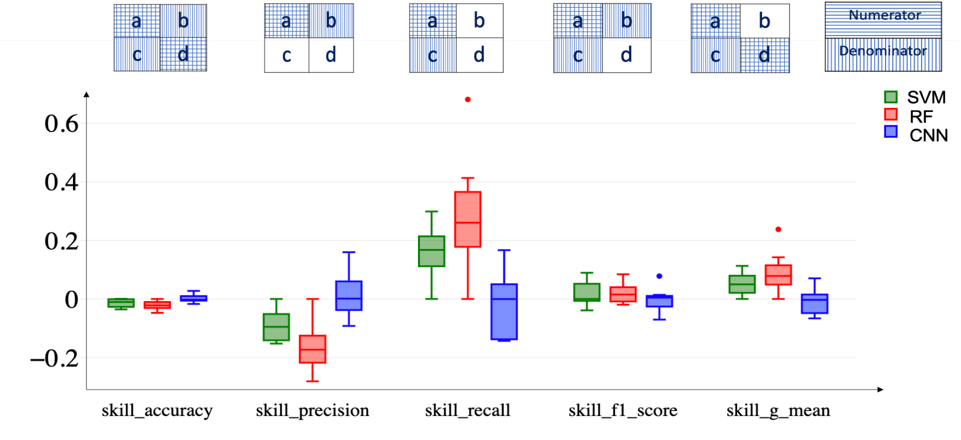
y_i are the prediction values, and $\hat{y}_i$ are the observed values.

Skill score

Use the skill score for comparing the target model *t* and reference model *ref* based on measurement *m*.

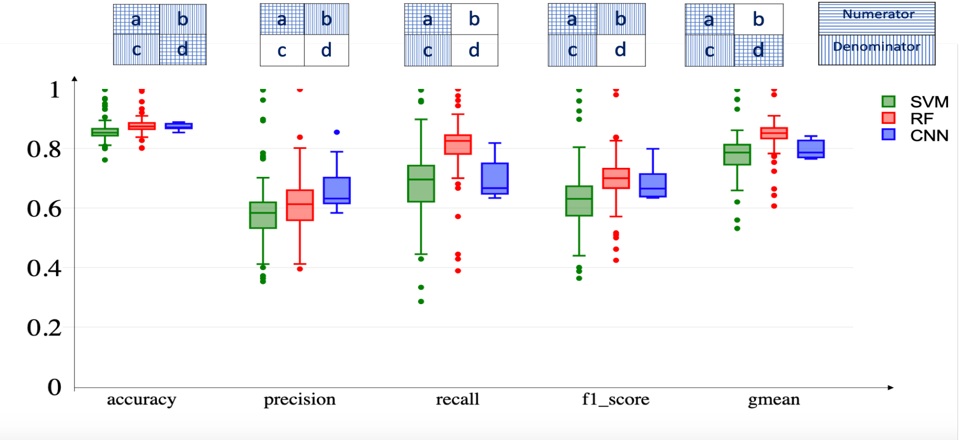
- $skill_m = (m - m_{ref})/m_{ref}$

Orig VS. SMOTE
- Does the SMOTE improve the model performance?



Skill scores ($skill_{SMOTE} = (m_{SMOTE} - m_{orig})/m_{orig}$) corresponding to different evaluation metrics for comparing on raw data and SMOTE data from all the monitoring stations

ML VS. CNN
- Does CNN improves the prediction accuracy?

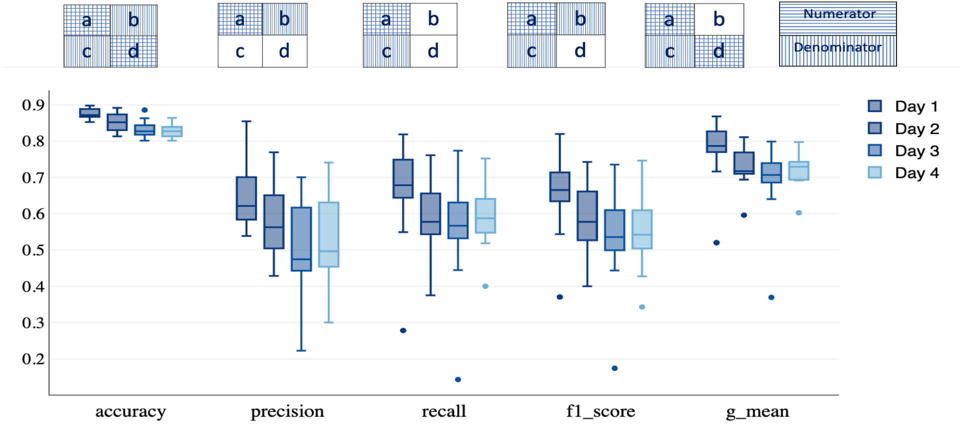


Evaluation metrics for SVM, RF, and CNN models on SMOTE dataset for all the monitoring stations

JÜLICH Forschungszentrum

Prediction accuracy on 4 leading days

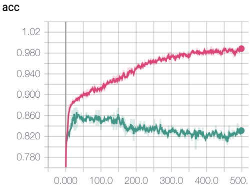
- Forecasting 4 leading days



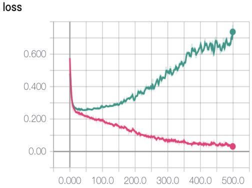
Evaluation metrics of CNN model on SMOTE to predict 4 days ahead for all the monitoring stations

JÜLICH Forschungszentrum

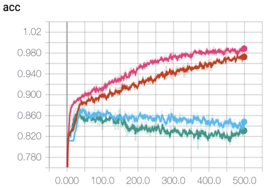
Deep learning improvement
- Deep Learning diagnosis



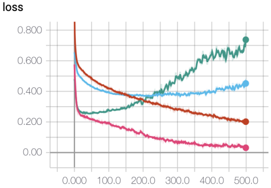
Accuracy-M1



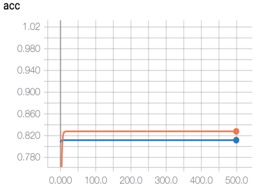
Loss-M1



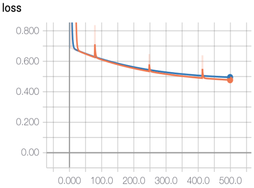
Accuracy-M1-Reg1-l-0.1



Loss-M1-Reg1-l-0.1



Accuracy-M1-Reg1-l-10



Loss-M1-Reg1-l-10

The reference 1D CNN was diagnosis as:

- Over-fitting
- Local minimal

Conclusions and Future studies
- Preliminary conclusions and Future studies

- **Preliminary conclusions:**
  - SMOTE can improve the classification performance for model SVM and RF.
  - With current set-up CNN is not better than traditional ML techniques; RF wins.
  - The prediction accuracy decrease significantly from 1 leading day to 2 day prediction.

- **Future studies:**
  - Deep learning structure
    - Cost-sensitive CNN for imbalanced data
    - Universal network for all the stations
    - LSTM
  - Data Level
    - Size of variables (Feature engineering, spatial factors etc.)
    - Heterogeneous data sources

Acknowledgement

IntelliAQ

http://www.IntelliAQ.eu

erc

European Research Council

Established by the European Commission

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

Thank You!

contact me: b.gong@fz-juelich.de

References I

- References:

Fowler, D., Amann, M., Anderson, F., Ashmore, M., Cox, P., Depledge, M., … others (2008). Ground-level ozone in the 21st century: future trends, impacts and policy implications. Royal Society Science Policy Report, 15(08).