

TEXTE

17/2023

Abschlussbericht

Weitergehende Untersuchungen zu Auswirkungen des Klimawandels auf die Ozonkonzentration in Deutschland (KliwO)

von:

Astrid Manders, Sadegh Mohammadi, Martijn Schaap
TNO, Utrecht

Herausgeber:

Umweltbundesamt

TEXTE 17/2023

Ressortforschungsplan des Bundesministeriums für
Umwelt, Naturschutz, nukleare Sicherheit und
Verbraucherschutz

Forschungskennzahl 3719 13 202 0
FB000972

Abschlussbericht

Weitergehende Untersuchungen zu Auswirkungen des Klimawandels auf die Ozonkonzentration in Deutschland (KliwO)

von

Astrid Manders, Sadegh Mohammadi, Martijn Schaap
TNO, Utrecht

Im Auftrag des Umweltbundesamtes

Impressum

Herausgeber

Umweltbundesamt
Wörlitzer Platz 1
06844 Dessau-Roßlau
Tel: +49 340-2103-0
Fax: +49 340-2103-2285
buergerservice@uba.de
Internet: www.umweltbundesamt.de

[f/umweltbundesamt.de](https://www.facebook.com/umweltbundesamt.de)

[t/umweltbundesamt](https://twitter.com/umweltbundesamt)

Durchführung der Studie:

TNO
Princetonlaan 6
3584CB Utrecht
die Niederlande

Abschlussdatum:

Oktober 2022

Redaktion:

Fachgebiet II 4.2 Beurteilung der Luftqualität
Andrea Mues

Publikationen als pdf:

<http://www.umweltbundesamt.de/publikationen>

ISSN 1862-4804

Dessau-Roßlau, Januar 2023

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren.

Kurzbeschreibung: Weitergehende Untersuchungen zu Auswirkungen des Klimawandels auf die Ozonkonzentration in Deutschland (KliwO)

Ozon ist ein Luftschadstoff mit negativen Auswirkungen auf die menschliche Gesundheit und auf die Vegetation. Seit den 1990er Jahre sind die Ozonkonzentrationen in Europa aufgrund von Emissionsreduzierungen zurückgegangen. Der Klimawandel wird jedoch die Bedingungen, die die Ozonbildung begünstigen, verstärken, was zu mehr Überschreitungen der Ozonrichtwerte führen kann. Wir haben den Einsatz von Methoden des Maschinellen Lernens (ML) untersucht, um Beziehungen zwischen jährlichen oder saisonalen Klimakennzahlen und der jährlichen Anzahl von Überschreitungen eines Ozonschwellenwertes für Messstandorte in Deutschland im Zeitraum 1995 – 2018 herzustellen, wobei Messungen aus den Messnetzen der Bundesländer und des UBAs und COSMO-REA6 meteorologische Daten verwendet wurden. Die Vorhersagewerte (Prädiktand) waren die Anzahl der Tage pro Jahr, an denen der höchste tägliche 8-Stunden-Mittelwert (MDA8) von $120 \mu\text{g}/\text{m}^3$ überschritten wurden (Ozonüberschreitungen), und die Anzahl der Episoden (zwei oder mehr aufeinander folgende Tage mit einer Ozonüberschreitung). Als Variablen (Prädiktoren) wurden unterschiedliche Klimakennzahlen (z. B. Anzahl Sommertage/Trockentage/tropische Nächte, Indikatoren zu Windrichtung und relativer Feuchte) und Stationsmerkmale verwendet. Für das Maschinelle Lernen wurde H2O AutoML verwendet. Es wurden getrennte ML-Modelle für normale (weniger als 55 Überschreitungen pro Jahr) und extreme Werte (55 oder mehr Überschreitungen) entwickelt.

Für die normalen Werte waren Temperaturindikatoren (Anzahl der tropischen Nächte/Sommertage) neben Informationen über den Breitengrad, die Stationshöhe und die Stationsklassifizierung die einflussreichsten Variablen. Bei den extremen Werten wurden die Auswirkungen von Strahlung, relativer Luftfeuchte und Wind dominanter. Für die normalen Werte war die Güte der ML-Modelle angemessen, für die extremen Werte waren die Ergebnisse jedoch weniger robust. Bei den ML-Modellen handelte es sich um sogenannte „relationship fitting“ Modelle, die auf neue Jahre und Stationen angewendet werden können, aber nicht für Emissionsszenarien. Sie sind weniger gut für die genaue Vorhersage extremer Bedingungen geeignet. Es ist zu empfehlen, die Klimadatensätze monatlich oder saisonal aggregiert statt jährlich aggregiert zu speichern. Es sollte zudem ein Indikator für die relative Luftfeuchte festgelegt werden, weil diese Variable eine wichtige Rolle in Prozessen im Bereich der Luftqualität spielt.

Abstract: Explorative study on quantifying the impact of climate change on ozone concentrations in Germany

Ozone is an air pollutant with negative impact on human health and on vegetation. Since the 1990s, ozone levels have decreased due to emission reductions. However, climate change will enhance the conditions favorable for ozone production, leading to more exceedances of ozone thresholds. We have explored the use machine learning methods to establish relationships between annual or seasonal climate indicators and the annual number of ozone exceedances for monitoring locations in Germany, covering 1995 –2018, using observations from the German networks and COSMO-REA6 meteorological data. The predictands were the number of days per year for which the daily running 8-hour average (MDA8) exceeded $120 \mu\text{g}/\text{m}^3$ and the number of episodes (two or more consecutive days with an exceedance). Variables (predictors) were several climate indicators (e. g. number of summer days/dry days/tropical nights, wind direction and relative humidity) and station characteristics. H2O AutoML was used for machine learning. Separate ML models were developed for normal (less than 55 exceedances per year) and extreme conditions (55 or more exceedances).

For the normal cases, temperature indicators (tropical nights, summer days) were dominant, next to information on latitude, station altitude and station classification. For the extreme cases, effects of radiation, relative humidity and wind became more dominant. The model performance was suitable for the normal conditions but the models were less robust for the extreme conditions. The models were relationship-fitting models that can be applied to new years and stations. They were not suitable for emission scenarios and less well suitable to predict extreme conditions accurately. It is advised to store climate data on a monthly or seasonal level instead of on annual level. It is also recommended to define an indicator for relative humidity as this variable is highly relevant for air quality processes.

Inhaltsverzeichnis

Abbildungsverzeichnis.....	10
Tabellenverzeichnis	15
Abkürzungsverzeichnis.....	17
Zusammenfassung.....	18
Summary	33
1 Einführung.....	49
1.1 Hintergrund und Ziel	49
1.2 Überblick der Arbeiten.....	50
2 Ozontrends und Klimawandel	51
2.1 Trends der Ozonkonzentration in Europa.....	51
2.1.1 Trend des Hintergrundozons	51
2.1.2 Trends der Ozonüberschreitungen	52
2.1.3 Einfluss des Emissionstrends	52
2.2 Ozontrends in Deutschland.....	54
2.2.1 Regionale und lokale Unterschiede	54
2.3 Ozontrends und Klimaszenarien	57
2.4 Jährliche Ozonüberschreitungen und Klimawandel	58
3 Auswahl der Ozonmetrik.....	60
3.1 Verfügbare Ozonmetriken	60
3.2 Auswahl der Metrik für den Ozonschwellenwert	62
3.2.1 Messstationen und Datenqualität	64
3.2.2 Überschreitung des Tageshöchstwertes MDH > 180 µg/m ³	66
3.2.3 Überschreitung des Schwellenwertes von MDH > 120 µg/m ³	67
3.2.4 Überschreitung des Schwellenwertes von MDA8 > 100 µg/m ³	69
3.2.5 Überschreitung des Schwellenwertes von MDA8 > 120 µg/m ³	71
3.2.6 Perzentilwerte der Ozonkonzentrationen	73
3.2.7 Zeitliche Variabilität der Anzahl der Überschreitungen und Episoden und deren Länge für ausgewählte Stationen	76
4 Für Ozon relevante meteorologische Variablen	82
4.1 KLENOS-Projekt.....	84
4.2 Studie für Bayern	85
4.3 Verfügbare Datensätze	85
4.3.1 ECAD-Daten.....	86

4.3.2	ReKliEs-De	87
4.3.3	Bewertung der Klimasimulationen im Projekt KLENOS	87
4.4	Auswahl der meteorologischen Daten und der Klimakennzahlen.....	88
4.5	Emissionsdaten	92
5	Regression und fortgeschrittenere Methoden des Maschinellen Lernens.....	93
5.1	Wesentliche ML Konzepte	94
5.1.1	Auswahl der Variablen.....	95
5.1.2	Abbruchkriterien und Bewertung.....	96
5.1.3	GAM (Generalized Additive Models)	97
5.1.4	Entscheidungsbaumbasierte Methoden.....	97
5.1.5	Bagging und Random Forest.....	98
5.1.6	Boosting und GBM	98
5.1.7	Neuronale Netze und andere Ansätze.....	99
5.1.8	Stacking oder Super Learning	100
5.2	Clusteranalyse.....	100
5.3	Datenanalyse und -visualisierung	101
6	Ergebnisse	105
6.1	Korrelationen zwischen Klimakennzahlen und Anzahl der Ozonüberschreitungen.....	105
6.2	Clustering	110
6.2.1	Clusterbildung nach zeitlich zusammenfallenden Ozonüberschreitungen, 2014-2018	111
6.2.2	Clusterbildung nach Anzahl der Überschreitungstage, gesamter Zeitraum.....	114
6.3	Ansätze für die Entwicklung von Algorithmen.....	117
6.4	Klassifizierung von Extrem- und Normalfällen.....	121
6.5	Zusammenfassung der Ergebnisse für Ozonüberschreitungen und Episoden	124
6.6	Modell A: normale Bedingungen, Anzahl der Ozonüberschreitungen	125
6.7	Modell A: normale Bedingungen, Anzahl der Ozonepisoden.....	132
6.8	Modell B: extreme Bedingungen, Anzahl der Ozonüberschreitungen	139
6.9	Modell B: extreme Bedingungen, Anzahl der Ozonepisoden.....	147
6.10	Ozonepisodenlänge.....	153
6.11	Reproduzierbarkeit der Ergebnisse.....	155
7	Schlussfolgerungen, Diskussion und Empfehlungen.....	157
8	Quellenverzeichnis	163
A	Benutzerhandbuch R-Skripte	167

Hauptebene: Library_Parameter.R	167
A.1 Verarbeitung von Ozondaten, Ozone_Exceedances.R	168
A.2 Stationsrasterindizes für COSMO Meteo_CoLocation.R.....	168
A.3 Extraktion von Klimakennzahlen: Meteo_Extraction.R	169
A.4 Konstruktion eines Datensatzes: Construct_Dataframe.R	169
A.5 Einstufung extreme und normale Jahre: Anomaly_Detection.R	170
A.6 Klassifizierung von Klimakennzahlen für extreme Jahre und normale Jahre Classification_Model.R.....	170
A.7 Training des Modells Regression_Model_Exceedances.R oder Regression_Model_Episodes.R	170
A.8 Visualisierung und Interpretation der Ergebnisse	171
B Zusätzliche Experimente	172
B.1 Berücksichtigung aller Daten	172
B.2 Verwendung von Klimazonen	175
B.2.1 Zone 1	175
B.2.2 Zone 2	178
B.2.3 Zone 3	181
C Alternative Ansätze: CTM und Training auf tägliche Daten	184

Abbildungsverzeichnis

Abbildung Z-1:	Anzahl der Überschreitungen des Ozon-MDA8-Wertes von 120 $\mu\text{g}/\text{m}^3$ pro Jahr, alle deutschen Stationen ohne Bergstationen.18
Abbildung Z-2:	Clusterbildung nach Anzahl der Ozonüberschreitungen (1995 – 2018, links) und nach zeitlich zusammenfallenden Ozonüberschreitungen (2014 – 2018, rechts).....23
Abbildung Z-3:	Flussdiagramm für die Entwicklung und Anwendung der ML-Modelle, unter Berücksichtigung der Klassifizierung der Werte in „normal“ und „extrem“.25
Abbildung Z-4:	SHAP-Plot (links) und Variable Importance Plot (rechts) für das GBM-Modell, das die Bedingungen für Extreme klassifiziert...26
Abbildung Z-5:	Korrelation der Prädiktoren und dem Prädiktand „Anzahl der Ozonüberschreitungen“ für „normale“ (links) und „extreme“ (rechts) Werte.27
Abbildung Z-6:	Modellvalidierung: Vorhergesagte gegenüber gemessene Anzahl von Ozonüberschreitungen pro Station und Jahr für „normale“ (links) und „extreme“ Werte (rechts).28
Abbildung Z-7:	Variable Importance Heatmap, Anzahl der Ozonüberschreitungen für „normale“ Werte. Die verschiedenen Modelle auf der x-Achse haben unterschiedliche Hyperparameter.29
Abbildung Z-8:	Partial Dependence Plots für die Anzahl der Ozonüberschreitungen für „normale“ Werte für die sechs wichtigsten Variablen.30
Abbildung 1:	Zusammenstellung des Jahresmittelwerts für Ozon (schwarz) und des vierthöchsten MDA8-Ozonwerts (rot), der zwischen 1990 und 2012 an 55 ländlichen EMEP-Messstationen gemessen wurde. Die dicke Linie ist der messnetzweite jährliche Median und die unteren/oberen Grenzen der schattierten Bereiche sind das 25. und 75. Perzentil. Die dünnen geraden Linien zeigen den linearen Trend über die Zeiträume 1990-2001 und 2002-2012 und die gestrichelten Linien zeigen die Referenzkonzentrationen der WHO-Luftqualitätsleitlinie (50ppb) und des EU-Langzeitzielwerts (60ppb). Quelle: TFMM-Bericht 2016 (Colette et al., 2017). ...52
Abbildung 2:	Trendzuweisung für 4MDA8 gemäß der TFMM/EURODELTA-TRENDS-Modellstudie (Colette et al., 2017)53

Abbildung 3:	Beobachtete Anzahl von Überschreitungen der Alarm- und Informationsschwellenwerte von 240 und 180 $\mu\text{g}/\text{m}^3$ in Deutschland (oberes Feld), beobachtete Jahresmittelkonzentrationen in Deutschland (unteres linkes Feld) und Emissionstrends in Deutschland (unteres rechtes Feld). Abbildungen entnommen aus Brasseur et al. (2017).....	55
Abbildung 4:	Obere Abbildung: Anzahl der Stunden mit Überschreitung der Informationsschwelle (180 $\mu\text{g}/\text{m}^3$) für ausgewählte Stationen. Untere Abbildung: Räumliche Verteilung der Anzahl der Tage mit Überschreitungen des MDA8 =120 $\mu\text{g}/\text{m}^3$ Schwellenwertes.	56
Abbildung 5:	Anzahl der gültigen Stationen pro Jahr nach der Datenprüfung	66
Abbildung 6:	Räumliche Verteilung der Anzahl der Überschreitungstage, MDH > 180 $\mu\text{g}/\text{m}^3$	67
Abbildung 7:	Räumliche Verteilung der Anzahl der Überschreitungstage, MDH > 120 $\mu\text{g}/\text{m}^3$	68
Abbildung 8:	Räumliche Verteilung der Anzahl der Episoden, MDH > 120 $\mu\text{g}/\text{m}^3$	68
Abbildung 9:	Mittlere Dauer der Episoden, MDH > 120 $\mu\text{g}/\text{m}^3$	69
Abbildung 10:	Räumliche Verteilung der Anzahl der Überschreitungstage, MDA8 > 100 $\mu\text{g}/\text{m}^3$	70
Abbildung 11:	Räumliche Verteilung der Anzahl der Episoden, MDA8 > 100 $\mu\text{g}/\text{m}^3$	70
Abbildung 12:	Mittlere Dauer der Episoden, MDA8 > 100 $\mu\text{g}/\text{m}^3$	71
Abbildung 13:	Räumliche Verteilung der Anzahl der Überschreitungstage, MDA8 > 120 $\mu\text{g}/\text{m}^3$	72
Abbildung 14:	Räumliche Verteilung der Anzahl der Episoden, MDA8 > 120 $\mu\text{g}/\text{m}^3$	72
Abbildung 15:	Mittlere Dauer der Episoden, MDA8 > 120 $\mu\text{g}/\text{m}^3$	73
Abbildung 16:	90. und 95. Perzentil der Ozonkonzentrationen, MDH	74
Abbildung 17:	90. und 95. Perzentil der Ozonkonzentrationen, MDA8	75
Abbildung 18:	Anzahl der Überschreitungen der MDH Schwellenwerte	76
Abbildung 19:	Anzahl und mittlere Dauer der Episoden, MDH > 120 $\mu\text{g}/\text{m}^3$.	76
Abbildung 20:	Anzahl der Überschreitungen der MDA8 Schwellenwerte	78
Abbildung 21:	Anzahl der Episoden der MDA8 Schwellenwerte	78
Abbildung 22:	Dauer der Episoden der MDA8 Schwellenwerte	78
Abbildung 23:	Anzahl der Ozonüberschreitungen von MDA8 > 120 $\mu\text{g}/\text{m}^3$, alle Stationen und alle Jahre	81

Abbildung 24:	Analyse der wichtigsten meteorologischen Einflussfaktoren für März-April-Mai und für Juni-Juli-August (Otero et al., 2016). Siehe Text für die wichtigsten Treiber für Deutschland. Die wichtigsten Faktoren für Deutschland sind Persistenz (LO ₃ , grau), relative Feuchte (RH, blau), Sonneneinstrahlung (SR, gelb) und Tagesmaximum der Temperatur (Tx, rot)84
Abbildung 25:	Jährliche Gesamtemissionen für Deutschland, wie in dieser Studie verwendet92
Abbildung 26:	Übersicht verschiedener Methoden zur Erstellung von Modellen für dieselben Datensätze.100
Abbildung 27:	Beispiel eines Partial Dependence Plots für die Anzahl der Überschreitungen und die normalisierte Anzahl der Sommertage102
Abbildung 28:	Beispiel eines SHAP-plots104
Abbildung 29:	Anzahl der Ozonüberschreitungen (oben), Anzahl der Sommertage (Mitte) und Anzahl der Hitzetage (unten) für ausgewählte Jahre105
Abbildung 30:	Korrelation der Eingangsvariablen und der Anzahl der Ozonüberschreitungen107
Abbildung 31:	Streudiagramme der Eingangsvariablen mit der Anzahl der Ozonüberschreitungen108
Abbildung 32:	Die drei Regionen mit Ozonüberschreitungen, die im KLENOS-Projekt ermittelt wurden.....110
Abbildung 33:	Centroide der Cluster von Überschreitungstagen für 2014 – 2018. Oberes Feld: MDA ₈ > 120 µg/m ³ , unteres Feld MDA ₈ > 100 µg/m ³ . Gelb = Überschreitung, lila = keine Überschreitung, auf der x-Achse sind die gesamten 5 Jahre aufgetragen, die y-Achse sind die Cluster.....112
Abbildung 34:	Cluster für den 5-Jahres-Zeitraum. Clustering an Tagen mit Überschreitungen für MDA ₈ > 100 µg/m ³ und MDA ₈ > 120 µg/m ³ . Obere Reihe: 6 Cluster in beiden Fällen, mittlere Reihe 4, 5, 7 Cluster MDA ₈ > 100 µg/m ³ und untere Reihe 4, 5, 7 Cluster MDA ₈ > 120 µg/m ³113
Abbildung 35:	Cluster Centroide für die jährlichen Überschreitungen des MDA ₈ > 120 µg/m ³115
Abbildung 36:	Clustering von MDA ₈ > 120 µg/m ³ , 1995 – 2018116
Abbildung 37:	Flussdiagramm für die Entwicklung und Anwendung der ML-Modelle, unter Berücksichtigung der Klassifizierung der Werte in „normal“ und „extrem“121

Abbildung 38:	Beiträge der Variablen für das Klassifizierungsmodell normale/extreme Anzahl von Ozonüberschreitungen (bestes GBM Modell), Variable Importance Plots (links) und SHAP Plot (rechts).....	123
Abbildung 39:	Partial Dependence Plots für das Klassifizierungsmodell.....	124
Abbildung 40:	Korrelation der Variablen, Anzahl der Ozonüberschreitungen für „normale“ Bedingungen, Modell A, Experiment 5	127
Abbildung 41:	Boxplot der gemessenen Anzahl der Ozonüberschreitungen (links) und vorhergesagte versus gemessene Anzahl der Ozonüberschreitungen (rechts) für „normale“ Bedingungen Modell A, Experiment 5	127
Abbildung 42:	Residuenanalyse für die Anzahl der Ozonüberschreitungen für „normale“ Bedingungen, als Streudiagramm mit vorhergesagten Werten (links) und als Histogramm (rechts)	128
Abbildung 43:	Model Prediction Correlation Heatmap, Anzahl der Ozonüberschreitungen für “normale“ Bedingungen, Modell A, Experiment 5	128
Abbildung 44:	Variable Importance Heatmap, Anzahl der Ozonüberschreitungen für „normale“ Bedingungen, Modell A, Experiment 5	129
Abbildung 45:	Partial Dependence Plots, Anzahl der Ozonüberschreitungen für „normale“ Bedingungen, Modell A, Experiment 5	129
Abbildung 46:	Variable Importance Plot (oben) und SHAP-Plot (unten) für das GBM-Modell mit der besten Performance, Anzahl der Ozonüberschreitungen für „normale“ Bedingungen, Modell A, Experiment 5	132
Abbildung 47:	Korrelation der Variablen, Anzahl der Episoden für „normale“ Bedingungen.....	133
Abbildung 48:	Boxplot (links) und vorhergesagte gegenüber der gemessenen Anzahl der Episoden (rechts) für „normale“ Bedingungen	134
Abbildung 49:	Residuenanalyse für die Anzahl der Episoden für „normale“ Bedingungen, als Streudiagramm mit vorhergesagten Werten (links) und als Histogramm (rechts).....	134
Abbildung 50:	Model Prediction Correlation Heatmap, Anzahl der Episoden für „normale“ Bedingungen	135
Abbildung 51:	Variable Importance Heatmap, Anzahl der Episoden für „normale“ Bedingungen.....	135
Abbildung 52:	Partial Dependence Plots, Anzahl der Episoden für „normale“ Bedingungen.....	136
Abbildung 53:	Variable Importance Plot (oben) and SHAP-Plot (unten) für das beste GBM-Modell, Anzahl der Episoden für „normale“ Bedingungen.....	139

Abbildung 54:	Korrelation der Variablen, Anzahl der Überschreitungen für „extreme“ Bedingungen	140
Abbildung 55:	Boxplot der Anzahl der Ozonüberschreitungen (links) und Scatterplot der vorhergesagten gegenüber gemessenen Anzahl der Überschreitungen (rechts) für „extreme“ Bedingungen..	141
Abbildung 56:	Residuenanalyse für die Anzahl der Ozonüberschreitungen für „extreme“ Bedingungen, als Streudiagramm mit vorhergesagten Werten (links) und als Histogramm (rechts), Modell B, Experiment 6	141
Abbildung 57:	Model Prediction Correlation Heatmap, Anzahl der Überschreitungen für „extreme“ Bedingungen, Modell B, Experiment 6	142
Abbildung 58:	Variable Importance Heatmap, Anzahl der Überschreitungen für „extreme“ Bedingungen	143
Abbildung 59:	Partial Dependence Plots, Anzahl der Überschreitungen für „extreme“ Bedingungen	144
Abbildung 60:	Variable Importance Plot (oben) and SHAP-Plot (unten) für das beste GBM-Modell, Anzahl der Überschreitungen für „extreme“ Bedingungen.....	147
Abbildung 61:	Korrelation der Variablen, Anzahl der Episoden für „extreme“ Bedingungen.....	148
Abbildung 62:	Boxplot (links) der Anzahl der Episoden für „extreme“ Bedingungen und Scatterplot (rechts) der vorhergesagten gegenüber der gemessenen Anzahl der Episoden für „extreme“ Bedingungen.....	149
Abbildung 63:	Residuenanalyse für die Anzahl der Episoden für „extreme“ Bedingungen, als Streudiagramm mit vorhergesagten Werten (links) und als Histogramm (rechts).....	149
Abbildung 64:	Model Prediction Correlation Heatmap, Anzahl der Episoden für „extreme“ Bedingungen	150
Abbildung 65:	Variable Importance Heatmap, Anzahl der Episoden für „extreme“ Bedingungen	151
Abbildung 66:	Partial Dependence Plots, Anzahl der Episoden für „extreme“ Bedingungen.....	152
Abbildung 67:	Variable Importance Plot (oben) und SHAP-Plot für das GBM-Modell mit der besten Performance (unten), Anzahl der Episoden für „extreme“ Bedingungen.....	153
Abbildung 68:	Korrelationen zwischen Anzahl der Ozonüberschreitungen, Anzahl der Ozonepisoden und maximale Ozonepisodendauer	154
Abbildung 69:	Korrelationen zwischen der Ozonepisodenlänge und der Anzahl der aufeinanderfolgenden Sommertage und Tropennächte .	155

Abbildung 70:	Korrelation von Eingangsvariablen/alle Klimakennzahlen, Emissionen und Anzahl der Ozonüberschreitungen, alle Daten	172
Abbildung 71:	Boxplot der Anzahl der Ozonüberschreitungen (links), vorhergesagte versus tatsächliche Anzahl der Überschreitungen (rechts) und Residuenanalyse (unten), alle Daten	173
Abbildung 72:	SHAP-Plot, Anzahl der Ozonüberschreitungen, alle Daten	174
Abbildung 73:	Korrelationsdiagramm Anzahl der Ozonüberschreitungen für Zone 1	175
Abbildung 74:	Boxplot der gemessenen Anzahl der Ozonüberschreitungen (links), vorhergesagte versus gemessene Anzahl der Ozonüberschreitungen (rechts) und Residuenanalyse (unten), Zone 1	176
Abbildung 75:	SHAP-Plot, Anzahl der Ozonüberschreitungen, Zone 1.....	177
Abbildung 76:	Korrelationsdiagramm Anzahl der Ozonüberschreitungen für Zone 2	178
Abbildung 77:	Boxplot der Anzahl der Ozonüberschreitungen (links), vorhergesagte versus tatsächliche Anzahl der Ozonüberschreitungen (rechts) und Residuenanalyse (unten), Zone 2	179
Abbildung 78:	SHAP-Plot, Anzahl der Ozonüberschreitungen, Zone 2.....	180
Abbildung 79:	Korrelationsdiagramm Anzahl der Ozonüberschreitungen Zone 3	181
Abbildung 80:	Boxplot der Anzahl der Ozonüberschreitungen (links), vorhergesagte versus tatsächliche Anzahl der Überschreitungen (rechts) und Residuenanalyse (unten), Zone 3	182
Abbildung 81:	SHAP-Plot, Anzahl der Ozonüberschreitungen, Zone 3.....	183
Abbildung 82:	Gemessene und vorhergesagte Anzahl von Überschreitungen für 2018 unter Verwendung eines „Relationship fitting“-Modells auf der Grundlage täglicher Daten oder der CTM LOTOS-EUROS.....	185

Tabellenverzeichnis

Tabelle Z-1:	Klimakennzahlen, die in den ML-Modellen verwendet wurden.	21
Tabelle Z-2:	Performance der endgültigen Modelle A und B.....	28
Tabelle 1:	Anzahl der Überschreitungstage für sechs Stationen.	77

Tabelle 2:	Anzahl der Episoden mit ihrer mittleren Dauer und Standardabweichung für sechs Stationen.....	79
Tabelle 3:	Leistungen der verschiedenen Metriken für verschiedene Kriterien.....	80
Tabelle 4:	Datenquellen meteorologische Variablen.....	89
Tabelle 5:	Klimakennzahlen, die in das Training der ML-Algorithmen einbezogen wurden.....	91
Tabelle 6:	Übersicht der Elemente im Datensatz.....	119
Tabelle 7:	Übersicht der Experimente	120
Tabelle 8:	Leistung der Modelle für Überschreitungen und Anzahl der Episoden, unter Verwendung separater Algorithmen für normale und extreme Bedingungen.....	125

Abkürzungsverzeichnis

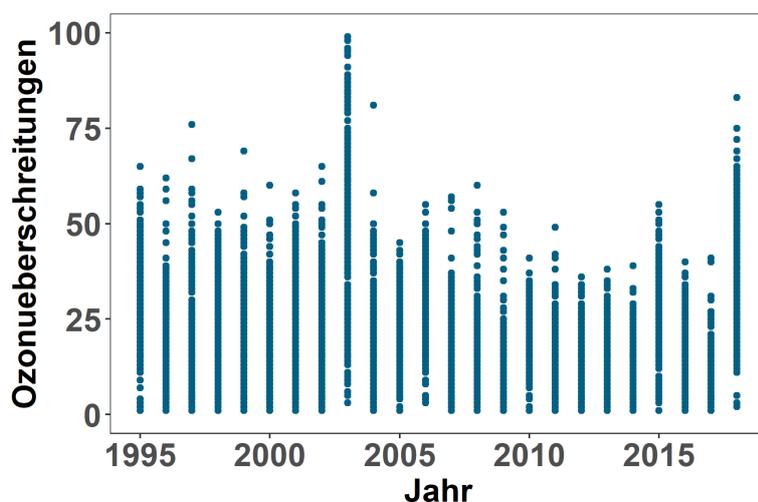
CH₄	Methan
DRF	Distributed Random Forest
ECAD	European Climate Assessment & Dataset
GAM	Generalized Additive Models
GBM	Gradient Boosting Machine
GLM	Generalized linear model
MDA8	Höchster täglicher 8-Stunden-Mittelwert
4MDA8	4. höchster MDA8
MDH	Höchster täglicher 1-Stunden-Mittelwert der Ozonkonzentration
MTDM	Mittelwert der zehn höchsten 1-Stunden-Mittelwerte eines Tages zwischen Mai und September
NMVOC	Flüchtige organische Verbindungen (Nicht-Methan)
N180	Anzahl der Stunden, in denen der Ozonwert die Informationsschwelle von 180 µg/m ³ überschreitet
NN	Neurales Netz
NO₂	Stickstoffdioxid
NO_x	Stickstoffoxide
O₃	Ozon
RF	Random Forest
Tmax	Maximale Temperatur
VOC	Flüchtige organische Verbindung
UBA	Umweltbundesamt, Dessau
XRT	eXtremely Randomized Trees

Zusammenfassung

Ozon ist ein Luftschadstoff mit negativen Auswirkungen auf die menschliche Gesundheit und die Vegetation. Zum Schutz der Gesundheit hat die Weltgesundheitsorganisation (World Health Organisation, WHO) daher Metriken und Richtwerte festgelegt. Auch in der europäischen Richtlinie zur Luftqualität (2008/50/EG) finden sich Schwellen- und Zielwerte für Ozon, die sich jedoch von den Richtwerten der WHO unterscheiden. Ozon ist ein sekundärer Luftschadstoff, der aus Stickoxiden (NO_x) und Emissionen von flüchtigen organischen Verbindungen (VOC) anthropogenen und natürlichen Ursprungs gebildet wird. Die Bildungsreaktionen werden durch die Sonneneinstrahlung angetrieben, daher werden hohe Ozonkonzentrationen im Sommer, insbesondere an warmen, sonnigen Tagen, erreicht. Auch andere meteorologische Parameter wie relative Luftfeuchte, Wind und weitere Transportgrößen spielen eine Rolle bei der Entstehung hoher Ozonkonzentrationen.

Seit den 1990er Jahren ist in Europa die Häufigkeit hoher Ozonkonzentrationen aufgrund von Emissionsreduktionen der Vorläuferstoffe zurückgegangen (Colette et al., 2017). Dieser Rückgang von hohen Ozonkonzentrationen wird jedoch teilweise durch die Wirkungen des Klimawandels ausgeglichen (Colette et al. 2017). Abbildung Z-1 zeigt den Verlauf der Anzahl der jährlichen Überschreitungen des höchsten täglichen 8-Stunden-Mittelwertes (MDA8 - Maximum Daily 8h Average) von $120 \mu\text{g}/\text{m}^3$ (EU-Zielwert) an Ozonmessstationen in Deutschland in den Jahren 1995 – 2018 (siehe auch UBA, 2020). Deutlich erkennbar ist, dass die außergewöhnlich heißen Sommer der Jahre 2003 und 2018 mit einer hohen Anzahl von Überschreitungen verbunden sind. In den vergangenen Jahrzehnten sind die Sommer in Deutschland wärmer geworden¹, und Klimaszenarien deuten auf einen weiteren Temperaturanstieg und ein häufigeres Auftreten von stagnierenden Wetterlagen hin. Dies begünstigt das Auftreten hoher Ozonkonzentrationen weiter, sofern die Emissionen der Vorläuferstoffe nicht weiter reduziert werden.

Abbildung Z-1: Anzahl der Überschreitungen des Ozon-MDA8-Wertes von $120 \mu\text{g}/\text{m}^3$ pro Jahr, alle deutschen Stationen ohne Bergstationen.



Quelle: Eigene Darstellung, TNO

¹ <https://www.umweltbundesamt.de/daten/klima/trends-der-lufttemperatur>

Um die zeitliche Veränderung der Ozonkonzentrationen zu quantifizieren, wurden beispielsweise im Projekt KLENOS (Pavlik et al., 2016) regionale Klimasimulationen mit anschließender Chemie-Transport-Modellierung für Deutschland durchgeführt. Da diese Simulationen sehr rechenintensiv sind, wurden sie in KLENOS auf nur wenige Klima- und Emissionsszenarien beschränkt. Es stellt sich in diesem Zusammenhang die Frage, ob es möglich wäre, mit der Methode des Maschinellen Lernens (ML) mehr Klimaszenarien zu berücksichtigen, da die Rechenzeit eines ML-Modells im Vergleich zu einem Chemie-Transport-Modell (CTM) geringer ist. In der Literatur wird über mehrere erfolgreiche Versuche berichtet, die Ozonkonzentration mit Hilfe Maschinellen Lernens auf Tagesbasis vorherzusagen. Dies geschieht meist mit dem Ziel, kurzfristige Ozonvorhersagen für das aktuelle Wetter (aktuelle Klimabedingungen) zu treffen. Für einen ähnlichen Ansatz auf Tagesbasis für ein künftiges Klima (Klimaszenarien) wären jedoch weiterhin Tagesdaten aus rechenintensiven Klimamodellen erforderlich. Bei vielen Klimasimulationen werden jedoch nur komprimierte Datensätze von Klimaparametern und -kennzahlen gespeichert, um das Datenvolumen zu verringern. Diese Daten eignen sich nicht zur Vorhersage von Ozonüberschreitungen auf Tagesbasis, können aber möglicherweise zur Vorhersage der Anzahl der Überschreitungen eines Ozonschwellenwertes auf Jahresbasis verwendet werden. Die zentrale Forschungsfrage, die im vorliegenden Bericht beantwortet werden soll, lautet daher:

Lassen sich mit Hilfe statistischer Regressionsverfahren Zusammenhänge zwischen der Überschreitung eines Schwellenwertes für die Ozonkonzentration und Klimakennzahlen für deutsche Ozonmessstationen finden?

Die Studie basiert auf Ozonmessungen an deutschen Stationen für den Zeitraum 1995 – 2018. Die Identifizierung einer geeigneten Metrik für den in dieser Studie zu verwendenden Ozonschwellenwert und geeigneter Klimakennzahlen ist Teil der Untersuchung. Dabei sind nicht nur die Anzahl der Überschreitungen des Ozonschwellenwertes, sondern auch Dauer und Anzahl von Ozonepisoden, also Schwellenwertüberschreitungen an mindestens zwei aufeinanderfolgenden Tagen, relevant.

Zur Beantwortung dieser Forschungsfrage wurden folgende Schritte bearbeitet:

1. Bereitstellung und Aufbereitung von Ozonmessungen für den Zeitraum 1995 – 2018 an deutschen Stationen der Bundesländer und des Umweltbundesamtes (UBA), die rund 200 aktive Stationen pro Jahr umfassen. Stationen die nicht das Datenqualitätsziel erfüllen, wurden für das jeweilige Jahr nicht berücksichtigt. Auch Bergstationen (> 900 m Höhe) wurden aus der Analyse ausgeschlossen
2. Identifizierung einer geeigneten Metrik für den Ozonschwellenwert. Die Metrik sollte Extremwerte erfassen, gleichzeitig sollten mit der Metrik eine so hohe Anzahl von Schwellenwertüberschreitungen in den zu analysierenden Daten gefunden werden, dass jede ML-Methode mit ausreichendem Vertrauen angewendet werden kann.
3. Identifizierung von Klimakennzahlen, die mit der Ozonbildung korrelieren. Eine Klimakennzahl ist eine berechnete Größe und für jede Klimakennzahl werden bestimmte Klimaparameter verwendet, z.B. Anzahl der Sommertage aus der Tageshöchsttemperatur.
4. Erstellung eines Datensatzes für das Maschinelle Lernen, der die Ozonmessungen, Informationen über die Messstationen, Klimakennzahlen an den Ozonmessstationen und Emissionswerte für Stickstoffoxide, VOC und Methan enthält.

5. Untersuchungen zu Stationsgruppen (Cluster), mit der Fragestellung, ob eine Clusterbildung der Ozonmessungen für Stationen mit ähnlichen Eigenschaften einen Mehrwert für die Ergebnisse der ML-Anwendungen darstellt.
6. Identifizierung geeigneter Methoden des Maschinellen Lernens (Algorithmen, die zu interpretierbaren Ergebnissen führen und mit Nichtlinearitäten und der Erfassung von Extremen umgehen können).
7. Entwicklung eines Tools zur Datenbearbeitung und Analyse. Das Tool trainiert relevante ML-Algorithmen, das heißt, der ML-Algorithmus wird auf den konkreten Datensatz angewendet (trainiert), um statistische Zusammenhänge zu beschreiben. Ein trainierter Algorithmus wird dann „ML-Modell“ genannt.
8. ML-Modelle auswerten, visualisieren und interpretieren.

Die aufgelisteten Schritte werden im Folgenden näher erläutert.

Auswahl der Ozonmetrik

Unter Verwendung aller verfügbaren Ozonmesswerte aus den Jahren 1995 - 2018 wurden mehrere mögliche Metriken untersucht. Diese beinhalten die WHO- und EU-Richtwerte bzw. Schwellen- und Zielwerte für den höchsten täglichen 8-Stunden-Mittelwert (MDA8), berechnet aus stündlich gleitenden 8-Stunden-Mittelwerten. Außerdem wurde der Tageshöchstwert basierend auf 1-Stunden-Mittelwerten (MDH – Maximum Daily 1h Average) in die Untersuchung einbezogen. Die endgültige Metrik, die in diesem Projekt verwendet wurde, ist $MDA8 > 120 \mu\text{g}/\text{m}^3$, d. h. eine Überschreitung des Schwellenwertes wird gezählt, wenn der höchste tägliche 8-Stunden-Mittelwert höher ist als $120 \mu\text{g}/\text{m}^3$. Die wichtigsten Argumente für diese Auswahl sind:

- ▶ MDA8 ist robuster gegenüber Ausreißern und Datenlücken im Vergleich zu Metriken, die den Tageshöchstwert verwenden.
- ▶ $MDA8 > 120 \mu\text{g}/\text{m}^3$ entspricht weitgehend dem 90. Perzentil der MDA8-Werte und ist ein guter Indikator für Extremwerte.
- ▶ $MDA8 > 120 \mu\text{g}/\text{m}^3$ weist eine ausreichende zeitliche und räumliche Variabilität auf, um Unterschiede in Bezug auf Ort und Jahr zu zeigen.
- ▶ $MDA8 > 120 \mu\text{g}/\text{m}^3$ ist eine bestehende Metrik, die in der Luftqualität-Community bekannt ist.

Diese Metrik, die dem EU-Zielwert entspricht, wird verwendet, um die Anzahl der Schwellenwertüberschreitungen, die Anzahl der Episoden und die Episodendauer zu quantifizieren. $MDA8 > 120 \mu\text{g}/\text{m}^3$ wird im weiteren Text als „Ozonüberschreitung“ bezeichnet.

Auswahl von Klimakennzahlen und Klimadaten

Der zweite Schritt besteht darin, Klimakennzahlen zu identifizieren, die mit hohen Ozonkonzentrationen in Zusammenhang stehen. Die Klimakennzahlen werden nur für den Zeitraum April – September berechnet, da sie nur für diese Monate eine sinnvolle Beziehung zu hohen Ozonkonzentrationen aufweisen. Tabelle Z-1 gibt einen Überblick über die verwendeten Klimakennzahlen. Da für die relative Luftfeuchte keine Klimakennzahl aus der Literatur verfügbar war, wurde eine eigene Definition erstellt. Es werden nur Klimakennzahlen verwendet, die direkt mit einzelnen meteorologischen Variablen zusammenhängen. Kombinierte Klimakennzahlen, wie zum Beispiel feuchtwarme Tage werden nicht berücksichtigt. Verschiedene etablierte Methoden des Maschinellen Lernens sollten von ihrer Konzeption her in der Lage sein, solche Kombinationen zu finden, wenn sie relevant sind. Als

Datenquelle für alle meteorologischen Daten wurden COSMO-Reanalysedaten mit einer Auflösung von 6x6 km² (COSMO-REA6, Bollmeyer et al., 2015) verwendet. Die Daten der Klimakennzahlen wurden aus den Gitterzellen entsprechend der Lage der Ozonmessstationen für den Zeitraum 1995 – 2018 extrahiert.

Tabelle Z-1: Klimakennzahlen, die in den ML-Modellen verwendet wurden.

Klimakennzahl	Verwendete meteorologische Variablen	Berechnung aus stündlichen Daten	Benennung der Klimakennzahlen in Abbildungen
Anzahl der Sommertage	Tagesmaximum Temperatur (T_{Max})	Tage mit $T_{Max} > 25\text{ °C}$	Sommertage
Anzahl der Hitzetage	Tagesmaximum Temperatur (T_{Max})	Tage mit $T_{Max} > 30\text{ °C}$	Hitzetage
Anzahl der tropischen Nächte	Tagesminimum Temperatur (T_{Min})	Tage mit $T_{Min} > 20\text{ °C}$	Tropennächte
Anzahl der Trockentage	Tagessumme Niederschlag	Tage mit Niederschlag $< 1\text{ mm}$	Trockentage
Anzahl der Tage mit RF $< 60\%$ (RF_trocken)	Tagesdurchschnitt relative Feuchte (RF)	Tage mit RF $< 60\%$	RF_trocken
Anzahl der Tage mit RF $> 80\%$ (RF_feucht)	Tagesdurchschnitt relative Feuchte (RF)	Tage mit RF $> 80\%$	RF_feucht
Anzahl der Tage mit östlicher Strömung	Windrichtung (WindRI)	$45^\circ < \text{WindRI} \leq 135^\circ$	Wind_Ost
Anzahl der Tage mit südlicher Strömung	Windrichtung (WindRI)	$135^\circ < \text{WindRI} \leq 225^\circ$	Wind_Süd
Anzahl der Tage mit westlicher Strömung	Windrichtung (WindRI)	$225^\circ < \text{WindRI} \leq 315^\circ$	Wind_West
Anzahl der Tage mit schwachem Wind	10m-Windgeschwindigkeit (WindGE)	WindGE $< 2\text{ m/s}$	Schwachwindtage
Mittlere einfallende Strahlung	Strahlung	Kumulative stündliche Mittelwerte der Strahlung	Strahlung

Aufbau des Datensatzes für das Maschinelle Lernen

Es wurde ein Datensatz für den Zeitraum 1995--2018 für alle Ozonmessstationen pro Jahr erstellt. Eine Datenzeile beinhaltet die folgenden Informationen:

- ▶ Standort der Station (Längengrad, Breitengrad, Höhe),
- ▶ Klassifizierung der Ozonstation (städtischer Hintergrund, vorstädtischer Hintergrund, Verkehr, Industrie, ländlicher Hintergrund, ländlich regional),
- ▶ Anzahl der Ozonüberschreitungen,
- ▶ Anzahl der Ozonepisoden,
- ▶ maximale Ozonepisodenlänge,

- ▶ jährlich gemeldete deutsche Gesamtemissionen von NO_x, VOC und Methan,
- ▶ Klimakennzahlen

Die NO_x-, VOC- und Methanemissionen spielen eine wichtige Rolle bei der Ozonbildung und bestimmen einen Teil des Ozontrends. Der endgültige Datensatz besteht aus mehr als 6000 Datenzeilen.

Clusterbildung von Stationen

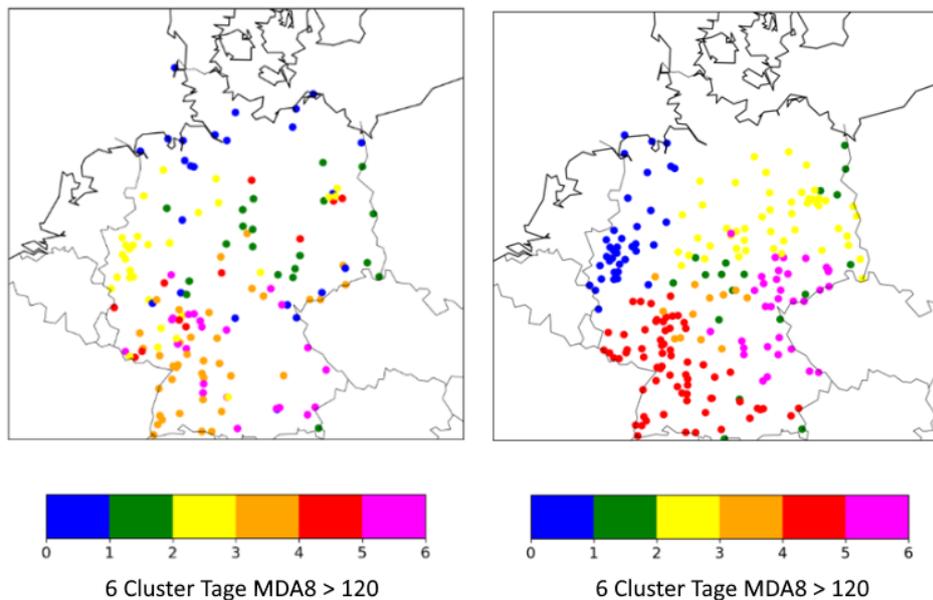
Die Idee der Clusterbildung besteht darin, Stationen mit ähnlichen Eigenschaften zu finden, die einen Datensatz für das Training eines ML-Algorithmus bilden können. Es wurde untersucht, ob Gruppen von Stationen (Cluster) identifiziert werden können, die unterschiedliche bzw. unterschiedlich starke statistische Beziehungen zwischen Ozonüberschreitungen und Klimakennzahlen aufweisen. Zur Gruppierung der Stationen können beispielsweise die Stationsklassifizierung oder die Klimazone der Station (West, Mitte, Ost, wie im KLENOS-Projekt abgeleitet) verwendet werden. Das Clustern der Stationen kann die Genauigkeit des Maschinellen Lernens für dieses Cluster erhöhen, führt jedoch auch zu einem kleineren Datensatz für das Training und die Validierung des ML-Algorithmus.

In dieser Studie wurde das Clustern mit dem K-Median-Algorithmus untersucht. Dies ist eine Methode, die solche Stationen gruppiert, die einander am ähnlichsten sind und sich gleichzeitig am stärksten von anderen Gruppen unterscheiden. Im Projekt wurden zwei Ansätze mit einer einfachen Definition dieses Unterschieds getestet. Im ersten Ansatz basierte das Clustern auf der jährlichen Anzahl der Ozonüberschreitungen im Zeitraum 1995 – 2018 und berücksichtigte auch das jeweilige Jahr, in dem die Überschreitungen festgestellt wurden. Der zweite Ansatz umfasste einen kürzeren Zeitraum (2014 – 2018), berücksichtigte jedoch den Tag des Jahres, an dem eine Überschreitung stattfand, und basierte somit auf täglichen statt auf jährlichen Daten. In beiden Experimenten wurde eine unterschiedliche Anzahl von Clustern zwischen 4 und 9 getestet. Ist die Anzahl der Cluster zu klein, können die Stationen in einem Cluster immer noch sehr unterschiedlich sein, und somit die Qualität der Aussage des endgültigen Modells zur Vorhersage von Ozonüberschreitungen mindern. Eine zu große Anzahl von Clustern kann jedoch dazu führen, dass zu wenig Daten für ein robustes Modell enthalten sind. Die Cluster aus beiden Experimenten weisen gewisse Ähnlichkeiten auf, dennoch heben die beiden Cluster-Experimente unterschiedliche Merkmale hervor. Experiment 1 hebt vor allem die Charakteristiken der Stationsklassifizierung hervor, während in Experiment 2 die Cluster vorwiegend stärker in Regionen eingegrenzt sind (Abbildung Z-2), wobei die Regionen allerdings nicht scharf begrenzt sind.

Als Alternative zu dem K-median-Algorithmus wurden die Stationen entsprechend der drei Klimazonen, wie sie in KLENOS definiert sind, gruppiert. Es zeigte sich, dass die Anzahl der Daten mit einer hohen Anzahl von Ozonüberschreitungen pro Cluster (Klimazone) sehr gering wurde, was zu weniger robusten Ergebnissen des Überschreitungsvorhersagemodells für diese Höchstwerte führte.

Im Ergebnis der Untersuchungen wurde das Clustern der Stationen verworfen und alle Stationen in den endgültigen Datensatz für die Modellentwicklung für ganz Deutschland aufgenommen. Methoden des Maschinellen Lernens können Informationen über den Standort und die Klassifizierung der Stationen direkt berücksichtigen. Eine vorherige Clusterbildung nach Region und Stationstyp ist daher nicht erforderlich. Wenn das Signal für bestimmte Stationscluster stark genug ist, wird es im ML-Ergebnis dargestellt.

Abbildung Z-2: Clusterbildung nach Anzahl der Ozonüberschreitungen (1995 – 2018, links) und nach zeitlich zusammenfallenden Ozonüberschreitungen (2014 – 2018, rechts).



Quelle: Eigene Darstellung, TNO

Auswahl der Methode des Maschinellen Lernens (Machine Learning)

Maschinelles Lernen ist ein noch recht neues wissenschaftliches Gebiet und stellt eine Verbindung von Statistik, Datenwissenschaft und computergestützten Verfahren dar. Die lineare Regression ist das einfachste Beispiel (Generalized Linear Model, GLM). Es sind jedoch bereits viele fortschrittlichere Methoden entwickelt worden, um einen so genannten Prädiktand (Erwartungswert, hier Ozonüberschreitungen) mit Prädiktoren (Vorhersagewerten, hier u.a. Klimakennzahlen) in Beziehung zu setzen, wie zum Beispiel Neuronale Netze und Random Forests. Es ist schwierig, im Voraus zu sagen, welche Methode für einen bestimmten Datensatz am besten geeignet ist. Für das zu bearbeitende Projekt wurden keine rein linearen Methoden berücksichtigt, da die Beziehung zwischen Klimakennzahlen und Ozonüberschreitungen nicht linear ist. Darüber hinaus sind die Klimakennzahlen untereinander nicht völlig unabhängig oder sogar stark korreliert, so dass Methoden benötigt werden, die dies berücksichtigen können. Das sind zum Beispiel Methoden wie das Gradient Boosting Model (GBM) und Random Forest (RF) die beide auf Entscheidungsbäume zurückgreifen, oder Neuronale Netze (NN), die auf einer Architektur von Knotenpunkten und ihren Verbindungen basieren. Diese Methoden kennen wiederum verschiedene Varianten, z. B. Distributed Random Forest (DRF), Extremely Randomized Trees (XRT), als Varianten der RF. Neuronale Netze mit komplexer Architektur werden auch als Deep Learning (DL) bezeichnet. Zusätzlich kann ein Ensemblemodell gebildet werden (Stacked Ensemble oder Super Learning), das aus den trainierten GBM-, RF- und NN-Modellen ein Modell konstruiert. Dieses Ensemble hat in der Regel eine bessere Aussagekraft als die individuellen Modelle.

Für die Entwicklung des Analysetools wurde das H2O-Paket² für Maschinelles Lernen verwendet. H2O ist ein Online-Tool, das über die Programmiersprache R (oder Python) aufgerufen werden kann. Es verfügt über die gängigsten und neuesten ML-Algorithmen sowie über gute vordefinierte Visualisierungen für die Analyse der Ergebnisse. Es ist kostenlos erhältlich und gut dokumentiert. Es wurde die AutoML-Option verwendet.

² <https://h2o.ai/>

AutoML testet eine Reihe von Algorithmen (GBM, RF, NN) und Hyperparametern (Steuerparameter für die Algorithmen, z. B. Anzahl der Bäume) auf den Datensatz. Die mit verschiedenen Hyperparametern trainierten einzelnen ML-Algorithmen bilden eine „Familie von Modellen“. Außerdem konstruiert es Stacked Ensembles aus den einzelnen Modellen. Neben dem generellen Stacked Ensemble, das alle Modelle aller Familien mit einbezieht, wird auch ein Stacked Ensemble gebildet, das nur die Modelle mit der besten Performance pro ML-Methode bzw. ML-Algorithmus berücksichtigt (Stacked Ensemble, „best of family“). Die von AutoML erstellten Modelle werden anhand der Performance für Validierungsdaten in eine Rangfolge gebracht. Das Modell mit der besten Performance kann somit ausgewählt werden.

Bei nichtlinearen Methoden ist es nicht mehr möglich, den Beitrag der verschiedenen Variablen vollständig zu isolieren. Bei GBM-, RF- und NN-Methoden ist es möglich, die Variablen nach ihrer Bedeutung in einem Variable Importance Plot zu ordnen, bei Stacked Ensemble-Modellen ist dies nicht möglich. Um den Einfluss eines Prädiktors (hier Klimakennzahl) zu untersuchen, können sogenannte Partial Dependence Plots erstellt werden. Eine weitere informationsreiche Visualisierung ist ein SHAP-Plot (Shapley Additive exPlanation plot), der Abhängigkeiten detailliert und in Reihenfolge darstellt. SHAP-Plots können nicht für alle ML-Modelle erstellt werden. Sie können für baumbasierte Modelle bestimmt werden, aber nicht für Neuronale Netze oder ein Ensemble von Modellen.

Analyse der Eingangsdaten und Vorgehen

Das Vorgehen zur Entwicklung und Anwendung der ML-Modelle ist in Abbildung Z-3 dargestellt. Ausgehend von dem vollständigen Datensatz wurde zunächst gefunden, dass die Sommer 2003 und 2018 eine viel höhere Anzahl von Ozonüberschreitungen aufweisen als andere Sommer (Abbildung Z-1). Es ist zu erwarten, dass solche Sommer aufgrund des Klimawandels in Zukunft häufiger auftreten werden. Daher ist es wichtig, dass das ML-Modell in der Lage ist, gerade diese Bedingungen richtig darzustellen. Der beste Weg, dies zu erreichen, bestand darin, zwei separate Modelle zu entwickeln: eines für eine sehr hohe Anzahl von Ozonüberschreitungen („extreme“ Werte) und eines für den Hauptanteil der Anzahl von Ozonüberschreitungen („normale“ Werte). Diese Einteilung wurde anhand einer statistischen Bewertung der Werteverteilung vorgenommen (Quartile und Interquartilsabstand). Dies führte zur Einstufung von Ozonüberschreitungen ab 55 Tagen pro Jahr als „extrem“, unabhängig von der Station und dem Jahr (siehe auch UBA, 2020). Es wurden ausschließlich im oberen Bereich der Verteilung „extreme“ Werte gefunden. Von den ursprünglich knapp 6600 Datenpunkten wurden 190 als „extrem“ klassifiziert.

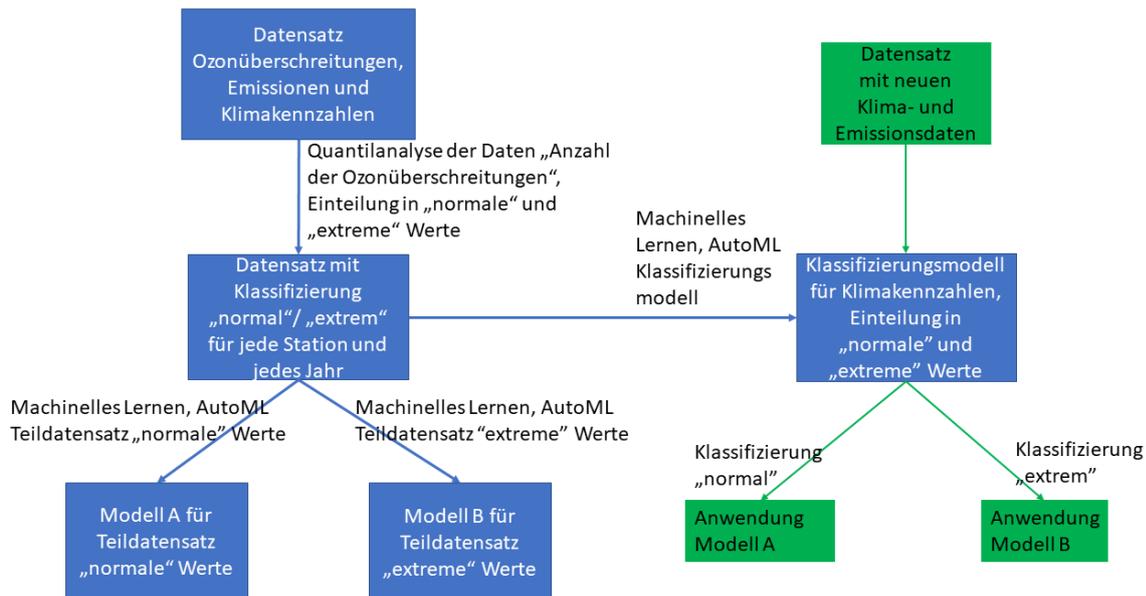
Im nächsten Schritt wurden für die zwei Datensätze die Algorithmen des Maschinellen Lernens getrennt angewendet und zwei verschiedene ML-Modelle erstellt (Modell A: „normal“, und Modell B: „extrem“).

Um eine Trainingsrunde zu starten, wurde der Datensatz in Trainingsdaten und Validierungsdaten aufgeteilt, wobei 80 % für das Training und 20 % für die Validierung verwendet wurden. Diese Aufteilung wird von H2O nach dem Zufallsprinzip durchgeführt. Der Trainingsdatensatz wird vom H2O-Tool weiter in Trainings- und Testdatensätze unterteilt, um eine Kreuzvalidierung durchzuführen. Der so trainierte Algorithmus wird auf die 20 % Daten angewendet, die für die unabhängige Validierung ausgelassen wurden. Die Trainingsprozeduren wurden für „normale“ und „extreme“ Werte, Überschreitungsanzahl und Episodenanzahl separat durchgeführt.

Nach der Klassifizierung der Anzahl der Ozonüberschreitungen (für jede Station und jedes Jahr) als „extrem“ oder „normal“ wurde ein Modell trainiert, um die Klimadaten für jedes Jahr und jede Station ebenfalls als „extrem“ oder „normal“ zu klassifizieren. Dies ist ein notwendiger

Schritt, wenn das Tool auf neue Stationen oder Jahre angewendet werden soll, denn es muss entschieden werden, ob Modell A oder B angewendet werden muss. Darüber hinaus zeigt dieses Klassifizierungsmodell, für welchen Bereich der Klimakennzahlen die Extremwerte auftreten.

Abbildung Z-3: Flussdiagramm für die Entwicklung und Anwendung der ML-Modelle, unter Berücksichtigung der Klassifizierung der Werte in „normal“ und „extrem“.

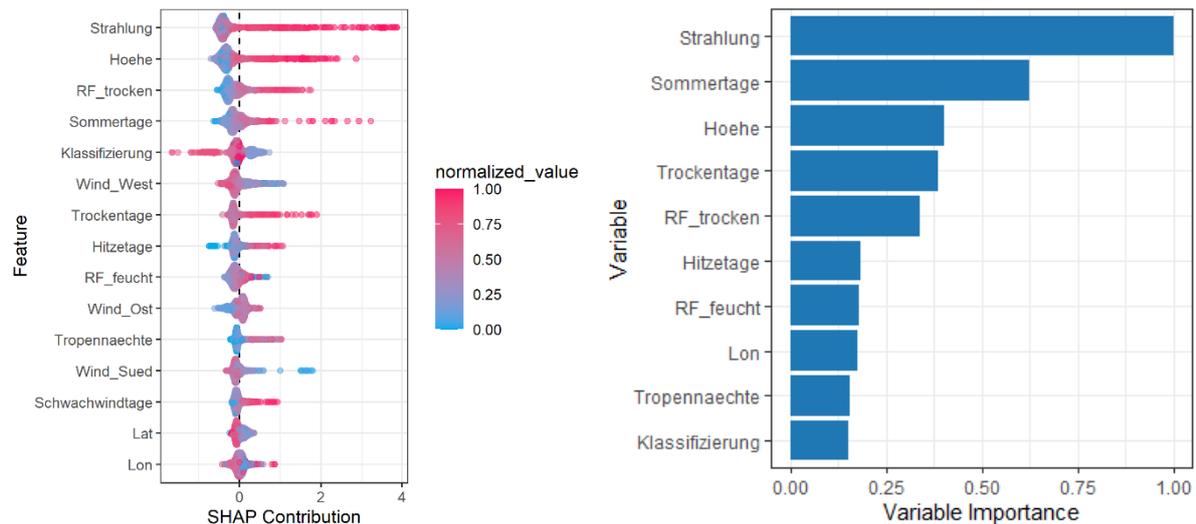


Quelle: Eigene Darstellung, TNO

Ergebnisse

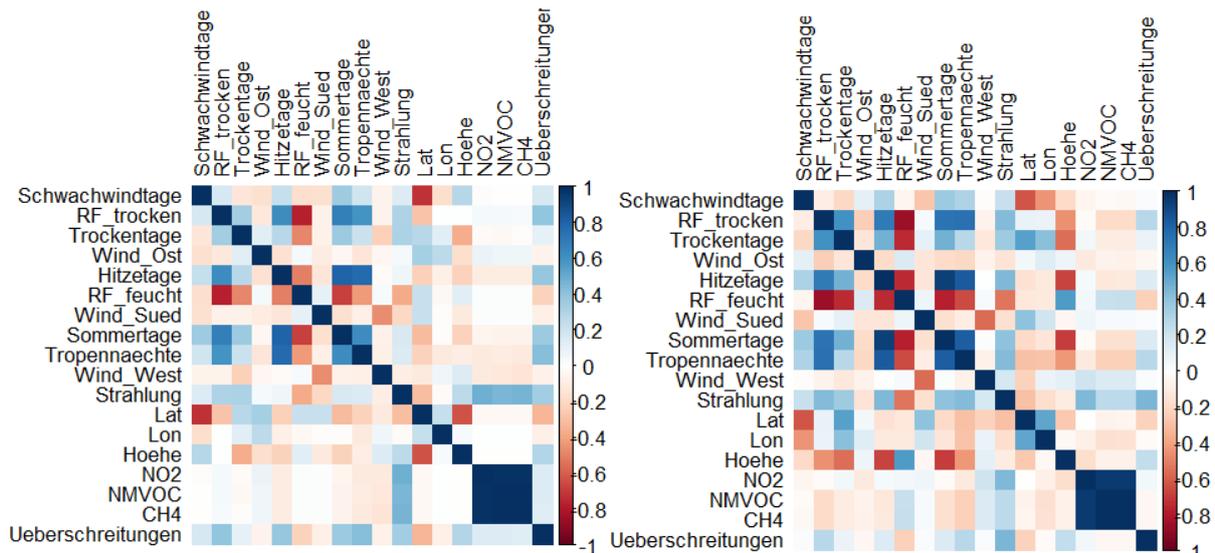
Zunächst werden die Ergebnisse des Klassifizierungsmodell, erstellt aus den Klima- und Stationsdaten, analysiert. Der SHAP-Plot und die Variable Importance Heatmap des besten GBM-Modells sind in Abbildung Z-4 dargestellt. Es ist zu beachten, dass die Reihenfolge der Variablen, die deren (relativen) Einfluss charakterisiert, in beiden Abbildungen wegen grundlegender Unterschiede in der Vorgehensweise bei der Berechnung der Variablenbedeutung nicht gleich ist. Die SHAP-Reihenfolge gibt einen vollständigeren Überblick. Es zeigt sich, dass die Strahlung der dominierende Klassifikator ist, wobei die Stationshöhe und die Luftfeuchte ebenfalls eine wichtige Rolle spielen, aber auch alle anderen Variablen - insbesondere Temperatur und Stationsklassifizierung - einen Einfluss haben.

Abbildung Z-4: SHAP-Plot (links) und Variable Importance Plot (rechts) für das GBM-Modell, das die Bedingungen für Extreme klassifiziert.



Nachfolgend werden die Ergebnisse für die Metrik Anzahl der Ozonüberschreitungen diskutiert. Zunächst werden die Korrelationen der Prädiktoren untereinander und mit den „normalen“ und „extremen“ Werte der Ozonüberschreitungen betrachtet (Abbildung Z-5). Diese Abbildung kann zur Unterstützung bei der Auswahl der Variablen und der Interpretation der Endergebnisse verwendet werden. Die Ergebnisse zeigen positive Korrelationen $> 0,4$ zwischen der Anzahl der Überschreitungen und der Anzahl der tropischen Nächten, der Anzahl der Sommertage und der Strahlung, mit Unterschieden in der Höhe der Korrelation für den „normalen“ und den „extremen“ Fall. Die stärkste negative Korrelation besteht mit dem Breitengrad. Auch Korrelationen zwischen den Klimakennzahlen sind erkennbar, wie z. B. eine negative Korrelation zwischen der Anzahl der Tage mit schwachem Wind und dem Breitengrad sowie zwischen den Klimakennzahlen bezogen auf die Temperatur und die Luftfeuchte. Sowohl für „normale“ als auch für „extreme“ Werte, haben die meisten Variablen eine Korrelation ungleich Null mit der Anzahl der Ozonüberschreitungen. Die Variablen korrelieren unterschiedlich stark miteinander. Die Korrelation übersteigt/unterschreitet jedoch nie den Wert von $+0,6$ bzw. $-0,6$. Die Stärke der Anwendung von Regressionen mit Methoden des Maschinellen Lernens wie Random Forest und GBM besteht darin, dass nicht nur die Korrelation jeder einzelnen Variablen, sondern auch die Korrelation zwischen Variablen berücksichtigt werden.

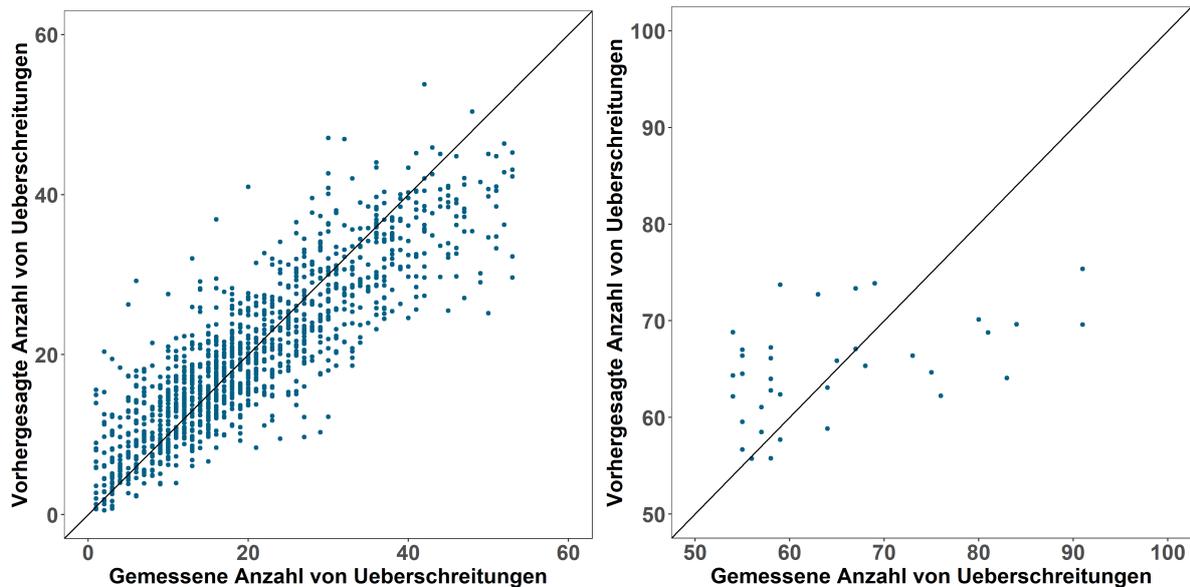
Abbildung Z-5: Korrelation der Prädiktoren und dem Prädiktand „Anzahl der Ozonüberschreitungen“ für „normale“ (links) und „extreme“ (rechts) Werte.



Quelle: Eigene Darstellung, TNO

Die Modellauswertung der Modelle A für „normale“ und B für „extreme“ Werte für die Anzahl der Ozonüberschreitungen ist in Abbildung Z-6 dargestellt. Es handelt sich um die Anwendung des endgültigen Modells auf die 20 % Validierungsdaten pro Teilmenge „normal“ bzw. „extrem“. Für die „normalen“ Werte zeigen die Ergebnisse eine gute Korrelation, wenn auch mit einer gewissen Streuung. Die höchsten Werte werden jedoch eher unterschätzt, was die Entscheidung bekräftigt, ein separates Modell für die extremen Werte zu erstellen. Die Modellqualität ist in Abbildung Z-6 zusammengefasst. Die Modellqualität ist mit RMSE-Werten in der Größenordnung von 10 % für die Anzahl der Überschreitungen und die Korrelationen zwischen modellierten und vorhergesagten Werten, wie in Abbildung Z-6 gezeigt, als gut einzustufen, und die Ergebnisse als robust. Für „extreme“ Werte ist die Anzahl der Daten gering und das Ergebnis ist sensibler gegenüber der für das Training gewählten Teilmenge. Dennoch weist das endgültige Modell eine beachtliche Performance auf (Tabelle Z-2). Die Trennung des Datensatzes in „extreme“ und „normale“ Werte führt zu besseren Ergebnissen für die „extremen“ Werte. Letzteres sollte jedoch im Hinblick auf die Variablenabhängigkeiten nicht überinterpretiert werden. Ein Stacked Ensemble-Modell, das alle Modelle umfasst, lieferte die beste Performance für „normale“ Werte, für „extreme“ Werte war dies das „Stacked Ensemble-best of family“ für Anzahl der Ozonüberschreitungen und Anzahl der Episoden.

Abbildung Z-6: Modellvalidierung: Vorhergesagte gegenüber gemessene Anzahl von Ozonüberschreitungen pro Station und Jahr für „normale“ (links) und „extreme“ Werte (rechts).



Quelle: Eigene Darstellung, TNO

Die Ergebnisse zeigen, dass die ML-Modelle fähig sind, die Ozonüberschreitungen mit den Klimakennzahlen und anderen Prädiktoren in Beziehung zu setzen. Diese Beziehung wurde weiter analysiert. Für das Stacked Ensemble-Modell ist es nicht möglich, die Korrelationen zu den einzelnen Prädiktoren zu spezifizieren oder gar die Beiträge der Prädiktoren zu ermitteln. Für die zugrundeliegenden Modelle kann zumindest die Rolle der einzelnen Variablen in Heatmaps und Partial Dependence Plots quantifiziert werden. Andererseits muss man teilweise akzeptieren, dass das Ensemble der Modelle das beste Modell ist, auch wenn es nicht vollständig interpretiert werden kann. Es werden hier die Ergebnisse für die Anzahl der Ozonüberschreitungen für die „normalen“ Werten detaillierter dargestellt.

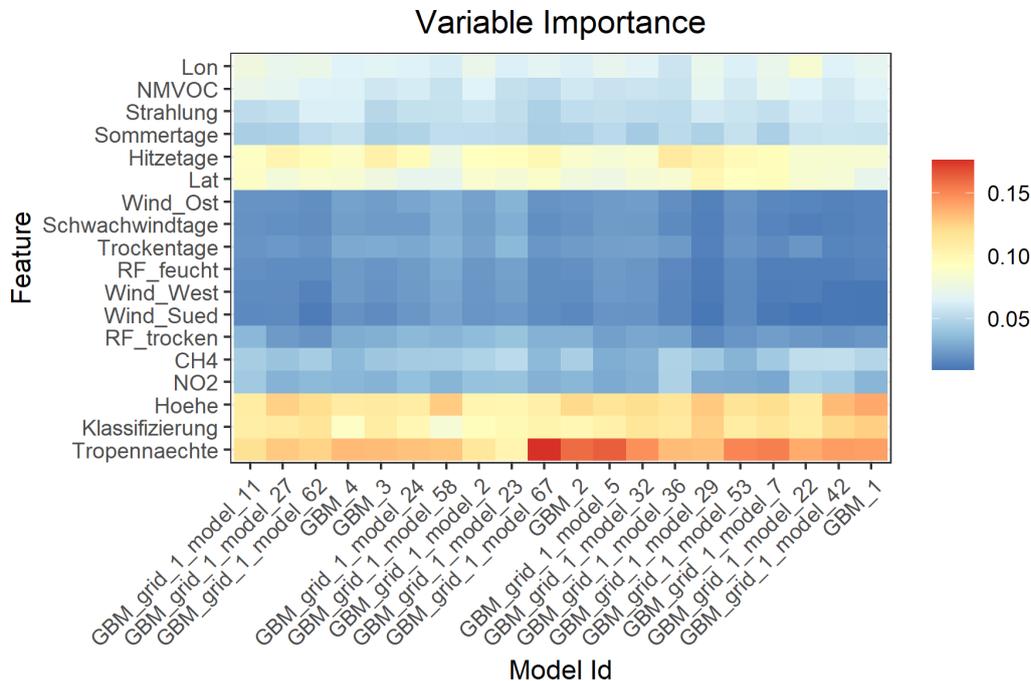
Tabelle Z-2: Performance der endgültigen Modelle A und B

Bedingung	Beste Performance	RMSE	Mittlerer absoluter Fehler
Normal, Überschreitungen	Stacked Ensemble, alle Modelle	5,99	4,42
Extrem, Überschreitungen	Stacked Ensemble, best of family	8,68	7,04
Normal, Episoden	Stacked Ensemble, alle Modelle	1,65	1,26
Extrem, Episoden	Stacked Ensemble, best of family	1,80	1,46

Die Variable Importance Heatmap in Abbildung Z-7 zeigt die Beiträge der Variablen für die 20 leistungsstärksten Modelle. In diesem Fall sind es immer die GBM-Modelle mit unterschiedlichen Hyperparameterwerten, die die beste Performance zeigen. Bei den „normalen“ Werten ist die Anzahl der tropischen Nächte bei weitem der stärkste Prädiktor. Stationshöhe, Stationsklassifizierung und Anzahl der Hitzetage sind ebenfalls von Bedeutung. Die Strahlung und der Breitengrad sind von zusätzlichem Wert. Feuchte, Niederschlag (hier Trockentage) sowie Windgeschwindigkeit und -richtung spielen in allen Fällen eine

untergeordnete Rolle. Bei den „extremen“ Werten (nicht gezeigt) sind die Ergebnisse aufgrund der geringen Datenmenge weniger zuverlässig. Die Strahlung ist hier ein starker Prädiktor und die Windrichtung ist wichtiger als bei „normalen“ Werten für Anzahl der Episoden.

Abbildung Z-7: Variable Importance Heatmap, Anzahl der Ozonüberschreitungen für „normale“ Werte. Die verschiedenen Modelle auf der x-Achse haben unterschiedliche Hyperparameter.



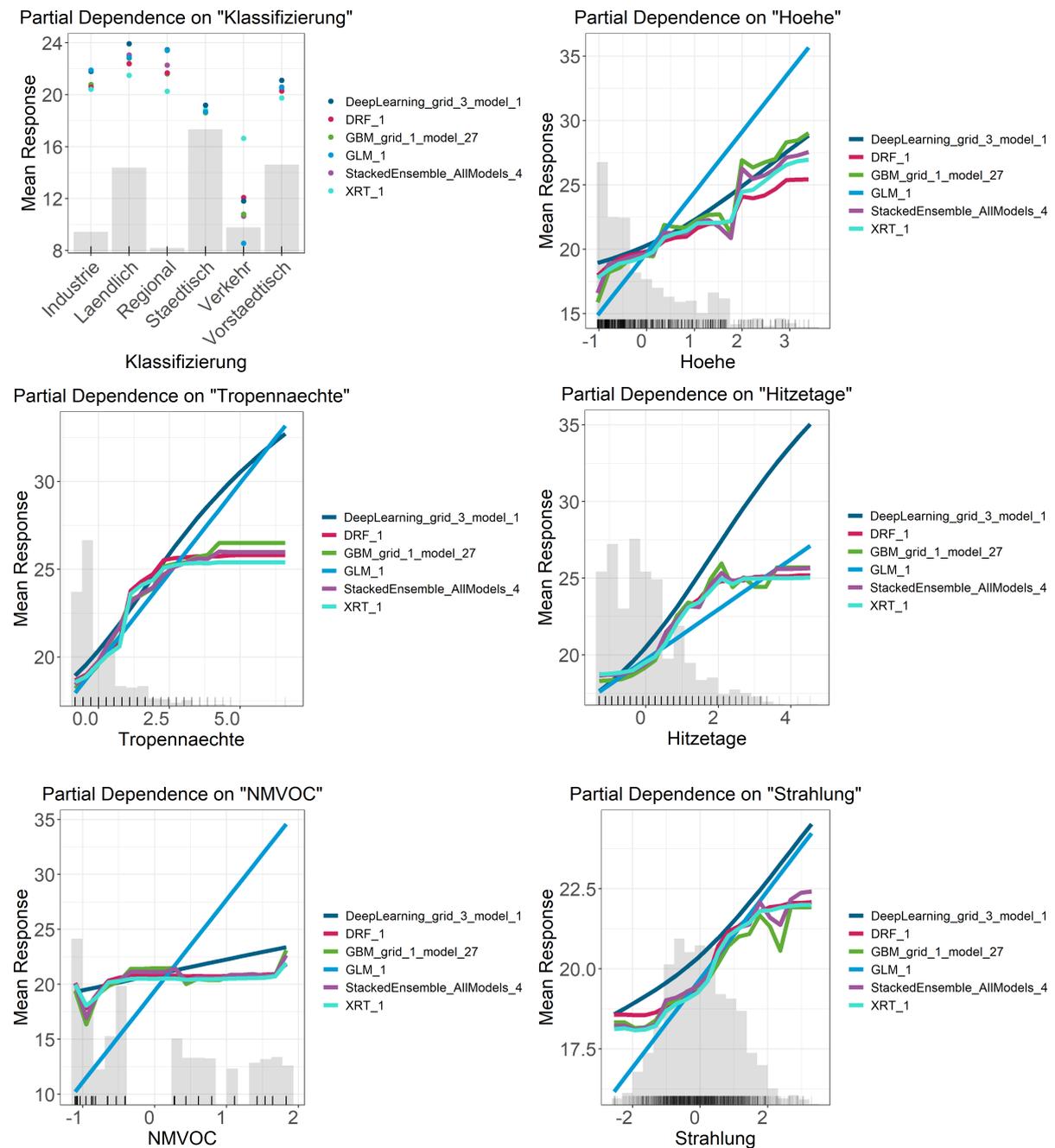
Quelle: Eigene Darstellung, TNO

Abbildung Z-8 zeigt die Partial Dependence Plots für die wichtigsten Prädiktoren. Sie werden für das leistungsstärkste Ensemble und das beste GBM-, Deep-Learning-, DRF- und XRT-Modell sowie das lineare GLM als farbige Linien dargestellt. Zusätzlich zu den Einflusskurven („response curves“) der verschiedenen Modelle sind die Daten (als vertikale schwarze Linien auf der x-Achse) und ein Histogramm der Daten (graue Schattierung) enthalten. Auf diese Weise wird veranschaulicht, in welchem Wertebereich die Daten den größten Einfluss haben und welches Vorzeichen die Abhängigkeit in diesem Bereich hat.

Mit Ausnahme des Deep-Learning-Modells und des linearen Modells stimmen die Modelle für den datenreichen Teil des Definitionsbereichs gut überein. Die Stationsklassifizierung spielt eine deutliche Rolle mit einer höheren Anzahl von Überschreitungen an ländlichen und suburbanen Stationen und niedrigeren Werten an den Verkehrsstationen. Die Abbildung für die Stationshöhe zeigen eine positive Korrelation mit der Höhe. Dies kann teilweise mit der Beziehung zwischen Breitengrad und Höhe zusammenhängen. In Mittel- und Süddeutschland ist die Topografie stärker ausgeprägt, wobei auch höhere Temperaturen erreicht werden und die Strahlung stärker ist. Bei der Strahlung nimmt die Zahl der Überschreitungen mit zunehmender Strahlung deutlich zu. Einige Modelle zeigen einen nahezu linearen Anstieg, andere weisen einen stärkeren Anstieg für die höheren Werte auf. Für die Temperaturindikatoren flachen die Kurven für die höheren Werte ab. Temperatur und Strahlung zeigen eine erhebliche Korrelation mit teilweise separaten Effekten (z. B. temperaturabhängige Ozonvorläuferemissionen und Depositionswerte der Vegetation). Da Ozon in photochemischen Reaktionen gebildet wird, ist dieser Zusammenhang zwischen Strahlung und Ozonüberschreitung sinnvoll. Die Abhängigkeiten von

den Emissionen sind sehr schwach. Die Diagramme enthalten die Antworten für verschiedene Teilmodelle sowie die Antwort für das endgültige Ensemble-Modell. Die x-Achse ist normiert (Z-Score-Normierung).

Abbildung Z-8: Partial Dependence Plots für die Anzahl der Ozonüberschreitungen für „normale“ Werte für die sechs wichtigsten Variablen.



Quelle: Eigene Darstellung, TNO

Schlussfolgerungen

Mithilfe des Maschinellen Lernens wurden Algorithmen entwickelt, die die jährliche Anzahl der Ozonüberschreitungen und die Anzahl der Ozonepisoden anhand von jährlich aggregierten meteorologischen Daten in Verbindung mit Klimakennzahlen vorhersagen. Eine

Ozonüberschreitung zählt, wenn der tägliche maximale 8-Stunden-Mittelwert der Ozonkonzentration $120 \mu\text{g}/\text{m}^3$ übersteigt.

Angesichts der begrenzten Datenmenge, die durch die Aggregation auf Jahresbasis entsteht (knapp 6600 Datenpunkte, 190 davon „extrem“) wurde eine brauchbare Performance erzielt. Die Ergebnisse für die „normalen“ Werte sind robuster, da für diesen Fall mehr Daten zur Verfügung standen. Für die „normalen“ Werte waren Temperaturindikatoren (tropische Nächte, Hitzetage, Sommertage) neben Informationen über den Breitengrad, die Stationshöhe und die Stationsklassifizierung die aussagekräftigsten Prädiktoren. Für die „extremen“ Fälle waren Strahlung und relative Luftfeuchte einflussreicher als für „normale“ Fälle. Die Korrelationen zwischen den Prädiktoren unterschieden sich bei den „extremen“ Werten ebenfalls von den „normalen“ Werten.

Der Vorteil des ML-Ansatzes ist, dass nach Erstellung des Datensatzes für ML der Rechenaufwand mit einer Größenordnung 10 Minuten im Vergleich zu CTMs (mehrere Tage pro Jahr) gering ist. Die Erstellung der Klimakennzahlen aus den COSMO-REA6 Daten dauerte insgesamt ein paar Tage für den Zeitraum 1995 - 2018 und ist somit der höchste zeitliche Aufwand. Der Ansatz des Maschinellen Lernens war in der Lage, die nichtlinearen Wechselwirkungen zwischen den Klimakennzahlen und den Ozonüberschreitungen so zu beschreiben, dass die Vorhersagen meist innerhalb von 10 % über oder unter der tatsächlich beobachteten Anzahl von Ozonüberschreitungen lagen. Sowohl CTMs als ML-Modelle sind in der Regel weniger robust im Bereich der höheren Anzahl der Ozonüberschreitungen wie in den Jahren 2003 und 2018.

Es gibt Studien in der Literatur, die auf einem Training für drei bis fünf Jahre auf der Grundlage von stündlichen oder täglichen Ozonwerten und meteorologischen Daten basieren. Diese Modelle können mit weniger Jahren trainiert werden, um ein robustes Ergebnis zu erzielen, haben aber den Nachteil, dass die meteorologischen Eingangsdaten auf Stunden- oder Tagesbasis benötigt werden. Auch kann der Einfluss von Emissionsänderungen für so kurze Zeiträume nicht gut dargestellt werden. Der Vorteil der im Projekt entwickelten Methode auf Jahresbasis ist ihre Dateneffizienz.

Der Nachteil der im Projekt entwickelten Methode auf Jahresbasis ist, dass die Übertragbarkeit zu Klima- und vor allem Emissionsszenarien limitiert ist. Bei den entwickelten Modellen handelt es sich um sogenannten „relationship-fitting“ Modelle. Diese können im Prinzip auch für künftige Bedingungen angewendet werden, wenn diese nicht zu weit von den Bedingungen entfernt sind, für die der Algorithmus trainiert wurde. Die Robustheit der aktuellen Modelle wurde anhand der Partial Dependence Plots für die Bereiche der Prädiktoren, für die ausreichende Datenmengen vorhanden waren, nachgewiesen.

In der Literatur wurden die Auswirkungen von Emissionsänderungen und Klima als getrennte Effekte festgestellt, mit unterschiedlichen Auswirkungen für städtische (hohe Emissionen) und ländliche (niedrigere Emissionen) Bedingungen, mit einem allgemeinen Trend hin zu sinkenden Ozonspitzenwerten, die im Einklang mit Emissionsrückgängen stehen. Es wurde in dieser Studie ein leichter Rückgang der Ozonüberschreitungen und der Anzahl der Ozonepisoden bei sinkenden Emissionen für „normale“ Werte gefunden. Für „extreme“ Werte sind die Ergebnisse nicht sehr robust, da die Jahre 2003 und 2018 in den Daten überrepräsentiert sind. Die begrenzte Auswirkung der Emissionen in unserer Studie kann mit der zwischenjährlichen Variabilität der Meteorologie zusammenhängen, die stärker ist als die jährliche Veränderung der Emissionen. Darüber hinaus wurden im Projekt nur landesweite Jahreswerte für die Emissionen für alle Stationen und pro Station nur jährliche Klimakennzahlen verwendet, was die Datenverfügbarkeit weiter einschränkt.

Empfehlungen

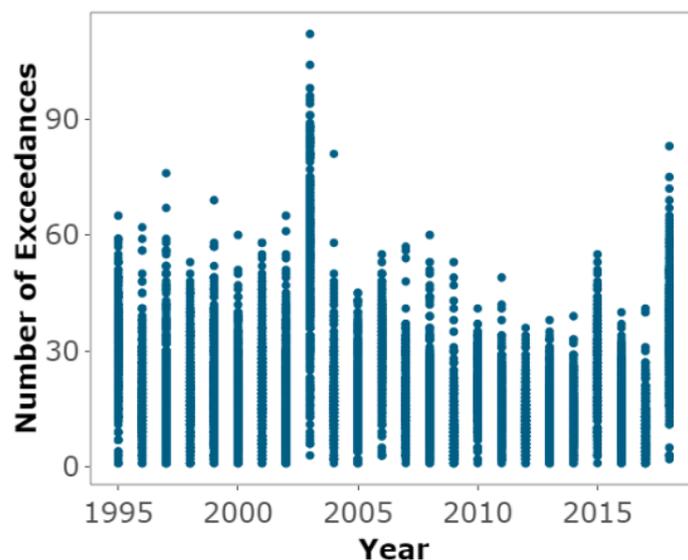
- ▶ Die Modelle sollten auf die Ozonmessungen der Jahre 2019, 2020 und 2021 angewendet werden, um somit eine Auswertung auf Jahresbasis zu evaluieren. Die aktuelle Modellbewertung wurde anhand einer Zufallsauswahl von 20 % der gesamten verfügbaren Daten für den Zeitraum 1995 – 2018 durchgeführt. In der Praxis würde man das Modell auf die Daten eines neuen Jahres anwenden. Die Daten für 2019 – 2021 sind erst seit kurzem verfügbar und konnten hier nicht berücksichtigt werden.
- ▶ Kennzahlen für Strahlung, Anzahl der Schwachwindtage und Windrichtungen sollten für den Zeitraum April – September oder sogar auf monatlicher Basis aus Klimasimulationen gespeichert werden. Für Klimamodelle werden die erforderlichen detaillierten Daten gar nicht oder nur auf einen Jahreswert aggregiert gespeichert. Monatliche Daten würden es ermöglichen, saisonale Effekte in den Beziehungen zwischen Luftqualität und meteorologischen Daten zu untersuchen und zu berücksichtigen.
- ▶ Eine Kennzahl für die relative Luftfeuchte sollte definiert und gespeichert werden. Solch eine Kennzahl fehlt in den bestehenden Definitionen der Klimakennzahlen. Diese Studie zeigt jedoch, dass eine solche Definition einen Mehrwert hat, insbesondere die Anzahl der Tage mit niedriger relativer Luftfeuchte (RF_trocken) im Klassifizierungsmodell für „normale“ und „extreme“ Werte und auch für das Modell für die Anzahl der Ozonüberschreitungen für die „extremen“ Werte.
- ▶ Es wird derzeit *nicht* empfohlen, das hier entwickelte Tool für Klimaszenarien als einzige Informationsquelle zu verwenden. Die Emissionen von NO_x, VOC und CH₄ werden in den kommenden Jahrzehnten voraussichtlich weiter sinken und unter den derzeit untersuchten Emissionsniveaus liegen. Eine Extrapolation in Bezug auf die Emissionen ist schwer zu rechtfertigen. Das Tool könnte jedoch sehr effektiv sein, wenn es darum geht, eine Bandbreite für ein Ensemble von Szenarien festzulegen, um teure Simulationen mit einem CTM zu ergänzen, die nur für ein oder zwei Szenarien durchgeführt werden können. CTM Simulationen könnten Daten für das Training neuer Emissionsszenarien liefern und das hier entwickelte ML-Modell könnte verwendet werden, um die Auswirkungen der Meteorologie zu untersuchen.

Summary

Ozone is an air pollutant with negative impact on human health and on vegetation. Therefore, metrics and target values have been defined by the World Health Organization (WHO). Also, in the European Air Quality Directive (2008/50/EG) threshold values and target values were defined, albeit different from those of the WHO for ozone. Ozone is a secondary pollutant that is formed from NO_2 and VOC emissions from anthropogenic and natural origin. Since the formation reactions are driven by incoming solar radiation, high ozone values are reached during summer conditions, in particular for warm sunny days. Also, other meteorological parameters like relative humidity, wind and transport-related quantities play a role in the build-up of ozone concentrations.

Since the 1990s, ozone peak levels have decreased due to emission reductions of precursor gases (Colette et al., 2017). However, this decrease is partly offset by climate change (Colette et al., 2017). Figure Z-1 shows the general decrease in number of exceedances of the Maximum Daily 8-hourly Average (MDA8) value of $120 \mu\text{g}/\text{m}^3$ (EU target value) for ozone measurements in Germany covering 1995-2018 (see also UBA, 2020). It is clearly visible that the exceptionally hot summers of the years 2003 and 2008 have resulted in higher exceedances. Over the past decades, summers in Germany have become warmer³. Climate scenarios point towards a further increase in temperature and more frequent occurrence of stagnant conditions, which further favors the occurrence of high ozone concentrations, unless precursor emissions are reduced further.

Figure Z-1: Number of ozone MDA8 exceedances of $120 \mu\text{g}/\text{m}^3$ per year, all German stations excluding mountain stations. Values of 55 and more are identified as extreme values.



Source: This project

³ <https://www.umweltbundesamt.de/daten/klima/trends-der-lufttemperatur>

To quantify the changes in ozone exceedances, computationally demanding simulations with chemical transport models coupled to climate models (on-line or off-line) were for example performed in the KLENOS project (Pavlik et al., 2016) for Germany. These simulations are therefore limited to a few climate and/or emission scenarios. The question arises whether machine learning (ML) methods could provide a numerically cheaper alternative to a chemical transport model (CTM), both using off-line data of climate models. Would such a method enhance the possibility to explore more climate scenarios because of the low computational costs of a ML model as compared to a CTM? Several successful attempts have been reported in literature to predict ozone concentration on a day-to-day basis using ML, mostly with the aim of short-time ozone predictions for the actual weather (current climate conditions). A similar day-to-day approach for a future climate (climate scenarios) however still would require daily data from computationally demanding climate models. For many climate simulations, only condensed sets of climate indicators are stored to reduce the data storage volumes. These data are not suitable to predict ozone exceedances on daily basis, but can potentially be used to predict number of ozone exceedances on an annual basis. So the main question that is to be answered in the current report is:

Can we find relationships between exceedances of ozone threshold values and climate indicators for German ozone measurement stations using statistical regression techniques?

The research is based on the German ozone measurements covering 1995 – 2018. The identification of a suitable metric and suitable climate indicators is part of the research. Not only the number of exceedances of the threshold value, but also duration and number of ozone episodes, which are defined as at least two subsequent days with exceedances, are relevant.

To answer the central research question the following steps were taken:

1. Provision and processing of ozone measurements for the period 1995 – 2018 at German stations of the federal states and the German Environment Agency (UBA), comprising around 200 active stations per year. Stations that do not meet the data quality criteria were not included for the respective year. Mountain stations (> 900 m altitude) were also excluded from the analysis.
2. Identification of a good metric for the ozone threshold value. The metric should capture extreme values (exceedances should not occur in more than 20 % of the year). At the same time, the metric should result in a number of threshold value exceedances in the data to be analysed, that is large enough to be able to apply any ML technique with sufficient confidence.
3. Identification of climate indicators that correlate with ozone formation. A climate indicator is a calculated quantity and for each climate indicator certain climate parameters are used, e.g. number of summer days from daily maximum temperature.
4. Construction of data frame for the machine learning exercises, containing ozone observations, information on measurement stations, climate indicators and emission values for NO_x, VOC.
5. Cluster analysis of measurement stations. Investigation whether a clustering of ground-based ozone measurements for stations with similar properties could add value to the results of the ML exercises.

6. Identification of most suitable machine learning approaches (algorithms that result in interpretable results and can handle non-linearities and are able to capture extremes).
7. Development of a tool for data processing and analysis. The tool trains relevant ML algorithms, i.e. the ML algorithm is applied (trained) to the concrete data set to describe statistical relationships. A trained algorithm is then called an "ML model".
8. Evaluate, visualize and interpret ML models

These steps will be elaborated further below.

Selection of relevant ozone metric

The first step is to define an appropriate metric that describes ozone exceedances of a threshold value (peak over threshold) in a meaningful way. Several metrics were explored using all available ozone measurement from the German Federal States and UBA, covering 1995 – 2018. Included metrics were WHO and EU target and threshold values of the MDA8 and Maximum Daily 1h average, MDH values. The final metric that was decided upon to be used in this project is $\text{MDA8} > 120 \mu\text{g}/\text{m}^3$, which means that when the maximum daily 8-hour running mean exceeded $120 \mu\text{g}/\text{m}^3$, the day was counted as an exceedance day. Main arguments are:

- ▶ MDA8 is more robust to outliers and loss of a few hours of data than daily maximum concentration
- ▶ $\text{MDA8} > 120 \mu\text{g}/\text{m}^3$ complies broadly with rule of thumb of 90th percentile of MDA values and is a good indicator of extreme values
- ▶ $\text{MDA8} > 120 \mu\text{g}/\text{m}^3$ exhibits enough variability in time and space to show differences in location and year
- ▶ $\text{MDA8} > 120 \mu\text{g}/\text{m}^3$ is an existing metric that is well understood in the air quality community

This metric is used to quantify number of exceedances, number of episodes and episode duration. The resulting episode number and duration are correlated to number of exceedances

Selection of climate indicators and climate data

The second step is to find climate indicators that are related to ozone exceedances. The selected indicators are based on existing climate metrics, but are now calculated only for the period April-September as values October-March will have no meaningful relationship to ozone exceedances that are a summer phenomenon. Only for relative humidity no climate indicator was available from literature. Therefore, we made a definition ourselves, counting the number of days with average relative humidity over 80 % (moist) or below 60 % (crisp). We decided to focus on direct climate indicators directly related to single meteorological variables rather than combined variables (like warm wet days). Several established machine learning methods would from their concept be well capable of finding such combinations when relevant. An overview of the used indicators is given in Table Z-1. COSMO reanalysis simulations of meteorology on a resolution of 6 km (COSMO-REA6, Bollmeyer et al., 2015) were used as data source for all meteorological data. These simulations have incorporated the available ground and satellite observations by data assimilation to be as accurate as possible. Time series were extracted for the ozone station locations for the period 1995 – 2018.

Table Z-1: Climate indicators included in ML training

Climate indicator	Involved pure Meteorological variables	Calculation from hourly data	Terminology in Figures
Number of summer days	Daily maximum temperature (T_{Max})	Days with $T_{Max} > 25\text{ °C}$	Sommertage
Number of hot days	Daily maximum temperature (M_{max})	Days with $T_{Max} > 30\text{ °C}$	Hitzetage
Number of tropical nights	Daily minimum temperature (T_{Min})	Days with $T_{Min} > 20\text{ °C}$	Tropennächte
Number of dry days	Daily cumulative rain	Days with Rain $< 1\text{ mm}$	Trockentage
RH_crisp days	Daily average relative humidity (RH)	Days with RH $< 60\%$	RF_trocken
RH_moist days	Daily average relative humidity (RH)	Days with RH $> 80\%$	RF_feucht
Days with easterly flow	Wind direction (Wdir)	$45^\circ < Wdir \leq 135^\circ$	Wind_Ost
Days with southerly flow	Wind direction (Wdir)	$135^\circ < Wdir \leq 225^\circ$	Wind_Süd
Days with westerly flow	Wind direction (Wdir)	$225^\circ < Wdir \leq 315^\circ$	Wind_West
Number of calm days	10m-Wind speed (Wspeed)	Wspeed $< 2\text{ m/s}$	Schwachwindtage
Mean incoming radiation	Radiation	Cumulative hourly mean radiation	Strahlung

Construction of data-frame for Machine Learning

A data frame was constructed for the period 1995 – 2018, containing for all valid stations per year:

- ▶ station location (longitude, latitude, altitude)
- ▶ Ozone station classification (urban background, suburban background, traffic, industrial, rural, rural background)
- ▶ the number of exceedances
- ▶ number of episodes
- ▶ maximum episode length.
- ▶ annual total reported German emissions of NO_x , VOC and methane
- ▶ climate indicators as sum April – September (Table Z-1)

The NO_x , VOC and methane emissions play a major role in ozone formation and determine part of the ozone trend. The data frame consisted of more than 6000 data lines

Clustering of stations

The idea of clustering is to find stations with similar signature, that form a data pool that can be used for training an algorithm. It was investigated whether groups of stations (clusters) could be

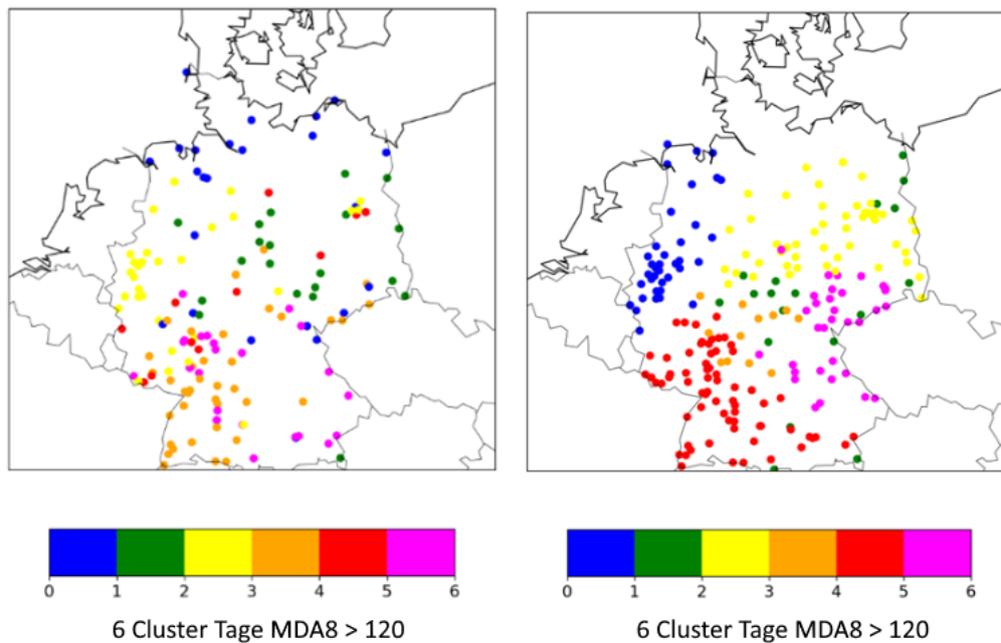
identified that have different statistical relationships between ozone exceedances and climate indicators. A way to cluster stations based on a priori knowledge would be to use the station classification (urban, traffic, regional) or the climate zone (west, middle, east, as derived in the KLENOS project) to define the cluster. Although clustering can make the machine learning more accurate for the cluster it was designed for, clustering also leads to a smaller data set for training and validation of the algorithm.

Clustering was explored with K-medians, a method that looks for groups of stations which are most similar to each other and most different to other groups. Here we tried two methods with a simple definition of this difference. In the first experiment the clustering was based on the annual number of exceedances for the period 1995 – 2018, and it also took into account the year in which the exceedances were found. The second experiment included a shorter time period (2014 – 2018) but included the day of the year on which an exceedance took place and thus was based on daily instead of annual data. Only stations covering all years in the clustering period were included. In both experiments a range of clusters between 4 and 9 was evaluated. If the number of clusters is too small, stations in a cluster may still be quite different which would lead to a less optimal result for the final algorithm for ozone exceedances prediction. If the number of clusters is too large, each individual cluster may contain too little data for a robust algorithm. The results of both experiments are shown in Figure Z-2. The method including only number of exceedances highlights the station conditions (remote, areas with higher emissions) with limited information on climate zones and less regional coherence of the stations in a cluster. For clustering including the timing of the exceedances, clusters are more confined in regions, except for the green cluster.

As an alternative to the K-median approach, the impact of using three climate zones as defined in KLENOS was explored, but the number of data for the extremes per cluster (zone) became very low, leading to less robust results for the extremes. Therefore it was decided to include all stations in the final dataset and develop models for the whole of Germany.

As a result of this analysis, clustering was discarded, and all stations were put together into one data frame for the model evaluation. Machine learning methods can include information on station location and classification. It is therefore not needed to cluster per region and station type. If the signal for specific groups of stations is strong enough, it will be presented in the ML results.

Figure Z-2: Clustering on number of exceedances (1995 – 2018, left) and on exceedances coinciding in time (2014 – 2018, right).



Source: This project, TNO

Selection of machine learning method

Machine learning is still a fairly new scientific field and represents a combination of statistics, data science and computer-aided methods. Linear regression is the simplest example (Generalised Linear Model, GLM). However, many more advanced methods have already been developed to relate a so-called predictand (expected value, in this case ozone exceedances) to predictors (predicted values, in this case climate indicators), such as neural networks and random forests. It is difficult to say in advance which method is best suited for a particular data set. For the project under consideration, purely linear methods were not considered because the relationship between climate indices and ozone exceedances is not linear. Furthermore, the climate indicators are not completely independent of each other or even strongly correlated, so that methods are needed that can take this into account. These are, for example, methods such as the Gradient Boosting Model (GBM) and Random Forest (RF), both of which make use of decision trees, or Neural Networks (NN), which are based on an architecture of nodes and their connections. These methods in turn know different variants, e.g. Distributed Random Forest (DRF), Extremely Randomised Trees (XRT) as variants of RF. Neural networks with complex architecture are also called Deep Learning (DL). In addition, an ensemble model can be formed (Stacked Ensemble or Super Learning), which constructs a model from the trained GBM, RF and NN models.

The H2O machine learning package was used to develop the analysis tool. H2O is an online tool that can be accessed via the R (or Python) programming language. It has the most common and latest ML algorithms as well as good predefined visualisation for analysing the results. It is available free of charge and well documented. The AutoML option was used.

AutoML tests a number of algorithms (GBM, Random Forest, Neural Network) and hyperparameters (control parameters for the algorithms, e.g. number of trees), resulting in a family of models per algorithm) on the dataset. The individual ML algorithms trained with different hyperparameters form a "family of models". It also constructs "stacked ensembles"

from the individual models. In addition to the general stacked ensemble, which includes all models of all families, a stacked ensemble is also formed, which only includes the models with the best performance per ML method or ML algorithm (stacked ensemble, "best of family"). The models created by AutoML are ranked based on the performance for validation data. The model with the best performance can thus be selected.

For non-linear methods, it is no longer possible to completely isolate the contribution of the different variables. With GBM, RF and NN methods, it is possible to rank the variables according to their importance in a Variable Importance Plot, but this is not possible with Stacked Ensemble models. In order to investigate the influence of a predictor (here climate index), so-called partial dependence plots can be created. Another information-rich visualisation is a SHAP plot (Shapley Additive exPlanation plot), which shows dependencies in detail and in sequence. SHAP plots cannot be created for all ML models. They can be determined for tree-based models, but not for neural networks or an ensemble of models.

Statistical analysis of input data and approach

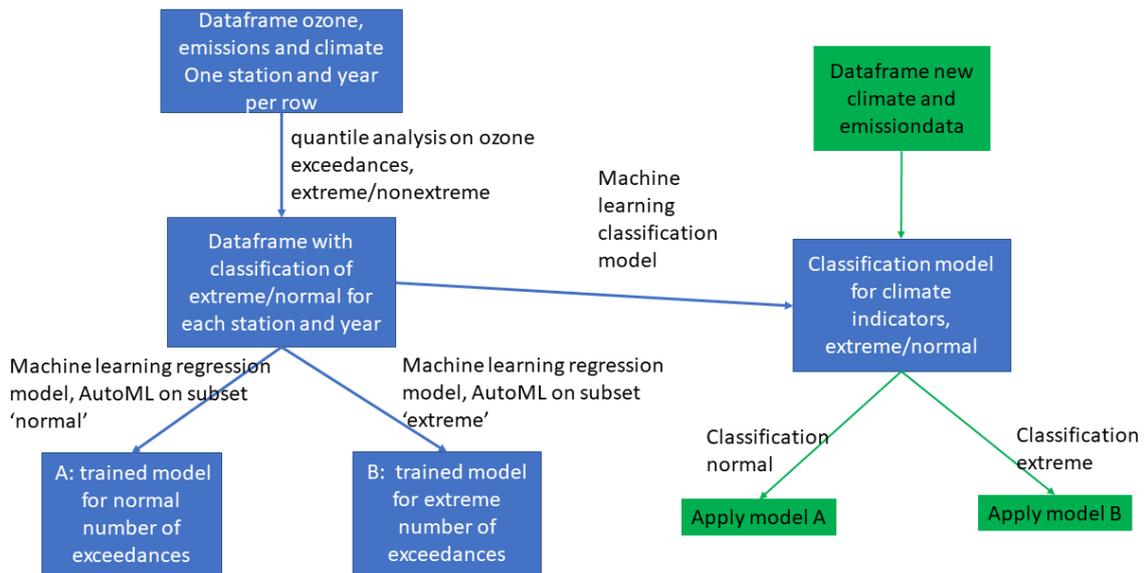
The procedure for developing and applying the ML models is shown in Figure Z-3. Based on the complete data set, it was first found that the summers of 2003 and 2018 have a much higher number of ozone exceedances than other summers (Figure Z-1). It is expected that such summers will occur more frequently in the future due to climate change. It is therefore important that the ML model is able to accurately represent these very conditions. The best way to achieve this was to develop two separate models, one for a very high number of exceedances ("extreme" values) and one for the bulk of the number of ozone exceedances ("normal" values). This classification was based on a statistical assessment of the distribution of values (quartile and interquartile values). This led to the classification of ozone exceedances of 55 days per year or more as "extreme", regardless of the station and the year (see also Figure Z-1).

Extreme values were found exclusively in the upper range of the distribution. Of the original almost 6600 data points, 190 were classified as extreme.

In the next step, the machine learning algorithms were applied separately for the two data sets and two different ML models were created (model A: "normal" and B: "extreme"). To start a training round, the dataset was split into training data and validation data, with 80 % used for training and 20 % for validation. This split is done randomly by H2O. The training dataset is further divided into training and testing datasets by the H2O tool to perform cross-validation. The algorithm thus trained is applied to the 20 % data omitted for independent validation. The training procedures were performed separately for "normal" and "extreme" values, exceedance number and episode number.

After classifying the number of ozone exceedances (for each station and year) as extreme or normal, a model was trained to classify the climate data for each year and station as extreme or normal as well. This is a necessary step if the tool is to be applied to new stations or years, because it must be decided whether model A or B has to be applied. Furthermore, this classification model shows for which range of climate indicators the extreme values occur.

Figure Z-3: Data flow for development of an application of models, taking into account the extreme or normal number of ozone exceedances.

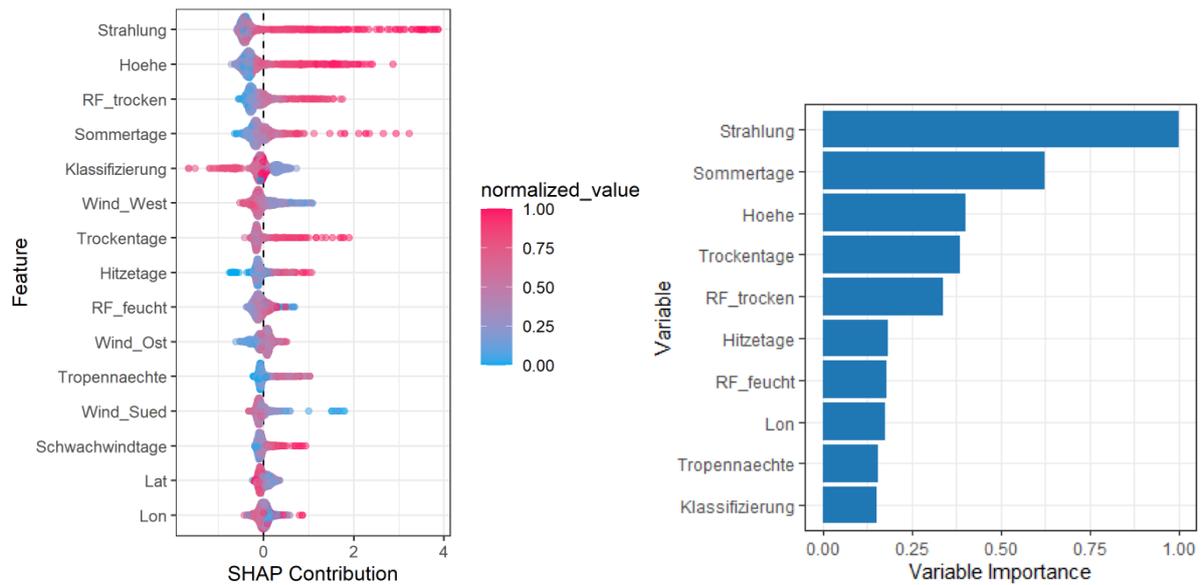


Source: This project, TNO

Results

First, the results of the classification model created from the climate and station data, are analysed. The SHAP plot and the Variable Importance Heatmap of the best GBM model are shown in Abbildung Z-4. Note that the order of the variables characterizing their (relative) influence is not the same in both figures, due to fundamental differences in the approach to calculating variable importance. The SHAP order gives a more complete overview. It shows that radiation is the dominant classifier, with station height and station classification also playing an important role, but all other variables - especially humidity and temperature - also have an influence.

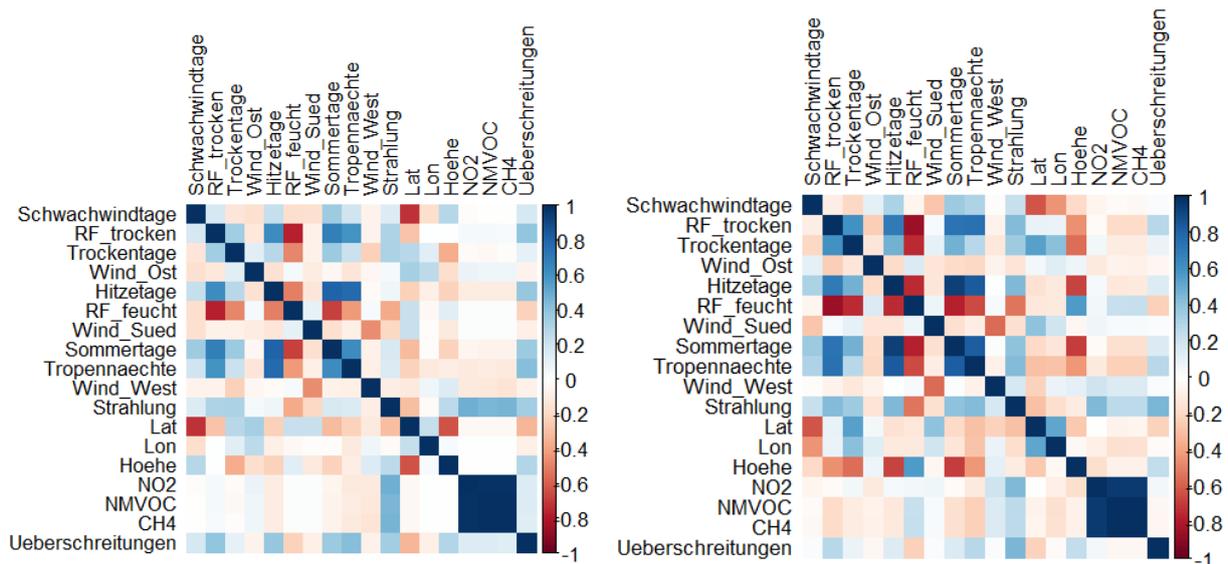
Figure Z-4: SHAP plot (left) and variable importance plot (right) for GBM model that classifies conditions for extremes.



Source: This project

Hereafter the results for number of ozone exceedances and number of episodes are discussed. First, the correlations of the predictors with each other and with the "normal" and "extreme" values of ozone exceedances are considered (Figure Z-5) This figure can be used to support selection of variables and interpretation of the final results (variable importance, SHAP). The figures show positive correlations > 0.4 between number of exceedances and number of tropical nights, number of summer days and radiation, with differences for the normal and the extreme cases. The strongest negative correlation is with latitude. Also the correlations between the climate indicators can be identified, like a negative correlation of number of calm days with latitude, and negative correlation of temperature indicator and moisture indices. For both normal and extreme cases, most indicators have nonzero correlation to number of exceedances and some variables correlate stronger than other variables but no single indicator has a correlation that exceeds/falls below the value of $+0.6$ or -0.6 . The power of the next step, applying machine learning regressions like random forest and GBM, is that they do not only take the correlation with each individual variable into account, but also accounts for correlation between variables.

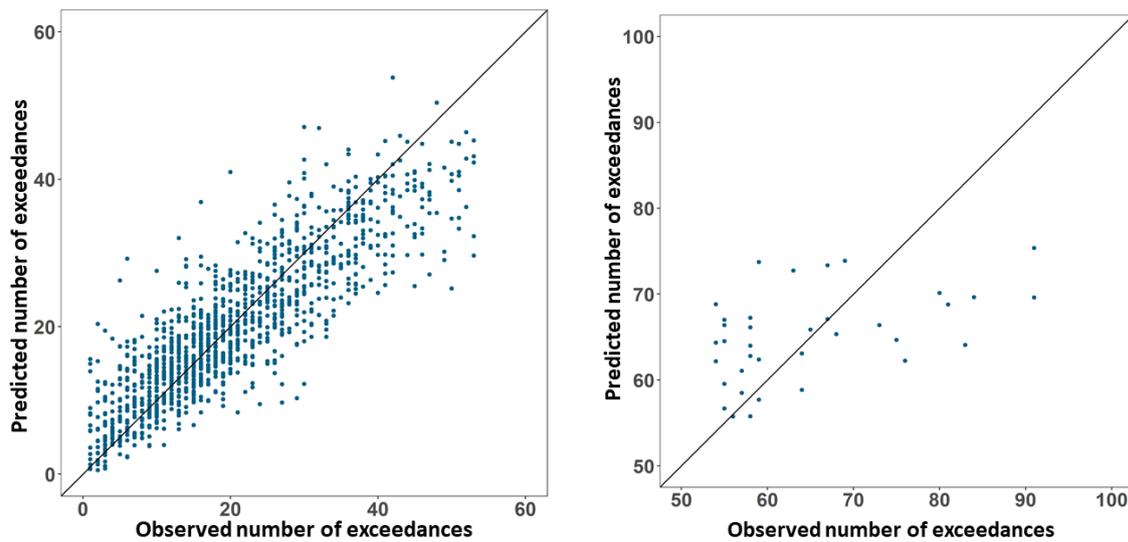
Figure Z-5: Correlation plots for normal (left) and extreme (right) values.



Source: This project

The model evaluation for the models A for normal values and B for extreme values for the number of exceedances is shown in Figure-Z 6. This is the application of the final models to the 20 % of data that were randomly selected out of the subsets for normal and for extreme conditions. For the normal values the results show good correlation, albeit with some scatter. It is also visible that the algorithm for normal values tends to underestimate the highest values, which was the motivation to make a separate model for the extreme conditions. The performance is also summarized in Table Z-2. The performance can be qualified as good with RMSE values in the order of 10 % of the range of the number of exceedances and correlations of modelled versus predicted values as shown in Figure-Z 6 and the results can be considered robust. For extreme cases the number of data points is small and the result is more sensitive to which subset was chosen for training. Still, the model showed considerable skill (Table Z-2). Also, due to the separation in extreme and normal cases, the results for the extreme cases are better than when they were included in one data set. The model for the extreme cases should however not be overinterpreted in terms of variable dependencies. The Stacked Ensemble Model, that included all models, resulted in the best performance for normal values, for the extreme values this was the „Stacked Ensemble-best of family“.

Figure-Z 6: Model validation: predicted versus true number of exceedances per station and year for normal (left) and extreme values (right)



Source: This project

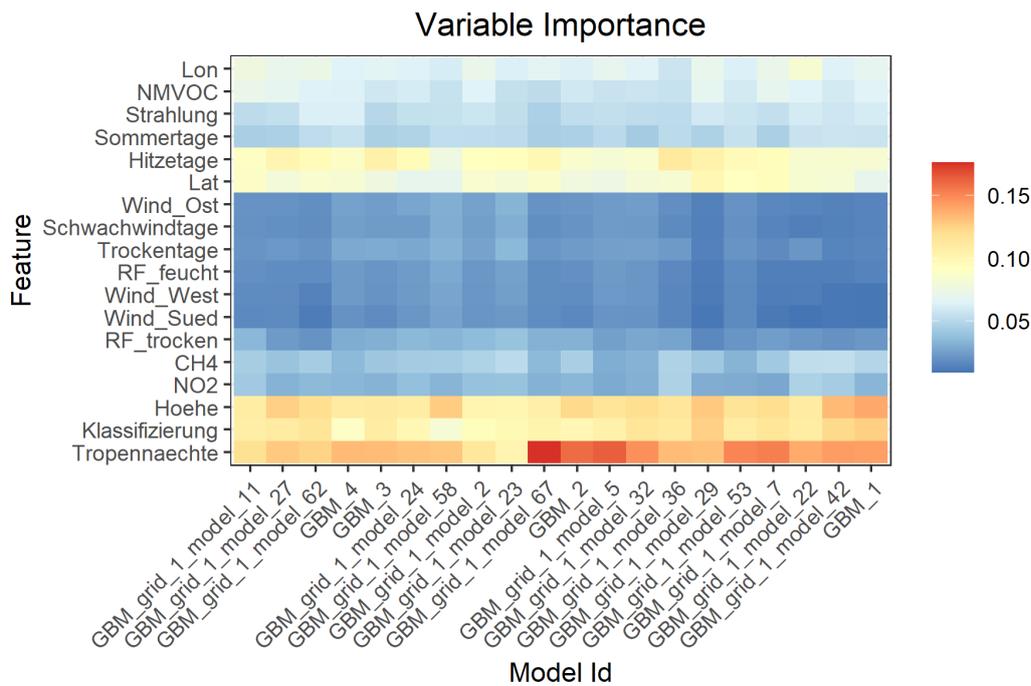
The results show that the ML models are able to relate the ozone exceedances to climate indicators and other predictors. These relationships were analysed further. For the stacked ensemble it is not possible to specify the correlations to the individual variables or even to find the variable contributions. For the underlying models at least the role of the individual variables can be quantified in variable importance heatmaps and partial dependence plots. On the other hand, one has to accept partly that the ensemble model is the best performing model, even if it cannot be interpreted fully. Only results for the more robust normal cases are presented in the summary.

Table Z-2: Performance of final algorithms

Case	Best performing	Root mean square error	Mean absolute error
Normal, exceedances	StackedEmsembe, all models	5.99	4.42
Extreme, exceedances	StackedEnsemble, best of family	8.68	7.04
Normal, episodes	StackedEnsemble, all models	1.65	1.26
Extreme, episodes	StackedEnsemble, best of family	1.80	1.46

The variable importance heatmap Figure Z-7 shows the variable contributions for the 20 best performing models. In all cases the GBM models with different hyperparameter values show the best performance. For the normal values, the number of tropical nights was by far the dominant variable. Station altitude, station classification and radiation were also relevant. Number of hot days and latitude were of additional value. Humidity, rain (dry days) and wind speed and direction played a minor role in all cases. For the extreme values, findings were less robust due to the small amount of data. Radiation was a strong predictor and wind direction was more important than for normal cases for the number of episodes.

Figure Z-7: Variable importance heatmap, number of exceedances for normal values.

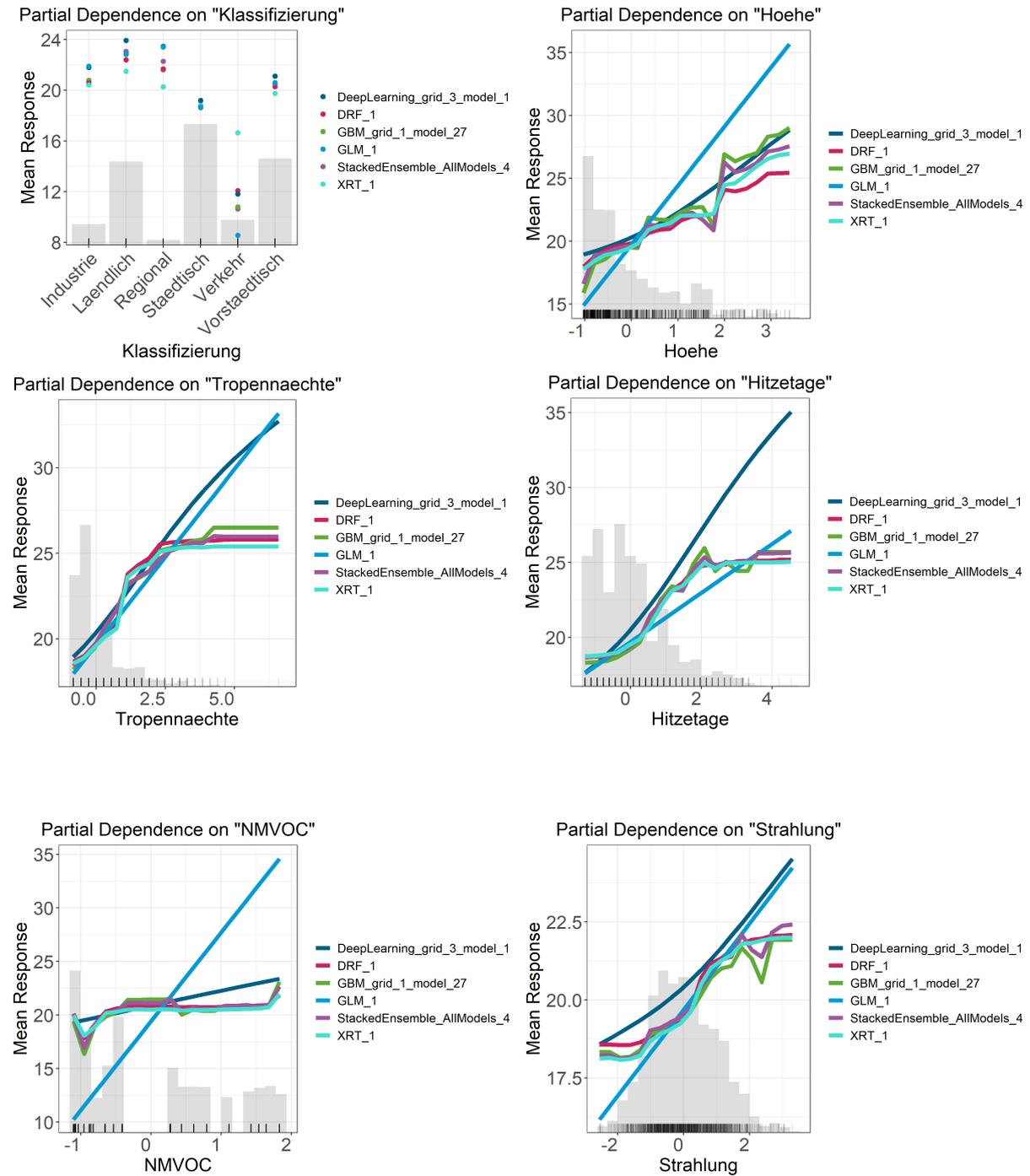


Source: This project

Figure-Z 8 shows the partial dependence plots for a number of predictors. Partial dependencies are shown for the best performing ensemble model and the best model of their class (GBM, Deep learning, Random Forest (DRF) and random trees (XRT)) as well as a generalized linear model (GLM) as colored lines. In addition to the response functions of various models, the data points (as vertical black lines on the x-Axis) and a histogram of the data (gray shading) is included. In this way, it is visualized in which range the data have most impact and what the sign of the response is in this range. The plots also show part of the difference between the data sets for the normal and extreme conditions. Note that the datasets for the normal cases and the extreme cases were normalized independently.

Except for the Deep learning and linear model, the models are in close agreement for the data-rich part. Station classification plays a clear role with higher number of exceedances for the rural and suburban stations and lower values for the traffic stations. The plots for station height (altitude) show a positive response with height. This can be partly related to the relationship between latitude and height: there is stronger topography in middle and southern Germany, where also higher temperatures are reached and radiation is stronger. For radiation, the number of exceedances clearly increases with radiation with limited leveling-off. Some models exhibit a nearly linear increase, others predict a sharp increase for the higher values. For the temperature indicators, the dependencies became nearly flat curves for the higher values. Temperature and radiation are not fully independent variables but show considerable correlation with partly separate effects (e.g. temperature-dependent ozone precursor emissions and deposition values over vegetation). As ozone is formed in photochemical reactions, this relationship between radiation and ozone exceedances is meaningful. The dependencies on emissions are weak and might not be robust. The diagrams contain the responses for different sub-models as well as the response for the final ensemble model. The x-axis is normalised (Z-score normalisation).

Figure-Z 8: Partial dependence plots for number of exceedances for normal cases for the six most important variables.



Source: This project

Conclusions

Machine learning was used to derive algorithms that predict the annual number of ozone exceedances and number of episodes using annually aggregated meteorological data, related to climate indicators. Exceedances were defined as days on which the daily maximum 8h average ozone concentration exceeded $120 \mu\text{g}/\text{m}^3$. This metric is suitable because it is related to the EU threshold value, is robust for small data losses and shows enough variability in space and time.

Given the limitations of the data amount, that results from annual aggregation (nearly 6600 points of which 190 extreme), suitable performance was achieved. Results are more robust for the normal cases as more data were available for that case. For the normal cases, temperature indicators (tropical nights, summer days) were dominant, next to information on latitude, station altitude and station classification. For the extreme cases, next to the effects of radiation, wind and relative humidity became more dominant. For the extreme cases, the correlations between the predictors were also different from those for the normal cases.

The advantage of the ML approach is that once the dataset for ML has been created, the computational effort is low, on the order of 10 minutes, compared to CTMs (several days per year). The creation of the climate metrics from the COSMO-REA6 data took a couple of days in total for the period 1995 – 2018, which is the highest in terms of time. The machine learning approach was able to describe the non-linear interactions between the climate metrics and ozone exceedances such that the predictions were mostly within 10 % of the number of ozone exceedances of the observed values. Both CTMs and ML models are generally less robust in the range of higher number of ozone exceedances as in 2003 and 2018. Also, compared to studies in the literature based on training for three to five years based on hourly or daily ozone values and meteorological data, our developed approach has an advantage in terms of data volume and data availability. These models are based on daily data and can be trained with fewer years to achieve a robust result, but again have the disadvantage that the meteorological input data are needed on an hourly or daily basis. Also, the influence of emission changes for such short time periods can only be represented in a very limited way.

The disadvantage of our method developed in the project on an annual basis is that the transferability to climate and especially emission scenarios is limited. The models developed are so-called "relationship-fitting" models. In principle, these can also be applied to future conditions if these are not too far removed from the conditions for which the algorithm was trained. The robustness of the current model was demonstrated using the partial dependence plots for the areas of the predictors for which sufficient data were available.

In the literature, the effects of emission changes and climate were found to be separate effects, with different impacts for urban (high emissions) and rural (lower emissions) conditions, with a general trend towards decreasing peak ozone levels consistent with emission decreases. A slight decrease in ozone exceedances and the number of ozone episodes with decreasing emissions was found in this study for "normal" values. For "extreme" values, the results are not very robust as the years 2003 and 2018 are overrepresented in the data. The limited impact of emissions in our study may be related to inter-annual variability in meteorology, which is stronger than the annual change in emissions. Furthermore, our approach in the project only used nationwide annual values for emissions for all stations and only annual climate metrics per station, which further limits data availability.

Recommendations

- ▶ The tool should be applied on the ozone data of 2019, 2020 and 2021, to evaluate the models on yearly basis. The current model evaluation was done on a random draw of the total available data covering 1995 – 2018, using 20 % of the data. In practice, one would apply the algorithm to data in a new year. The data for 2019, 2020 and 2021 became available only recently and could not be taken into account.
- ▶ Climate indicators for radiation, number of calm days and wind directions should be stored for the period April-September, or even on a monthly basis, while performing climate

simulations. For climate models, the necessary detailed information is not always stored, or is only aggregated to an annual value. Monthly data would allow to study and include seasonal effects in relationships between air quality and meteorological data.

- ▶ An indicator for relative humidity should be defined and stored. Such an indicator is missing in the existing definitions and this study shows that our simple definition has added value, in particular the number of crisp days in the classification model for normal and extreme cases, and also for the model for number of exceedances for the extreme cases.
- ▶ It is currently *not* recommended to use the tool for climate change scenarios as only source of information. Emission changes for the decades ahead will go beyond the currently investigated emission levels and extrapolation with respect to emissions is hard to justify. But the tool could be very effective in establishing a band width for an ensemble of scenarios to complement expensive simulations with a CTM that can only be executed for one or two scenarios. CTM simulations could provide data for training for new emission scenarios and the updated machine learning model could be used to explore the impact of meteorology.

1 Einführung

1.1 Hintergrund und Ziel

Die Konzentrationen von Luftschadstoffen sind in Deutschland in den letzten Jahrzehnten zurückgegangen. Dies kann direkt auf den Emissionsrückgang von Luftschadstoffen und deren Vorläufersubstanzen zurückgeführt werden. Die Luftschadstoffkonzentrationen sind jedoch auch von der Meteorologie abhängig. Durch den Klimawandel bedingte Veränderungen der Meteorologie können die Emissionsminderungen konterkarieren, so dass langfristig die Luftqualitätsgrenzwerte trotz dieser Emissionsminderungen nicht mehr eingehalten werden. Für Ozon wird dieser Effekt des Klimawandels als „ozone climate penalty“ bezeichnet. Insbesondere die Ozonbildung ist aufgrund mehrerer Mechanismen sehr empfindlich gegenüber den meteorologischen Bedingungen. Für Projektionen der Luftqualität in die Zukunft wurden Modellsimulationen mit Klimamodellen verwendet, die an Chemie-Transport-Modelle (CTM) gekoppelt sind, wie die Kopplung zwischen COSMO und REM-CALGRID (RCG) für Deutschland (KLENOS-Projekt, Pavlik et al., 2016), RACMO und LOTOS-EUROS (Manders et al. 2012) und viele andere Studien, wie von Colette et al., (2015). Alle diese Studien weisen darauf hin, dass aufgrund des Klimawandels die Ozonüberschreitungen und die Dauer der Episoden wegen des Temperaturanstiegs und des häufigeren Auftretens von stagnierenden Wetterbedingungen zunehmen werden, sofern die Emissionen von Vorläuferstoffen nicht reduziert werden.

Diese Studien haben jedoch auch gezeigt, dass die Ergebnisse unsicher sind, da einerseits die von Klimamodellen erzeugte Meteorologie in Bezug auf Zirkulationsmuster und systematischen Fehlern bei modellierten Variablen unsicher ist (Pavlik et al. 2016, Manders et al. 2012) und andererseits kein Luftqualitätsmodell perfekt ist, wie verschiedene Modellvalidierungs- und Vergleichsstudien zeigen (z. B. Otero et al., 2018, Im et al., 2015). Daher sind Ensemble-Ansätze erforderlich, um robuste Ergebnisse zu erzielen. Solche Studien erfordern einen hohen Rechenaufwand.

Die Fülle der Ozonmessungen ermöglicht einen anderen Ansatz: die Untersuchung des Zusammenhangs zwischen Ozonkonzentrationen und meteorologischen Variablen. Diese Zusammenhänge können mit Hilfe von Regressionsalgorithmen und fortgeschrittenen Maschinellen Lernverfahren hergestellt werden. Diese Methoden werden derzeit erfolgreich eingesetzt, um Ozonkonzentrationen anhand von meteorologischen Variablen zu analysieren und vorherzusagen, ohne atmosphärische Prozesse explizit zu berücksichtigen. Solche Zusammenhänge wurden bereits in den 1990er Jahren für operationelle Ozonsmogvorhersagen aufgestellt (z. B. Noordijk (2003) für die Niederlande). Die Datensätze und die Rechenleistung haben sich jedoch erheblich gesteigert und Maschinelles Lernen hat sich als robustes Instrument erwiesen, um Muster in großen Datensätzen von Ozonmessungen zu finden (Eskes et al., 2013, Otero et al., 2016, Boleti et al., 2018, 2019, Feng et al., 2019).

Ziel des hier beschriebenen Projekts ist die Entwicklung eines statistischen Modells, das das Auftreten von Ozonepisoden (Anzahl, Dauer) auf **Jahresbasis** mit einer begrenzten Anzahl **jährlich aggregierter Klimakennzahlen** für Deutschland in Beziehung setzt. Wenn ein solches robustes Modell erstellt werden kann, können rechenzeitintensive Simulationen mit gekoppelten Klima-Luftqualitäts-Modellen durch ein zeiteffektiveres Modell ersetzt werden. Dies würde die Möglichkeit verbessern, die Klimawandelszenarien in Bezug auf die Ozonbelastung schnell zu untersuchen. Darüber hinaus kann eine begrenzte Anzahl von

Klimavariablen an Messstationen verwendet werden, anstatt ein Luftqualitätsmodell mit großen meteorologischen Datensätzen zu füttern. Der erste Schritt besteht darin, ein statistisches Modell zu erstellen, das das Auftreten (Anzahl) und die Dauer von Ozonepisoden mit Klimavariablen für das aktuelle Klima in Beziehung setzt, wobei Daten über einen ausreichenden Zeitraum (1995 – 2018) verwendet werden. Ein solcher Ansatz ist unseres Wissens nach in der Literatur nicht verfügbar, was das Ziel zu einer Herausforderung macht.

1.2 Überblick der Arbeiten

In diesem Bericht werden mit den verfügbaren Daten einige Algorithmen trainiert, um Modelle zu erstellen, die Klimakennzahlen mit beobachteten Überschreitungen von einem hier definierten Ozonschwellenwert in Beziehung setzt. Der Bericht beginnt mit einer Literaturstudie zu verschiedenen Aspekten.

Die Ozontrends werden sowohl durch meteorologische Veränderungen als auch durch Emissionsveränderungen beeinflusst. Dieser Aspekt wird in Kapitel 2 behandelt, wobei der Schwerpunkt auf den Ergebnissen und Auswirkungen für Deutschland liegt.

Der erste Schritt besteht darin, eine nützliche Metrik für Ozonüberschreitungen zu ermitteln. Daher wird in Kapitel 3 ein Überblick über gängige Ozonmetriken gegeben. In diesem Kapitel werden auch die Auswirkungen der Verwendung einiger dieser Metriken untersucht, um zu einer endgültigen Entscheidung über die in dieser Arbeit zu verwendenden Metrik zu gelangen.

Der nächste Schritt besteht darin, die Beziehung zwischen Meteorologie und Ozon, wie sie in der wissenschaftlichen Literatur zu finden ist, als Grundlage für die Entwicklung des Algorithmus zusammenzufassen. Dies wird in Kapitel 4 zusammen mit einer Beschreibung der verfügbaren Klimakennzahlen aus Klimadatenbanken und Klimasimulationen niedergeschrieben. Am Ende des Kapitels werden die Klimadaten ausgewählt, die für das Maschinelle Lernen verwendet werden sollen.

Kapitel 5 bietet eine allgemeine Einführung in Methoden des Maschinellen Lernens. Die Methoden des Maschinellen Lernens gehen über die klassischen Regressionsmethoden hinaus und verwenden fortgeschrittenere Ansätze. Die wichtigsten Methoden und Konzepte werden vorgestellt, und es werden Anwendungen für Luftqualitätsvorhersagen aufgezeigt.

In Kapitel 6 werden die wichtigsten Ergebnisse des Maschinellen Lernens vorgestellt und erörtert, einschließlich Clustering, Klassifizierung, Training und Validierung der verschiedenen Modelle sowie eine Diskussion über die Robustheit der Ergebnisse. Abschließend werden die Schlussfolgerungen zusammengefasst und im Hinblick auf die beabsichtigte Anwendung erörtert; außerdem werden Empfehlungen ausgesprochen.

Anhang A beschreibt die Datenströme und das R-Tool (eine Reihe von R-Skripten), die zur Verarbeitung der Ozondaten und meteorologischen Daten, zur Durchführung des Maschinellen Lernens und zur Visualisierung der Ergebnisse entwickelt wurden. Anhang B zeigt die Ergebnisse von Experimenten mit Maschinellern Lernen bei unterschiedlicher Auswahl von Datenclustern (alle Daten, Klimazonen). Anhang C zeigt die Ergebnisse einer alternativen Methode, die formal nicht Teil der aktuellen Studie war, aber dieselben Daten verwendet. Er veranschaulicht die Verwendung täglicher Daten für ein Beziehungsmodell für die täglichen Ozonkonzentrationen, aus dem die Anzahl der Überschreitungen berechnet wurde.

2 Ozontrends und Klimawandel

Die Trends in den Ozonkonzentrationen werden durch Trends in der Meteorologie und durch Trends in den Emissionen bestimmt. Neben den saisonalen Effekten sind auch Trends bei den Hintergrundkonzentrationen zu beobachten, die mit dem Klima und den Emissionen zusammenhängen. Emissionsreduzierungen haben zu einem Rückgang der Ozonüberschreitung in Europa geführt, der jedoch teilweise durch die steigenden Sommertemperaturen kompensiert wird. In diesem Abschnitt stellen wir die beobachteten und modellierten Ozontrends für Europa und Deutschland im Zusammenhang mit den Emissions- und Temperaturtrends vor.

2.1 Trends der Ozonkonzentration in Europa

2.1.1 Trend des Hintergrundozons

Ein wichtiger Einflussfaktor für die Ozontrends ist das so genannte Hintergrundozon, d. h. Konzentrationen, die nicht durch die jüngste, lokal emittierte oder produzierte anthropogene Verschmutzung beeinflusst werden, sondern von anderswo herantransportiert werden (Dentener et al., 2010). Ozon kann lokal, zwischen Ländern, innerhalb eines Kontinents oder sogar über Kontinente hinweg (hemisphärisch) transportiert werden. Das Verhalten von Ozon im Hintergrund kann anhand von Messungen an entfernten Stationen, z. B. in großer Höhe, untersucht werden. Für Europa wird die Station Mace Head herangezogen, um den hemisphärischen Transport und langfristige Trends zu untersuchen. Dennoch spiegeln die Spitzenkonzentrationen meist lokale photochemische Produktionsmechanismen wider, die auf günstige Wetterbedingungen und lokale Emissionen von Ozonvorläuferkonzentrationen zurückzuführen sind. Andererseits werden die durchschnittlichen Ozonkonzentrationen durch eine Kombination von Hintergrundozon und regionaler photochemischer Ozonproduktion beeinflusst. Chemie-Klima-Modelle wurden zur Quantifizierung von Hintergrundozon verwendet (Dentener et al., 2010). Die Unterschiede zwischen den Modellen sind jedoch aufgrund von Unsicherheiten in der Darstellung verschiedener Prozesse immer noch groß (z. B. biogene Emissionen, Fiore et al., 2014).

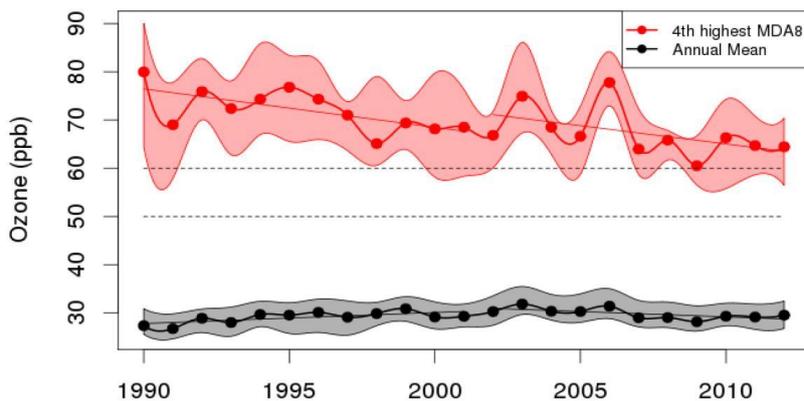
Die mittleren Ozonkonzentrationen sind seit den 1990er Jahren und bis Mitte der 2000er Jahre in den meisten Teilen Europas entweder auf einem ähnlichen Niveau geblieben oder haben in ländlichen Hintergrundgebieten zugenommen wie in Abbildung 1 gezeigt sowie in z. B. Boleti et al., (2018), Paoletti et al., (2014), Querol et al., (2016) und Anttila und Tuovinen (2010) beschrieben, während an städtischen Standorten ein Anstieg der durchschnittlichen Ozonkonzentrationen zu verzeichnen war (z. B. Boleti et al., 2018). Nach Mitte der 2000er Jahre wurde jedoch ein Rückgang der mittleren Ozonkonzentrationen beobachtet (Boleti et al., 2018; Munir et al., 2013). Die durchschnittlichen Ozontrends sind auf Veränderungen vieler Faktoren zurückzuführen, wie z. B. Veränderungen der hemisphärischen anthropogenen und biogenen NO_x - und VOC-Emissionen sowie Veränderungen der Zirkulationsmuster (stagnierende Bedingungen, stratosphärische Intrusionen).

In diesem Projekt liegt der Schwerpunkt auf den Ozonspitzenkonzentrationen, daher wird der Einfluss der Ozonhintergrundkonzentration nicht untersucht. Dies ist auch vor dem Hintergrund der Modellstudie von Giordano et al. (2015) gerechtfertigt, die feststellte, dass der Einfluss von Hintergrundozon im Winter am größten war und relativ wenig Einfluss auf die sommerlichen Spitzenkonzentrationen hatte.

2.1.2 Trends der Ozonüberschreitungen

In Europa sind verschiedenen Metriken und Grenzwerte für Ozon definiert (EC, 2008, siehe auch Abschnitt 3.1 für einen Überblick von Metriken). In den meisten europäischen Ländern kommt es in den warmen Monaten des Jahres regelmäßig zu Ozonüberschreitungen der EU Grenzwerte (z. B. Querol et al., 2016). In diesen Monaten verstärkt eine Kombination aus günstigen Wetterbedingungen und hohen Emissionen von Vorläufern die Mechanismen der Ozonbildung. Genauer gesagt haben sich die Ozonkonzentrationen in abgelegenen, ländlichen Regionen in Europa zwischen den 1950er und 1990er Jahren fast verdoppelt, was auf die zunehmenden Emissionen von NO_x und VOC zurückzuführen ist (Staehelin et al., 1994), die als die wichtigsten Ozonvorläuferstoffe gelten. In den letzten Jahrzehnten sind die Ozonhöchstwerte an städtischen und ländlichen Standorten jedoch zurückgegangen (z. B. Boleti et al., 2019b, Colette et al., 2017). Abbildung 1 zeigt Trends in MDA8 (höchster täglicher 8-Stunden-Mittelwert) anhand des vierthöchsten MDA8-Wertes als Indikator für den Trend an ländlichen Standorten. Allerdings kann die hohe interannuelle Variabilität der gemessenen Ozonkonzentrationen aufgrund des starken meteorologischen Einflusses zu großen Unsicherheiten bei den langfristigen Trends führen.

Abbildung 1: Zusammenstellung des Jahresmittelwerts für Ozon (schwarz) und des vierthöchsten MDA8-Ozonwerts (rot), der zwischen 1990 und 2012 an 55 ländlichen EMEP-Messstationen gemessen wurde. Die dicke Linie ist der messnetzweite jährliche Median und die unteren/oberen Grenzen der schattierten Bereiche sind das 25. und 75. Perzentil. Die dünnen geraden Linien zeigen den linearen Trend über die Zeiträume 1990-2001 und 2002-2012 und die gestrichelten Linien zeigen die Referenzkonzentrationen der WHO-Luftqualitätsleitlinie (50ppb) und des EU-Langzeitzielwerts (60ppb). Quelle: TFMM-Bericht 2016 (Colette et al., 2017).



Quelle: Colette et al., 2017

2.1.3 Einfluss des Emissionstrends

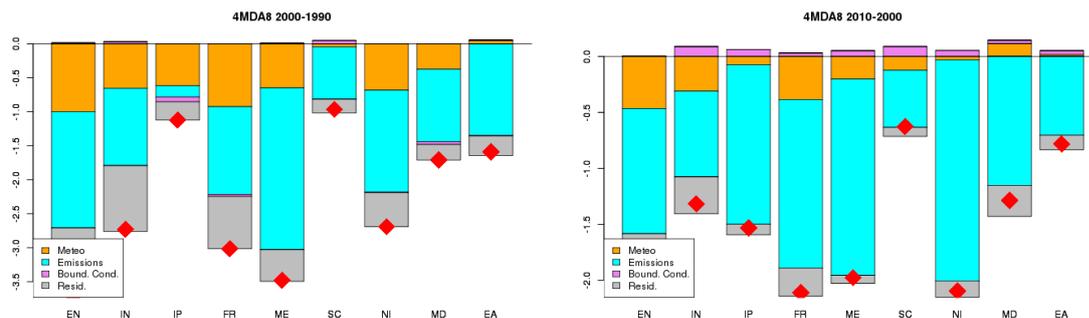
Seit Anfang der 1990er Jahre sind die NO_x - und VOC-Emissionen in Europa aufgrund der Umsetzung von Emissionsreduktionsmaßnahmen deutlich zurückgegangen (Colette et al., 2011, Guerreiro et al., 2014, Henschel et al., 2015, Colette et al., 2017), beispielsweise wurden in den Städten Katalysatoren für Benzinfahrzeuge eingeführt. Das EMEP-Emissionsinventar (Cooperative Programme for Monitoring and Evaluation of the Long-range Transmission of Air

Pollutants in Europe) berichtet, dass die Emissionen von NO_x und Nicht-Methan-VOC (NMVOC) in Europa zwischen 1990 und 2012 deutlich zurückgegangen sind. Somit sind die maximalen Ozonkonzentrationen in ganz Europa zurückgegangen (Colette et al., 2017; Yan et al., 2018; Fleming et al., 2018).

Andererseits spiegeln die Ozonspitzenkonzentrationen hauptsächlich Veränderungen bei den lokalen bis regionalen Vorläuferemissionen wider. Es wurde festgestellt, dass mehrere Metriken, die die Ozonspitzenwerte repräsentieren, von den 1990er Jahren bis heute kontinuierlich abgenommen haben (Paoletti et al., 2014). Dies ist hauptsächlich auf eine Verringerung der NO_x-Emissionen zurückzuführen. Für Standorte in der Nähe von Verkehrsemissionen wurde ein Anstieg des Mittelwerts der zehn höchsten 1-Stunden-Mittelwerte eines Tages zwischen Mai und September (MTDM) und ein Anstieg der vierthöchsten MDA8-Werte beobachtet (Boletti et al., 2019), da die Ozonkonzentrationen in diesen Gebieten durch lokale Titration beeinflusst werden.

Im Rahmen der TFMM/EURODELTA-Trendstudie wurden Modellsimulationen von 1990 bis 2010, die Emissionsänderungen und Klimaänderungen einschließen, mit Simulationen kombiniert, bei denen entweder die Emissionen, die Meteorologie oder die Randbedingungen konstant gehalten wurden. Dadurch konnte der Einfluss von Klimaänderungen und Emissionsänderungen getrennt abgeschätzt werden. Die Ergebnisse sind in Colette et al. (2017) dargestellt, siehe auch Abbildung 2. Die Emissionsänderungen sind eindeutig der wichtigste Faktor.

Abbildung 2: Trendzuweisung für 4MDA8 gemäß der TFMM/EURODELTA-TRENDS-Modellstudie (Colette et al., 2017)



Quelle: Colette et al., 2017

Dies bedeutet, dass Änderungen der Emissionen bei der Bewertung von Ozonspitzen berücksichtigt werden müssen. NO₂-, Methan- und NMVOC-Emissionen müssen als Variablen in einem Ansatz des Maschinellen Lernens berücksichtigt werden, um Ozonüberschreitungen mit Klimavariablen in Beziehung zu setzen, entweder durch die Berücksichtigung von zeitgleich durchgeführten NO_x-Messungen oder durch die Berücksichtigung von Emissionskatastern.

Biogene Emissionen aus Böden und Vegetation, insbesondere NO_x und Isopren, werden sich als Folge des Klimawandels verändern. Dies kann in dieser Studie jedoch nur als direkt mit der Temperatur und der Trockenheit zusammenhängende Faktoren berücksichtigt werden.

2.2 Ozontrends in Deutschland

Für die Ozontrends in Deutschland wurden kürzlich mehrere Studien durchgeführt. Die Trends der Jahresmittelwerte für den Zeitraum 1990 – 2017 sind in Schmidtke und Schmidt (2019) angegeben und betragen durchschnittlich $0,12 \mu\text{g}/\text{m}^3$ pro Jahr für ländliche Stationen und $0,23 \mu\text{g}/\text{m}^3$ pro Jahr für städtische Hintergrundstationen. Für die vorliegende Studie sind die Trends der Spitzenwerte relevanter als die Trends der Jahresmittelwerte.

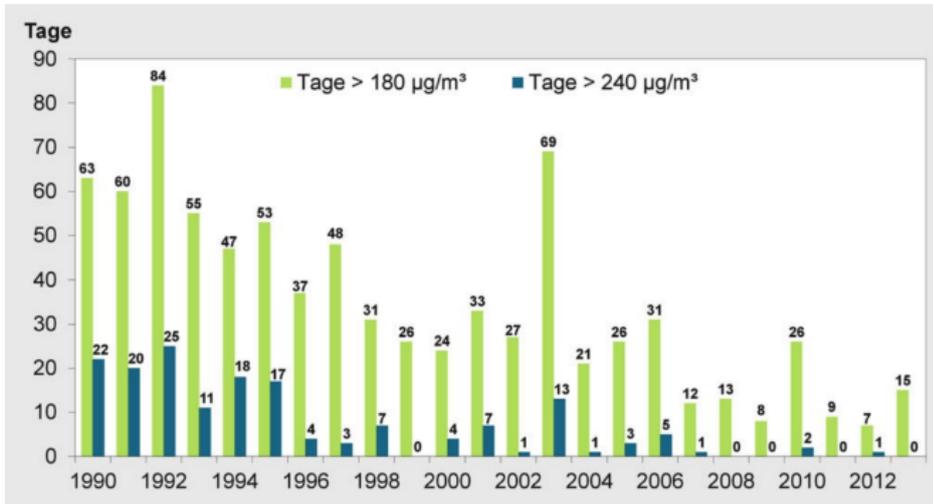
In ihrem Kapitel im Buch von Brasseur et al. (2017) präsentieren Schultz et al. die bisherigen Beobachtungen von Ozonkonzentrationen und -überschreitungen sowie die Emissionsreduktionen (Abbildung 3), die einen deutlichen Rückgang der Anzahl der Überschreitungstage zeigen. Die Informationsschwelle von $180 \mu\text{g}/\text{m}^3$ wird in den meisten Jahren nach 2006 weniger als 15 Mal pro Jahr überschritten, die Alarmschwelle von $240 \mu\text{g}/\text{m}^3$ wird nur noch selten überschritten. Das UBA stellt in seinem Jahresbericht (z.B. UBA, 2020) die Trends in Anzahl der Stunden, in denen der Ozonwert die Informationsschwelle von $180 \mu\text{g}/\text{m}^3$ überschreitet (N180-Trends) sowie die Anzahl der Überschreitungen des MDA8-Wertes von $120 \mu\text{g}/\text{m}^3$ dar (Abbildung 4). Alle Messwerte zeigen eine große interannuelle Variabilität, mit hohen Werten in den Jahren 2003 und 2006 mit extremen Sommern. Der Trend für den Zeitraum 1990 – 2000 scheint viel deutlicher zu sein als für den Zeitraum 2000 – 2019.

2.2.1 Regionale und lokale Unterschiede

Obwohl Ozon hauptsächlich ein großräumiger Luftschadstoff ist, unterscheiden sich die Ozonkonzentrationen für verschiedene Standorte. Die Klassifizierungen für die EU-Berichterstattung (2011/850/EU) beziehen sich auf das Gebiet (städtisch, vorstädtisch, regional) mit weiteren lokalen Spezifikationen (Straße, Industrie und abgelegene Gebiete). Auf Straßenebene und in städtischen Gebieten sind die Ozonkonzentrationen im Allgemeinen niedriger als in der Umgebung, was auf die Ozontitration durch lokale Emissionen zurückzuführen ist, während in vorstädtischen Gebieten die Emissionen der Stadt zur Ozonbildung beigetragen haben.

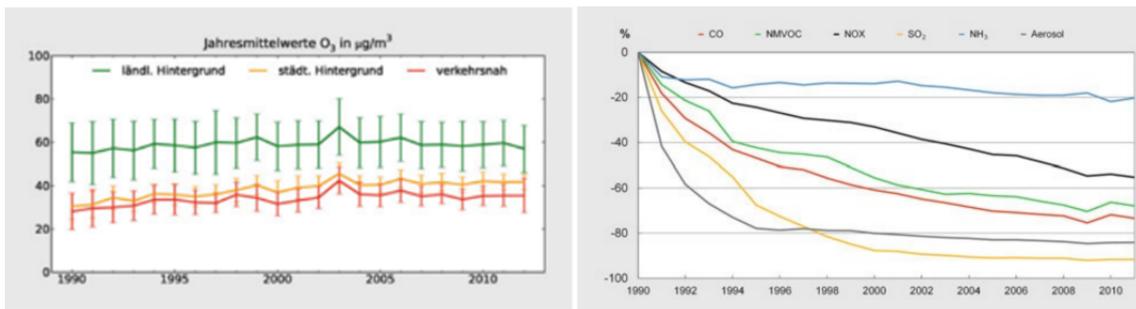
Neben dieser allgemeinen Klassifizierung könnte man nach geographischen Regionen verfeinern. In mehreren deutschen Projekten wurde der Zusammenhang zwischen Ozon und Klima durch weitere Analysen der verfügbaren Ozondaten untersucht. Im Projekt KLENOS (Pavlik et al., 2016) wurde für Deutschland nach Stationstyp und geographischen Regionen unterschieden. Die Regionen wurden anhand der Verteilung der Überschreitungen mit einem Ost-West- und Süd-Nord-Gefälle bestimmt. Als Überschreitungswert wurde $\text{MDA8} > 120 \mu\text{g}/\text{m}^3$ verwendet.

Abbildung 3: Beobachtete Anzahl von Überschreitungen der Alarm- und Informationsschwellenwerte von 240 und 180 $\mu\text{g}/\text{m}^3$ in Deutschland (oberes Feld), beobachtete Jahresmittelkonzentrationen in Deutschland (unteres linkes Feld) und Emissionstrends in Deutschland (unteres rechtes Feld). Abbildungen entnommen aus Brasseur et al. (2017).



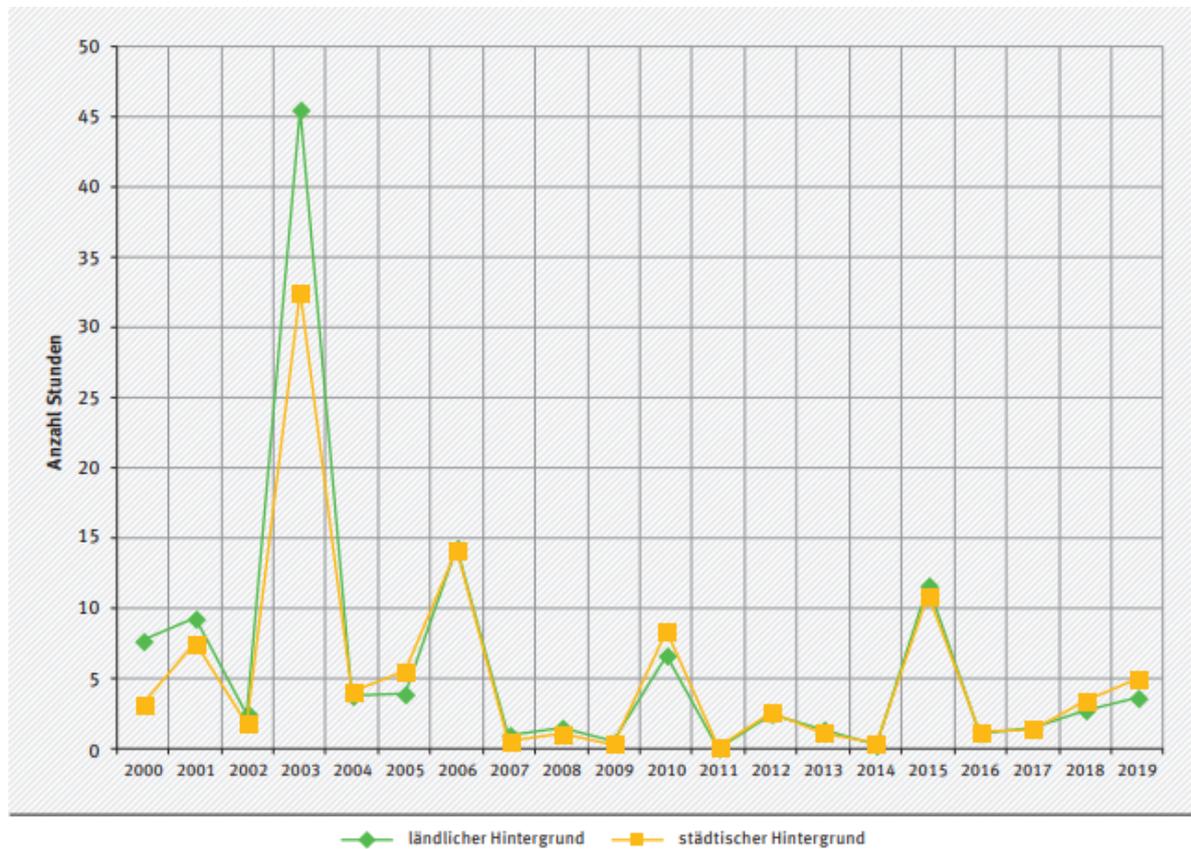
■ Abb. 13.5 Anzahl der Tage mit Überschreitungen der maximalen stündlich gemittelten Ozonkonzentration von 180 und 240 $\mu\text{g}/\text{m}^3$ an deutschen Stationen. (Datenquelle: Umweltbundesamt)

Y-Achse: Ozonkonzentration in $\mu\text{g}/\text{m}^3$, X- Achse Jahre 1990-2013. In grün ländlicher Hintergrund (Werte um 60), in gelb städtischer Hintergrund (Werte um 40), in rot verkehrsnah (2-5 $\mu\text{g}/\text{m}^3$ niedriger als städtisch), als Linien mit Standardabweichung in vertikalen Strichen (um 15 $\mu\text{g}/\text{m}^3$ für ländlich, 5-10 $\mu\text{g}/\text{m}^3$ für städtisch und verkehrsnah).

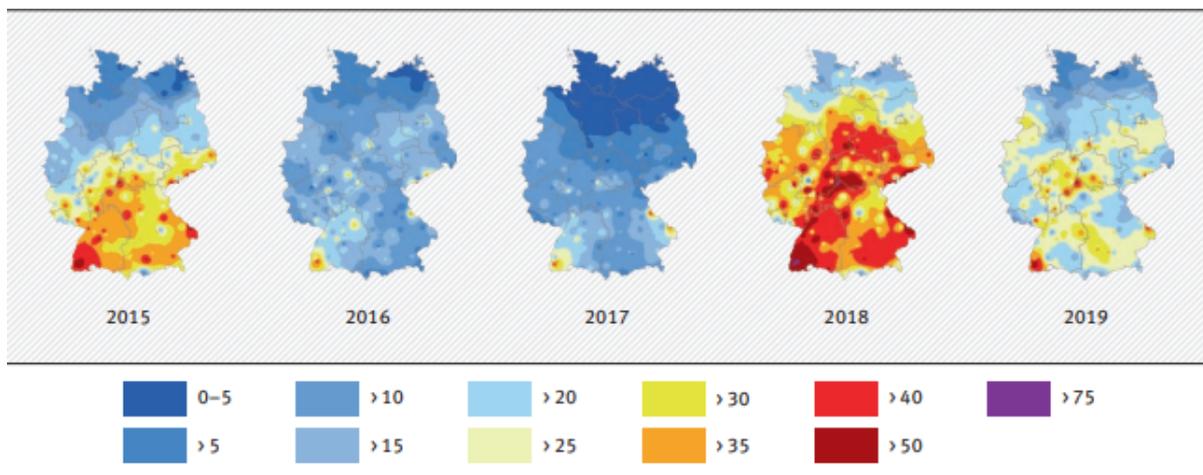


Quelle: Brasseur et al, 2017

Abbildung 4: Obere Abbildung: Anzahl der Stunden mit Überschreitung der Informationsschwelle ($180 \mu\text{g}/\text{m}^3$) für ausgewählte Stationen. Untere Abbildung: Räumliche Verteilung der Anzahl der Tage mit Überschreitungen des MDA8 = $120 \mu\text{g}/\text{m}^3$ Schwellenwertes.



Quelle: Umweltbundesamt 2020



Quelle: UBA, 2020

2.3 Ozontrends und Klimaszenarien

Die Auswirkungen des Klimawandels und der Emissionsänderungen in den letzten 20 Jahren wurden in der bereits genannten TFMM/EURODELTA-Trendstudie (Collette et al., 2017) mit einem Ensemble von Modellen untersucht. Neben der Simulation des Szenarios mit Klimawandel und Emissionsänderungen, die anhand von EMEP-Beobachtungen ausgewertet wurde, konnten auch die Auswirkungen des Klimawandels und der Emissionsänderungen einzeln untersucht werden. Dafür wurden zusätzliche Simulationen durchgeführt, bei denen entweder die Emissionen oder die Meteorologie für den Zeitraum konstant gehalten wurden.

Für die Bewertung des künftigen Klimas müssen Klimaszenarien definiert werden. Im Rahmen der CMIP-Projekte (Coupled Model Intercomparison Project⁴) wurden detaillierte Emissionsszenarien mit „business as usual“, „maximal realisierbarer Technologie“ und verschiedenen plausiblen Zwischenformen dieser Szenarien skizziert. Diese Szenarien wurden mit Zahlen und Buchstaben kodiert. In neueren Studien wurden repräsentative Konzentrationspfade (Representative Concentration Pathway, RCP) verwendet, die für den fünften Sachstandbericht des Intergovernmental Panel on Climate Change (IPCC) definiert wurden. Der Begriff Konzentrationspfad weist darauf hin, dass sich die Szenarien an ausgewählten zeitlichen Verläufen der Treibhausgas-Konzentrationen orientieren und weniger an Emissionen. Es werden vier Szenarien ausgewiesen, die entsprechend dem angenommenen Bereich des Strahlungsantriebs im Jahre 2100 (z. B. 2,6 W/m²) als RCP2.6, RCP4.5, RCP6.0 und RCP8.5 bezeichnet werden. Das mildeste Szenario RCP2.6 bedeutet einen Anstieg der Globalstrahlung um 2,6 W/m², das RCP 8.5 ist das extremste Szenario. Diese Szenarien wurden verwendet, um globale Zirkulationsmodelle für langfristige Simulationen anzutreiben. Die daraus resultierenden Klimaszenarien werden mit regionalen Klimamodellen auf eine höhere räumliche Auflösung heruntergerechnet. Das Problem bei globalen Klimamodellen ist, dass sie, abhängig vom Modell, systematische Fehler in ihren meteorologischen Variablen aufweisen. Dies hängt mit den Präferenzen der Zirkulationsmuster in den verschiedenen Modellen zusammen. Bei gleichem Jahresmittelwert für die Temperatur kann ein Modell höhere Temperaturen im Winter und das andere im Sommer aufweisen, was sich deutlich auf die modellierten Ozon- und Feinstaubkonzentrationen auswirkt (z. B. Manders et al., 2012, Pavlik et al., 2016). Eine Fehlerkorrektur der einzelnen Variablenfelder kann jedoch nicht für die in der Luftqualität relevanten Variablen vorgenommen werden. Diese weisen eine nichtlineare Korrelation auf und sollten daher nicht einzeln korrigiert werden.

Es gibt gekoppelte Klima-Luftqualität-Modelle, bei denen Chemie-Transport-Modelle die von Klimamodellen erzeugte Meteorologie nutzen, um die Auswirkungen von Klimaänderungen auf die Luftschadstoffkonzentrationen zu untersuchen. Diese gekoppelte Modellsysteme sind jedoch sehr rechen- und zeitintensiv in der Anwendung. Daher werden sie entweder mit niedriger räumlicher Auflösung oder für kurze Zeiträume verwendet. Die für die Modellierung der Luftqualität verwendeten Emissionsszenarien sind häufig unabhängig von den Szenarien, die für Studien zur Klimaänderung verwendet werden, da sich die Zielsetzungen weitgehend nicht überschneiden und für die Luftqualität häufig regionale Szenarien verwendet werden. Für solche gekoppelten Modellsysteme gibt es mehrere Trendanalysestudien, in denen die Auswirkungen von Emissionsänderungen und Klimaänderungen untersucht wurden.

Collette et al., (2015) bewerteten ein Ensemble von Klimasimulationen und gaben eine robuste Schätzung des Anstiegs der sommerlichen Ozonkonzentrationen in Europa um 1 – 1,5 ppbv bis

⁴ <https://www.wcrp-climate.org/wgcm-cmip>

2100 aufgrund des Klimawandels an. Dieser Anstieg könnte jedoch durch Emissionsmaßnahmen kompensiert werden.

Pavlik et al., (2016) bewerteten explizit die Auswirkungen des Klimawandels auf die Ozonkonzentrationen für Deutschland. In einer Simulation mit dem RCG Chemie-Transport-Modell, das von ECHAM6 unter dem RCP8.5-Szenario angetrieben wird, stieg die Anzahl der Überschreitungstage ($MDA8 > 120 \mu\text{g}/\text{m}^3$) im Mittel der Jahre 2020-2050 um 30 % im Vergleich zum Referenzzeitraum 1976-2005. Dies könnte eine Unterschätzung des Anstiegs sein, da das treibende Klimamodell auf einen geringeren Anstieg der Sommertemperaturen ausgerichtet ist.

Hertig (2020) untersuchte Ozonüberschreitungen in Bayern. Mit Hilfe von Methoden des Maschinellen Lernens wurde eine Beziehung zwischen meteorologischen Variablen und Ozon hergestellt. Diese wurde verwendet, um die Auswirkungen des Klimawandels unter dem RCP8.5-Szenario mit Eingangsdaten basierend auf Ergebnissen von sechs CMIP5-Erdsystemmodellen zu untersuchen. Es wurde geschätzt, dass der Schwellenwert eines täglichen Ozonmaximums von $>120 \mu\text{g}/\text{m}^3$ am Ende des Jahrhunderts an 45 % aller Tage von April bis September überschritten wird, im Vergleich zu 15 % im Referenzzeitraum.

2.4 Jährliche Ozonüberschreitungen und Klimawandel

Ziel dieses Projekts ist es, den statistischen Zusammenhang von jährlichen Ozonüberschreitungen mit jährlichen Klimakennzahlen und Emissionen zu analysieren. Dieser statistische Zusammenhang sollte robust sein, und es werden Methoden des Maschinellen Lernens eingesetzt, um diesen Zusammenhang zu ermitteln. Es wurde keine Literatur gefunden, die die jährliche Anzahl der $MDA8$ -Überschreitungen und die Episodenlänge mit jährlichen Klimakennzahlen in Beziehung setzt.

Dieser statistische Zusammenhang sollte zunächst für die beobachtete Meteorologie und das beobachtete Ozon entwickelt werden. Dies bedeutet, dass diese Analyse sowohl Änderungen der Emissionen also auch Änderungen des Klimas einschließt, da diese gleichzeitig stattfanden. Es gibt Methoden, um sie teilweise zu entkoppeln und Trends in Ozonmaxima als Folge von Klimaänderung zu bewerten, wenn stündliche/tägliche Ozonmessungen und tägliche meteorologische Daten analysiert werden. Wir verwenden hier jedoch ausdrücklich die Einschränkung, dass die Eingangsvariablen jährliche aggregierte Klimakennzahlen sind. Die Philosophie unseres Ansatzes besteht darin, dass die Auswirkungen von Emissionsänderungen und Klimawandel durch einen guten Algorithmus für Maschinelles Lernen ordnungsgemäß getrennt und berücksichtigt werden können.

Wenn robuste statistische Zusammenhänge gefunden werden, könnte die Methode in Zukunft verwendet werden, um die Auswirkungen von Klimaszenarien auf Ozonüberschreitungen zu untersuchen. Es ist zu beachten, dass sich die statistischen Zusammenhänge in der Zukunft aufgrund von Änderungen der großräumigen Zirkulationsmuster ändern können.

Meteorologische Variablen ändern sich nicht unabhängig voneinander, außerdem hängen ihre Änderungen und Korrelationen teilweise von dem verwendeten Klimamodell ab (z. B. Manders et al., 2012, Pavlik et al., 2016). Dies ist bei der Auswahl von Klimakennzahlen, wie sie in den nächsten Abschnitten vorgestellt werden, von Bedeutung, insbesondere wenn einige von ihnen sehr dominant zu sein scheinen. In der realen Welt können Änderungen der Bodenbedeckung, einschließlich Änderungen der Vegetation, einen zusätzlichen Einfluss haben (z. B. Hendriks et al., 2016). Darüber hinaus steigen die Methankonzentrationen, was zu einem Anstieg der Ozonkonzentrationen beitragen kann (Van Dingenen et al., 2018), die Auswirkungen auf die

Spitzenkonzentrationen sind jedoch nicht sehr deutlich. Dies bedeutet, dass obwohl die Methode generell für Szenarien angewandt werden kann, kann das für das aktuelle Klima trainierte Modell nicht einfach auf ein zukünftiges Szenario angewandt werden. Es kann ein neues Training erforderlich sein, zum Beispiel basierend auf modellierten Ozonwerten aus einem gekoppelten Klima-Luftqualitätsmodell.

3 Auswahl der Ozonmetrik

3.1 Verfügbare Ozonmetriken

Ziel des Projektes ist es, Veränderungen bei den Ozonüberschreitungen in Bezug auf die Anzahl der Tage und die Episodendauer zu ermitteln. Es muss eine ausreichende Anzahl von Episoden ermittelt werden, um einen Algorithmus zu trainieren, der robuste und physikalisch aussagekräftige Ergebnisse in Bezug auf regionale Unterschiede und Trends liefert. Daher besteht einer der ersten Schritte darin, einen geeigneten Schwellenwert für Ozon zu ermitteln. Dieser Schwellenwert bezieht sich vorzugsweise auf international anerkannte Schwellenwerte und Konventionen wie die EU-Richtlinie 2008/50/EG und den EU-Beschluss 2011/850/EU, USEPA- oder WHO-Normen. Eine umfangreiche Liste von Metriken für Ozon und eine Diskussion über ihre Relevanz für die Gesundheit und die Auswirkungen auf Ökosysteme findet sich in Lefohn et al. (2018), teilweise in Verbindung mit dem IGAC TOAR-Projekt (Tropospheric Ozone Assessment). Sie kommen zu dem Schluss, dass für dieselbe Ozonzeitreihe unterschiedliche Trends für die verschiedenen Metriken gefunden werden. Die Auswahl der Metriken ist also von Bedeutung, wenn Änderungen in Klima und Emissionen mit Änderungen der Konzentrationen in Verbindung gebracht werden sollen. In diesem Abschnitt werden die wichtigsten Metriken vorgestellt und diskutiert. Wir beschränken uns auf die basalen Metriken und konzentrieren uns auf die Werte, die für Expositions- und Trendbewertungen relevant sind.

Ozonüberschreitungen lassen sich anhand der folgenden offiziellen Metriken quantifizieren:

- ▶ Täglicher Schwellenwert, der auf folgenden Werten basiert:
 - stündliche Informationsschwelle von $180 \mu\text{g}/\text{m}^3$ und Alarmschwelle von $240 \mu\text{g}/\text{m}^3$, basierend auf 1-Stunden-Mittelwerten. Die Gesamtzahl der Stunden pro Jahr, in denen die Informationsschwelle überschritten wurde, wird angegeben (N180). (EU, 2008). Tageshöchstwerte der 1-Stunden-Mittelwerte werden als MDH (Maximum Daily 1-Hourly average) gekennzeichnet.
 - Tägliche Ozonwerte über $120 \mu\text{g}/\text{m}^3$ (EU, 2008; Zielwert / Langfristziel zum Schutz der menschlichen Gesundheit) oder $100 \mu\text{g}/\text{m}^3$ (WHO, 2021; WHO-Richtwert basierend auf dem höchsten täglichen 8-Stunden-Mittelwert, MDA8 - Maximum Daily 8h Average), berechnet aus stündlich gleitenden 8-Stunden-Mittelwerten. Der EU-Zielwert soll langfristig nicht überschritten werden, zurzeit gilt er als überschritten, wenn er an mehr als 25 Tagen im Kalenderjahr auftritt, gemittelt über drei Jahre. Für den WHO-Richtwert gilt, dass das 99. Perzentil der täglichen MDA8-Werte des Jahres den Wert von $100 \mu\text{g}/\text{m}^3$ nicht überschreiten soll.
- ▶ Kumulierte Werte über ein Jahr:
 - SOMO35, die jährliche Summe der täglichen MDA8-Werte über 35 ppb über ein Jahr, ausgedrückt in $\text{ppb} \cdot \text{Tage}$ (WHO, 2008)
 - AOT40, Summe der Differenzen zwischen den 1-Stunden-Mittelwerten von Ozon über 40 ppb ($\sim 80 \mu\text{g}/\text{m}^3$) und dem Wert 40 ppb zwischen 8 Uhr morgens und 20 Uhr abends von Mai bis Juli, ausgedrückt in $\text{ppb} \cdot \text{Stunde}$ (EU2008).

Die Überschreitungswerte für den 1-Stunden-Mittelwert ($180, 240 \mu\text{g}/\text{m}^3$) und den MDA8 ($100, 120 \mu\text{g}/\text{m}^3$) beruhen auf Studien über die Auswirkungen auf die Gesundheit. Es wurden neue

Hinweise darauf gefunden, dass auch bei niedrigeren Konzentrationen noch ein Gesundheitsrisiko besteht. Orellano et al. (2020) fassen in ihrer Metaanalyse die Ergebnisse von elf Arbeiten zusammen, die sich mit Ozon befassen. Werte von $60 \mu\text{g}/\text{m}^3$ für Saisonmittelwert (Peak Season) und $100 \mu\text{g}/\text{m}^3$ für MDA8 wurden als aussagekräftige Schwellenwerte für gesundheitliche Auswirkungen ermittelt. Diese Ergebnisse könnten zu niedrigeren Überschreitungsschwellenwerten in neuen oder überarbeiteten EU-Luftqualitätsrichtlinien führen, aber es gibt noch keine offizielle Entscheidung über die zu verwendenden Werte. Eine Metrik, die wie MDA8 auf Tageswerten basiert, ist robuster als eine Metrik, die wie Informations- und Alarmschwelle auf 1-Stunden-Mittelwerten basiert. MDA8 wird daher in den meisten Ozonstudien verwendet (z. B. Otero et al., 2016, Pavlik et al., 2016).

Weitere Metriken für die jährliche Darstellung hoher Ozonkonzentrationen sind:

- ▶ Tageshöchstwert basierend auf 1-Stunden-Mittelwerten (MDH – Maximum Daily 1h Average)
- ▶ Jährlicher maximaler 1-Stunden-Mittelwert. Dieser Wert ist nicht sehr robust und wird kaum verwendet.
- ▶ Mittelwert der zehn höchsten 1-Stunden-Mittelwerte eines Tages zwischen Mai und September (Mean of the ten highest mean data – MTDM)
- ▶ 90. oder 95. Perzentil der täglichen maximalen 1-Stunden-Mittelwerte
- ▶ 3-Monats-Mittelwert der täglichen maximalen 1-Stunden-Mittelwerte (TOAR, bezogen auf Studien zur globalen Krankheitslast)
- ▶ Jährliches Maximum des MDA8
- ▶ Sommerlicher Mittelwert des MDA8 (TOAR)
- ▶ Vierthöchster jährlicher MDA8 (4-MDA8). Entspricht dem 98. Perzentil des 6-Monats-Sommer-MDA8.
- ▶ Vierthöchster gewichteter 5h kumulativer Expositionsindex. Er ist mit dem MDA8 verwandt, verwendet jedoch einen Gewichtungsfaktor, der Werte über 90 ppb (W90) mehr Gewicht verleiht als Werten darunter. Für Details siehe Lefohn et al. (2018).

Die jährlichen Werte werden alle in verschiedenen Studien verwendet. MTDM ist robuster als z. B. das 95. Perzentil oder 4-MDA8, da es sich um ein Mittel von 10 Werten handelt. Die Metrik 4-MDA8 wurde mit mehreren Expositionsmetriken in Verbindung gebracht und wird z. B. in Collette et al. (2017) verwendet. Die jährlichen Metriken sind für eine Trendanalyse nützlich, können aber nicht verwendet werden, um Veränderungen in der Saisonalität oder in der Episodendauer zu erkennen. Das Gleiche gilt für SOMO35 und AOT40. In TOAR wird zusätzlich der Tag angegeben, an dem das Maximum des gleitenden Dreimonatsmittelwerts erreicht wird, um dies zu beheben.

Die EEA (2020) beschreibt eine Episode als einen Zeitraum von in der Regel einigen Tagen bis zu 2 – 3 Wochen mit täglichen Überschreitungen der zum Schutz der menschlichen Gesundheit festgelegten Schwellenwerte. Dies ist keine scharfe Definition und es muss ein geeigneter unterer Schwellenwert gewählt werden. In diesem Bericht ist eine Ozonepisode definiert als

zwei oder mehr aufeinanderfolgende Tage mit Überschreitungen. Die täglichen Überschreitungen können aggregiert werden, um folgende Größen zu bestimmen:

- ▶ Anzahl der Episoden
- ▶ Dauer der Episoden (Anzahl der aufeinanderfolgenden Tage mit Überschreitung)
- ▶ Tag im Jahr der Ozonüberschreitung zur Untersuchung möglicher Veränderungen der Saisonalität
- ▶ Geographie der Episoden (betroffene Stationen).

Obwohl viele Studien dem Verständnis spezifischer Episoden gewidmet sind (z. B. die sommerlichen Hitzewellen in den Jahren 2003 und 2006, z. B. Vautard et al. (2006), Solberg et al. (2006), werden Anzahl und Dauer der Episoden unseres Wissens nach nicht systematisch erfasst. Collette et al. (2015) geben beispielsweise einen Überblick über europäische Studien zu Ozontrends, von denen keine explizit die Episodendauer und die Anzahl der Episoden berücksichtigt. Die Anzahl und Dauer der Episoden sind nicht Teil der offiziellen Berichterstattung der Länder zur Luftqualität.

3.2 Auswahl der Metrik für den Ozonschwellenwert

In diesem Projekt geht es um die Analyse von Ozonextremen, wobei nur die Werte über einem Ozonschwellenwert von Interesse sind (Peaks over Threshold, POT). Bei der Analyse von Spitzenwerten ist es von entscheidender Bedeutung, einen geeigneten Schwellenwert zu wählen. Es gibt jedoch keine einheitliche Regel für die Festlegung eines solchen Schwellenwertes. Die Anzahl der Überschreitungen sollte groß genug sein, um das Ende der Verteilung der Werte abzudecken. Gleichzeitig sollte die Zahl der Überschreitungen klein genug sein, um nur das Ende abzudecken und die Extrema zu repräsentieren. Das Ende der Verteilung kann mit der aus der Statistik bekannten verallgemeinerten Pareto-Verteilung (GPD) beschrieben werden. Faustregeln zur Definition der Extrema sind das 90. Perzentil oder die Quadratwurzel (\sqrt{n}) der n höchsten Werte. Es kann aber auch praktisch relevant sein, sich an eine bestehende Metrik zu halten, die zu einer aussagekräftigen Beschreibung des Schwellenwertes führt. Im nächsten Abschnitt wird eine Reihe von Schwellenwerten untersucht und im Hinblick auf eine gute Metrik diskutiert. Folgende Kriterien für die Auswahl einer Metrik für den Ozonschwellenwert sind zu beachten. Die Anzahl der Überschreitungen:

- ▶ sollte nicht zu groß sein (< 25% der Tage → weniger als 45 Tage im Sommer),
- ▶ sollte eine interannuelle Variabilität aufweisen, die die meteorologischen Schwankungen und die Emissionsschwankungen widerspiegelt,
- ▶ sollte regionale Schwankungen aufweisen, die die Variabilität von Meteorologie und Emissionen widerspiegeln,
- ▶ sollte hinsichtlich der Datenverfügbarkeit robust sein,
- ▶ muss die Möglichkeit bieten, die Anzahl der Episoden und die Episodenlänge zu bestimmen.

Die Wahl der Metriken ist wichtig, da verschiedene Metriken eine unterschiedliche Bedeutung haben (z. B. Collette et al., 2017). Einige Metriken oder Indikatoren sind für die Untersuchung der **Auswirkungen auf die Gesundheit der Menschen oder der Vegetation** relevanter. Diese

Indikatoren, wie AOT40 und die Anzahl der Überschreitungen, sind z. B. in den Luftqualitätsrichtlinien definiert. Andere Metriken oder Indikatoren können sich auf **Prozesse beziehen**, z. B. ist 4-MDA8 ein Indikator für das Ausmaß der Sommerepisoden.

Um die Wahl eines relevanten Ozonschwellenwertes zu unterstützen, wurden die folgenden Metriken untersucht:

1. Anzahl der Tage, an denen eine Überschreitung des Informationsschwellenwertes von $180 \mu\text{g}/\text{m}^3$ stattfindet ($\text{MDH} > 180 \mu\text{g}/\text{m}^3$),
2. Anzahl der Tage, an denen der 1-Stunden-Mittelwert den Schwellenwert von $120 \mu\text{g}/\text{m}^3$ überschreitet ($\text{MDH} > 120 \mu\text{g}/\text{m}^3$, keine EU/WHO Metrik),
3. Anzahl der Tage, an denen der höchste tägliche 8-Stunden-Mittelwert (MDA8) den Schwellenwert von $120 \mu\text{g}/\text{m}^3$ überschreitet ($\text{MDA8} > 120 \mu\text{g}/\text{m}^3$),
4. Anzahl der Tage, an denen der höchste tägliche 8-Stunden-Mittelwert (MDA8) den Schwellenwerte von $100 \mu\text{g}/\text{m}^3$ überschreitet ($\text{MDA8} > 100 \mu\text{g}/\text{m}^3$),
5. Perzentilwerte (90. oder 95.) der Tageshöchstwerte und der höchsten täglichen 8-Stunden-Mittelwerte.

Die erste Metrik steht im Zusammenhang mit der Regulierung der Luftqualität und bezieht sich auf kurzzeitige Überschreitungen. Die zweite Metrik bezieht sich ebenfalls auf kurzfristige Überschreitungen und steht nicht in direktem Zusammenhang mit den der Luftqualitätsrichtlinie, wurde aber von Hertig (2020) als die nützlichste Kennzahl angesehen. In dieser Studie lieferte sie die zuverlässigsten Ergebnisse in Bezug auf die Anzahl der Überschreitungen. Überschreitungen des MDH-Wertes von $100 \mu\text{g}/\text{m}^3$ traten häufig in den Monaten April bis September auf (40 % der Tage), Überschreitungen des MDH-Wertes von $180 \mu\text{g}/\text{m}^3$ an weniger als 1 % der Tage und der Wert von $120 \mu\text{g}/\text{m}^3$ wurde an etwa 15 % der Tage überschritten. Für Deutschland ist die Zahl der Überschreitungen des Wertes von 180 oder $240 \mu\text{g}/\text{m}^3$ auf eine Zahl zurückgegangen, die zu gering ist, um zuverlässige Ergebnisse für Anwendungen des Maschinellen Lernens zu liefern.

Die dritte und vierte Metrik sind robuster als die ersten beiden, da es sich um Durchschnittswerte handelt. Sie reagieren weniger empfindlich auf Ausreißer und Datenverluste in dem Sinne, dass der Verlust von ein paar Stunden pro Tag weniger Auswirkungen auf das Ergebnis hat. Beide Messgrößen beziehen sich auf die Schwellenwerte der Luftqualitätsvorschriften für die langfristigen gesundheitlichen Auswirkungen von Ozon (EU $120 \mu\text{g}/\text{m}^3$ und WHO $100 \mu\text{g}/\text{m}^3$).

Darüber hinaus werden die Perzentilwerte untersucht. Das 90. Perzentil über das Jahr würde ungefähr dem 95. Perzentil über den Sommer entsprechen. Der Perzentilwert ist auch ein Hinweis auf einen sinnvollen Schwellenwert, da er in einem robusten Verhältnis zu den Extremen steht und in Trendanalysen verwendet wird. Jedoch kann er nicht zur Quantifizierung der Anzahl von Überschreitungen und Episoden oder der Episodendauer verwendet werden.

Im Hinblick auf die Forschungsfrage sind andere Metriken weniger nützlich, da sie weniger direkt mit der maximalen Ozonkonzentration zusammenhängen (z. B. SOMO, AOT), weniger eindeutig mit den Auswirkungen verbunden sind (z. B. MTDM) oder nur relativ wenige Ereignisse auswählen (z. B. stündliche Werte $> 180 \mu\text{g}/\text{m}^3$).

Für jede obengenannte Metrik wurde die Anzahl der Überschreitungstage für jedes Jahr für alle Stationen berechnet, die die Screening-Kriterien für das jeweilige Jahr erfüllten. Das heißt, Stationen, für die nicht genügend Daten vorliegen, werden nicht in die Analyse einbezogen. Auch die Anzahl der Episoden sowie die Episodenlänge wurden berechnet. Zur Untersuchung der Anzahl der Episoden wurden alle Ereignisse mit mehr als einem Tag Überschreitung der Metrik als eine Episode gezählt. Die Länge jeder gefundenen Episode und die maximale Episodenlänge pro Jahr wurden ebenfalls berechnet.

Zur Unterstützung der Wahl der Metrik werden auch Metriken wie Perzentilwerte herangezogen, für die jedoch keine Aggregation auf Episoden möglich ist.

3.2.1 Messstationen und Datenqualität

Als Eingangsdaten werden für die vorliegende Studie die stündlichen Ozonmessungen der Bundesländer und des UBA in Deutschland verwendet, wie sie vom UBA bereitgestellt werden. Diese Stationen sind in Bezug auf ihre weitere und unmittelbare Umgebung (direkte Beiträge von Straßen, Industrie oder Hintergrund) klassifiziert, sowie nach einer EEA/Airbase-Ozonklassifizierung (städtisch, vorstädtisch, ländlicher Hintergrund, ländlich regional). Diese beiden Klassifizierungen wurden zu einer Klassifizierung je Station kombiniert:

- ▶ Straße: Messungen, die direkt von Verkehrsemissionen beeinflusst werden,
- ▶ Industrie: Messungen, die direkt von Industrieemissionen beeinflusst werden,
- ▶ Städtisch: Messungen in einem dicht besiedelten Gebiet, aber abseits von Hauptverkehrsstraßen oder anderen wichtigen Quellen,
- ▶ Vorstädtisch: Messungen in weniger dicht besiedelten städtischen Gebieten abseits des Zentrums; für Ozon können dies die Orte mit den höchsten Konzentrationen in Verbindung mit einer starken Exposition sein,
- ▶ Ländlicher Hintergrund: Messungen, die für Hintergrundkonzentrationen repräsentativ sind,
- ▶ Ländlich regional: Messungen, die für saubere Hintergrundbedingungen weit entfernt von direkten Quellen repräsentativ sind.

Bevor die Daten analysiert werden, müssen die Stationen zunächst auf ihre Gültigkeit für die Analyse überprüft werden.

Gemäß den EU-Verordnungen gelten folgende Kriterien für die Gültigkeit eine Station:

- ▶ mindestens 18 gültige stündliche Messungen pro Tag,
- ▶ zur Berechnung von MDA8 müssen mindestens 6 von 8 Stunden im Mittelungsfenster abgedeckt werden,
- ▶ mindestens 27 gültige Tage pro Monat,
- ▶ 5 von 6 gültige Monate in der Sommersaison (April bis September).

Unter Anwendung dieser Kriterien zeigt das Screening, dass die Zahl der gültigen Stationen pro Jahr zwischen 250 und 350 liegt (Abbildung 5). Die Gesamtzahl der in die Analyse einbezogenen

Stationen beträgt 595. Jedoch waren nicht alle Stationen während des gesamten Zeitraums aktiv, da sich das Messnetz im Laufe der Zeit verändert hat, da Stationen geschlossen und andere neu eingerichtet wurden. So gibt es zum Beispiel 109 Stationen für den Zeitraum 1995 – 2018 und 160 Stationen für den Zeitraum 2000 – 2018. Dieses Screening lässt uns immer noch genügend robuste Daten, um Methoden des Maschinellen Lernens anzuwenden, da die Anzahl der Prädiktoren (Klimakennzahlen, Stationskennzeichnung) und damit die Anzahl von Freiheitsgraden begrenzt ist. Je nachdem, ob das Training für den gesamten Zeitraum oder nur für einen begrenzten Zeitraum von einem oder fünf Jahren durchgeführt wird, wird eine einzelne Station in das Training einbezogen oder ausgeschlossen, wenn an dieser Station nicht die vollen 24 Jahre verfügbar sind.

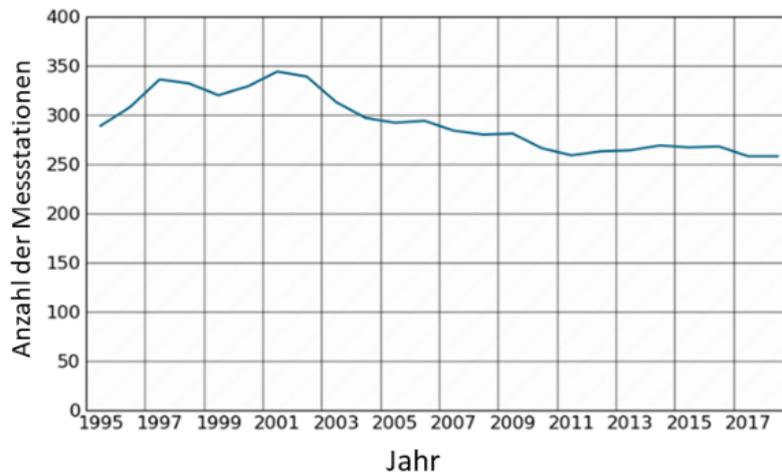
Drei Stationen werden zur Veranschaulichung der interannuellen Variabilität und des Trends herangezogen: Waldhof, Radebeul und Spessart. Sie repräsentieren Stationen in verschiedenen geografischen Regionen und decken den gesamten Zeitraum 1995 – 2018 ab. Weitere Stationen, die detaillierter ausgewertet wurden, sind Potsdam, Eggenstein und Mülheim-Styrum, um andere geografische Gebiete, einschließlich dicht besiedelter Gebiete, mit langfristiger Datenerfassung zu repräsentieren.

Einige Stationen befinden sich auf isolierten Berggipfeln und unterliegen einem anderen Ozonregime, da sie empfindlicher auf großräumige Hintergrundkonzentrationen als auf regionale Emissionsänderungen reagieren und auch durch die vertikale Schichtung der Atmosphäre anders beeinflusst werden. Diese Stationen mit einer Höhe von über 900 m sind in dieser Studie bei der Entwicklung des finalen Modells nicht berücksichtigt. Die Stationen sind:

- ▶ DEUB004 (Schauinsland)
- ▶ DEUB043, DEUB044, DEBY084 (alle Zugspitze)
- ▶ DEUB046 (Forellenbach)
- ▶ DEBW031 (Scharzwald-Süd)
- ▶ DEBY082 (Wankgipfel)
- ▶ DEHE51 (Wasserkuppe)
- ▶ DESN053 (Fichtelberg)
- ▶ DEST039 (Brocken)
- ▶ DETH040 (Großer Eisenberg)
- ▶ DEUB029 (Schmücke)

Für die Bestimmung der Ozonmetrik wurden die vom UBA bereitgestellten stündlichen Daten verwendet inklusive Bergstationen.

Abbildung 5: Anzahl der gültigen Stationen pro Jahr nach der Datenprüfung



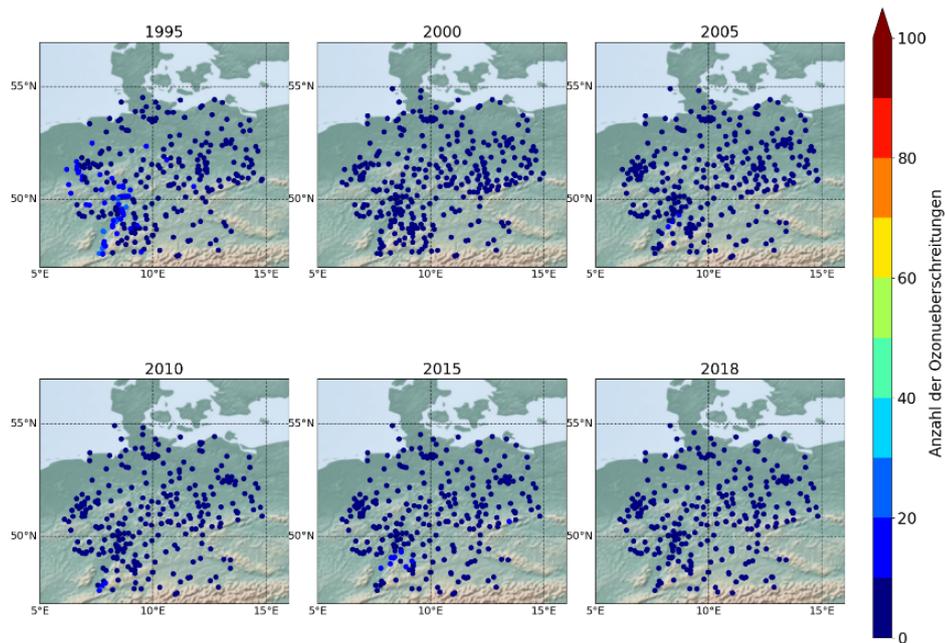
Quelle: Eigene Darstellung, TNO

In den folgenden Abschnitten werden zunächst die verschiedenen Metriken mit Zahlen pro Jahr für 6 verschiedene Jahre vorgestellt, um die geografische Verteilung zu zeigen. Danach werden die Trends für 3-6 Referenzstationen dargestellt.

3.2.2 Überschreitung des Tageshöchstwertes MDH > 180 $\mu\text{g}/\text{m}^3$

Abbildung 6 zeigt die räumliche Verteilung der Anzahl der Überschreitungstage mit MDH > 180 $\mu\text{g}/\text{m}^3$. Bei dieser Metrik ist die Anzahl der Überschreitungstage sehr gering (< 20 Tage im Jahr 1995, < 10 Tage in den anderen Jahren). Außerdem ist die zeitliche und räumliche Verteilung der Überschreitungen ziemlich gleichmäßig. Da die Zahl der Überschreitungen sehr gering ist und die zwischenjährlichen und regionalen Schwankungen gering sind, ist diese Metrik für den Zweck der Studie nicht geeignet. Sie wird daher nicht näher untersucht.

Abbildung 6: Räumliche Verteilung der Anzahl der Überschreitungstage, MDH > 180 µg/m³



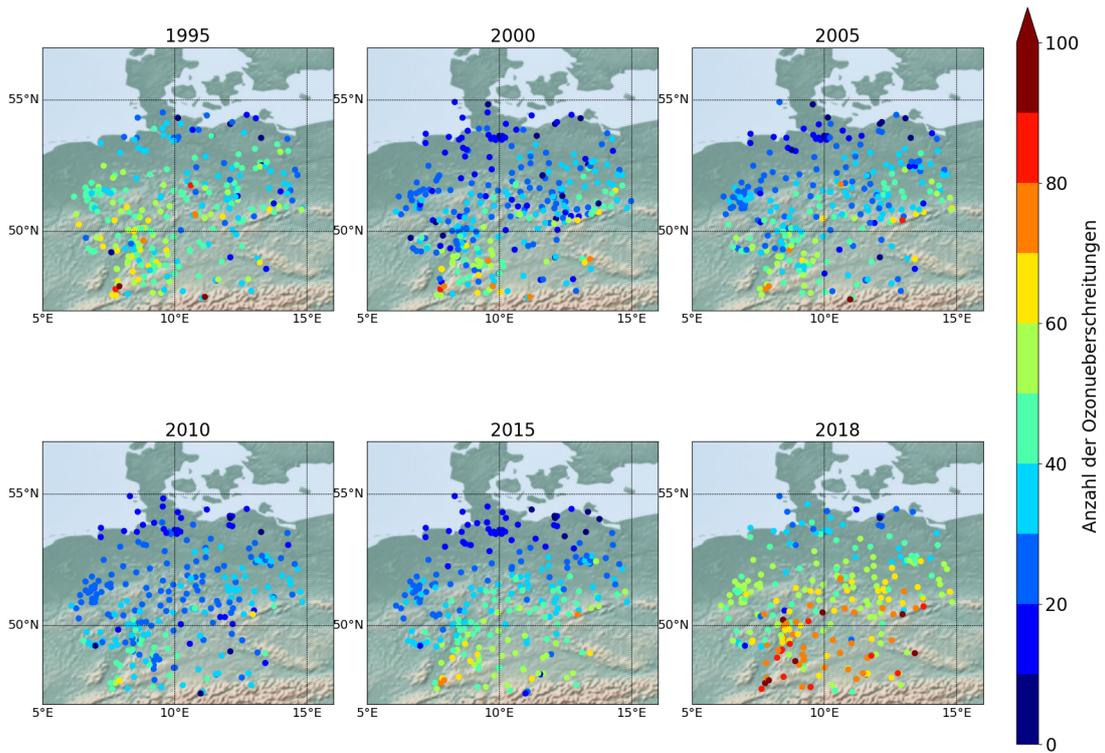
Quelle: Eigene Darstellung, TNO

3.2.3 Überschreitung des Schwellenwertes von MDH > 120 µg/m³

Abbildung 7 zeigt die räumliche Verteilung der Anzahl der Überschreitungstage für den Tageshöchstwert > 120 µg/m³. Die Anzahl der Überschreitungen liegt zwischen 0 und 100. In dieser Abbildung sehen wir eine deutliche Variabilität in der zeitlichen und räumlichen Verteilung der Überschreitungen und insbesondere einen Nordwest-Südost-Gradienten.

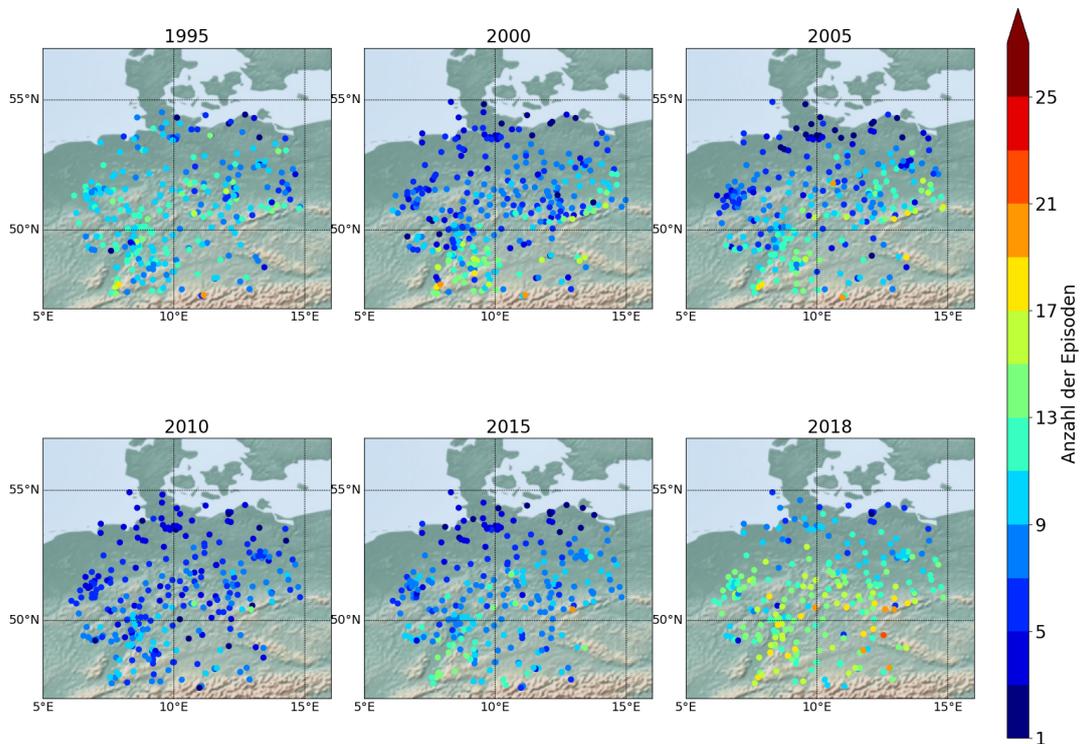
Abbildung 8 und Abbildung 9 zeigen, dass die durchschnittliche Anzahl der Episoden zwischen 2 und 8 pro Jahr liegt und ihre Dauer im Durchschnitt zwischen 3 und 5 Tagen beträgt.

Abbildung 7: Räumliche Verteilung der Anzahl der Überschreitungstage, MDH > 120 µg/m³



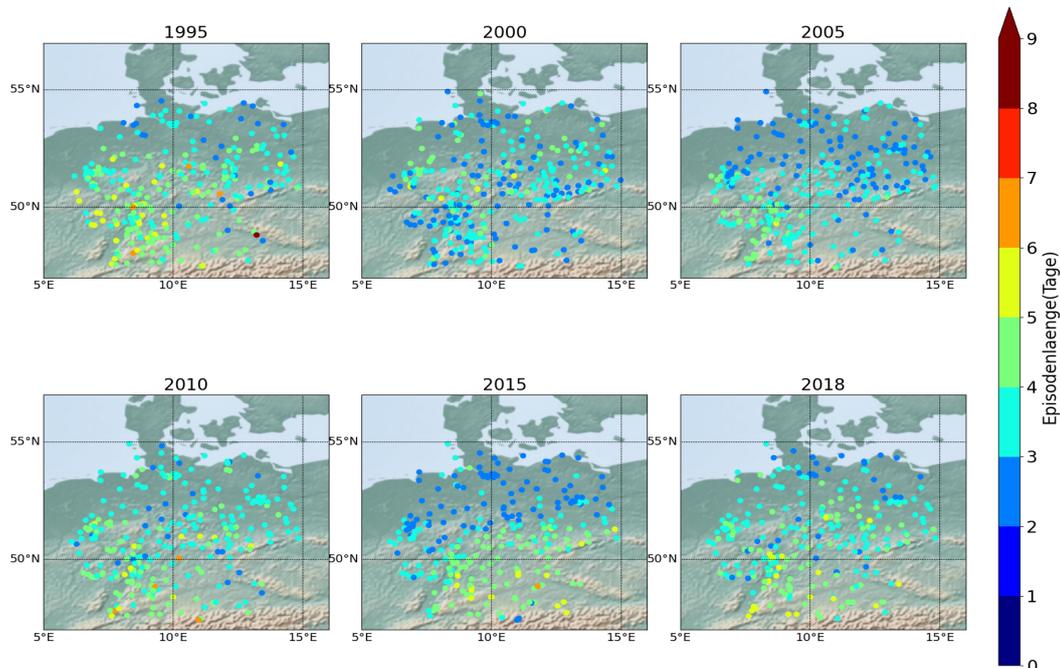
Quelle: Eigene Darstellung, TNO

Abbildung 8: Räumliche Verteilung der Anzahl der Episoden, MDH > 120 µg/m³



Quelle: Eigene Darstellung, TNO

Abbildung 9: Mittlere Dauer der Episoden, MDH > 120 µg/m³

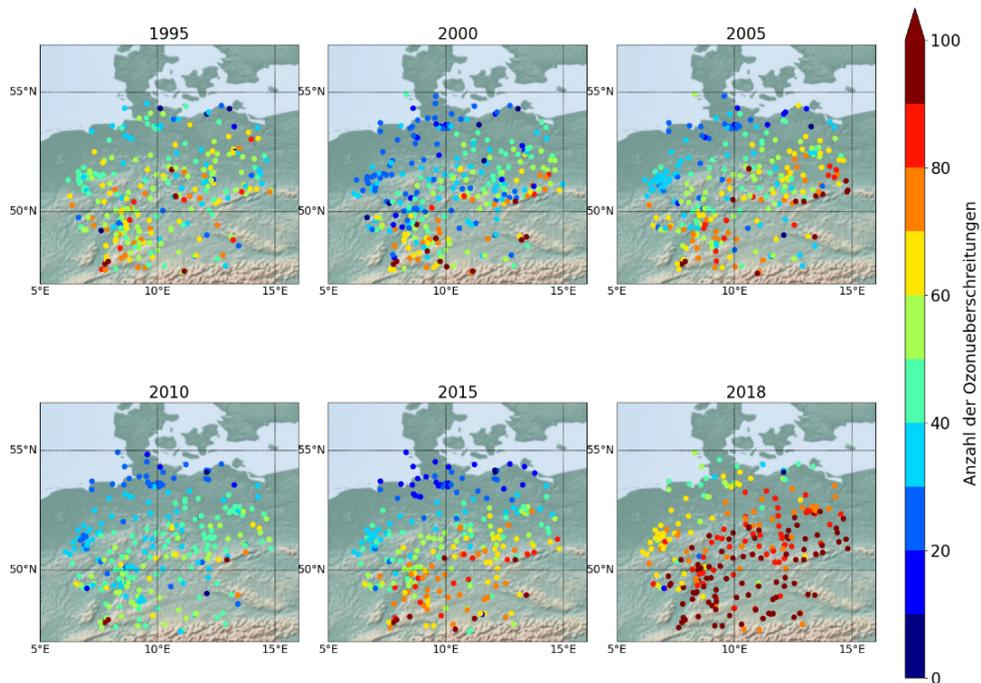


Quelle: Eigene Darstellung, TNO

3.2.4 Überschreitung des Schwellenwerts von MDA8 > 100 µg/m³

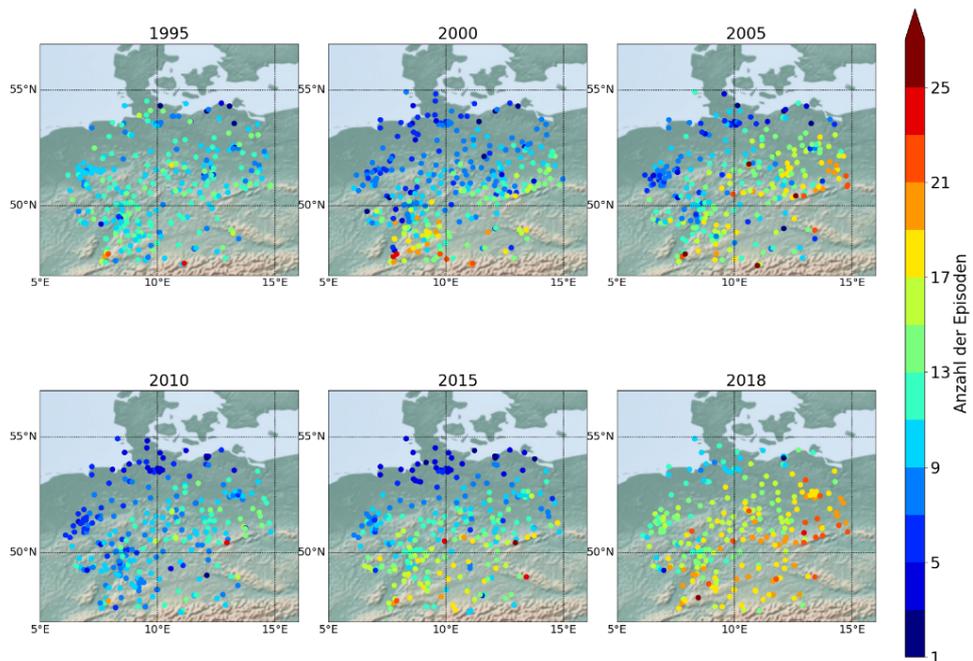
Für den MDA8 > 100 µg/m³ reicht die Zahl der Überschreitungen von 10 bis 100. Abbildung 10 zeigt eine deutliche Variabilität in der zeitlichen und räumlichen Verteilung und insbesondere einen Nordwest-Südost-Gradienten. Darüber hinaus wird der Schwellenwert an hoch gelegenen Stationen (freie Troposphäre) an vielen Tagen erreicht. Abbildung 11 und Abbildung 12 zeigen, dass die jährliche Anzahl der Episoden zwischen 9 und 15 liegt, mit einer mittleren Dauer von 5 bis 6 Tagen.

Abbildung 10: Räumliche Verteilung der Anzahl der Überschreitungstage, $MDA8 > 100 \mu\text{g}/\text{m}^3$



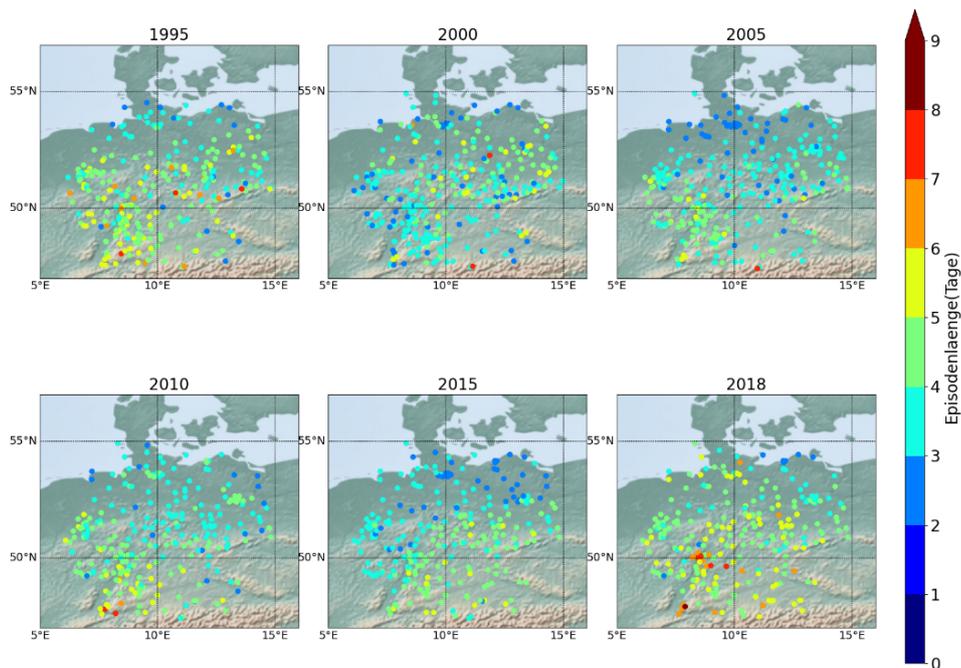
Quelle: Eigene Darstellung, TNO

Abbildung 11: Räumliche Verteilung der Anzahl der Episoden, $MDA8 > 100 \mu\text{g}/\text{m}^3$



Quelle: Eigene Darstellung, TNO

Abbildung 12: Mittlere Dauer der Episoden, MDA8 > 100 µg/m³

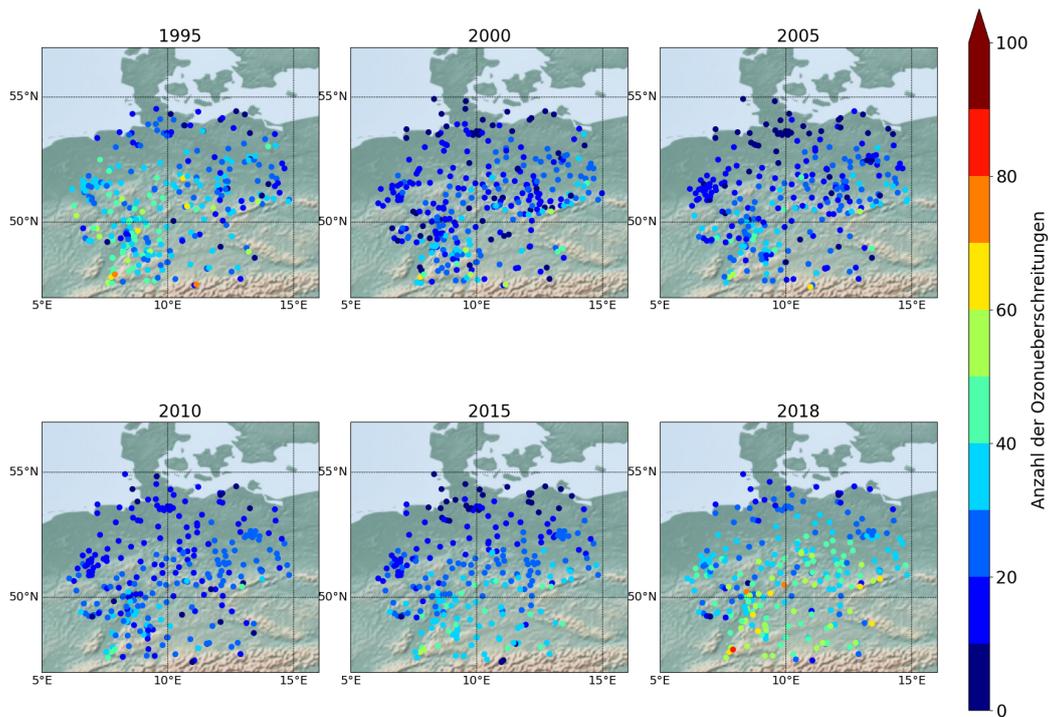


Quelle: Eigene Darstellung, TNO

3.2.5 Überschreitung des Schwellenwertes von MDA8 > 120 µg/m³

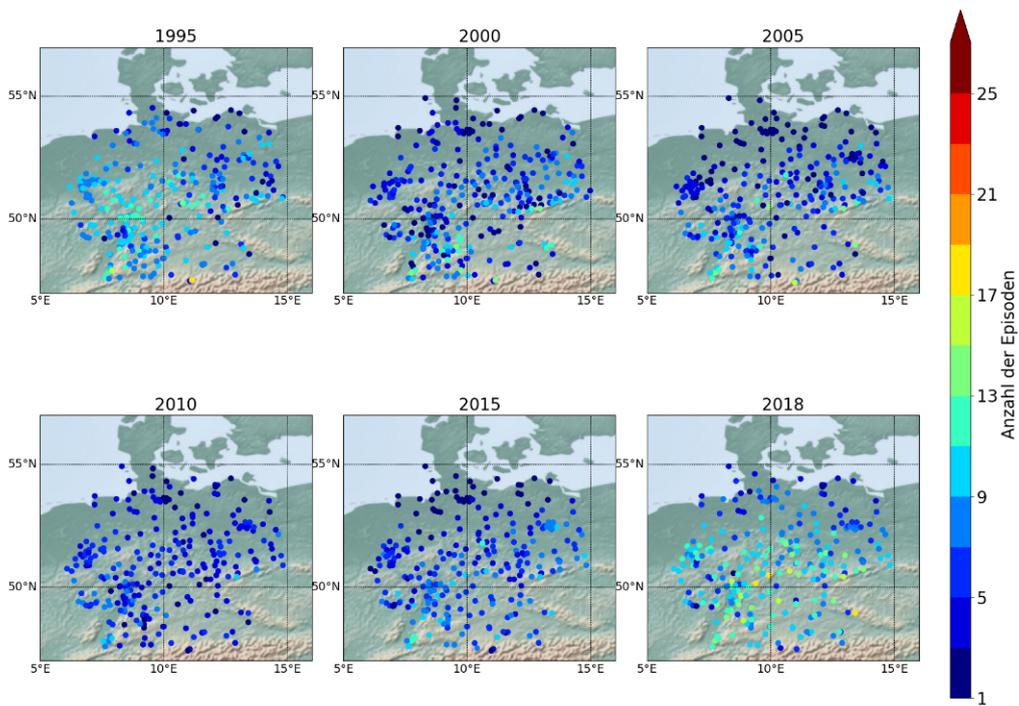
Für MDA8 > 120 µg/m³ reicht die Anzahl der Überschreitungen von 0 bis 60. In Abbildung 13 sehen wir eine deutliche Variabilität in der zeitlichen und räumlichen Verteilung und insbesondere einen Nordwest-Südost-Gradienten. Die durchschnittliche Anzahl der Episoden pro Jahr liegt zwischen 4 und 8 Tagen, mit einer durchschnittlichen Dauer von 2 bis 4 Tagen (Abbildung 14 und Abbildung 15).

Abbildung 13: Räumliche Verteilung der Anzahl der Überschreitungstage, $MDA8 > 120 \mu\text{g}/\text{m}^3$



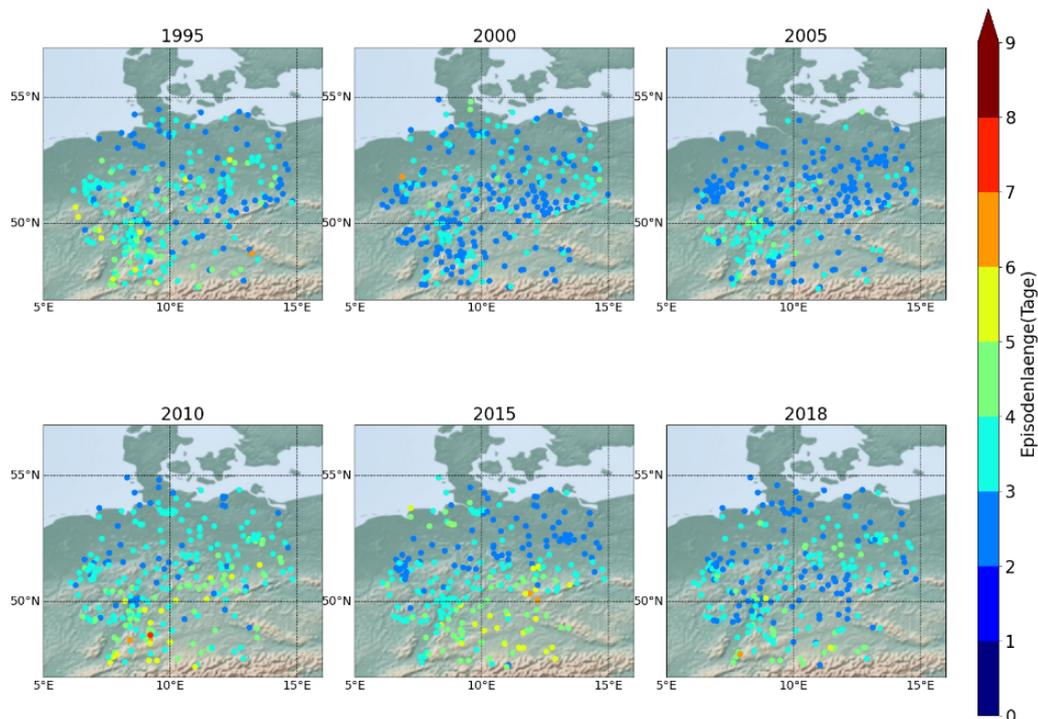
Quelle: Eigene Darstellung, TNO

Abbildung 14: Räumliche Verteilung der Anzahl der Episoden, $MDA8 > 120 \mu\text{g}/\text{m}^3$



Quelle: Eigene Darstellung, TNO

Abbildung 15: Mittlere Dauer der Episoden, MDA8 > 120 µg/m³

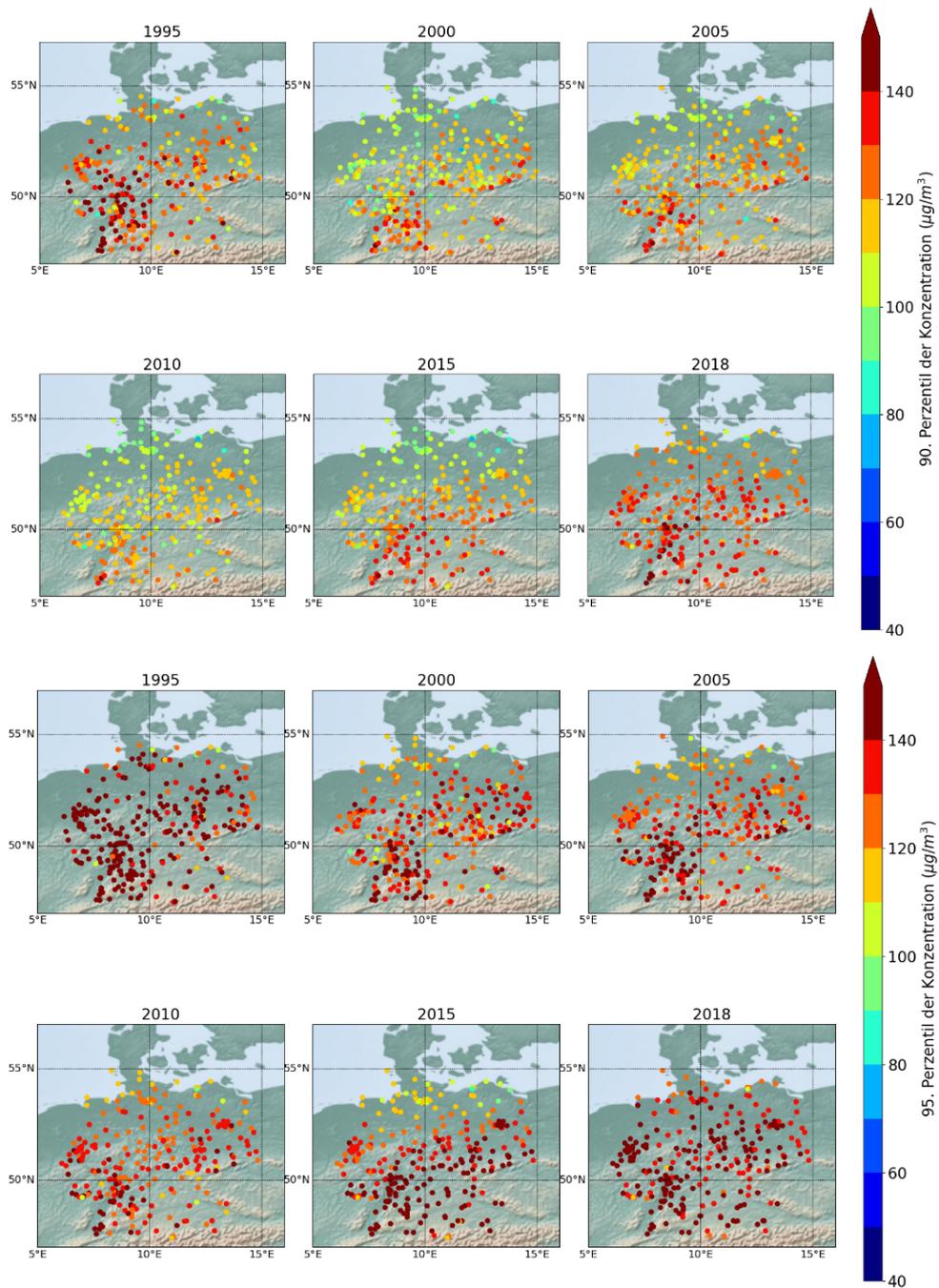


Quelle: Eigene Darstellung, TNO

3.2.6 Perzentilwerte der Ozonkonzentrationen

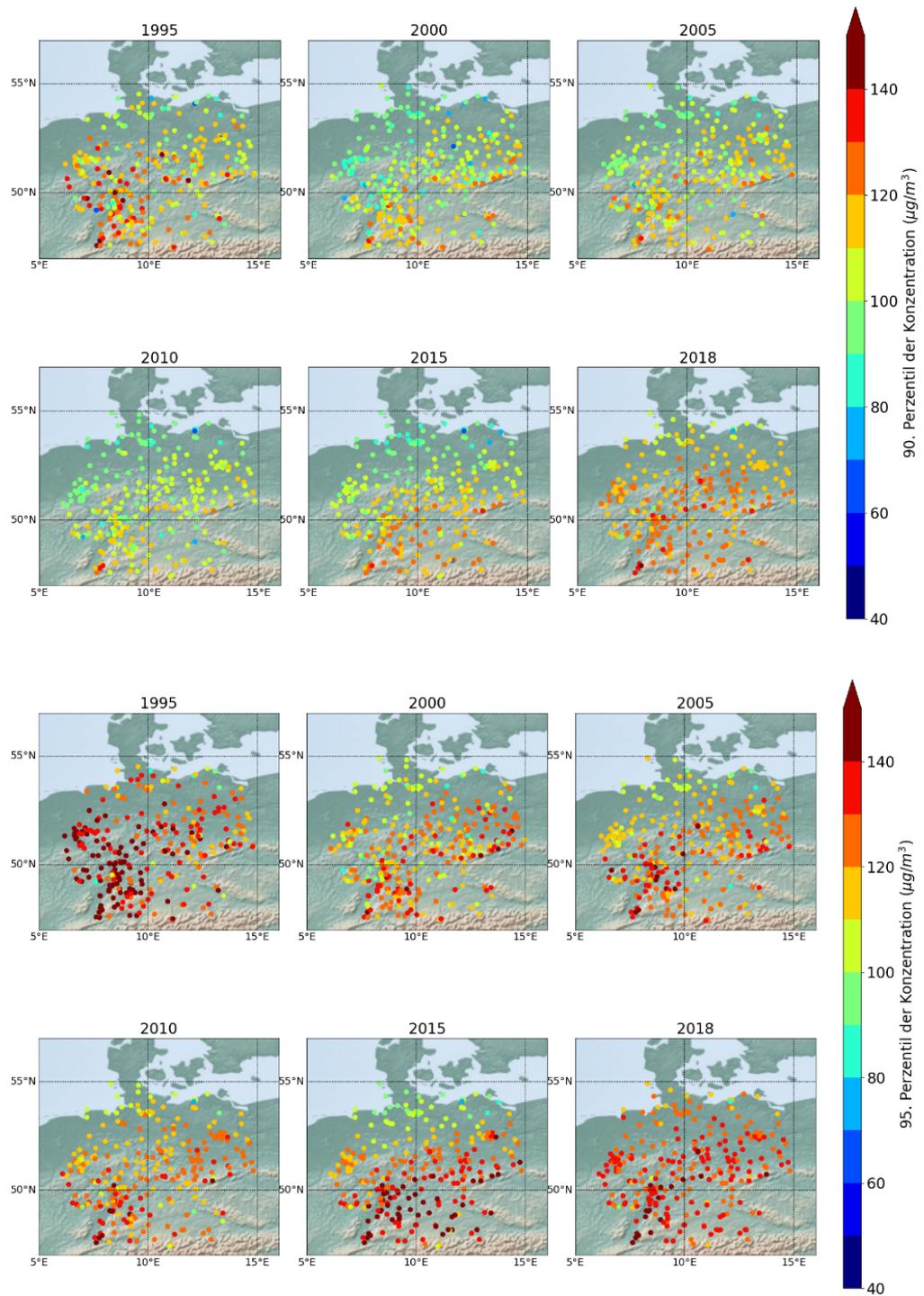
Extremwerte können auch in Form von Perzentilwerten für den Jahresbereich ausgedrückt werden. Das 95. Perzentil bezogen auf das Jahr entspricht dem 90. Perzentil in der Ozonsaison April-September. Die 90. und 95. Perzentile des MDH und des MDA8 sind in Abbildung 16 und Abbildung 17 dargestellt. Für den MDH liegen die 90. Perzentile in einem Bereich von 100 - 140 µg/m³, wobei für die meisten Stationen die Werte im oberen Bereich liegen. Die 95. Perzentile liegen für die meisten Stationen über 140 µg/m³. Weil diese 90. und 95. Perzentilwerte Indikatoren für Extremwerte sind, deuten diese Perzentilwerte darauf hin, dass der untere Wert für die Schwellenwertmetriken für die meisten Standorte und Jahre bei 120 µg/m³ liegen sollte. Es ist ein Nord-Süd-Gefälle festzustellen. Für MDA8 liegen die 90. Perzentile im Bereich von 80 - 130 µg/m³ und die 95. Perzentile im Bereich von 100 - 150 µg/m³ mit einem Nordwest-Südost-Gefälle. Das Muster steht im Einklang mit Abbildung 10 für die Anzahl der Tage mit MDA8 > 120 µg/m³. Die Abbildung zeigt, dass der untere Wert für die MDH-Überschreitung im Bereich 120 - 140 µg/m³ liegen sollte und für die MDA8-Überschreitung im Bereich 100 - 120 µg/m³, je nach Standort, um eine Anzahl von Überschreitungen zu haben, die Ozonspitzenwerte gut wiedergeben in dem Sinne, dass sie nicht zu viel oder zu wenig Ozonüberschreitungen zeigen. Es soll sowohl genügend Ereignisse für das Maschinelle Lernen geben, als auch eine Darstellung von Ozonhöchstwerten ermöglichen.

Abbildung 16: 90. und 95. Perzentil der Ozonkonzentrationen, MDH



Quelle: Eigene Darstellung, TNO

Abbildung 17: 90. und 95. Perzentil der Ozonkonzentrationen, MDA8



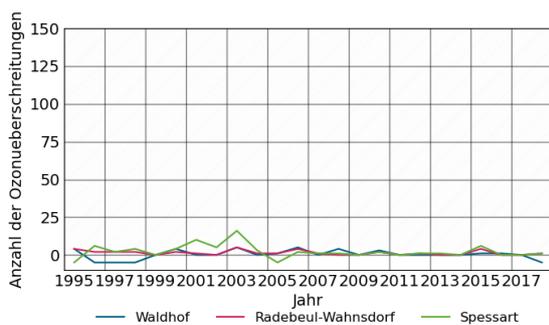
Quelle: Eigene Darstellung, TNO

3.2.7 Zeitliche Variabilität der Anzahl der Überschreitungen und Episoden und deren Länge für ausgewählte Stationen

Um die Trends und die zwischenjährliche Variabilität der Metriken zu untersuchen, wurden drei Stationen ausgewählt, für die die Anzahl der Überschreitungen, die Anzahl der Episoden und die mittlere Episodendauer über die Zeit aufgetragen wurden. Für alle hier gezeigten Zeitreihen (Abbildung 18, Abbildung 19, Abbildung 20, Abbildung 21 und Abbildung 22) stehen die negativen Werte für Jahre, die das Datenqualitätsscreening für die betreffende Station nicht bestanden haben.

Abbildung 18: Anzahl der Überschreitungen der MDH Schwellenwerte

a MDH > 180 µg/m³



b MDH > 120 µg/m³

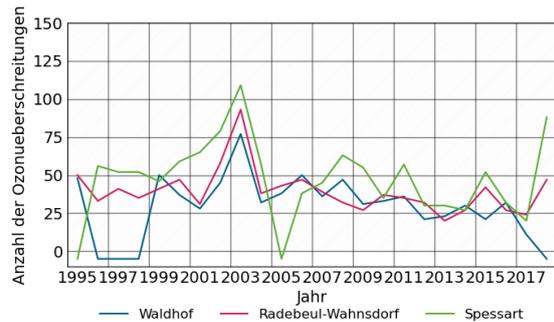
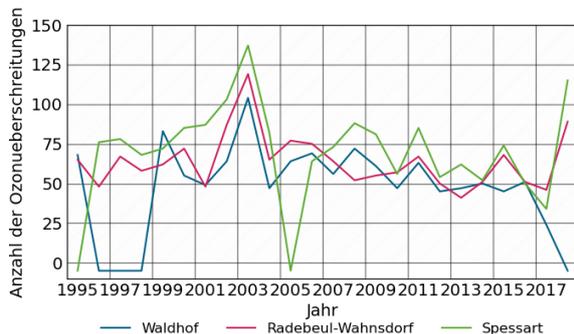
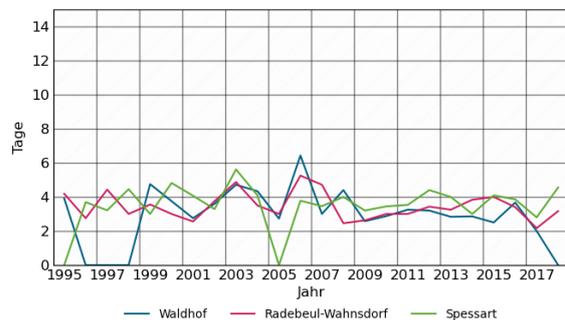


Abbildung 19: Anzahl und mittlere Dauer der Episoden, MDH > 120 µg/m³

a Anzahl



b Mittlere Dauer



Quelle: Eigene Darstellung, TNO

Für MDH > 180 µg/m³ ist die Zahl der Überschreitungen gering, in der Regel weniger als 10. Daher ist dieser Wert nicht als relevanter Schwellenwert geeignet. Für MDH > 120 µg/m³ wird eine gute Korrelation zwischen der Anzahl der Überschreitungen und der Anzahl der Episoden beobachtet (Abbildung 18 und 19a). Die mittlere Dauer der Episoden schwankt zwischen 3 und 7 Tagen (Abbildung 19b).

Für den MDA8 wird eine gute Korrelation zwischen der Anzahl der Überschreitungen und der Anzahl der Episoden für beide Schwellenwerte festgestellt (Abbildung 20 und Abbildung 21). Für MDA8 > 100 µg/m³ gibt es viel mehr Überschreitungen (zwischen 8 und 22) als für MDA8 > 120 µg/m³ (zwischen 3 und 18) (Abbildung 20). Die mittlere Episodendauer ist für diese Metrik zeitlich sehr stabil, mit einer Verschiebung in Abhängigkeit vom Schwellenwert (zwischen 3 und 7 für MDA8 > 100 µg/m³, zwischen 2 und 5 für MDA8 > 120 µg/m³) (Abbildung 21). Diese Beobachtungen zeigen, dass MDA8 > 100 µg/m³ weniger gut mit extremen Konzentrationen verbunden ist. Außerdem scheint die mittlere Episodendauer für MDA8 > 120 µg/m³ stabiler zu sein (Abbildung 22).

Wir haben die Anzahl der Ozonüberschreitungen und -episoden für die 3 oben genannten Stationen und für 3 weitere Stationen in Tabelle 1 zusammengestellt, die ebenfalls städtische/dicht besiedelte Gebiete und andere Regionen in Deutschland repräsentieren.

Tabelle 1: Anzahl der Überschreitungstage für sechs Stationen⁵.

Tabelle 1a: MDH > 120 µg/m³

Jahr	Waldhof	Radebeul-Wahnsdorf	Spessart	Potsdam Zentrum	Eggenstein	Mülheim-Styrum
2000	37	47	59	26	50	21
2015	21	42	52	23	58	27
2018	nv	47	88	37	84	52

Tabelle 1b: MDA8 > 100 µg/m³

Jahr	Waldhof	Radebeul-Wahnsdorf	Spessart	Potsdam Zentrum	Eggenstein	Mülheim-Styrum
2000	55	72	85	42	59	23
2015	45	68	74	30	73	38
2018	nv	89	115	81	100	68

Tabelle 1c: MDA8 > 120 µg/m³

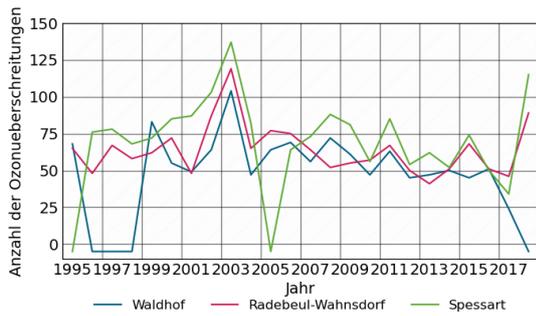
Jahr	Waldhof	Radebeul-Wahnsdorf	Spessart	Potsdam Zentrum	Eggenstein	Mülheim-Styrum
2000	28	30	37	18	30	14
2015	15	31	42	14	37	17
2018	nv	31	63	26	58	31

Quelle: Eigene Darstellung, TNO

⁵ : Für Waldhof wurde das Datenqualitätskriterium für 2018 nicht erfüllt (nicht valide, nv).

Abbildung 20: Anzahl der Überschreitungen der MDA8 Schwellenwerte

a MDA8 > 100 µg/m³



b MDA8 > 120 µg/m³

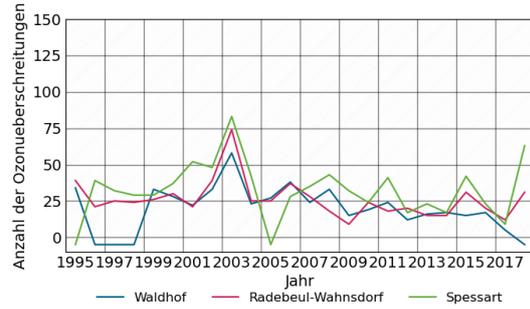
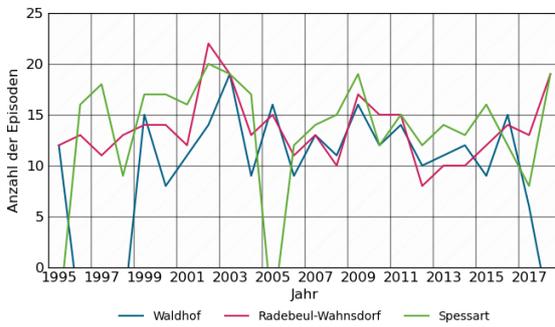


Abbildung 21: Anzahl der Episoden der MDA8 Schwellenwerte

a MDA8 > 100 µg/m³



B MDA8 > 120 µg/m³

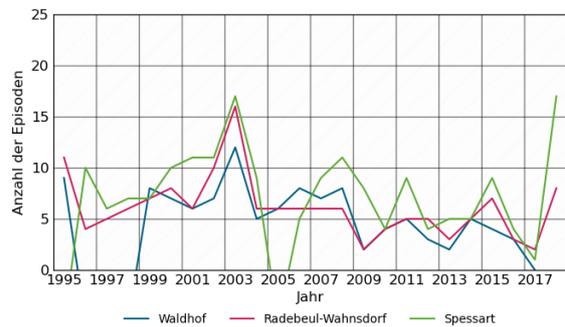
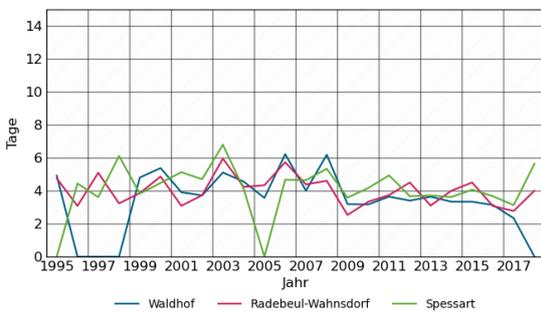


Abbildung 22: Dauer der Episoden der MDA8 Schwellenwerte

a MDA8 > 100 µg/m³



b MDA8 > 120 µg/m³

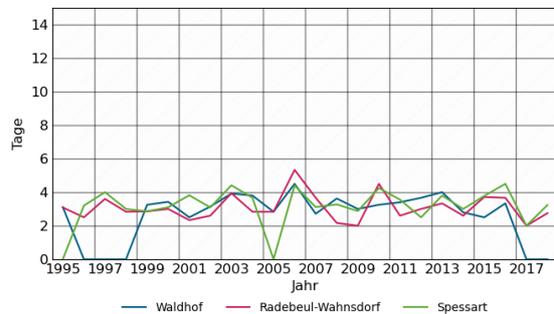


Tabelle 2: Anzahl der Episoden mit ihrer mittleren Dauer und Standardabweichung für sechs Stationen.

Tabelle 2a MDH > 120 µg/m³

Jahr	Waldhof	Radebeul-Wahnsdorf	Spessart	Potsdam Zentrum	Eggenstein	Mülheim-Styrum
2000	8 (3,8±1,7)	14 (3,0±1,2)	11 (4,8±2,9)	5 (4,2±2,5)	11 (3,5±1,8)	7 (2,7±0,9)
2015	6 (2,5±0,8)	8 (4,0±2,3)	11 (4,1±3,7)	7 (2,7±0,5)	12 (4,3±3,4)	7 (2,7±1,4)
2018	nv	12 (3,2±1,7)	18 (4,6±2,9)	8 (3,1±1,5)	15 (5,0±2,7)	10 (4,3±3,2)

Tabelle 2b MDA8 > 100 µg/m³

Jahr	Waldhof	Radebeul-Wahnsdorf	Spessart	Potsdam Zentrum	Eggenstein	Mülheim-Styrum
2000	8 (5,4±4,2)	14 (4,9±2,7)	17 (4,5±2,9)	8 (4,6±2,2)	14 (3,6±1,7)	7 (2,9±0,8)
2015	9 (3,3±1,6)	12 (4,5±3,5)	16 (4,1±3,2)	8 (2,8±0,7)	15 (4,2±3,5)	8 (3,4±2,0)
2018	nv	19 (4,0±2,7)	19 (5,6±4,7)	20 (3,7±3)	18 (5,1±4,1)	14 (4,4±3,4)

Tabelle 2c MDA8 > 120 µg/m³

Jahr	Waldhof	Radebeul-Wahnsdorf	Spessart	Potsdam Zentrum	Eggenstein	Mülheim-Styrum
2000	7 (3,4±2)	8 (3±1)	10 (3,1±1,4)	6 (2,8±1,1)	7 (3,1±1,2)	3 (3,3±0,9)
2015	4 (2,5±0,9)	7 (3,7±2,7)	9 (3,8±2,9)	5 (2,2±0,4)	6 (5±2,9)	3 (3,0±1,4)
2018	nv	8 (2,8±1,3)	17 (3,2±1,3)	7 (2,7±1)	12 (4,1±2,2)	9 (2,9±2,2)

Die Ergebnisse der einzelnen Metriken wurden in Tabelle 3 zusammengefasst. Aufgrund der obigen Analyse schlagen wir als endgültige Metrik für den Ozonschwellenwert, die in diesem Projekt verwendet werden soll, MDA8 > 120 µg/m³ vor. Die wichtigsten Argumente sind:

- ▶ Die Metrik ist weniger empfindlich als der Tageshöchstwert gegenüber Datenlücken in den stündlichen Zeitreihen und ist robuster gegenüber Ausreißern.
- ▶ Die Metrik ist realistisch in Bezug auf die Anzahl der Überschreitungen und entspricht im Großen und Ganzen der Faustformel für das 90. Perzentil.
- ▶ Die Metrik zeigt zeitliche und räumliche Variabilität.
- ▶ Es handelt sich um eine bestehende Metrik, die in der Community bekannt ist.
- ▶ Darüber hinaus ermöglicht diese Metrik die Quantifizierung der Episodenzahl und der Episodendauer.
- ▶ Die sich daraus ergebende Anzahl und Dauer der Episoden ist mit der Anzahl der Überschreitungen korreliert.

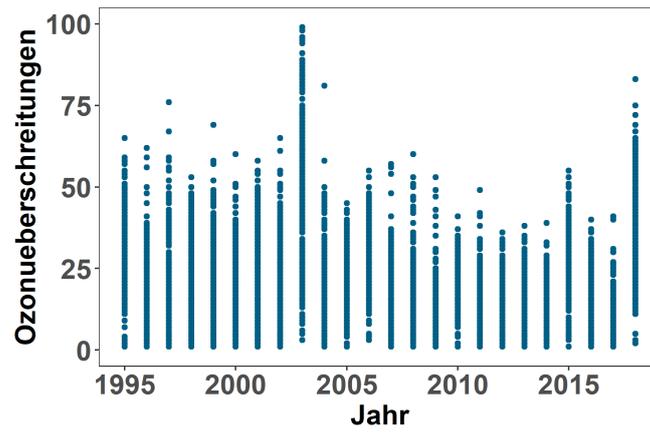
Für das endgültige Training des Algorithmus wurde ein Datensatz mit täglichen MDA8-Werten verwendet, die direkt aus der UBA-Datenbank für die EU-Berichterstattung extrahiert wurden. Die Anzahl der Überschreitungen für alle Jahre und alle gültigen Stationen (ohne Bergstationen) sind in Abbildung 23 gezeigt. Diese Abbildung zeigt einen abnehmenden Trend bei der Anzahl der Überschreitungen, aber auch eine hohe Anzahl von Überschreitungen für 2003 und 2018. Die hohe Zahl der Überschreitungen im Jahr 2018 spiegelte sich nicht in den Überschreitungen der täglichen maximalen Stundenkonzentration von 180 µg/m³ in Abbildung 4 wider.

Um in diesem Bericht Missverständnissen vorzubeugen, werden die Schwellenwertüberschreitungen, wie sie hier definiert sind (MDA8 > 120 µg/m³), im weiteren Text als „Ozonüberschreitung“ bezeichnet.

Tabelle 3: Leistungen der verschiedenen Metriken für verschiedene Kriterien

Metrik	Anzahl Überschreitungen > 10	Anzahl Überschreitungen < 80	Interannuelle Variabilität	Regionale Variabilität	Robust gegenüber Datenlücken	Erfassung Länge und Anzahl der Episoden
MDH >180 µg/m ³	-	y	-	-	-	-
MDH > 120 µg/m ³	y	-	y	y	-	y
MDA8 > 100 µg/m ³	y	-	-	-	y	y
MDA8 > 120 µg/m ³	y	y	y	y	y	y
90., 95. Perzentile	-	-	y	y	y	-

Abbildung 23: Anzahl der Ozonüberschreitungen von MDA8 > 120 µg/m³, alle Stationen und alle Jahre



Quelle: Eigene Darstellung, TNO

4 Für Ozon relevante meteorologische Variablen

Die bodennahen Ozonkonzentrationen weisen starke Schwankungen auf: Schwankungen der Ozonkonzentration von Tag zu Tag und einen starken Tagesgang, der mit den meteorologischen Bedingungen und den Emissionen von Ozonvorläuferstoffen zusammenhängt. Zu den meteorologischen Bedingungen, die die Bildung von bodennahem Ozon fördern, gehören: hohe Lufttemperatur, intensive Sonneneinstrahlung und stagnierende Zirkulationsmuster (Ordóñez et al., 2005; Pearce et al., 2011, Otero et al., 2016). Die Mechanismen werden hier kurz erläutert. Auf längeren Zeitskalen zeigen die Ozonkonzentrationen saisonale Schwankungen mit niedrigeren Konzentrationen im Winter, wo der Ferntransport einen größeren Einfluss hat als die lokale Ozonbildung. Auch in größeren Höhen sind die Ozonkonzentrationen weniger schwankend, da die Emissionen von Vorläufersubstanzen meist bodennah stattfinden. Während Hitzewellen steigen die Ozonspitzenkonzentrationen oft an aufeinanderfolgenden Tagen an. Auch die Persistenz ist ein wichtiger Faktor.

Die Temperatur gilt als Schlüsselparameter für Ozon in Mitteleuropa, insbesondere während Ozonspitzenkonzentrationen, da sich Peroxyacetylnitrate (PAN) bei hohen Temperaturen thermisch zersetzen und dadurch mehr NO_x entsteht. Die thermische Zersetzung von PANs erhöht die Ozonbildungsrate (Sillman und Samson, 1995; Vogel et al., 1999; Bärtsch-Ritter et al., 2004). Zunehmende Trockenheit, die oft mit hohen Temperaturen einhergeht, kann zu verringerten Depositionsgeschwindigkeiten führen, wenn Pflanzen ihre Spaltöffnungen schließen. Dies führt zu höheren Ozonspitzenkonzentrationen (Andersson und Engardt, 2010). Außerdem führen höhere Temperaturen zu erhöhten biogenen VOC-Emissionen, die die Ozonproduktion verstärken (Andersson und Engardt, 2010, Doherty et al., 2013, Jacob und Winner, 2009).

Hohe Luftfeuchtigkeit in der Atmosphäre kann die Lebensdauer von Ozon verringern, was zu niedrigeren Hintergrundkonzentrationen führt. Gleichzeitig erhöht sie die Produktion von Hydroxylradikalen (OH), was zu höheren Ozonkonzentrationen in stark verschmutzten Umgebungen wie Städten und durch Verkehr beeinflusste Standorte führt (Vogel et al., 1999).

Windgeschwindigkeit und -richtung hängen mit der Mischung und dem Transport von sauberer oder verschmutzter Luft zusammen. Auch die Stabilität der Atmosphäre und die **Höhe der Grenzschicht** bestimmen, inwieweit Vorläuferemissionen und erzeugtes Ozon verdünnt werden.

Die einfallende kurzweilige Strahlung und die **Wolkenbedeckung** sind teilweise miteinander korreliert und bestimmen die Photolyseraten und die Ozonproduktion.

Die Temperatur weist immer die stärkste Korrelation auf, ist aber mit anderen Variablen untrennbar verbunden. Die Mechanismen des Beitrags der Variablen wurden oben beschrieben. Um ihren relativen Beitrag zu untersuchen, wurden multivariate lineare Regression (Otero et al. 2016, 2018) und verallgemeinerte additive Modelle (Boleti et al., 2018, 2019) auf beobachtete und modellierte Ergebnisse angewendet. Die wichtigsten zehn Variablen waren:

1. Tägliche Höchsttemperatur (2 m), tägliche Durchschnittstemperatur
2. Relative/spezifische Luftfeuchtigkeit (Tagesdurchschnitt)
3. Abwärts gerichtete kurzweilige Sonnenstrahlung
4. Windgeschwindigkeit

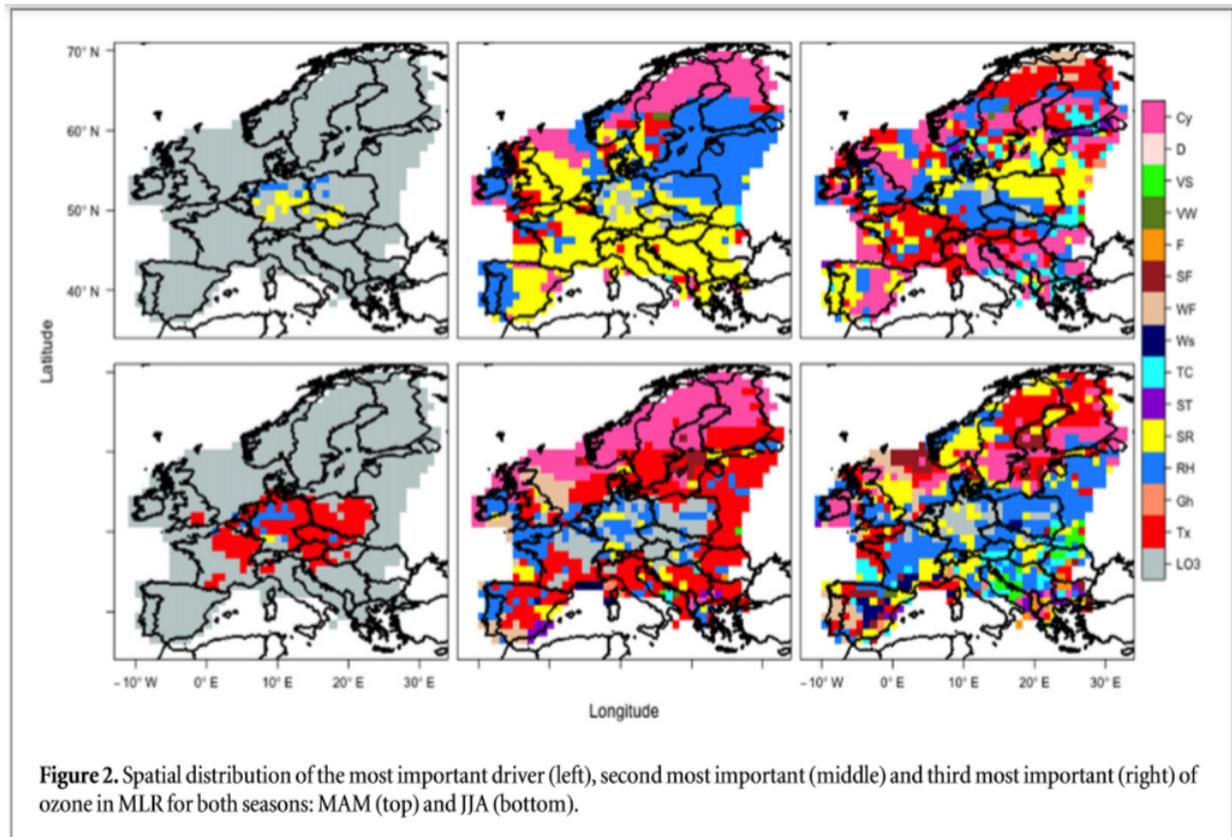
5. Windrichtung (oder Nord-Süd-“Surface stress“, West-Ost-“Surface stress“, Ost-/Westströmung)
6. Grenzschichthöhe
7. Konvektiv verfügbare potenzielle Energie (Ende des Vormittags)
8. Wolkenbedeckung
9. Luftdruck auf Meereshöhe (Änderungen)
10. Geopotentielle Höhe

Die Punkte 9 und 10 sind repräsentativ für synoptische Situationen (Großwetterlage). Es wurden mehrere alternative Klassifizierungen von synoptischen Situationen gefunden, wobei einige der oben genannten Parameter zu einem einzigen Parameter kombiniert wurden. Bedingungen auf verschiedenen Höhenniveaus wurden verwendet, um eine Klassifizierung der Wetterlagen zu bestimmen: 950 und 500 hPa (Pavlik et al., 2016), 850 hPa (Hertig, 2020). Feng et al. (2019) fanden in einer Studie über Hangzhou (China) eine starke Beziehung zwischen Ozon und Taupunktdefizit.

Welche meteorologischen Variablen den größten Beitrag leisten, hängt von der Jahreszeit und dem Standort ab (Abbildung 24). Für Deutschland sind dieser Abbildung zufolge die 24-Stunden-Verzögerung von Ozon (Persistenz), die relative Luftfeuchtigkeit und die solare Einstrahlung die wichtigsten Faktoren im Frühjahr. Im Sommer scheint die tägliche Höchsttemperatur der wichtigste Faktor für die Ozonspitzenkonzentrationen zu sein, während die anderen für das Frühjahr genannten Faktoren weiterhin relevant sind.

Während Hitzewellen können die Beziehungen etwas anders sein, da sich die Ozonkonzentrationen in solchen Perioden normalerweise erhöhen. Varotsos et al. (2019) analysierten die europäischen Hitzewellen von 2003 und 2014 und stellten fest, dass die Beziehung zwischen Ozon und Temperatur während Hitzewellen um etwa 1 ppb/°C höher ist. Dies sollte beim Training von Algorithmen berücksichtigt werden.

Abbildung 24: Analyse der wichtigsten meteorologischen Einflussfaktoren für März-April-Mai und für Juni-Juli-August (Otero et al., 2016). Siehe Text für die wichtigsten Treiber für Deutschland. Die wichtigsten Faktoren für Deutschland sind Persistenz (LO3, grau), relative Feuchte (RH, blau), Sonneneinstrahlung (SR, gelb) und Tagesmaximum der Temperatur (Tx, rot)



Quelle: Otero et al. (2016)

4.1 KLENOS-Projekt

Im UBA-Projekt KLENOS (Pavlik et al., 2016) wurden die Auswirkungen des Klimawandels auf die Ozon- und PM10-Konzentrationen anhand von Modellsimulationen mit an Klimamodelle gekoppelten Chemie-Transport-Modellen untersucht. Im Rahmen dieses Projektes wurden die aktuellen Zusammenhänge zwischen hohen Ozonkonzentrationen und meteorologischen Bedingungen analysiert. Die Eingangsdaten für diese Analyse waren:

- ▶ Ozonmessungen, Datensatz vom UBA
- ▶ Datensatz mit Zirkulationsmuster-Klassifizierung (Wetterlagenklassen) vom DWD

Die Klassifizierung der Zirkulationsmuster definiert 40 verschiedene Klassen, und diese Klassifizierung basiert auf:

- ▶ großräumige Anströmungsrichtung (Nordost, Südost, Südwest, Nordwest, undefiniert)
- ▶ Zirkulationsmuster (zyklonal/antizyklonal) auf 950 und 500 hPa
- ▶ Feuchtegehalt der Atmosphäre (feucht/trocken).

Für Deutschland wurde nach Stationstyp und geographischen Regionen unterschieden. Basierend auf der Verteilung der Überschreitungen (täglicher 8-Stunden-Mittelwert $>120 \mu\text{g}/\text{m}^3$) wurden drei Regionen festgelegt.

Ein Drittel der Überschreitungstage entfiel auf die Klassifikationen SWAAF (Südwestliche Anströmung, antizyklonal bei 950 und 500 hPa, feucht) und SWZAF (Südwestliche Anströmung, zyklonal bei 950 hPa, antizyklonal bei 500 hPa, feucht). Dies gilt insbesondere für den Osten Deutschlands (40 % der Überschreitungen), während im Westen Deutschlands die antizyklonalen Bedingungen in Oberflächennähe eine größere Rolle spielen.

4.2 Studie für Bayern

Eine weitere Studie von besonderer Relevanz ist die Untersuchung eines Zusammenhangs zwischen großräumigen meteorologischen Prädiktoren mit Ozonüberschreitungen und mit Tageshöchsttemperaturen für Bayern (Hertig, 2020). Datenquellen sind Beobachtungen aus dem offiziellen Luftqualitätsnetz des Bundeslandes Bayern und offiziellen meteorologische Stationen. Da die Stationen für die Luftqualität und die Meteorologie nicht an denselben Orten liegen, wurde die nächstgelegene meteorologische Station ohne weitere Interpolationstechniken verwendet. Für städtische Gebiete wurde angenommen, dass es keine Auswirkungen der städtischen Wärmeinsel gibt, da dies hauptsächlich für die nächtliche Temperatur gilt. Neben den beobachteten meteorologischen Variablen wurden auch Zirkulationsklassifizierungen verwendet. Zu diesem Zweck wurden ERA5-Tagesmittelwerte des Luftdrucks auf Meereshöhe, der Temperatur und der relativen Luftfeuchtigkeit auf 850hPa sowie Luftströmungsindizes (Richtung, Intensität, Verwirblung) zu Klassifizierungsparametern kombiniert. Diese unterscheiden sich geringfügig von denen, die in KLENOS verwendet werden.

Statistische Modelle wurden für die Tageshöchsttemperatur (Prädiktor) und Tageshöchstwerte der Ozonkonzentrationen (Prädiktand) separat auf der Grundlage desselben Satzes von meteorologischen Variablen unter Verwendung von verallgemeinerten additiven Modellen (Generalized Additive Models, GAMs, siehe 5.1.3 zur Erläuterung) erstellt. Die Verwendung der Tageshöchstwerte der Ozonkonzentrationen als Prädiktand führte zu einer besseren Leistung der statistischen Modelle als die maximalen 8-Stunden-Mittelwerte oder die Tagesmittelwerte. Überschreitungen von $100 \mu\text{g}/\text{m}^3$ traten von April bis September häufig auf (40 % der Tage), Überschreitungen von $180 \mu\text{g}/\text{m}^3$ weniger als 1 %, der Wert von $120 \mu\text{g}/\text{m}^3$ wurde an etwa 15 % der Tage überschritten. Diese Überschreitungen fielen häufig mit Tageshöchsttemperaturen von über $30 \text{ }^\circ\text{C}$ zusammen. Für Ozon funktionierte die Gamma- oder Normalverteilung für die meisten Stationen besser als die Alternativen (Lognormal-, Weibull-Verteilung). Auf der Grundlage des GAM wurde eine nichtlineare Beziehung zwischen den Tageshöchstwerten der Ozonkonzentrationen und der Tageshöchsttemperaturen abgeleitet. Schließlich wurde das trainierte Modell mit meteorologischen Variablen aus Erdsystemmodellldaten unter dem Klimaszenario RCP8.5 mit statistischem Downscaling verwendet, um die Auswirkungen des Klimawandels auf die Ozonkonzentrationen zu untersuchen.

4.3 Verfügbare Datensätze

In dieser Studie wird die Entwicklung eines statistischen Modells erzielt, das das Auftreten von Ozonepisoden (Anzahl, Dauer) auf Jahresbasis mit einer begrenzten Anzahl jährlich aggregierter Klimakennzahlen für Deutschland in Beziehung setzt. Die Klimakennzahlen, die letztlich in dieser Studie verwendet werden, sollten aus vorhandenen gerasterten Datensätzen stammen.

Die wichtigsten Datenquellen für das aktuelle Projekt sind die Daten des European Climate Assessment & Dataset (ECAD, <https://www.ecad.eu/>) und die Daten aus dem ReKliEs-Projekt (Regionale Klimaprojektionen Ensemble für Deutschland, ReKliEs-De, 2017). Die Daten decken den gesamten geografischen Bereich als Raster ab, was ein großer Vorteil gegenüber Messstationsdaten nationaler Wetterdienste ist. Darüber hinaus enthalten sie fertig aufbereitete Klimakennzahlen, die weithin akzeptiert sind.

4.3.1 ECAD-Daten

Die ECAD-Daten bilden die Grundlage für die Klimadaten für Europa der World Meteorological Organization (WMO). Die E-OBS-Daten aus ECAD sind ein europaweiter gerasterter Datensatz, der auf qualitativ hochwertigen meteorologischen Beobachtungen aus dem Jahr 1950 basiert. E-OBS wird als Ensemble-Datensatz geliefert und ist auf einem regelmäßigen 0,1- und 0,25-Grad-Gitter für die Elemente Tagesmitteltemperatur **TG**, Tagesminimumtemperatur **TN**, Tageshöchsttemperatur **TX**, Tagesniederschlagssumme **RR**, täglich gemittelter Meeresspiegeldruck **PP** und Globalstrahlung **QQ** verfügbar. Sie decken folgendes Gebiet ab: 25° N-71,5° N x 25° W-45° E. Die Datendateien liegen im NetCDF-4-Format vor.

Zusätzlich zu den täglichen Daten werden auch jährliche Daten (Klimakennzahlen) angegeben. Das aktuelle Projekt möchte auf solchen jährlichen Klimakennzahlen basieren. Allerdings sind nicht alle von diesen Klimakennzahlen relevant für Ozonkonzentrationen. Aufgrund der direkten Beziehung zu Temperatur, Strahlung, Feuchtigkeit und Zirkulationsmustern könnten die folgenden Indikatoren relevant sein:

- ▶ Anzahl der Sommertage: Tage mit Tageshöchsttemperatur $TX > 25\text{ °C}$,
- ▶ Anzahl der tropischen Nächte: Tage mit Tagesminimumtemperatur $TN > 20\text{ °C}$,
- ▶ Anzahl der warmen Tage: Tage mit Tagesmitteltemperatur $TG > 90.$ Perzentil der Tagesmitteltemperaturen im Referenzzeitraum,
- ▶ Anzahl der warmen Nächte: Tage mit Tagesminimumtemperatur $TN > 90.$ Perzentil der Tagesminimumtemperaturen im Referenzzeitraum
- ▶ Anzahl der warmen Tage: Tage mit Tageshöchsttemperatur $> 90.$ Perzentil der Tageshöchsttemperatur im Referenzzeitraum
- ▶ maximale Anzahl von aufeinanderfolgenden Sommertagen
- ▶ Anzahl der Tage mit östlichen Winden ($45^\circ < \text{Windrichtung} < 135^\circ$)
- ▶ Anzahl der Tage mit Westwind ($225^\circ < \text{Windrichtung} < 315^\circ$)
- ▶ Anzahl der Tage mit südlichen Winden ($135^\circ < \text{Windrichtung} < 225^\circ$)
- ▶ Anzahl der windstillen Tage (mittlere Windgeschwindigkeit $< 2\text{ m/s}$)
- ▶ Anzahl der warmen, trockenen Tage: Tage mit Tagesmitteltemperatur $> 75.$ Perzentil der Tagesmitteltemperatur und Tagesniederschlagssumme $< 25.$ Perzentil der Tagesniederschlagssumme im Referenzzeitraum

- ▶ Anzahl der feuchtwarmen Tage: Tage mit Tagesmitteltemperatur > 75. Perzentil der Tagesmitteltemperatur und RR > 75. Perzentil der Tagesniederschlagssumme im Referenzzeitraum
- ▶ Anzahl der meist sonnigen Tage: Anzahl der Tage mit einer Bewölkung ≤ 2 Oktas
- ▶ Sonnenscheindauer insgesamt
- ▶ Tourismus-Klima-Index/Universal Thermal Climate Index: Kombination von Temperatur, Wind, Luftfeuchtigkeit und Strahlung zur Bestimmung der Hitzebelastung des Menschen, siehe www.utci.org)

4.3.2 ReKliEs-De

Eine weitere Quelle für Klimadaten stammt aus den ReKliEs-Projekten, in denen Klimasimulationen durchgeführt wurden. Um die Datenmenge für diese Klimasimulationen zu komprimieren, wurden jährliche Kennzahlen gespeichert, wie die jährlichen Kennzahlen in der ECAD-Datenbank. Die potenziell relevanten Klimakennzahlen sind der Tabelle 4.1 im ReKliEs-Bericht entnommen. Wo sich die Indikatoren von den ECAD-Parametern unterscheiden, werden sie beschrieben:

- ▶ Anzahl der Sommertage
- ▶ Anzahl der Hitzetage (Tage mit Tageshöchsttemperatur > 30 °C)
- ▶ Anzahl der tropischen Nächte
- ▶ Anteil warmer Tage pro Jahr (Definition ETCCDI: Prozent der Tage, an denen die Maximaltemperatur $T_{max} > T_{x90}$ ist. T_{x90} ist das 90. Perzentil der Tagesmaximaltemperatur im Referenzzeitraum (1971 – 2000))
- ▶ Dauer der Warmzeiten (Definition ETCCDI: Summe der zusammenhängenden Tage (mindestens 6), an denen die Tagesmaximaltemperatur $T_{max} > T_{x90}$ ist. T_{x90} ist das 90. Perzentil der Tagesmaximaltemperatur im Referenzzeitraum (1971-2000))
- ▶ Mittlere bodennahe Windgeschwindigkeit (Monatsmittel)
- ▶ Mittlere Globalstrahlung am Boden W/m^2 jährlich, saisonal und monatlich

Beide Datensätze enthalten auch eine Kennzahl für die Andauer von Wärmeperioden. Diese Kennzahl bezieht sich auf Perioden im Jahr, die wärmer sind als der Durchschnitt für diesen Kalenderzeitraum. Es werden so auch Perioden gezählt, die im Winter wärmer sind als üblich, aber damit nicht zu Ozonüberschreitungen beitragen und deshalb hier nicht relevant sind. Die Dauer der Wärmeperiode und die Perzentile für warme Tage sind in ECAD und ReKlies-DE ähnlich definiert, jedoch wird in ECAD das Tagesmittel und in ReKlies-DE das Tagesmaximum verwendet.

4.3.3 Bewertung der Klimasimulationen im Projekt KLENOS

Die im Projekt KLENOS erstellten Klimasimulationen werden auf ihre Anwendbarkeit in diesem Projekt überprüft. Außerdem diskutieren wir kurz die Klimamodellsimulationen in Bezug auf das beobachtete Klima.

Die KLENOS-Daten wurden mit den folgenden Datensätzen ausgewertet:

- ▶ Temperatur und Niederschlag von ECA&D (E-OBS) (von [www. Ecad.eu](http://www.Ecad.eu)),
- ▶ HYRAS-Daten für die Jahresmitteltemperatur (Frick et al., 2014), Auflösung 1 km (hier nicht relevant),
- ▶ Niederschlag aus REGNIE-Datensatz (DWD, REGNIE, 2013), Auflösung 1 km (hier nicht relevant),
- ▶ Windgeschwindigkeitsdaten sind nicht als Raster verfügbar, verfügbare Beobachtungen wurden mit den Modellwerten im nächstgelegenen Rasterpunkt und den umliegenden Rasterpunkten verglichen.

Bei den Klimasimulationen im Rahmen des KLENOS-Projekts wurden für das COSMO-Klimamodell für den Zeitraum 1976-2005 Abweichungen festgestellt, die vom treibenden globalen Modell (ERA-Interim Reanalysis oder ECHAM6) abhängen. Für Deutschland wurden die Jahresmitteltemperaturen um 0,8-1 K unterschätzt, wobei die Simulationen, die von ERA-Interim angetrieben wurden, im Sommer besser abschnitten und die Simulationen, die von ECHAM6 angetrieben wurden, im Winter besser abschnitten. Der Niederschlag wurde bei beiden Modellen um 22 – 56 % überschätzt. Bei der Windgeschwindigkeit schnitten die Modelle in den flachen Gebieten im Osten Deutschlands am besten ab, wobei die ECHAM6-gesteuerte Simulation tendenziell größere Unterschiede aufwies. Für einige Anwendungen können aus solchen Klimasimulationen fehlerkorrigierte Zeitreihen erzeugt werden. Wenn nur die Temperatur oder der Niederschlag benötigt werden, ist dies machbar, aber wenn mehrere zusammenhängende Variablen benötigt werden (Temperatur, Luftfeuchtigkeit, Windgeschwindigkeit), wie bei der Bewertung der Luftqualität, ist dies nicht möglich, da sie auf nichtlineare Weise miteinander verbunden sind.

4.4 Auswahl der meteorologischen Daten und der Klimakennzahlen

Die meteorologischen Variablen, die nachweislich den größten Einfluss auf die Ozonhöchstwerte haben, sind die tägliche Höchsttemperatur, die relative Luftfeuchtigkeit, die einfallende Sonnenstrahlung und eine Klassifizierung der synoptischen Situation. Es ist jedoch zu bedenken, dass aus Klimadatensätzen nicht alle diese Parameter zur Verfügung stehen. Das Weiteren sollen nur die Parameter ausgewählt werden, die als Jahreszahlen verfügbar sind, wie z. B. die Anzahl der Sommertage.

Wir haben mehrere Datensätze mit meteorologischen Daten aus Beobachtungen und Reanalysedaten geprüft, um den nützlichsten Datensatz auszuwählen. Die Merkmale verschiedener Optionen werden in Tabelle 4 aufgelistet. Die E-OBS-Daten basieren auf konsistent gerasterten Beobachtungsdaten und werden z. B. in Hertig (2020) und KLENOS verwendet, enthalten aber nicht alle relevanten meteorologischen Variablen. Auf die DWD-Beobachtungen kann zugegriffen werden, und sie enthalten alle relevanten meteorologischen Variablen. Allerdings befinden sich die meteorologischen Stationen nicht am Ort der Ozonmessungen. Daher wäre eine Interpolation erforderlich, oder es müsste die nächstgelegene meteorologische Station für jede Ozonstation genommen werden (wie von Hertig (2020) für die Windgeschwindigkeit angewandt). COSMO-REA6⁶ (Bollmeyer et al., 2015) bietet einen konsistenten Satz aller verfügbaren meteorologischen Daten, basierend auf Modelldaten mit

⁶ <https://reanalysis.meteo.uni-bonn.de/?COSMO-REA6>

Datenassimilation, um Beobachtungen von meteorologischen Stationen einzubeziehen. Die Reanalyse ERA5⁷ (Hersbach et al., 2020) ist vom Ansatz her mit COSMO-REA6 vergleichbar, deckt aber einen längeren Zeitraum ab und bietet eine globale Abdeckung, wenn auch mit einer größeren Auflösung.

Tabelle 4: Datenquellen meteorologische Variablen

Variable	E-OBS (täglich), 1km	DWD-CDC Beobachtungen	Variabel	E-OBS (täglich), 1km
Temperatur	TMax, TMittelwert, TMin	y	y	y
Relative Luftfeuchtigkeit	-	y	y	y
Niederschlag	täglich	y	y	y
Einfallende Sonnenstrahlung	täglich	Dauer des Sonnenscheindauer	y	y
Windgeschwindigkeit und -richtung	-	y	y	y
Höhe der Grenzschicht	-	-	y	y
Wolkendecke	-	octa	y	y
Bodendruck	täglich	y	y	y

Die COSMO-Reanalysen mit einer Auflösung von 6 km (COSMO-REA6) wurden als der beste Kompromiss zwischen Auflösung, Verfügbarkeit und Konsistenz der relevanten Parameter ausgewählt. Die verwendeten Variablen waren 2m-Temperatur, relative Feuchte, einfallende kurzwellige Strahlung und 10m Windgeschwindigkeit (u- und v-Komponenten). Folgende Klimadaten wurden verarbeitet, um Daten für jeden Messort zu erhalten:

- ▶ Tagesmaximum Temperatur
- ▶ Tagesdurchschnitt relative Luftfeuchte
- ▶ Summe der täglich eingehenden kurzwelligen Gesamtstrahlung
- ▶ Tagesdurchschnitt Windgeschwindigkeit
- ▶ Tägliche vorherrschende Windrichtung
- ▶ Tägliche Niederschlagssumme

Diese Parameter haben eine bekannte Beziehung zu Ozonspitzenkonzentrationen und können auf die gleiche Ebene aggregiert werden wie die jährlich aggregierten Daten (Klimakennzahlen), die von ECAD und REKLIES zur Verfügung gestellt werden. Diese sind:

- ▶ Anzahl der Sommer- und Hitzetage,

⁷ <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>

- ▶ 90. Perzentil der Temperatur
- ▶ Index der Dauer der Wärmeperiode (warm spell duration index)
- ▶ Tage mit östlichen Winden, Tage mit westlichen Winden
- ▶ Tage mit südlichen Winden, Tage mit nördlichen Winden

Andere Parameter wie die Höhe der Grenzschicht, die geopotentielle Höhe und die potentielle Konvektionsenergie stehen mit Ozon in Zusammenhang, sind aber nicht standardmäßig in jedem meteorologischen Modell verfügbar und lassen sich nur sehr schwer mit Parametern aus aggregierten Daten von Klimasimulationen in Beziehung setzen.

Wir wissen, dass die relative Luftfeuchte für Ozon wichtig ist. Von allen Indikatoren bezieht nur der Universal Thermal Climate Index (UTCI) die relative Luftfeuchte direkt mit ein. Der UTCI ist jedoch ein relativ komplexer Indikator, der auf die thermische Behaglichkeit des Menschen abzielt und auch den Windchill einbezieht, so dass dieser Indikator im Projekt nicht verwendet wird. Daher wurde hier ein eigener Indikator definiert, um die Rolle der relativen Luftfeuchte zu testen.

- ▶ RF_feucht: Tage mit einer mittleren täglichen relativen Luftfeuchte von > 80 % werden als feucht angesehen,
- ▶ RF_trocken: Tage mit einer mittleren täglichen relativen Luftfeuchte von < 60 % werden als trocken betrachtet

Eine Größe, die mit Feuchte und Temperatur zusammenhängt, ist die Taupunkttemperatur. Allerdings wird dieselbe Taupunkttemperatur bei verschiedenen Kombinationen von Temperatur und relativer Luftfeuchte erreicht und sie wird nicht als Klimakennzahl berechnet. Außerdem sollte ein Algorithmus für Maschinelles Lernen wie GBM oder RF in der Lage sein, diese Beziehung und ihre Auswirkungen auf die Ozonüberschreitung wiederzugeben, wenn sie besteht. Daher wurde diese Größe nicht als Variable verwendet.

Niederschlag ist weniger gut mit Ozon korreliert als Luftfeuchte. Die Variablen "warme, trockene Tage" und "warme, feuchte Tage" sind zwar Klimakennzahlen, basieren aber auf der Tagesmitteltemperatur und nicht auf der Tageshöchsttemperatur, so dass wir erwarten, dass Ozon weniger empfindlich darauf reagiert.

Die Klimakennzahlen werden für ein Kalenderjahr ermittelt, während für die Ozonepisoden nur die Saison April-September relevant ist. Nur für die temperaturbezogenen Variablen sind die Klimakennzahlen für den Sommer repräsentativ. Um das Potenzial der Klimakennzahlen für Ozonüberschreitungen zu untersuchen, wurden alle Klimakennzahlen für den Zeitraum April-September berechnet.

Zur Bestimmung der Episodenlänge wurden für die Variablen, die gut mit den Ozonüberschreitungen korrelieren (tropische Nächte, Sommer- und Hitzetage), die maximale Anzahl aufeinanderfolgender Tage berechnet.

Die Klimakennzahlen, die im endgültigen Datensatz verwendet wurden, und ihre Definition sind in Tabelle 5 zusammengefasst.

Tabelle 5: Klimakennzahlen, die in das Training der ML-Algorithmen einbezogen wurden

Klimakennzahl	Verwendete meteorologische Variablen	Benennung der Klimakennzahlen wie in Abbildungen	Berechnung aus COSMO Reanalyse 6km 1995-heute, stündlich	Direkt verfügbar in ReKlies	Direkt verfügbar in ECAD
Anzahl der Sommertage	Tagesmaximum der Temperatur (T_{\max})	Sommertage	Tage mit $T_{\max} > 25 \text{ °C}$	y	y
Anzahl der Hitzetage	Tagesmaximum der Temperatur (T_{\max})	Hitzetage	Tage mit $T_{\max} > 30 \text{ °C}$	y	n
Anzahl der tropischen Nächte	Tagesminimum Temperatur (T_{\min})	Tropennächte	Tage mit $T_{\min} > 20 \text{ °C}$	y	y
Anzahl der Trockentage	Tagessumme Niederschlag	Trockentage	Tage mit Niederschlag $< 1 \text{ mm}$	y	y
Anzahl der Tage mit RF $< 60 \%$ (RF_trocken)	Tagesdurchschnitt relative Feuchte (RF)	RF_trocken	Tage mit RF $< 60\%$	-	-
Anzahl der Tage mit RF $> 80 \%$ (RF_feucht)	Tagesdurchschnitt relative Feuchte (RF)	RF_feucht	Tage mit RF $> 80 \%$	-	-
Anzahl der Tage mit östlicher Strömung	Windrichtung (WindRI)	Wind_Ost	$45^\circ < \text{WindRI} \leq 135^\circ$	n	y
Anzahl der Tage mit westlicher Strömung	Windrichtung (WindRI)	Wind_West	$225^\circ < \text{WindRI} \leq 315^\circ$	n	y
Anzahl der Tage mit südlicher Strömung	Windrichtung (WindRI)	Wind_Süd	$135^\circ < \text{WindRI} \leq 225^\circ$	n	y
Anzahl der Tage mit schwachem Wind	10m-Windgeschwindigkeit (WindGE)	Schwachwindtage	WindGE $< 2 \text{ m/s}$	n	y
Mittlere einfallende Strahlung	Strahlung	Strahlung	Kumulative stündliche Mittelwerte der Strahlung	(jährlich, monatlich, saisonal)	n

4.5 Emissionsdaten

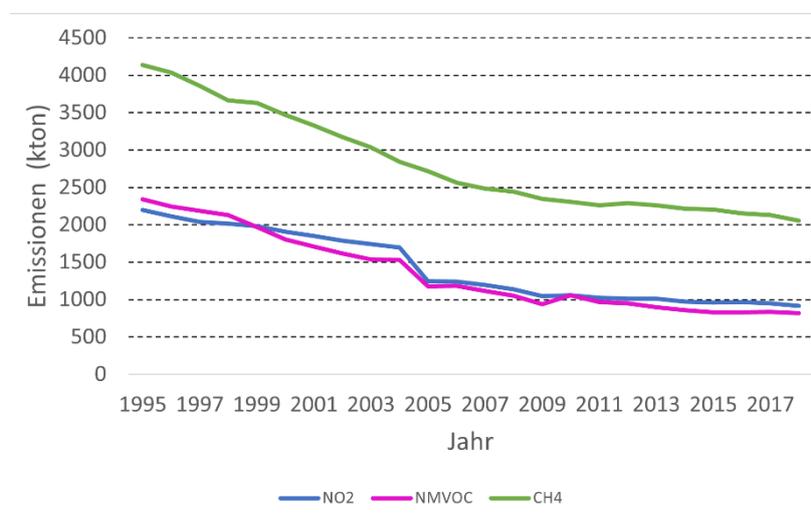
Natürliche Emissionen von Ozonvorläuferstoffen werden unter anderem durch die Meteorologie gesteuert. Sie werden daher implizit durch die Meteorologie und die Ozonkonzentrationen berücksichtigt.

Trends bei anthropogenen Emissionen können auf verschiedene Weise berücksichtigt werden. Die einfachste Möglichkeit besteht darin, einen Indikator für die Emissionen als Variable in den Algorithmus für Maschinelles Lernen einzubeziehen. Dies hat den Vorteil, dass diese Variable dann direkt in das Tool zur Vorhersage von Ozonepisoden verwendet werden kann, so dass die Auswirkungen von Änderungen der Klimakennzahlen und Änderungen der Emissionen unterscheidbar sind.

In diesem Projekt wurden die Emissionstrends berücksichtigt, indem die Jahressummen der Vorläuferemissionen (NO_x , CH_4 , NMVOC) für das gesamte Gebiet, wie sie vom UBA für die Jahre 1995 – 2018 geliefert wurden, als Variablen verwendet wurden (Abbildung 25). Es besteht die Möglichkeit, diese Gesamtemissionen mithilfe eines Tools spezifisch für kleinere Regionen anzugeben, so dass die Emissionen an den einzelnen Stationen verfügbar sind. Dieser Detailgrad wird in der vorliegenden Studie jedoch nicht vorgenommen. Zunächst ist die Klassifizierung der Ozonstationen entsprechend der Unterscheidung zwischen Verkehrsstationen und Hintergrundstationen relevanter als gerasterte Emissionen. Und zweitens hat sich die räumliche Verteilung der Emissionen im Laufe der Zeit verändert. Es ist eine separate, umfangreiche Aufgabe, diese Karten genau zu erstellen, was nicht in den Rahmen dieses Projekts fällt.

Ein weiterer einfacher Ansatz ist die Einbeziehung der beobachteten NO_2 -Konzentrationen, die für die meisten Stationen verfügbar sind. Dies ist ein direkter Indikator für das Emissionsregime. Für künftige Klimaexperimente werden solche lokalen NO_2 -Konzentrationen jedoch nicht verfügbar sein, so dass es nicht sinnvoll ist, diese Informationen in einen Algorithmus einzubeziehen.

Abbildung 25: Jährliche Gesamtemissionen für Deutschland, wie in dieser Studie verwendet



Quelle: Eigene Darstellung, TNO

5 Regression und fortgeschrittenere Methoden des Maschinellen Lernens

Physikalische und chemische Prozesse können mit Hilfe deterministischer Naturgesetze beschrieben werden. Diese Gesetze können für Vorhersagen verwendet werden, wenn die Eingangsdaten hinreichend gut bekannt sind, und werden z. B. zur Erstellung komplexer Computermodelle wie numerische Wettervorhersagemodelle und Chemie-Transport-Modelle genutzt. Die Eingangsdaten sind jedoch nie genau bekannt. Daher funktionieren die aktuellen Luftqualitätsmodelle, die die Auswirkungen von Emissionen und Meteorologie berücksichtigen, recht gut, stimmen aber nie genau mit den beobachteten Daten überein. Ein völlig anderer Ansatz besteht darin, von Beobachtungen auszugehen und Muster zu finden, indem mehrere beobachtete Eingangsvariablen mit der Zielvariablen korreliert werden. Dies ist ein Beispiel für Maschinelles Lernen, bei dem ein Algorithmus zur Erstellung von Regeln herangezogen wird. Aus Sicht der Luftqualität hat der letztgenannte Ansatz den Vorteil, dass sehr lokale Informationen berücksichtigt werden, die in auf Gitterboxen basierende CTMs nicht in Betracht gezogen werden können. Trotz des deterministischen Wissens, das in die auf Naturgesetzen basierenden Modelle einfließt, übertreffen die statistischen Modelle oft die CTMs in ihrer Leistung und benötigen nur weitaus weniger Input, z. B. nur die lokale Temperatur anstelle eines Temperaturfeldes über ein großes geografisches Gebiet. Außerdem sind die Berechnungsregeln, sobald sie trainiert sind, weniger rechenintensiv.

Maschinelles Lernen (ML) ist ein noch recht neues wissenschaftliches Gebiet und stellt eine Verbindung von Statistik, Datenwissenschaft und computergestützten Verfahren dar, mit zahlreichen Anwendungen (Smola und Vishwanathan, 2008; Kuhn, 2013). Die Regression ist das elementarste Beispiel für Maschinelles Lernen und eine der am weitesten verbreiteten Methoden zur Beschreibung der Beziehung zwischen Ozon und meteorologischen Variablen sowie für Vorhersagen von Ozon mit unterschiedlichen Komplexitätsgraden. Die chemischen und physikalischen Prozesse, die am Verhalten von Ozon in Bezug auf die Meteorologie beteiligt sind, sind jedoch nicht linear und additiv. Die lineare Regression ist daher unzureichend, um Wechselwirkungen zwischen Variablen und Nichtlinearitäten in der Reaktion des Ozons zu erfassen (Thompson, 2001). So wird beispielsweise die Abhängigkeit des Ozons vom Wind (Geschwindigkeit und Richtung) von der Topographie des Ortes sowie von der Schadstoffbelastung der umliegenden Gebiete beeinflusst, wodurch diese Beziehung räumlich begrenzt und zeitlich variabel ist. Die Beziehung zwischen Luftschadstoffen - insbesondere Ozon als Sekundärschadstoff - und meteorologischen Variablen lässt sich daher am besten durch nichtlineare statistische Modelle beschreiben (Thompson et al., 2001).

Maschinelle Lernverfahren können diese nichtlinearen Beziehungen berücksichtigen. ML-Methoden können nicht-parametrisch sein und haben den entscheidenden Vorteil, dass sie viele der für statistische Modelle erforderlichen Annahmen wie Normalität der Stichprobe, Homoskedastizität, Unabhängigkeit, Einhaltung anderer strenger parametrischer Annahmen und den sorgfältigen Umgang mit Interaktionseffekten zwischen Variablen nicht berücksichtigen müssen (Immitzer et al., 2012). ML-Vorhersagemodelle haben das Potenzial, klassische statistische Verfahren zu ergänzen, was zu einer verbesserten Analyse von Trends in der Luftqualität führen kann. Darüber hinaus können ML-Modelle eine bessere Vorhersagegenauigkeit aufweisen und große und komplexe Datensätze besser verarbeiten als traditionelle statistische Modellierungsansätze (Breiman, 2001).

In den letzten Jahren wurde eine Fülle von Methoden, Algorithmen und Ansätzen entwickelt, die auf verschiedenen Überlegungen zur Nichtlinearität, der Größe des Parameterraums, der Anzahl der Beobachtungen, der Empfindlichkeit gegenüber Ausreißern, der Recheneffizienz und der Frage beruhen, ob der Schwerpunkt auf der Ermittlung der erklärenden Variablen und ihrer relativen Bedeutung oder auf der Erstellung genauer Vorhersagen liegt. Es ist eine unmögliche Aufgabe, alle möglichen Methoden des Maschinellen Lernens im Detail zu behandeln. Wir verweisen hierzu auf Bücher wie Hastie et al. (2001). In diesem Abschnitt werden die wichtigsten Aspekte und die am weitesten verbreiteten Methoden hervorgehoben, die für die aktuelle Anwendung als geeignet angesehen werden. Es wurden auch verschiedene Algorithmen für spezielle Fälle entwickelt, um die Leistung der grundlegenden Methoden für die jeweilige Anwendung zu verbessern. Die Vor- und Nachteile der Methoden für unsere aktuelle Anwendung werden aufgezeigt. Es ist oft nicht möglich, im Voraus genau zu entscheiden, welcher Ansatz zu den besten Ergebnissen führen wird, und im Allgemeinen werden mehrere Methoden auf dasselbe Problem angewandt, um herauszufinden, welche am besten funktioniert (Hastie et al., 2001). Darüber hinaus sind die richtigen Überlegungen in der Auswahl der Variablen und eine sorgfältige Behandlung der Eingangsdaten mindestens genauso wichtig wie die Auswahl der Methode des Maschinellen Lernens.

Bevor wir ins Detail gehen, ist es wichtig, die folgende Formulierung in Hinblick auf das derzeitige Ziel, ein Modell zu entwickeln, das Ozonüberschreitungen in Abhängigkeit von Klimakennzahlen vorhersagt, zu realisieren. Es werden Methoden des Maschinellen Lernens verwendet, um ein Modell mit Vorhersagefähigkeiten zu entwickeln. Zu diesem Zweck wird ein **Algorithmus** entwickelt, der aus einer Reihe von Rechenregeln besteht. Diese Regeln werden mit einem Datensatz trainiert, um die Beobachtungen mit einem angemessenen Detaillierungsgrad und einer angemessenen Genauigkeit darzustellen, d. h. die richtigen Variablen auszuwählen und den Beziehungen die entsprechenden Gewichtungsfaktoren zu geben. Der trainierte Algorithmus ist das **endgültige Modell**. Wenn ein neuer Satz von Beobachtungen zur Verfügung steht, kann derselbe Algorithmus erneut trainiert werden, um zu einer Aktualisierung des Modells zu gelangen.

Weit verbreitete ML-Vorhersagemodelle sind z. B. Entscheidungsbaummethoden wie Boosted Regression Trees und Random Forest, die Kernel-Methoden, zu denen Support Vector Machines gehören, und schließlich künstliche neuronale Netze (Friedman, 2006). Diese ML-Methoden können neben der Vorhersage auch als multiple Regressionsmodelle (z. B. GAMs) angewendet werden. In diesem Abschnitt heben wir die wichtigsten Aspekte und die am häufigsten verwendeten Methoden hervor. Abbildung 26 gibt einen Überblick über die verschiedenen Ansätze.

5.1 Wesentliche ML Konzepte

In diesem Abschnitt werden in Anlehnung an Hastie et al. (2001) einige wesentliche Begriffe und Konzepte erläutert.

Das Konzept des maschinellen Lernens besteht darin, ein Modell zu erstellen, das Eingangsvariablen mit einer Ausgangsvariablen in Beziehung setzt. Dies kann durch **überwachtes Lernen geschehen**, wenn ein Wert einer Ergebnisgröße (als **Prädiktand** oder **abhängige Variable** bezeichnet) auf der Grundlage einer Reihe von Eingangsgrößen (**Prädiktoren** oder **unabhängige Variablen**) vorhergesagt werden soll. Im Gegensatz dazu

kann **unüberwachtes Lernen** zur Beschreibung von Assoziationen und Mustern zwischen einer Reihe von Eingabewerten verwendet werden, ohne dass es einen Ergebniswert gibt. Für ein **Regressionsproblem** wird ein quantitativer Output verlangt, während für ein **Klassifizierungsproblem** ein Maß für Ähnlichkeit/Distanz benötigt wird und eine Anzahl von Klassen identifiziert werden muss. Sowohl bei der Klassifizierung als auch bei der Regression geht es darum, eine annähernd funktionale Beziehung zu finden, und es handelt sich um überwachtes Lernen.

Die Methoden des Maschinellen Lernens sind darauf ausgelegt, die optimale funktionale Beziehung im Sinne des kleinsten Fehlers oder der maximalen Wahrscheinlichkeit in statistischer Hinsicht zu finden. Bei vielen Problemen gibt es keine eindeutige Lösung, da es viele Freiheitsgrade gibt und die Eingabedaten verrauscht sein können. Daher kann die Lösung mit Hilfe von Nebenbedingungen **reguliert** werden, wobei die Wahl der Nebenbedingung vom jeweiligen Problem abhängt. Einige Algorithmen können mehrmals mit verschiedenen Anfangswerten gestartet werden, um ein globales Minimum zu finden, anstatt nur ein lokales Minimum. Außerdem ist oft im Voraus nicht bekannt, welches die relevanten Variablen sind oder wie viele Klassen identifiziert werden können. Es wurden Kriterien ermittelt, um ein gutes Gleichgewicht zwischen der Komplexität und der Verwendung einer ausreichenden Anzahl von Variablen zur Beschreibung der nichtlinearen Wechselwirkungen zwischen den Variablen zu erreichen, aber einen Freiheitsgrad zu vermeiden, der mit der Anzahl der Trainingsdaten nicht gelöst werden kann, d. h. Lösungen in einem spärlich abgedeckten hochdimensionalen Raum zu finden. Letzteres kann auch zu einer Überanpassung führen, bei der das Modell die Trainingsdaten sehr gut abbildet, aber bei den Validierungsdaten schlecht abschneidet. Die richtige Anzahl von Variablen oder Clustern zu finden und zu bestimmen, wann das Training beendet werden soll, sind daher Schlüsselemente des maschinellen Lernens.

5.1.1 Auswahl der Variablen

Ein wichtiges Thema bei der Erstellung statistischer Modelle ist die Auswahl der Prädiktoren, die in das Modell aufgenommen werden sollen. Die Modelle müssen genügend Informationen aus den erklärenden Variablen enthalten, um das Modell bestmöglich zu trainieren, aber gleichzeitig muss eine Überanpassung vermieden werden, da dies zu falschen Vorhersagen führt. Das Akaike-Informationskriterium (AIC) ist ein Maß, das ein Regressionsmodell benachteiligt, wenn mehr erklärende Variablen hinzugefügt werden, und daher sollte das Modell mit dem niedrigsten AIC das einfachste Modell sein, das genügend Freiheit in Bezug auf Variablen und Interaktionen zulässt. Es ist wie folgt definiert:

$$AIC = -2 \cdot \log(L) + k \cdot n$$

wobei L der Maximalwert der Likelihood-Funktion für das Modell, k der Strafparameter für die Anzahl der erklärenden Variablen und n die Anzahl der erklärenden Variablen ist.

Darüber hinaus könnte Kollinearität, d. h. Wechselwirkungen zwischen den erklärenden Variablen, zu einer falschen Darstellung der Korrelation zwischen der Antwortvariablen und den erklärenden Variablen und folglich zu einer falschen Interpretation der Ergebnisse führen. Der Effekt der Kollinearität kann durch die Festlegung des maximal zulässigen Varianzinflationsfaktors (VIF) bekämpft werden. Der VIF zwischen zwei erklärenden Variablen kann wie folgt berechnet werden:

$$VIF_j = 1 / (1 - R_j^2)$$

wobei R_j^2 das entsprechende Bestimmtheitsmaß für einen Parameter j ist, dass die Varianz der Antwortvariablen darstellt, die durch die erklärenden Variablen erklärt wird.

5.1.2 Abbruchkriterien und Bewertung

Es kann passieren, dass das Modell sehr genau die Daten des Trainingsdatensatzes abbildet, aber damit zu wenig generalisierend ist. Das führt dazu, dass wenn das Model Vorhersagen für einen neueren Datensatz macht, die Performance viel niedriger ist als die Performance für die Trainingsdaten. Bei der Entwicklung einer funktionalen Beziehung auf der Grundlage beobachteter Werte, die für die Vorhersage verwendet werden soll, muss diese sogenannte Überanpassung oder ein Übertraining vermieden werden. Der einfachste Ansatz besteht darin, den Datensatz in Trainings-, Validierungs- und Testdaten aufzuteilen, wobei eine Aufteilung von 50 % – 25 % – 25 % als Faustregel gilt.

Während des Trainings wird die Fehlerquote über die Trainingsmenge überwacht. Dieser **Trainingsfehler** ist jedoch oft keine gute Schätzung des **Testfehlers**, also des Fehlers über die Teststichprobe. Woran man wirklich interessiert ist, ist der **erwartete Vorhersagefehler**.

In vielen Fällen reicht die Anzahl der Daten nicht aus, um den Datensatz in einen Trainings-, Validierungs- und Testdatensatz aufzuteilen. In solchen Fällen können andere Methoden angewandt werden.

Bei der **K-fachen Kreuzvalidierung** werden die Daten zufällig in $K-1$ Teile für das Training und 1 Teil für die Validierung aufgeteilt. Das Verfahren wird wiederholt, wobei jedes Mal ein weiterer Teil für die Validierung und der Rest für das Training verwendet wird, typischerweise werden $K=5$ oder $K=10$ gewählt. In vielen Fällen reichen die Daten nicht aus, um nur 50 % als Trainingsdaten zu verwenden, und schließlich werden alle Daten für das Training verwendet, wobei die K -fach kreuzvalidierten Ergebnisse als endgültige Leistungsmetriken präsentiert werden (z. B. De Hoogh et al., 2019).

Bootstrapping ist eine Methode, bei der eine Reihe von Stichproben durch zufällige Ziehungen aus den Trainingsdaten generiert werden, wobei jede Stichprobe die gleiche Größe hat (die gleiche Größe wie der Trainingsdatensatz). Diese Stichprobendatensätze können dann verwendet werden, um das Modell zu trainieren und das Verhalten des Modells zu bewerten. Diese Methode hat den Nachteil, dass die Bootstrap-Stichprobe und der ursprüngliche Datensatz gemeinsame Daten haben. Das Modell sollte daher mit Beobachtungen, die nicht Teil der Bootstrap-Stichprobe waren, bewertet werden.

Die effektive Anzahl der Prädiktoren kann, wie bereits erläutert, mit Hilfe des AIC (oder ähnlicher Kriterien wie dem Bayes'schen Informationskriterium (BIC), auch bekannt als Schwarz-Kriterium) ausgewählt werden. Für die Auswahl des "besten Modells" gilt die Faustregel, dass das einfachste ML-Modell gewählt wird, bei dem der Fehler nicht mehr als eine Standardabweichung über dem Fehler des Modells mit der besten Leistung liegt. Damit wird ein Gleichgewicht zwischen Komplexität und Leistung hergestellt. AIC/BIC ist keine gute Methode zur Schätzung des Test- und Vorhersagefehlers, Kreuzvalidierung und Bootstrapping sind bessere Methoden.

5.1.3 GAM (Generalized Additive Models)

In jüngster Zeit wurden die sogenannten Generalized Additive Models (GAMs, Hastie und Tibshirani, 1990; Wood, 2006) in verschiedenen Studien zur Luftqualität verwendet (z. B. Boleti et al., 2018, 2019; Barmpadimos et al., 2011; Carslaw et al., 2007). Sie berücksichtigen die Beziehung zwischen den Prädiktoren, d. h. den Ozonkonzentrationen, und den erklärenden Variablen (Prädiktoren) oder Kovariaten, d. h. den Klimakennzahlen, und können wie folgt beschrieben werden:

$$C_i = \sum_{j=1}^n s_j(x_{ij}) + \varepsilon_i$$

wobei C_i die i -te Konzentration der Zeitreihe ist, $s_j(x_{ij})$ eine glatte Funktion einer Kovariaten p ist, n die Gesamtzahl der Kovariaten und ε_i das i -te Residuum ist; es wird angenommen, dass die Residuen normalverteilt sind (Carslaw, 2007). Die Freiheitsgrade der glatten Funktionen müssen kalibriert werden.

Ein wichtiges Merkmal der GAMs bei der Modellierung von Luftschadstoffen ist die Fähigkeit, die komplexe Wechselwirkung zwischen Windgeschwindigkeit und Windrichtung zu modellieren. Nach Carslaw et al. (2007) könnten GAMs besonders in Situationen nützlich sein, in denen die Windströmungsmuster variabel und schwer vorherzusagen sind, was bei den meisten Messorten für Luftschadstoffe der Fall ist. GAM können interaktiv eingesetzt werden, Variablen können einbezogen oder ausgeschlossen werden und die Auswirkungen können überwacht werden. Auf diese Weise kann eine Reihe von Modellen – von linearen bis hin zu hochgradig nichtlinearen – erforscht und der beste Kompromiss ermittelt werden. GAMs funktionieren nicht gut bei großen Data-Mining-Problemen mit einer großen Anzahl von Prädiktoren, aber für Ozon ist die Anzahl der Prädiktoren begrenzt und die Methode hat sich bewährt. GAM wurde zum Beispiel in den Studien von Pearce et al. (2011), Boleti et al. (2018, 2019) und Hertig et al. (2020) verwendet, um die oben beschriebenen Beziehungen zwischen Ozon und meteorologischen Variablen zu untersuchen und zu nutzen. Weitere Vorteile von GAMs sind, dass das Modell gut interpretiert werden kann und dass glatte Funktionen abgeleitet werden, was das Modell als Vorhersagemodell robust macht.

5.1.4 Entscheidungsbaumbasierte Methoden

Entscheidungsbäume basieren auf einem rekursiven Klassifizierungsalgorithmus, der die Beobachtungen an jedem Knoten in zwei homologe Gruppen (Gruppen von Elementen mit ähnlichen Eigenschaften, z. B. eine Gruppe mit Tageshöchsttemperaturen über 25 °C und eine Gruppe mit Tageshöchsttemperaturen unter 25 °C) aufteilt, bis reine Knoten erreicht sind. Reine Knoten sind definiert als Knoten mit homogenen Elementen. Die rekursive Natur des Algorithmus bedeutet, dass die Aufteilung so lange wiederholt wird, bis die Reinheit des Knotens erreicht ist (zum Beispiel Ozonwerte über 180 µg/m³). Die gesamte Reihe der Aufteilungen, die einzeln als Knoten oder Zweige bezeichnet werden, wird als Baum bezeichnet. Der rekursive Algorithmus wird die Eingabedaten immer korrekt klassifizieren, wenn man die Bäume bis zu ihrer maximalen Tiefe wachsen lässt. Der Nachteil von Entscheidungsbäumen besteht im Allgemeinen darin, dass sie dazu neigen, die zum Trainieren des Modells verwendeten Daten zu stark anzupassen (Kotsiantis, 2013), sodass sie auf neue Datensätze, die während der Trainingsphase des Algorithmus nicht verwendet wurden, nicht angewendet werden können.

Das bedeutet, dass Entscheidungsbäume hilfreich sind, um die Relevanz von Variablen zu interpretieren, aber weniger nützlich für Vorhersagen sind.

5.1.5 Bagging und Random Forest

Zur Behandlung von Overfitting wird das sogenannte **Bagging** (Bootstrap-Aggregation) verwendet, d. h. es werden viele einzelne Entscheidungsbäume aus einem Trainingsatz erstellt, die dann zu einem Ensemble aggregiert werden. Das Ensemble-Modell besteht aus vielen einzelnen Bäumen/Modellen, die aus aggregierten Daten zusammengesetzt wurden, und bei der Verwendung für die Vorhersage werden die Ergebnisse aller Bäume zusammen verwendet (Grange, 2018). GAM kann auch als ein Bagging-Ansatz angesehen werden, bei dem unabhängige Modelle erstellt werden.

Random Forests (RFs) sind ein neuartiger Regressionsansatz (auch bekannt als Entscheidungswälder) (Breiman, 2001) und eine Bagging-Methode. Sie basieren auf dem Ansatz der Regressions- (oder Entscheidungs-) Bäume, verwenden aber unkorrelierte Bäume anstelle von korrelierten Bäumen. Das Endergebnis ist eine Mehrheitsentscheidung über die Ergebnisse der einzelnen Bäume. Dies führt zu einer besseren Varianzreduzierung als bei den anderen baumbasierten Methoden. Sie sind einfacher zu trainieren als Boosting-Algorithmen (siehe nächster Absatz) und die Leistung ist oft ähnlich. RF berücksichtigt im Gegensatz zu Boosting alle Variablen, ist aber auch gut in der Lage, aus einer großen Menge die relevantesten zu identifizieren. Übertraining tritt kaum auf. RFs funktionieren gut bei fehlenden Daten und können im unüberwachten, im Regressions- oder im Klassifikationsmodus eingesetzt werden. Dies hat die Methode sehr populär gemacht. Der Beitrag der einzelnen Variablen kann bis zu einem gewissen Grad quantifiziert werden.

RF wird z. B. in Zhan et al. (2018) zur Vorhersage der täglichen Ozonwerte in China verwendet, basierend auf den guten Erfahrungen mit RF für PM_{2,5} in China (Zhan et al., 2017) und für PM_{2,5} in Verbindung mit einem Waldbrandereignis (Reid et al., 2015). Watson et al. (2019) fanden eine gute Leistung von RF für die Vorhersage von Ozonkonzentrationen im Zusammenhang mit demselben Waldbrandereignis.

Die erwiesene Robustheit und gute Leistung der Methode in verschiedenen Anwendungen im Zusammenhang mit der Luftqualität und anderen Bereichen macht die Methode zu einem guten Kandidaten, um ein Modell mit den erforderlichen Eigenschaften zu finden.

5.1.6 Boosting und GBM

Eine weitere Möglichkeit, die Leistung von Entscheidungsbäumen zu verbessern, ist die Verwendung von **Boosting**. Dabei wird eine Reihe von Variablen mit jeweils relativ geringer Auswirkung zusammen verwendet, um eine kombinierte Variable mit größerer Auswirkung zu erhalten. Das Modell wird sequentiell aufgebaut, so dass die Regressionsvariablen nicht unabhängig voneinander ausgewählt werden, sondern die Informationen aus dem vorherigen Schritt berücksichtigt werden. Dies führt in der Regel zu sehr guten Ergebnissen. Es werden verschiedene Boosting-Ansätze verwendet, wie AdaBoost und GBM (Gradient Boosting Machine), die sich darin unterscheiden, wie sie Unterschiede zwischen beobachteten und angepassten Ergebnissen behandeln, um sie weiter zu optimieren. Boosting ist für Variablen geeignet, die Korrelationen und nichtlineare Probleme aufweisen. Darüber hinaus kann das

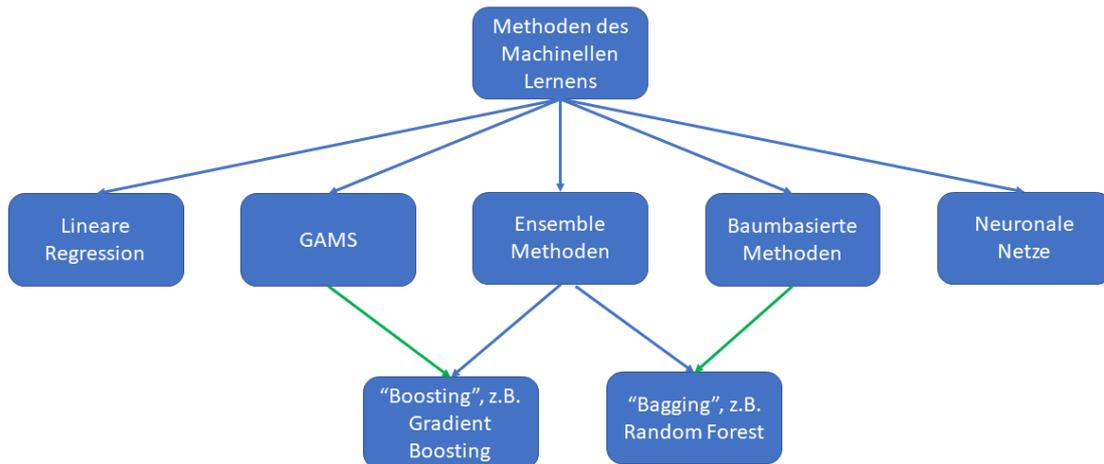
Modell den absoluten und relativen Beitrag der einzelnen Variablen zum endgültigen Modell besser als andere Methoden aufzeigen. Die Methode ist relativ rechenaufwändig.

Gradient Boosting war einer der erfolgreichsten Algorithmen in Watson et al. (2019) zur Vorhersage von Ozon bei Waldbränden im Jahr 2007, wobei Random Forest an zweiter Stelle lag. Reid et al. (2015) fanden zuvor bei demselben Waldbrandereignis schon das für PM_{2,5} Gradient Boosting am Erfolgreichsten war.

5.1.7 Neuronale Netze und andere Ansätze

In den letzten Jahren wurden mehrschichtige neuronale Netze mit Rückkopplung (z. B. Rumelhart et al., 1986), die auch mit dem Begriff "Deep Learning" in Verbindung gebracht werden, mit großem Erfolg bei der Bilderkennung und bei Computerspielen eingesetzt. Angeregt durch den Erfolg in der Computer-Vision und bei Computerspielen begannen die Forscher auch, das Potenzial neuronaler Netze bei der Vorhersage zeitlich variierender Probleme zu erkunden. Neuronale Netze (NN) basieren auf einer Eingabe- und einer Ausgabeschicht mit Neuronen (oder Knoten) und einer Reihe von internen Schichten mit Neuronen. Diese Neuronen der verschiedenen Schichten können vollständig verbunden sein, aber auch andere Konstruktionen sind möglich, und die Gewichte dieser Verbindungen werden trainiert. Neuronale Netze bergen das Risiko der Überanpassung und sind weniger geeignet, um die Auswirkungen einzelner Vorhersagevariablen aufzuzeigen. Es gibt viele verschiedene NN-Methoden, die alle teilweise die Aspekte berücksichtigen, die für die jeweilige Anwendung am wichtigsten sind. Feng et al. (2019) untersuchten beispielsweise die Leistung von neuronalen Netzen bei der Vorhersage von Ozonkonzentrationen in Hangzhou, China. Sie verglichen die Leistung von Extreme Learning Machine (ELM), Multi-layer Perceptron (MLP), Random Forest (RF) und Recurrent Neural Network (RNN) mit dem Community Multi-scale Air Quality (WRF-CMAQ). Die Ergebnisse der neuronalen Netzwerkmethoden sind vielversprechend und weisen eine höhere Vorhersagegenauigkeit auf als WRF-CMAQ. Feng et al. (2019) haben festgestellt, dass das RNN (Liu und Sullivan, 2019) in dieser ersten Studie die beste Leistung erzielte. Eslami et al. (2020) verwendeten ein faltendes neuronales Netzwerk (CNN, Convolutional Neural Network) zur Vorhersage der Ozonkonzentrationen in Seoul mit überwachtem Training. Obwohl die Methode vielversprechende Ergebnisse liefert, geben neuronale Netzwerkansätze weniger Einblick in die zugrundeliegenden Mechanismen und können bei Änderungen einer der Variablen weniger robust sein, was die Methode weniger robust für Vorhersagen bei sich ändernden Eingangsbedingungen (Klima und Emissionsänderungen) macht.

Abbildung 26: Übersicht verschiedener Methoden zur Erstellung von Modellen für dieselben Datensätze.



Quelle: Eigene Darstellung, TNO

5.1.8 Stacking oder Super Learning

Stacking ist eine Methode zum Trainieren eines Second-Level-Modells, um die optimale Kombination eines Ensembles von Modellen zu finden, also eine optimale Kombination der trainierten GBM-, RF- und/oder NN-Algorithmen. Diese „gestapelten Modelle“ werden als „Stacked Ensembles“ gekennzeichnet.

Jeder Algorithmus (GBM, Random Forest, Neuronales Netz) enthält eine Anzahl von Hyperparametern (Steuerparameter für die Algorithmen, z. B. Anzahl der Bäume) die spezifiziert werden sollten und es gibt eine optimale Auswahl der Steuerparameter. Eine Gruppe von mit verschiedenen Hyperparametern trainierten einzelnen ML-Algorithmen einer Methode bildet eine „Familie von Modellen“.

Beim Konstruieren eines Stacked Ensembles können die besten Modelle aus ihrer Familie einbezogen werden („best of family“), aber man könnte auch mehrere Modelle der jeweiligen Familien gleichzeitig einbeziehen (generelles Stacked Ensemble).

5.2 Clusteranalyse

In diesem Projekt wurde der Cluster-Ansatz angewendet, um Gruppen von Stationen mit ähnlichen räumlich-zeitlichen Merkmalen zu ermitteln. Bei der Analyse der Luftqualität werden die Stationen standardmäßig in die Kategorien "ländlich", "städtischer Hintergrund", "städtischer Verkehr" oder "Industrie" eingeteilt. Diese Einteilung kann jedoch zu begrenzt sein, da auch eine regionale Variabilität besteht, die durch Unterschiede in den regionalen Wettermustern bedingt ist. Dies geht aus der Analyse räumlicher Muster (z. B. Otero et al., 2016), der Beziehung zur synoptischen Wetterklassifikation (Pavlik et al., 2016) und der expliziten Clusteranalyse (z. B. Boleti et al., 2019b) hervor. Die Clusterbildung kann in einer Dimension erfolgen (z. B. Stationstyp, Landnutzung, Klimazone, Anzahl der Überschreitungen), aber auch in mehreren Dimensionen (z. B. Zeitreihen oder Kombination von Anzahl und Lage).

Die Clusteranalyse bezieht sich auf die Mustererkennung in mehrdimensionalen Daten. Die Hauptidee besteht darin, Datengruppen (Cluster) auf der Grundlage von Ähnlichkeitsgraden bezogen auf bestimmte Merkmale zu identifizieren. Elemente, die zum selben Cluster gehören, müssen den höchsten Ähnlichkeitsgrad aufweisen, während Objekte aus verschiedenen Clustern einen niedrigen Ähnlichkeitsgrad haben müssen (Jain, 2010).

Es wurden mehrere Clusteralgorithmen entwickelt, die auf einer Abstandsschätzung zwischen den Objekten (Merkmale der Daten) der Cluster basieren. Cluster-Verfahren können in hierarchische (zentroid-basierte) und partitionierende (k-means) unterteilt werden. Beim hierarchischen Clustern wird der Datensatz schrittweise in immer feinere Cluster aufgeteilt oder es wird von einzelnen Daten ausgegangen und diese schrittweise gruppiert. Die Anzahl der Cluster und die anfängliche Schätzung der Clustermittelpunkte (Merkmale, die den Schwerpunkt eines Clusters definieren) müssen nicht im Voraus festgelegt werden, aber es ist ein Maß für den Ähnlichkeitsgrad erforderlich. Das partitionierende Clustern gruppiert die Daten in eine vorher festgelegte Anzahl von Clustern. Cluster sind nicht überlappende Gruppen, so dass am Ende der Berechnung jedes Objekt zu genau einem Cluster gehört (Lyapina, 2016). Dies ist erforderlich, um die Ozonstationen auf eine begrenzte Anzahl von Klimazonen und lokalen Merkmalen aufzuteilen. Für jedes dieser Cluster kann mithilfe anderer ML-Techniken eine funktionale Beziehung abgeleitet werden, um die Eingangsdaten mit den Ozonüberschreitungen zu verknüpfen.

Das gängigste Verfahren ist der k-means-Algorithmus, der die durchschnittlichen euklidischen Abstände zwischen den einzelnen Objekten und der vorgegebenen Anzahl von Cluster-Mittelpunkten (Zentroids) minimiert. Ein Zentroid ist ein künstliches Objekt, das sein Cluster repräsentiert und das arithmetische Mittel aller Eigenschaften der Clustermitglieder ist.

Der Cluster-Algorithmus Partitioning Around Medoids (PAM) basiert auf k-means (MacQueen, 1967; Hartigan und Wong, 1979). PAM ist robuster als k-means und weniger empfindlich gegenüber Ausreißern, da es Medoide (tatsächliche Punkte in den Daten) anstelle von Zentroiden (normalerweise künstliche Punkte) verwendet (R Development Core Team, 2017). Die Methode wurde z. B. in Boleti et al. (2019) verwendet.

5.3 Datenanalyse und -visualisierung

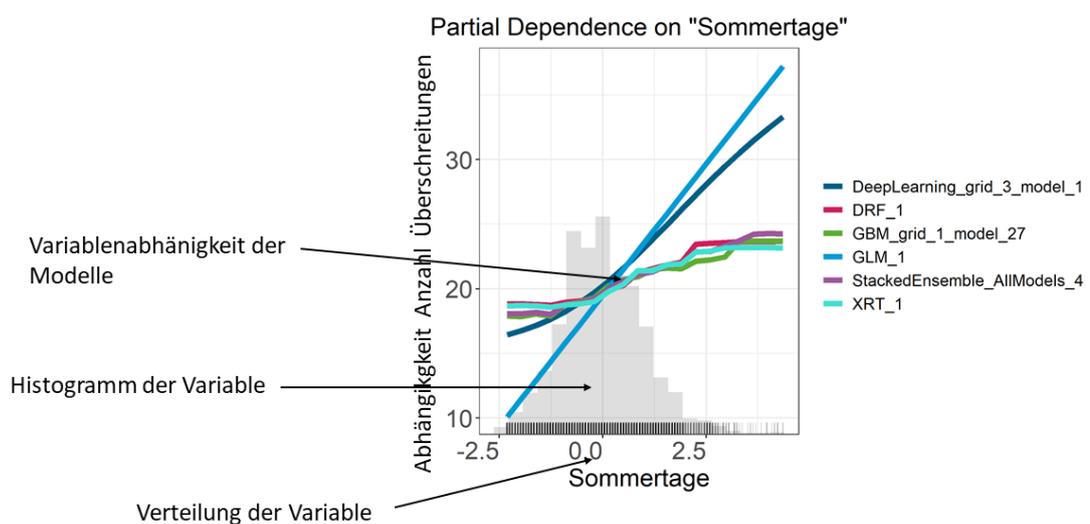
Eine erste Prüfung der Ergebnisse besteht darin, die Abhängigkeiten der Variablen in einem Korrelationsdiagramm zu untersuchen, um einige der Variablen auszuwählen oder auszuschließen. Bei linearen Regressionen haben Variablen mit schwacher Korrelation nur geringe Auswirkungen, aber bei Ensemble-Methoden können sie dennoch einen positiven Beitrag zu den Endergebnissen leisten.

Sobald ein Algorithmus trainiert ist, werden zur Bewertung der Modellperformance üblicherweise der mittlere quadratische Fehler (Root Mean Square Error, RMSE) und der mittlere absolute Fehler (Mean Absolute Error, MAE) für die Trainingsdaten und die (vom Training ausgeschlossenen) Testdaten separat gemessen. Normalerweise ist die Leistung bei den Trainingsdaten besser als bei den Testdaten, und große Diskrepanzen deuten auf ein Übertraining des Modells hin. Übertraining bedeutet, dass jedes Detail des Trainingsatzes angepasst wird, was nicht unbedingt repräsentativ ist. Dies kann durch die Verwendung von weniger Daten für das Training verringert werden, aber es kann auch eine Redundanz bei den Variablen bestehen. Residualanalyse-Diagramme zeigen die Residuen als Funktion der

angepassten Werte (z. B. Abbildung 42). Die Residuen sollten zufällig verteilt sein, aber es können streifige Muster auftreten, wenn die Antwortvariable ganzzahlig ist, wie im Fall der Anzahl der Ozonüberschreitungen oder Episoden. Die Verteilung der Residuen kann auch in Form eines Histogramms dargestellt werden.

Das Modell kann als eine annähernde Beschreibung einer Oberfläche in mehreren Dimensionen, bestimmt durch die Anzahl von Variablen, betrachtet werden. Diese Oberfläche definiert die Beziehung zwischen Variablen (Prädiktoren) und Prädiktand und ist die Lösung des Optimierungsproblems des Maschinellen Lernens. Eine Möglichkeit, den Beitrag einzelner Variablen zu untersuchen, besteht darin, **Partial Dependence Plots** (partielle Abhängigkeitsdiagramme) zu erstellen (Abbildung 27). Dies sind Darstellungen der Abhängigkeit der Lösung von diesem Prädiktor, wobei die Auswirkungen anderer Variablen minimiert (aber nicht ignoriert!) werden. Sie zeigen die Empfindlichkeit der Lösung für die Variable über den Variablenbereich. Sie können nicht als einfache Querschnitte der mehrdimensionalen Oberfläche betrachtet werden, sondern geben einen Einblick in den Gültigkeitsbereich und die Empfindlichkeit. Für einige Algorithmen des maschinellen Lernens, wie lineare Regression oder GAM, kann der Einfluss einzelner Variablen auf das Endergebnis quantifiziert werden. Bei stark nicht-linearen Algorithmen wie Random Forests und GBM kann die Korrelation zu einzelnen Prädiktoren (Variablen) nicht explizit berechnet werden, da der Algorithmus eine hochdimensionale Fläche beschreibt. In diesen Diagrammen der Variablenabhängigkeit werden in der Regel auch Histogramme der Variablen aufgetragen, sowie die Variablenwerte (als Markierungen auf der x-Achse). Diese zeigen an, welche Daten der Algorithmus gesehen hat, so dass man leicht erkennen kann, für welchen Bereich von Variablenwerten der Algorithmus voraussichtlich robuster ist und wo man mit der Interpretation vorsichtig sein muss, da in dem Bereich nur wenige Daten verfügbar waren.

Abbildung 27: Beispiel eines Partial Dependence Plots für die Anzahl der Überschreitungen und die normalisierte Anzahl der Sommertage



Quelle: Eigene Darstellung, TNO

Variable importance plots zeigen, welche Variablen die stärkste Vorhersagekraft haben. Sie können für einzelne trainierte Modelle erstellt werden. Für einen einzelnen Entscheidungsbaum ist die quadrierte relative Wichtigkeit einer Variable die Summe der quadrierten Verbesserungen über alle internen Knoten, für die die Variable als Splitting-Variable gewählt wurde (Gini-Splitting-Index). Bei additiven Methoden wie GBM und Random Forest ist es der Durchschnitt dieses Wertes über alle Bäume und er ist stabiler als der Wert für einzelne Bäume. Die Werte für die Wichtigkeit der Variablen werden skaliert, um einen relativen Zusammenhang mit dem Wert herzustellen. Ein Beispiel ist Abbildung 46. Neben dieser Methode wird in Random Forest auch eine andere Methode zur Berechnung der Variablenbedeutung verwendet, die im Grunde zum gleichen Ranking führt, aber mit einer gleichmäßigeren Verteilung über die Variablen. Die letztgenannte Methode verwendet eine "Out-of-Bag"-Randomisierung, bei der Daten verwendet werden, die der Baum beim Training nicht gesehen hat, und zeichnet den Genauigkeitsverlust des Endergebnisses für zufällige Permutationen der Variablen j nach. Auch hier wird die durchschnittliche Verringerung der Genauigkeit verwendet, um die Variablen in eine Rangfolge zu bringen, wobei diese in % des Maximums ausgedrückt wird. Die Mittelwertbildung über die Permutationen führt zu einer homogeneren Verteilung. Für ein Ensemble von Modellen kann die Variablenbedeutung in Form von **variable importance heatmaps** dargestellt werden (z. B. Abbildung 44). Diese Heatmaps können interpretiert werden, um zu sehen, welche Variablen im Allgemeinen am Wichtigsten sind und auch, ob sie in einigen Modellen mehr oder weniger dominant sind. In den Heatmaps für die Variablenbedeutung werden die Werte für den wichtigsten Wert nicht auf 100 % skaliert, wie in der Balkengrafik für die Variablenbedeutung, so dass ein Vergleich zwischen verschiedenen Modellen besser möglich ist.

Shapley Additive ExPlanation plots (SHAP-plots) gehören zu den neuesten und komplexeren Visualisierungen der ML-Modellerklärung. Sie stammen aus der Spieltheorie, bei der der Beitrag jedes einzelnen Spielers quantifiziert wurde. Ein SHAP-Plot quantifiziert den Beitrag jeder der Variablen zur Vorhersage des Modells zum Prädiktanden in Verbindung mit den Beiträgen der anderen Variablen. Sie stellt keine kausalen Beziehungen dar, sondern visualisiert die Beiträge, um die Interpretation des Modellverhaltens zu unterstützen. Eine ausgezeichnete Einführung für SHAP-plots ist unter folgendem Link zu finden:

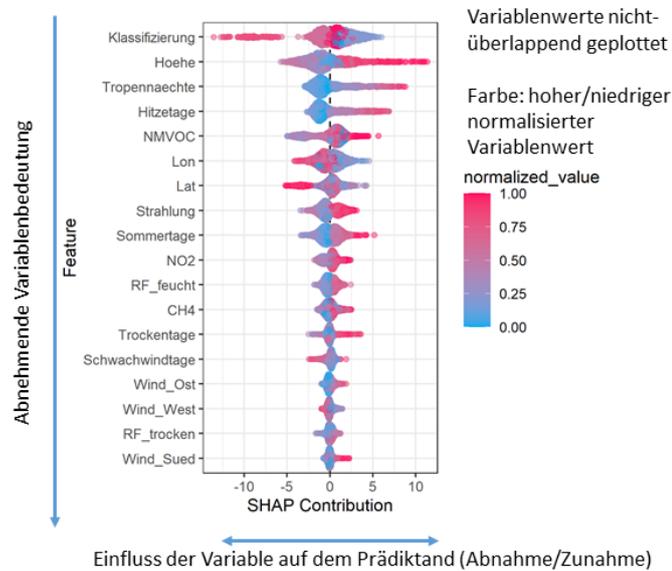
https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html

Abbildung 28 veranschaulicht die Verwendung eines SHAP-plots. In der Vertikalen sind die Variablen nach ihrer Merkmalsbedeutung geordnet. In der Horizontalen wird die Auswirkung der Variable auf den Prädiktand angegeben (der Prädiktand nimmt mit steigendem Prädiktorwert ab/zu). Die Farbe zeigt an, ob die Variable hoch oder niedrig ist (Verteilung der ursprünglichen Werte). Die einzelnen Daten werden in einem nicht vollständig überlappenden Modus aufgetragen, so dass eine Streuung in der Vertikalen für eine einzelne Variable die Anzahl der Daten in diesem Bereich angibt. SHAP-plots können nicht für jedes ML-Modell erstellt werden, sie können für baumbasierte und GBM-Modelle bestimmt werden, aber nicht für ein Ensemble von Modellen. Die Abbildung zeigt, dass, abgesehen von der Auswirkung der Klassifizierung der Ozonstationen, in diesem Fall die Stationshöhe (höher für mehr rosafarbene Werte) positiv zur Anzahl der Überschreitungen beiträgt (positive Abweichung vom Erwartungswert für hohe normalisierte Werte), und auch, dass der Großteil der Daten in geringerer Höhe liegt (dicker Fleck) mit geringerem (und leicht negativem) Einfluss auf den

Erwartungswert. Der Einfluss der Tropennächte ist geringer, ebenfalls mit dem größten positiven Effekt für die größten Werte.

Wenn ein Ensemble von Modellen erstellt wurde, kann eine **Modellkorrelations-Heatmap** erstellt werden, um die Ähnlichkeit der Modelle zu visualisieren.

Abbildung 28: Beispiel eines SHAP-plots



Quelle: Eigene Darstellung, TNO

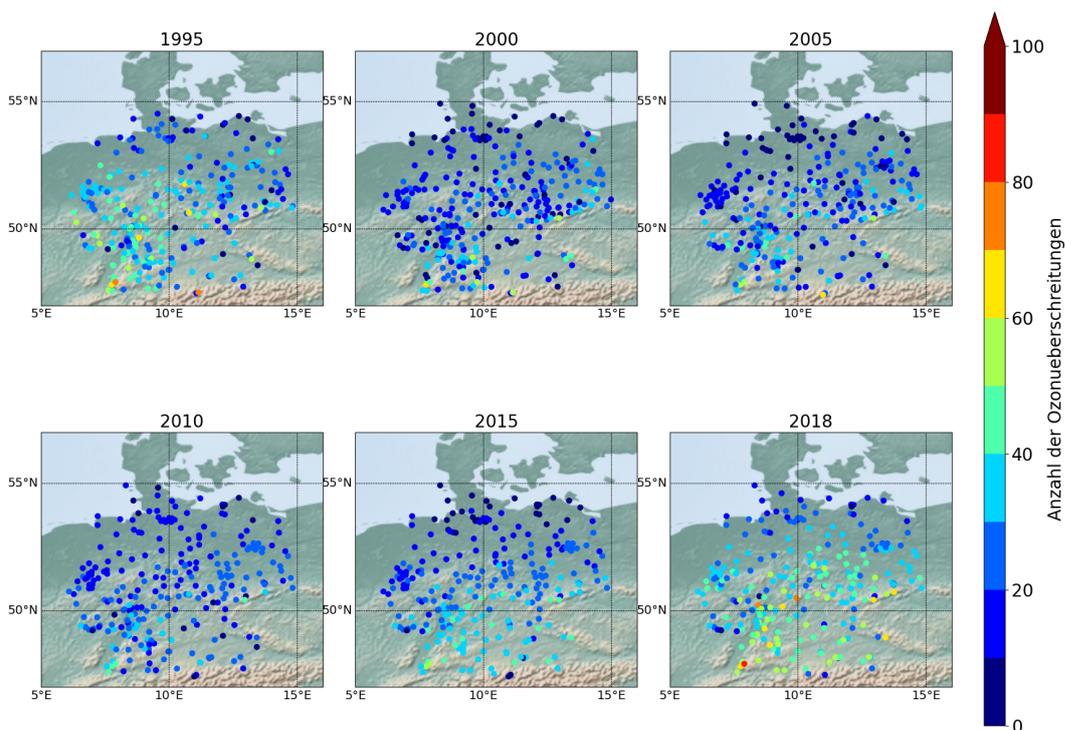
6 Ergebnisse

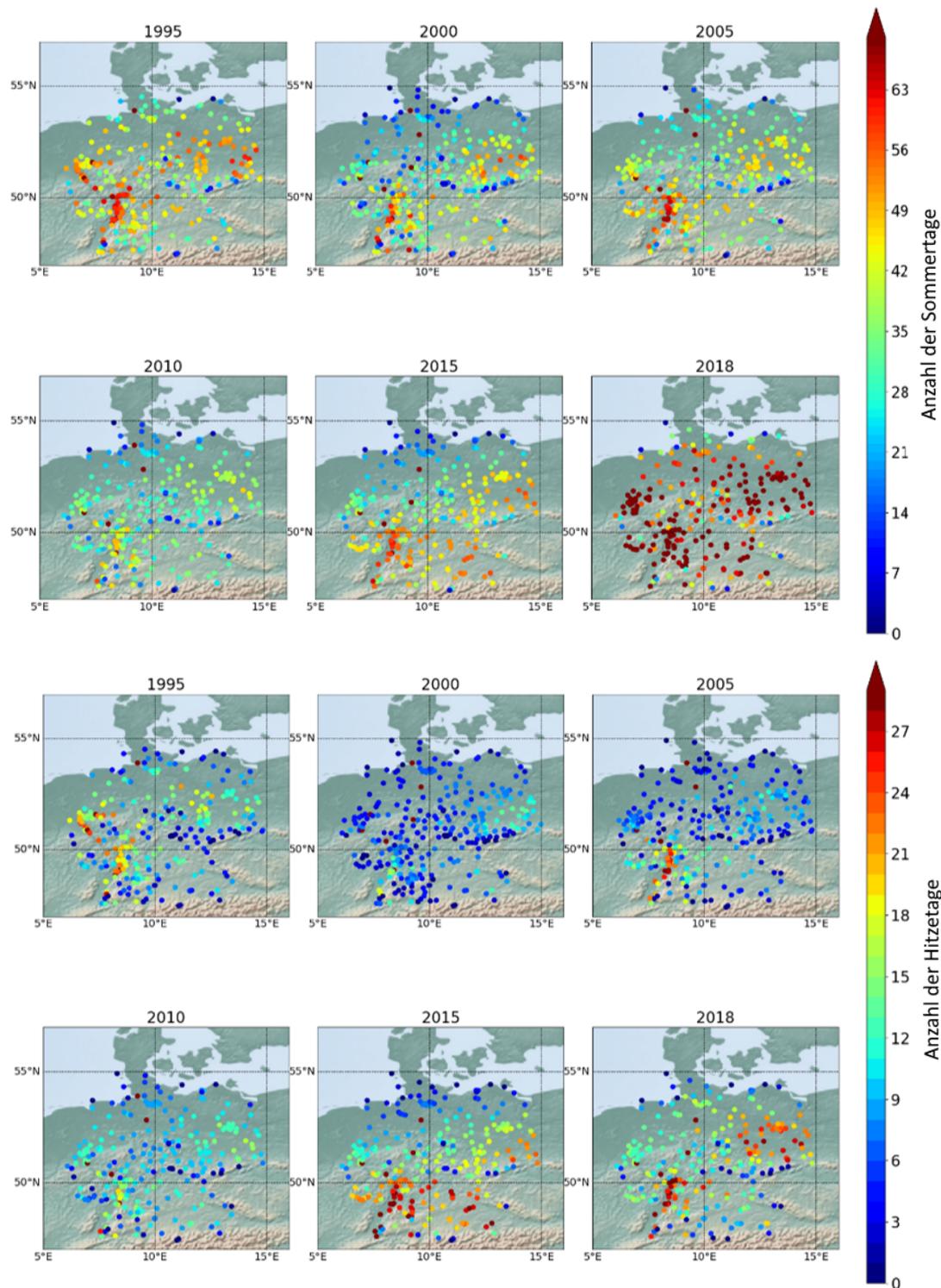
Zunächst werden die Korrelationen zwischen den Klimakennzahlen, den geografischen Informationen und der Anzahl der Ozonüberschreitungen untersucht, um die Auswahl und die Interpretation der Ergebnisse zu unterstützen. Dann wird die Clusteranalyse der Daten beschrieben. Anschließend wird der endgültige Ansatz für die Entwicklung der ML-Algorithmen vorgestellt. Schließlich werden die ML-Modelle und ihre Performance dargestellt und diskutiert.

6.1 Korrelationen zwischen Klimakennzahlen und Anzahl der Ozonüberschreitungen

Die Temperatur wird allgemein als der direkteste Zusammenhang mit Ozonüberschreitungen angesehen. Abbildung 29 zeigt die Anzahl der Hitze- und Sommertage für einige verschiedene Jahre, um die geografische Verteilung zu verdeutlichen und einen visuellen Vergleich mit der Anzahl der Ozonüberschreitungen zu ermöglichen. Die Gebiete und Jahre mit einer höheren Anzahl von Sommertagen und Hitzetagen entsprechen im Großen und Ganzen den Gebieten und Jahren mit einer höheren Anzahl von Überschreitungstagen. Allerdings spiegelt sich zum Beispiel die hohe Anzahl an Hitze- und Sommertagen im Rhein-Main-Gebiet (im Vergleich zur Umgebung) nicht direkt in der Anzahl der Überschreitungstage im Vergleich zur Umgebung wider.

Abbildung 29: Anzahl der Ozonüberschreitungen (oben), Anzahl der Sommertage (Mitte) und Anzahl der Hitzetage (unten) für ausgewählte Jahre





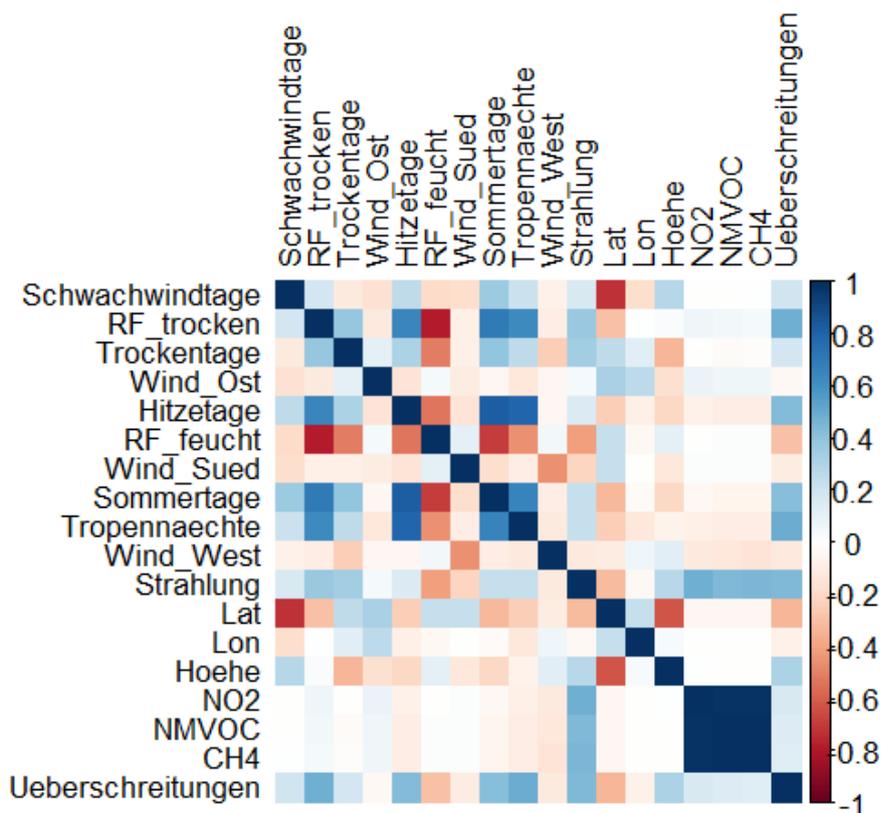
Quelle: Eigene Darstellung, TNO

Eine Zusammenfassung der Korrelationen zwischen allen Klimakennzahlen und der Ozonüberschreitungen für alle Standorte findet sich in Abbildung 30. Die Anzahl der Ozonüberschreitungen steht erwartungsgemäß in einem positiven Zusammenhang mit der Anzahl der Tropennächte, der Sommertage und der Strahlung und in einem negativen Zusammenhang mit dem Breitengrad, der Anzahl der RF_feuchten Tage und der Anzahl der Tage mit Westwind (Wind_West). Aber auch mit der Anzahl der RF_trockenen Tage und in

geringerem Maße mit der Anzahl der Trockentage und Schwachwindtage werden positive Korrelationen festgestellt. Die Abbildung zeigt auch, dass diese Variablen nicht völlig unabhängig sind. Die Anzahl der Schwachwindtage ist stark negativ mit dem Breitengrad korreliert, wobei es im Norden Deutschlands weniger Schwachwindtage gibt. Der Breitengrad ist negativ mit der Stationshöhe korreliert, was mit der Tatsache übereinstimmt, dass die Berggebiete sich überwiegend in der Mitte und im Süden Deutschlands befinden. Die Anzahl der Hitze-/Sommertage, die Strahlung und die Tage mit niedriger relativer Luftfeuchte (RF_trocken) sind erwartungsgemäß positiv korreliert.

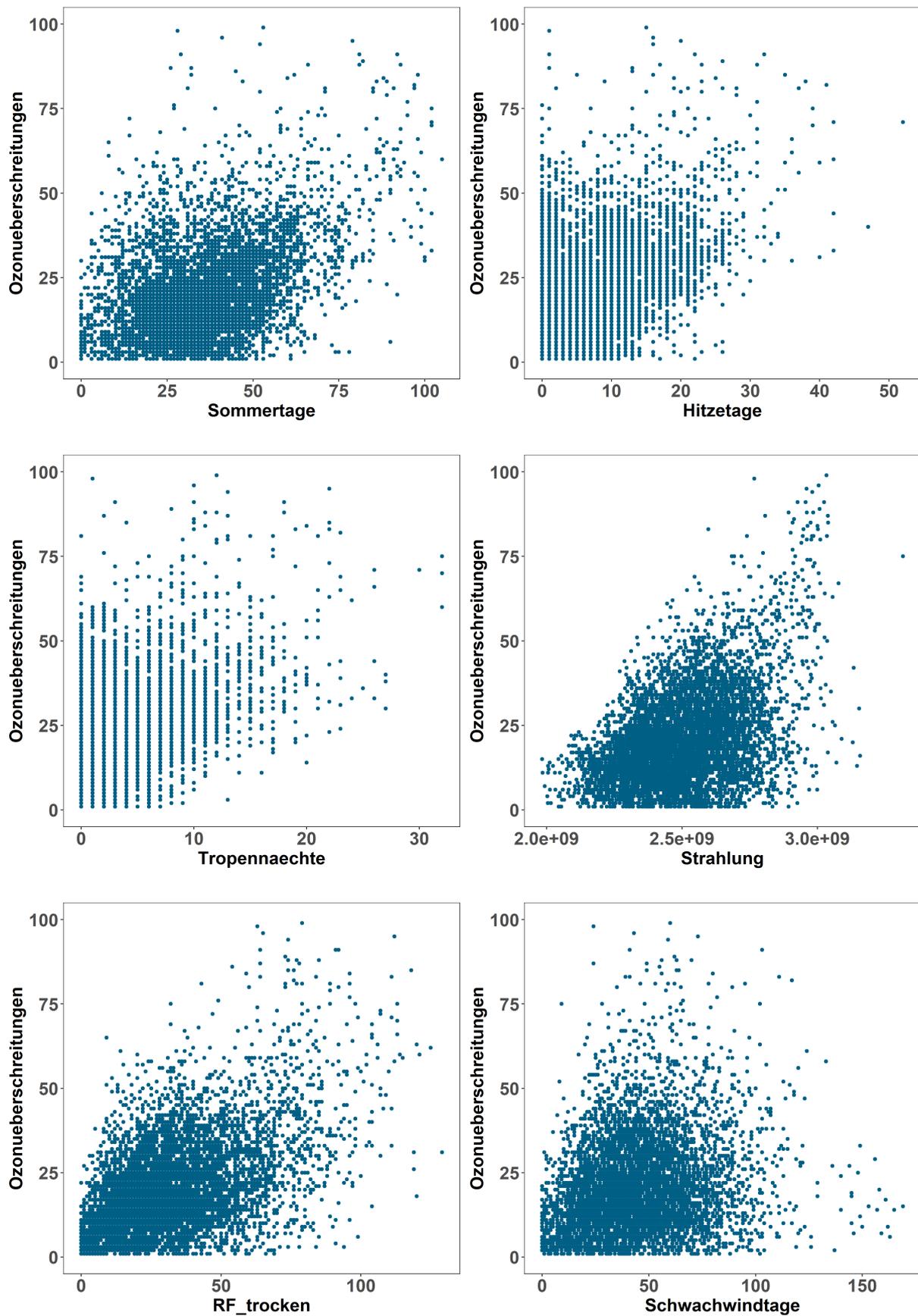
In Abbildung 31 sind Streudiagramme dargestellt, die verschiedene Klimakennzahlen mit der Anzahl der Überschreitungen in Beziehung setzen. Sie zeigen, dass für jede Klimakennzahl eine beträchtliche Streuung festzustellen ist und dass es keine einzige Variable gibt, die für sich genommen eine hohe Vorhersagekraft hat. Für fast alle Variablen, mit Ausnahme der Anzahl der Tage mit südlicher Strömung, kann eine (schwache) Korrelation mit der Anzahl der Ozonüberschreitungen beobachtet werden. Für den Längengrad und die Höhe ist die Korrelation nicht offensichtlich, obwohl die Anzahl der Überschreitungen bei den Stationen in großer Höhe höher ist als bei den Stationen in geringer Höhe.

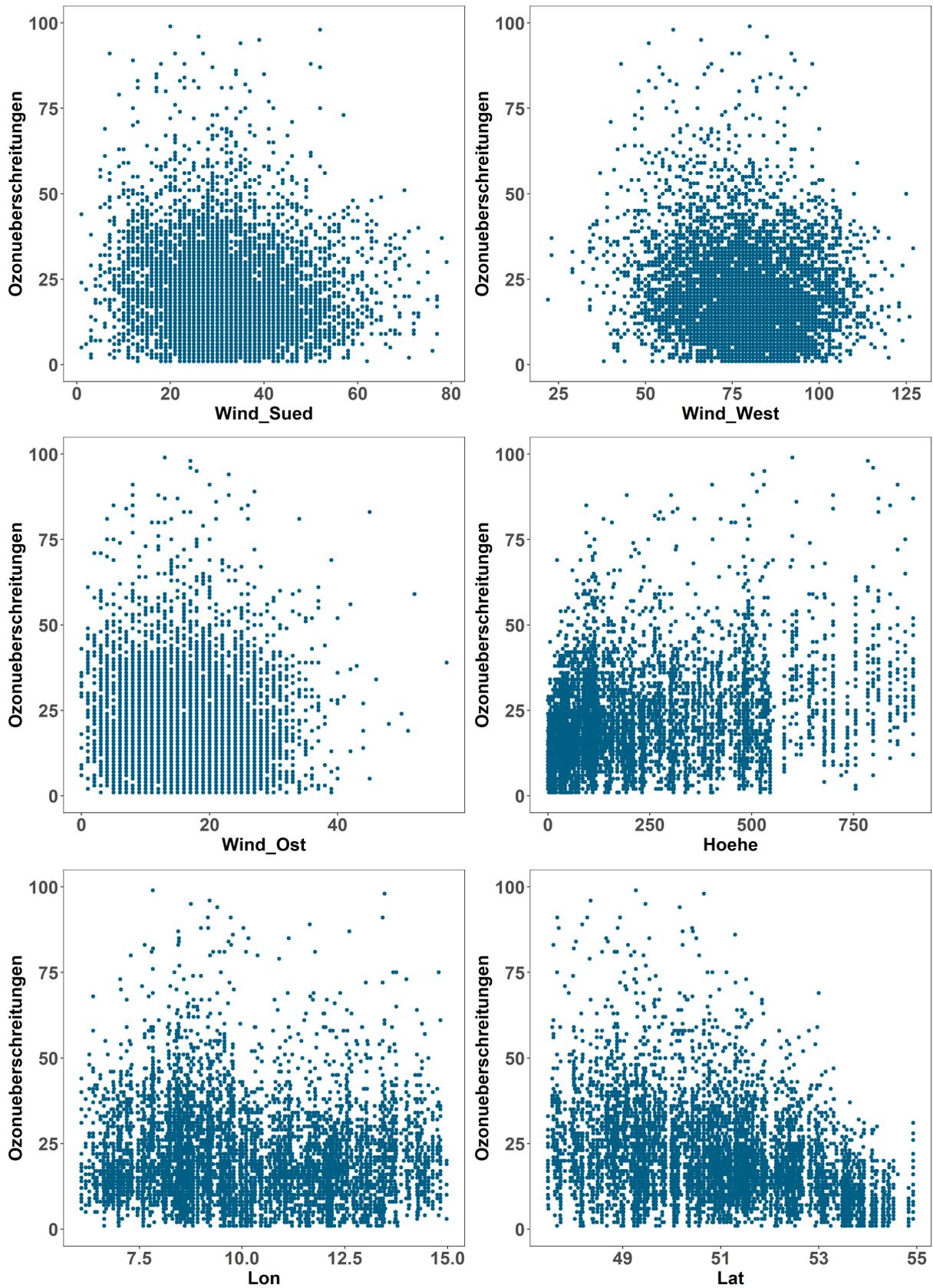
Abbildung 30: Korrelation der Eingangsvariablen und der Anzahl der Ozonüberschreitungen



Quelle: Eigene Darstellung, TNO

Abbildung 31: Streudiagramme der Eingangsvariablen mit der Anzahl der Ozonüberschreitungen





Quelle: Eigene Darstellung, TNO

6.2 Clustering

In diesem Abschnitt werden die Experimente beschrieben, die durchgeführt wurden, um herauszufinden, wie das Clustering funktionieren könnte und welche Konsequenzen die verschiedenen Ansätze haben. Die einfachsten Möglichkeiten zum Clustern der Daten sind:

- ▶ die Definition von drei geografischen Regionen (wie in KLENOS, siehe Abbildung 32)
- ▶ die Auswahl eines Stationstyps (Verkehr, städtisch, vorstädtisch, ländlich).

Wenn genügend Daten vorhanden sind, könnte die Aufteilung nach Stationstypen sogar nach Regionen erfolgen. Dies ist die Ausweichoption, für die kein Clustering-Algorithmus entwickelt werden muss.

Wir haben jedoch auch einen flexiblen Ansatz mit einem Clustering-Algorithmus analysiert, der Cluster auf unüberwachte Weise definiert. Auf diese Weise können Muster gefunden werden, die sich von der Ausweichoption unterscheiden und neue Informationen hinzufügen können. K-median (PAM) ist der am häufigsten verwendete Ansatz (siehe Abschnitt 5.2). Wir haben dieses Verfahren direkt auf die Anzahl der Überschreitungen pro Jahr angewandt, um Klima- und Emissionszonen zu ermitteln. Darüber hinaus wurde die Clusterbildung für kürzere Zeiträume auf vollständige tägliche Zeitreihen von Ozonüberschreitungen angewendet. Dies ermöglichte einen besseren Einblick in Stationen, die Ozonüberschreitungen und Episoden in größeren Gebieten ähnlich erfassen. Aber Abbildung 13 zeigte bereits, dass die Unterschiede zwischen den Jahren beträchtlich sind und sich auch die Muster von Jahr zu Jahr ändern.

Für das Endergebnis stellte sich heraus, dass der gesamte Datensatz ohne regionale Gruppierung verwendet werden sollte, um genügend Daten für die höchsten Überschreitungswerte zu erhalten. Stationsinformationen (Längengrad, Breitengrad, Höhe, Stationsklassifizierung) können direkt in den endgültigen Trainingsdatensatz aufgenommen werden. Die Stationsklassifizierung kann als kategoriale Variable aufgenommen werden.

Abbildung 32: Die drei Regionen mit Ozonüberschreitungen, die im KLENOS-Projekt ermittelt wurden



Quelle: KLENOS (Pavlik et al., 2016)

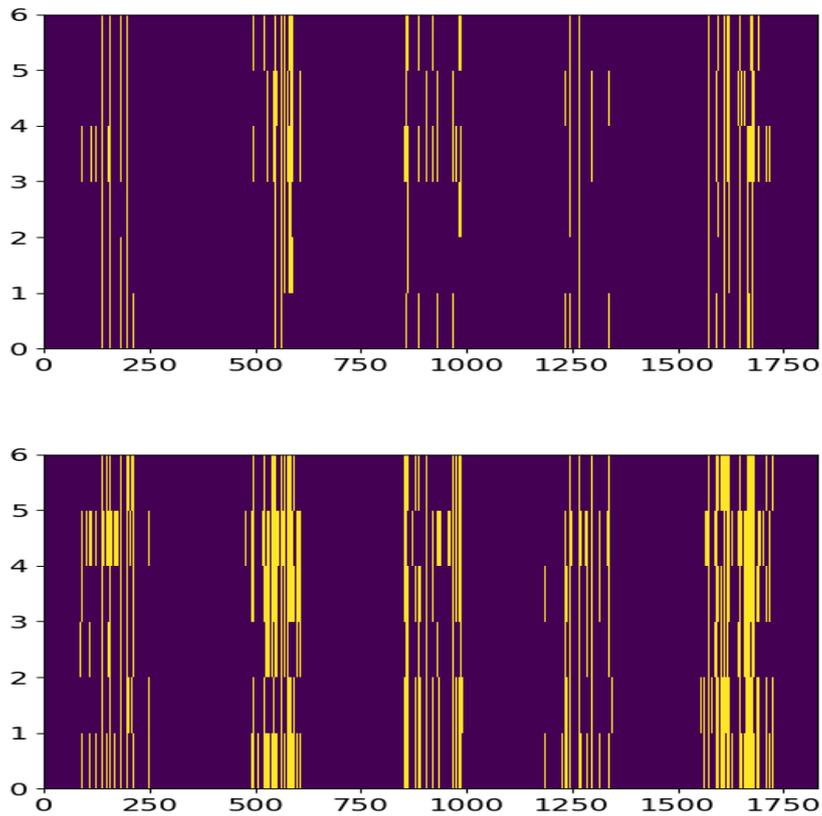
6.2.1 Clusterbildung nach zeitlich zusammenfallenden Ozonüberschreitungen, 2014-2018

Bei diesem Experiment wird bei der Clusterbildung der Tag des Jahres, an dem eine Überschreitung stattfand, berücksichtigt. Die Clusterbildung basiert somit auf täglichen und nicht jährlichen Daten. Außerdem wird hier ein kürzerer Zeitraum (2014 – 2018) berücksichtigt. Dies berücksichtigt sowohl die Anzahl der Überschreitungen als auch die geografische Kohärenz, die durch lokale Emissionen und großräumige meteorologische Bedingungen beeinflusst wird. Die Daten der einzelnen Stationen wurden so strukturiert, dass sie für jeden Tag des Jahres ein Feld mit dem Wert 1 enthielten, wenn der MDA8 > 120 µg/m³ überschritten wurde, und andernfalls 0. Die Anzahl der Cluster wurde zwischen 4 und 9 variiert. Wir gehen davon aus, dass die ideale Anzahl von Cluster bei etwa 6 liegt, um die Stationstypen und Klimazonen mit den Clustern zu repräsentieren. Für einzelne Jahre weisen einige Stationen relativ geringe Überschreitungen auf, so dass die Clusterbildung weniger stabil ist für kurze Perioden. Wenn man mehr als 4 Cluster nimmt, führt dies zu Warnungen für MDA8 > 120 µg/m³ für die Jahre mit mäßigen Ozonüberschreitungen. Bei der Auswahl von 5-Jahres-Perioden wird die Clusterbildung stabiler. Bei einigen Stationen in Norddeutschland kann die Zahl der Überschreitungen in einem Jahr gleich Null sein, und daher werden diese Stationen bei der Datenprüfung für die Clusterbildung nicht berücksichtigt. Aus diesem Grund wird auch die Clusterbildung für MDA8 > 100 µg/m³ angewendet, um die Robustheit der Ergebnisse zu untersuchen.

Die Methode ist in Abbildung 33 und Abbildung 34 für die Jahre 2014 – 2018 dargestellt. Abbildung 33 zeigt die Mittelpunkte der Cluster, d. h. in diesem Fall die Tage mit Überschreitungen, die typisch für diesen Cluster sind (Median). In der Horizontalen sind die Tage mit Überschreitungen gelb markiert und in der Vertikalen sind die verschiedenen Cluster eingezeichnet. In der Tat treten die Überschreitungen zu unterschiedlichen Zeitpunkten für die verschiedenen Cluster auf, so dass die Streifen nicht alle gleich ausgerichtet sind und auch die Anzahl der Überschreitungen unterschiedlich ist. Dies wird durch Abbildung 34 bestätigt, die zeigt, dass die Stationscluster tatsächlich unterschiedliche Regionen aufweisen, mit einigen Ausnahmen. Bei dem Vergleich der beiden Bilder in Abbildung 34 sollte beachtet werden, dass die Gruppen von Clustern verglichen werden sollten, nicht die Clusternummer (Farbe) selbst. Zum Beispiel sind die blauen und gelben Cluster im Westen sehr ähnlich, ebenso wie der Cluster im Süden. Bei MDA8 > 100 µg/m³ gibt es jedoch zwei Cluster im Nordosten, während es bei MDA8 > 120 µg/m³ aufgrund der geringen Anzahl von Überschreitungen nur einen Cluster gibt, der dieses Gebiet abdeckt. Der grüne Cluster für MDA8 > 120 µg/m³ weist keine räumliche Kohärenz auf. Es handelt sich um eine Ansammlung von Tagen mit relativ wenigen Überschreitungstagen, wie in Abbildung 34 dargestellt. Wenn die Ergebnisse für 5 Cluster untersucht werden, werden diese Stationen Teil eines größeren Clusters, aber bei mehr Clustern scheint es relativ robust zu sein.

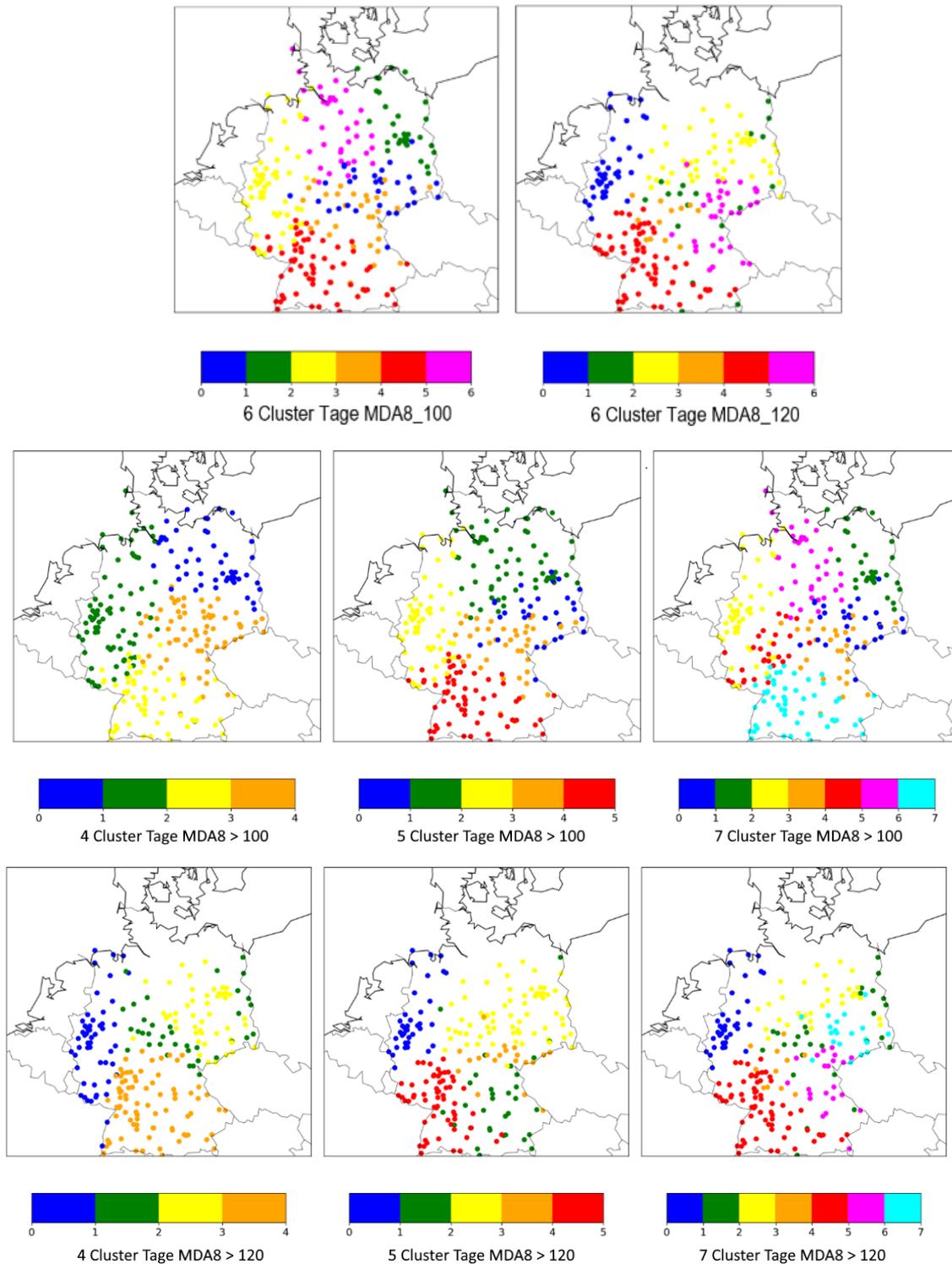
Beim Vergleich der Cluster mit den KLENOS-Zonen zeigt sich, dass die Überschreitungen anders verteilt sind und es eine klare Nord-Süd-Trennung gibt, die in den KLENOS-Zonen nicht zu erkennen war. Die Ost-West-Zonen werden jedoch weitgehend von unseren Clustern nachgeahmt, insbesondere für MDA8 > 100 µg/m³. Unsere Nord-Süd-Gruppierung spiegelt die Korrelationen zwischen der geografischen Breite und z. B. der Anzahl der Tage mit niedriger relativer Feuchte, der windstillen Tage, der Anzahl der Sommertage, der Strahlung und der Stationshöhe wider (Abbildung 30).

Abbildung 33: Centroide der Cluster von Überschreitungstagen für 2014 – 2018. Oberes Feld: MDA8 > 120 µg/m³, unteres Feld MDA8 > 100 µg/m³. Gelb = Überschreitung, lila = keine Überschreitung, auf der x-Achse sind die gesamten 5 Jahre aufgetragen, die y-Achse sind die Cluster.



Quelle: Eigene Darstellung, TNO

Abbildung 34: Cluster für den 5-Jahres-Zeitraum. Clustering an Tagen mit Überschreitungen für MDA8 > 100 µg/m³ und MDA8 > 120 µg/m³. Obere Reihe: 6 Cluster in beiden Fällen, mittlere Reihe 4, 5, 7 Cluster MDA8 > 100 µg/m³ und untere Reihe 4, 5, 7 Cluster MDA8 > 120 µg/m³.



Quelle: Eigene Darstellung, TNO

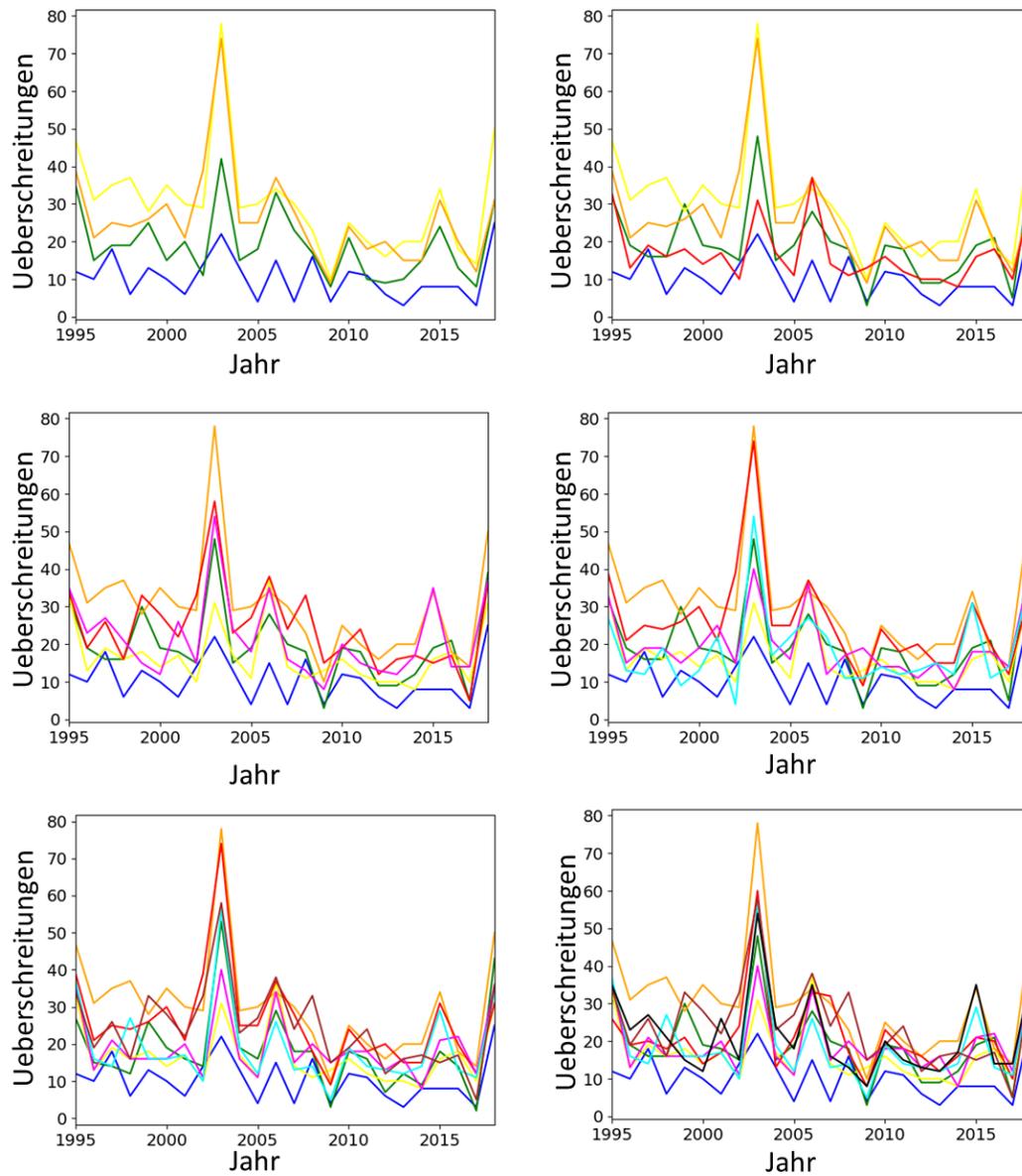
6.2.2 Clusterbildung nach Anzahl der Überschreitungstage, gesamter Zeitraum

In dieser Analyse wurde die Anzahl der Überschreitungen pro Jahr, unabhängig vom Zeitpunkt, über die Stationen geclustert, die die gesamten 24 Jahre der Daten abdecken. Diese Clusterbildung umfasst Reaktionen auf meteorologische Bedingungen und Emissionen. Die Stationen, die miteinbezogen wurden, sind nicht identisch zu denen der vorherigen Methode, weil nicht alle Stationen 24 Jahre aktiv waren. Außerdem werden hier Stationen mit Null-Überschreitungen einbezogen. Die Anzahl der Cluster wurde zwischen 4 und 9 variiert, wobei wir davon ausgehen, dass die ideale Anzahl von Clustern bei etwa 6 liegt, um Stationstypen und Klimazonen zu repräsentieren. Der Datensatz für die Clusterbildung war kleiner als im vorherigen Abschnitt. Geografische Zufälligkeiten werden nicht mehr berücksichtigt. Dennoch ist die Methode in der Lage, einige robuste Merkmale zu finden, wie in Abbildung 35 für die Centroids, d. h. die Anzahl der Überschreitungstage in einem bestimmten Jahr, und Abbildung 36 für die Cluster. Wir können die Gruppe von Stationen in Norddeutschland mit relativ wenigen Ozonüberschreitungen deutlich von dem Cluster im Ruhrgebiet unterscheiden. Bei 6 Clustern ist eine stärkere regionale Differenzierung im Süden Deutschlands zu erkennen, bei sieben und mehr Clustern gibt es jedoch kaum neue Merkmale. Für Berlin und sein direktes Umland gehören Stationen zu mehreren Clustern, da alle Stationstypen beteiligt sind und vom Cluster-Algorithmus aufgrund ihrer Beschaffenheit erkannt werden. Es sind sowohl Verkehrsstationen enthalten, die weniger Überschreitungen aufgrund direkter lokaler Ozontitration aufweisen, als auch Vorortstationen, an denen die Ozonüberschreitungen in der Regel stärker sind, da das in der Stadt emittierte NO₂ Ozon bildet, das schließlich durch lokale biogene VOC-Emissionen verstärkt wird.

Die Ähnlichkeit mit den KLENOS-Klimazonen ist viel geringer als bei der Clusterbildung mit Überschreitungstagen. Auch hier gibt es ein starkes Nord-Süd-Gefälle, um die Auswirkungen der Temperatur zu berücksichtigen. Andererseits berücksichtigt diese Clusterbildung die Art einer Station (Stadt, Straße) auf natürlichere Weise als die Clusterbildung mit Überschreitungstagen.

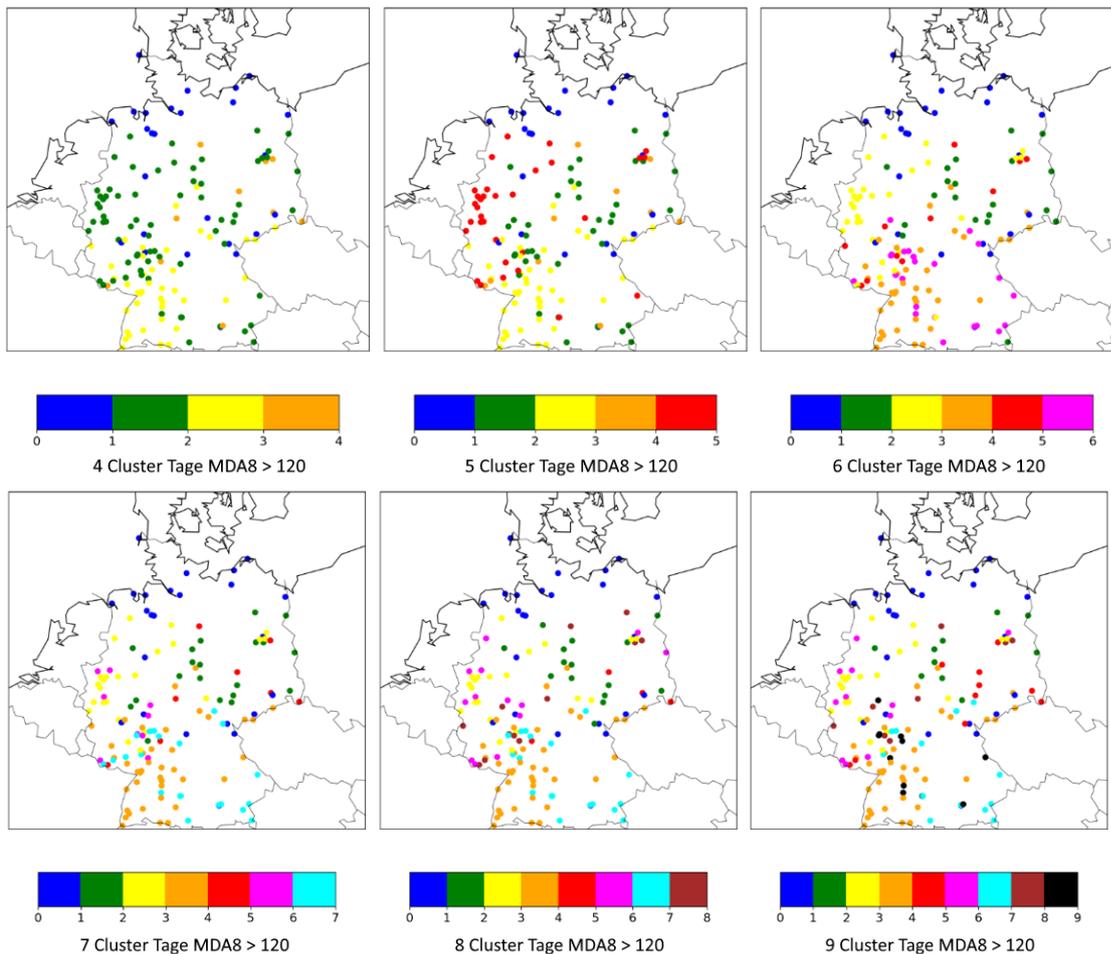
Beim Training des Algorithmus werden die meteorologischen Variablen und die lokalen Emissionen explizit berücksichtigt. Wir werden untersuchen, ob die Einbeziehung dieser Informationen zu einem anderen Ansatz in Bezug auf die Clusterbildung führt.

Abbildung 35: Cluster Centroide für die jährlichen Überschreitungen des MDA8 > 120 µg/m³



Quelle: Eigene Darstellung, TNO

Abbildung 36: Clustering von MDA8 > 120 µg/m³, 1995 – 2018



Quelle: Eigene Darstellung, TNO

Bisher wurde das Clustering auf zwei Dimensionen durchgeführt (Anzahl der Überschreitungen pro Jahr und Zeit oder tägliche Überschreitungen und Zeit). Es ist auch möglich, zusätzliche Informationen in die Clusterbildung einzubeziehen. Dies würde das Clustering zusätzlich auf reine Überschreitungswerte beschränken, was das Clustering stabiler macht, aber auch einige stationsbezogene Signaturen unterdrücken würde. Ein einfach einzubeziehender Parameter wäre der Längen- und Breitengrad der Station. Die Einbeziehung dieser Parameter würde die räumliche Kohärenz des Clusters verbessern, da diese Informationen explizit enthalten sind.

Auch lokale NO_x-Konzentrationen (z. B. als Jahresmittelwert) könnten einbezogen werden. Dies würde die Auswirkungen der ländlichen/städtischen/vorstädtischen Signatur einer Station verstärken, aber die räumliche Korrelation könnte verloren gehen, da der Algorithmus dann dazu neigt, z. B. vorstädtische Stationen in einem Cluster zusammenzufassen. Statistische Parameter wie Mittelwert und Standardabweichung der täglichen MDA8-Werte pro Station würden helfen, Muster zu erkennen, was jedoch bedeutet, dass man über die Beschränkung hinausgeht, nur Überschreitungen zu zählen. Es wäre auch möglich, Klimavariablen, wie die Anzahl der Sommertage, in die Clusterbildung einzubeziehen. Die Wirkung der Klimavariablen könnte jedoch in den verschiedenen Regionen unterschiedlich sein. Der Ansatz, die Klimazonen von KLENOS zu verwenden, ist eine Möglichkeit, Klimacluster zu nutzen.

Es ist zu bedenken, dass diese Informationen, wenn sie in die Clusterbildung einbezogen werden, weniger Einfluss auf die Ausbildung von Ozonüberschreitungen in dem endgültigen Training der Daten innerhalb eines Clusters haben, da die Daten bereits aufgrund ähnlicher Reaktionen auf die Variable ausgewählt wurden. Andererseits würde eine gute Clusterbildung die Aufgabe der Imputation von Daten erleichtern, da die Kriterien, nach denen Daten von anderen Stationen verwendet werden können, klarer sind.

Im endgültigen Ansatz wurde beschlossen, keine Clusterbildung vorzunehmen, sondern eine Trennung der Daten auf der Grundlage der Anzahl der Überschreitungen vorzunehmen. Dies wird im nächsten Abschnitt erläutert.

6.3 Ansätze für die Entwicklung von Algorithmen

Für die Entwicklung des Analysetools wurde das H2O-Paket für Maschinelles Lernen verwendet. H2O ist ein Online-Tool, das über die Programmiersprache R (oder Python) aufgerufen werden kann. Es verfügt über die gängigsten und neuesten ML-Algorithmen sowie über gute vordefinierte Visualisierungen für die Analyse der Ergebnisse. Es ist kostenlos erhältlich und gut dokumentiert. Bei der Erstellung des Tools muss zunächst ein ML-Algorithmus ausgewählt und zugehörige Hyperparameter angegeben werden. Ein Hyperparameter ist ein Parameter, der zur Steuerung des Trainingsalgorithmus verwendet wird und dessen Wert im Gegensatz zu anderen Parametern vor dem eigentlichen Training des Modells festgelegt werden muss.

Hyperparameter sind zum Beispiel die Anzahl der einbezogenen Bäume und die Anzahl der Optimierungszyklen. H2O hat die Möglichkeit, einen Bereich des Hyperparameterraums zu scannen (z. B. Anzahl der Random Trees) und die optimalen Ergebnisse mit der AutoML-Option zu präsentieren. AutoML testet eine Reihe von Algorithmen (GBM, Random Forest, Neuronales Netz) und Hyperparametern (z. B. Anzahl der Bäume) auf den Datensatz. Außerdem konstruiert es Stacked Ensembles aus entweder allen Modellen („Stacked Ensemble“) oder nur aus den besten Modellen ihrer Art („Stacked Ensemble, best of family“). Einzelne Modelle werden anhand der Modellabweichung basierend auf den Validierungsdaten in eine Rangfolge gebracht. Das Modell mit der besten Performance kann ausgewählt werden. Der Vorteil ist, dass auch Personen mit begrenzter Erfahrung im Bereich des Maschinellen Lernens verwendbare Ergebnisse erzielen können, ohne viel Zeit für die Auswahl einer ML-Methode und die Abstimmung der Hyperparameter einer bestimmten Methode aufwenden zu müssen, wenn nicht von vornherein klar ist, welche Methode optimal ist.

Der endgültige Datensatz enthielt die Spalten in Tabelle 6. Für jedes Jahr und jede Station wurde eine Zeile erzeugt. Es wurden keine Imputationen vorgenommen, d. h. es wurden ausschließlich alle vorhandenen gültigen Daten berücksichtigt, was zu mehr als 6000 Daten führte. Die Datenstruktur ermöglicht eine einfache Auswahl von Stationen und Jahren. Bergstationen (Höhe über 900 m) wurden ausgeschlossen. Es ist auch einfach, Stationen aus bestimmten Bundesländern auszuwählen, indem man auf die Identifikatoren im Stationscode klickt. Die Ozondaten enthalten mehrere Stationsklassifizierungen wie in Abschnitt 3.2.1 beschrieben. Die Einbeziehung all dieser Klassifizierungen führte zu Ergebnissen, die im Hinblick auf den Beitrag der Stationsumgebung schwer zu interpretieren waren. In den abschließenden Simulationen wurde beschlossen, die Stationsklassifizierung auf die sechs Kategorien Verkehr, Industrie, Stadt, Vorstadt, ländlicher Hintergrund, ländlich regional zu vereinfachen.

Vor dem Training müssen die Daten normalisiert und kategorische Werte wie die Stationsklassifizierung in numerische Werte umgewandelt werden. Die Normalisierung erfolgt durch Subtraktion des Mittelwerts und Division durch die Standardabweichung. Dieses Verfahren wird auch als Z-Score-Normalisierung oder Standardisierung bezeichnet. Auf diese Weise haben alle Variablen einen Null-Mittelwert und eine Einheitsvarianz. Dadurch erhält jedes Merkmal das gleiche Gewicht in der Regressionsanalyse und kann zu einer schnelleren Minimierung der Kostenfunktion (Konvergenz) einiger Algorithmen für Maschinelles Lernen beitragen.

Um den Algorithmus zu trainieren, wurde der Datensatz der Teilmenge in Trainingsdaten und Validierungsdaten aufgeteilt, wobei 80 % für das Training und 20 % für die Validierung verwendet wurden. Dies wird von H2O nach dem Zufallsprinzip durchgeführt. Der Trainingsdatensatz wird von H2O weiter in Trainings- und Testdatensätze unterteilt, um eine Kreuzvalidierung durchzuführen. Es wird eine zehnfache Kreuzvalidierung verwendet, d. h. der Algorithmus wird zehnmal für verschiedene Teilmengen des ursprünglichen Trainingsdatensatzes trainiert, und die Validierung wird über die Trainings- und Bewertungsrunden gemittelt. Der so trainierte Algorithmus kann auf die 20 % Daten angewendet werden, die für die unabhängige Validierung übrigbleiben.

Tabelle 6: Übersicht der Elemente im Datensatz

Informationen zur Station	Emissionen	Klimakennzahlen	Prädiktoren
Stationscode	Jährliche NO _x -Emission als Summe NO ₂ für ganz Deutschland	Anzahl der Sommertage	Anzahl der Ozonüberschreitungen
Längengrad	Jährliche NMVOC - Emission als Summe für ganz Deutschland	Anzahl der Hitzetage	Anzahl der Ozonepisoden
Breitengrad	Jährliche CH ₄ -Emission als Summe für ganz Deutschland	Anzahl der tropischen Nächte	Maximale Episodendauer
Stationshöhe		Mittlere einfallende kumulative Strahlung	
Stationstyp		Anzahl der Tage mit durchschnittlicher Luftfeuchte < 60%	
Jahr		Anzahl der Tage mit durchschnittlicher Luftfeuchte > 80%	
		Anzahl der Trockentage	
		Anzahl der Tage mit schwachem Wind	
		Anzahl der Tage mit westlicher Strömung	
		Anzahl der Tage mit südlicher Strömung	
		Anzahl der Tage mit östlicher Strömung	

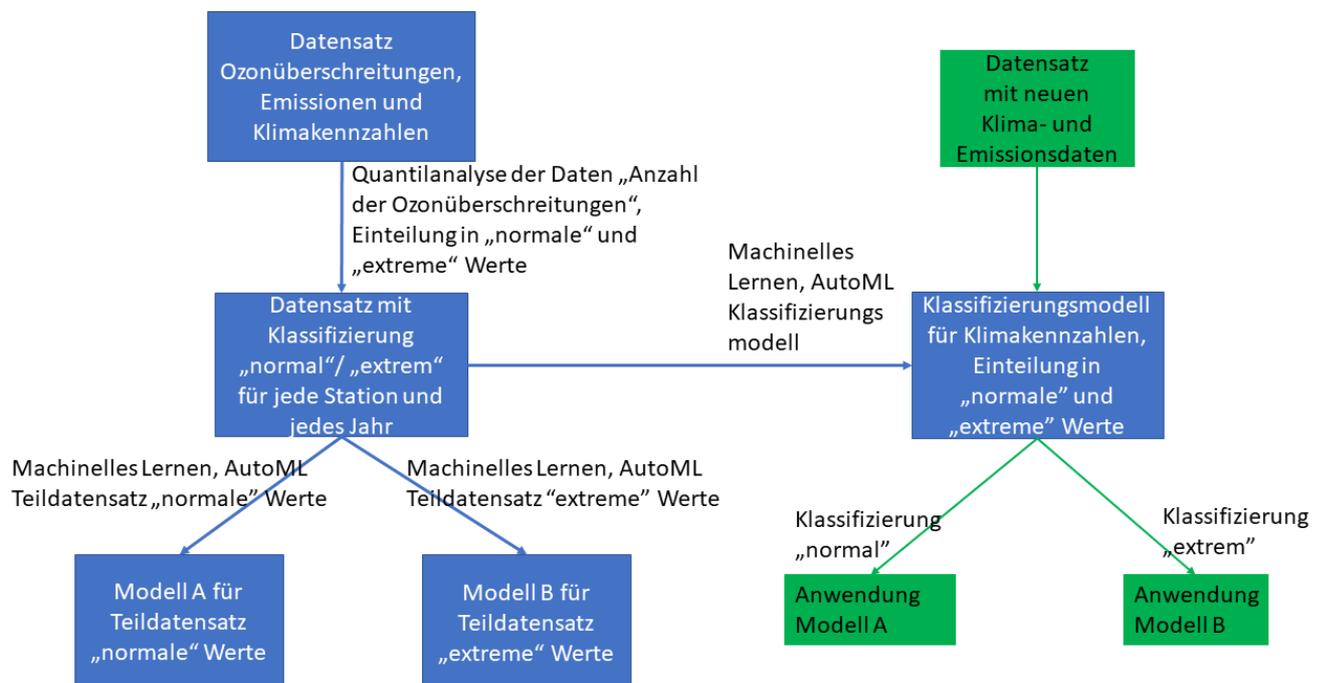
Tabelle 7 zeigt die Experimente, die in dieser Studie durchgeführt wurden. Die ersten beiden Experimente (1, 2) wurden mit mehreren AirBase-Stationsklassifizierungen (jeweils eine der Kategorie ländlich, vorstädtisch, städtisch und eine der Kategorie Hintergrund, Verkehr, Industrie) anstelle der in Abschnitt 3 beschriebenen Klassifizierung einzelner Ozonstationen durchgeführt. Es stellte sich heraus, dass die Einbeziehung aller Jahre in das Training in den Experimenten 1, 2 und 3 zu einer schlechten Performance für die Jahre 2003 und 2018 mit einer sehr hohen Anzahl von Ozonüberschreitungen (Abbildung 23) führte. Es wurden mehrere Ansätze untersucht, um einen Kompromiss zwischen der Genauigkeit im Allgemeinen und einer guten Leistung für diese extremen Jahre zu finden. Beim Training der Algorithmen überwiegt bei der Minimierung der Kostenfunktion (Anpassung der Kurve) die große Anzahl von Trainingsdaten für die „normalen“ Werte gegenüber der geringen Anzahl von „extremen“ Werten (hohe Anzahl von Ozonüberschreitungen pro Jahr). Als Lösung kann man das Sampling der Daten von einer einfachen zufälligen Ziehung von 80 % aller Daten zu einer Auswahl mit

Überrepräsentation von Daten aus 2003 und 2018 ändern (Experiment 4, mehrere Strategien wurden getestet). Auf diese Weise sank die allgemeine Performance des Algorithmus. Dies führte zu der Entscheidung, ein separates Modell für die „normalen“ Werte und eines für die „extremen“ Werte (Experimente 5 und 6) sowie ein Klassifizierungsmodell (Experiment 7) zu erstellen, um neue Daten in die richtige Klasse für die Vorhersage von Ozonüberschreitungen einzuteilen. Dies wird im nächsten Abschnitt näher erläutert. Die endgültige Methode für die Datenauswahl und das Training des Modells ist in Tabelle 7 gezeigt.

Tabelle 7: Übersicht der Experimente

	Experiment	Kommentar	Schlussfolgerungen
1	1995 – 2018 alle Daten, vollständige Stationsklassifizierung	Vollständige Stationsklassifizierung berücksichtigt (Art der Umgebung, Art des Gebiets, Ozonklassifizierung)	Gute Performance während des Trainings, schlechte Performance bei „extremen“ Werten, schwierige Interpretation der Rolle der Stationsklassifizierung
2	2014 – 2018 alle Daten, vollständige Stationsklassifizierung	Geringer Einfluss von Trends in Klima und Emissionsdaten, wenige Stationswechsel	Gute Performance während des Trainings, schlechte Performance für „extreme“ Werte
3	1995 – 2018 alle Daten	Grundlage	Angemessene Performance während des Trainings, weniger gute Performance für „extreme“ Werte
4	1995 – 2018 mit Überstichproben für 2003	Aufnahme mehr hoher Werte für die Anzahl der Ozonüberschreitungen zur besseren Darstellung von „extremen“ Werten	Kompromiss mit besserer Performance für 2003, aber schlechtere Performance im Allgemeinen. Es wurden mehrere Ansätze für Überstichproben von 2003 untersucht, keine war zufriedenstellend
5	1995 – 2018 "normal", Modell A	Ausschluss von sehr hohen Werten für die Anzahl der Ozonüberschreitungen, Datensätze mit „normalen“ Werten im statistischen Sinne	Gute Leistung für „normale“ Werte
6	1995 – 2018 "extrem", Modell B	Nur „extreme“ Werte für die Anzahl der Ozonüberschreitungen im statistischen Sinne	Gute Leistung für „extreme“ Werte
7	Klassifizierung normal/extrem	Modell zur Klassifizierung neuer Daten in „normale“ oder „extreme“ Werte, entscheidet, welches Model (Normal- oder Extremwertmodell) für die Vorhersage angewendet wird	Zuverlässige Trennung der Daten in „normale“ und „extreme“ Werte
8	Klimazonen 1, 2, 3	Modell pro Klimazone (KLENOS)	Angemessene Performance während des Trainings, weniger gute Performance für „extreme“ Werte

Abbildung 37: Flussdiagramm für die Entwicklung und Anwendung der ML-Modelle, unter Berücksichtigung der Klassifizierung der Werte in „normal“ und „extrem“



Quelle: Eigene Darstellung, TNO

6.4 Klassifizierung von Extrem- und Normalfällen

Es wurde beschlossen, anhand der Anzahl der Ozonüberschreitungen automatisch festzustellen, ob eine Station und ein Jahr extrem sind oder nicht. Die Einteilung in extrem/normal erfolgte auf der Grundlage von Statistiken unter Verwendung von Quartilsverhältnissen. Sowohl niedrige als auch hohe Extremwerte wurden in die Klasse der Extremfälle aufgenommen. Alle Stationen und Jahre des Datensatzes wurden einbezogen, und die Quantile der Anzahl der Extreme für den gesamten Datensatz wurden berechnet. Extremwerte bezogen auf die Anzahl der Ozonüberschreitungen wurden definiert als

- Werte über $Q3 + 1,5 \times IQR$
- Werte unter $Q1 - 1,5 \times IQR$

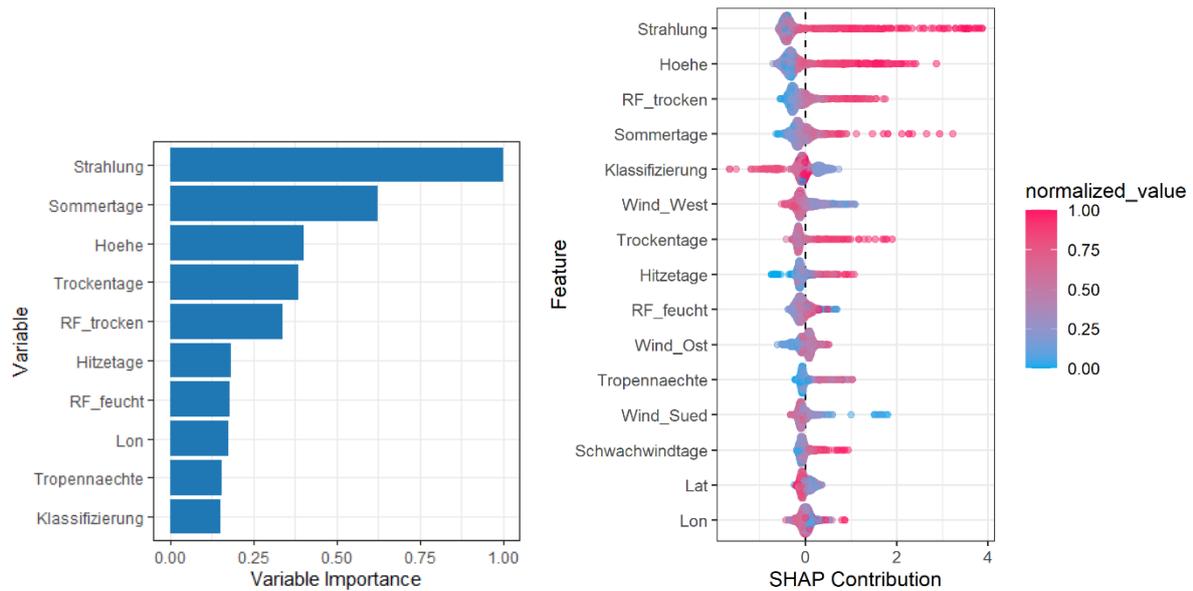
mit $Q1$ erstes Quartil, $Q3$ drittes Quartil und IQR Interquartilsabstand. Dies ist eine weithin akzeptierte Methode zur Definition von Extremen und die Standardeinstellung in der R-Statistikbibliothek RStatix.

Diese Analyse ergab, dass Werte über 54 Ozonüberschreitungen als „extrem“ eingestuft wurden, alle im oberen Teil der Häufigkeitsverteilung gefunden. Es wurden keine Extremwerte gefunden im niedrigen Wertebereich. Der vollständige Datensatz besteht aus 6589 Daten (ohne die Bergstationen), der Datensatz für „normale“ Werte aus 6399 Daten und der Datensatz für „extreme“ Werte aus 190 Daten. Die begrenzte Anzahl von „extremen“ Werten macht es schwieriger, robuste statistische Beziehungen zu finden. Andererseits erhöht das Wissen, dass ein Datenpunkt zu den „extremen“ Werten gehören würde, bereits die Genauigkeit der Vorhersage erheblich im Vergleich zu einem allgemeinen Modell, das alle Daten umfasst.

Wenn für künftige Jahre eine Vorhersage der Ozonüberschreitung gemacht werden soll, ist eine Trennung auf der Grundlage der Anzahl der Ozonüberschreitungen nicht möglich, da dies die Vorhersagegröße ist. Keine einzelne Klimakennzahl (z. B. die Anzahl der Sommertage) stand in eindeutigem Zusammenhang mit den Extremwerten. Mit Hilfe des Maschinellen Lernens kann interpretiert werden, welche Klimakennzahlen dazu beitragen, dass ein Stationswert „extrem“ wird. Daher muss ein Klassifizierungsmodell für die Klimakennzahlen erstellt werden (Experiment 7), bevor der entsprechende trainierte Algorithmus zur Vorhersage der Ozonüberschreitungen unter Verwendung dieser Klimakennzahlen angewendet werden kann. Im vorliegenden Bericht wurde dieser Ansatz entwickelt, aber nicht auf die Daten angewandt, da alle Daten in das Training und Bewertung der Modelle für „normale“ und „extreme“ Werte einbezogen wurden.

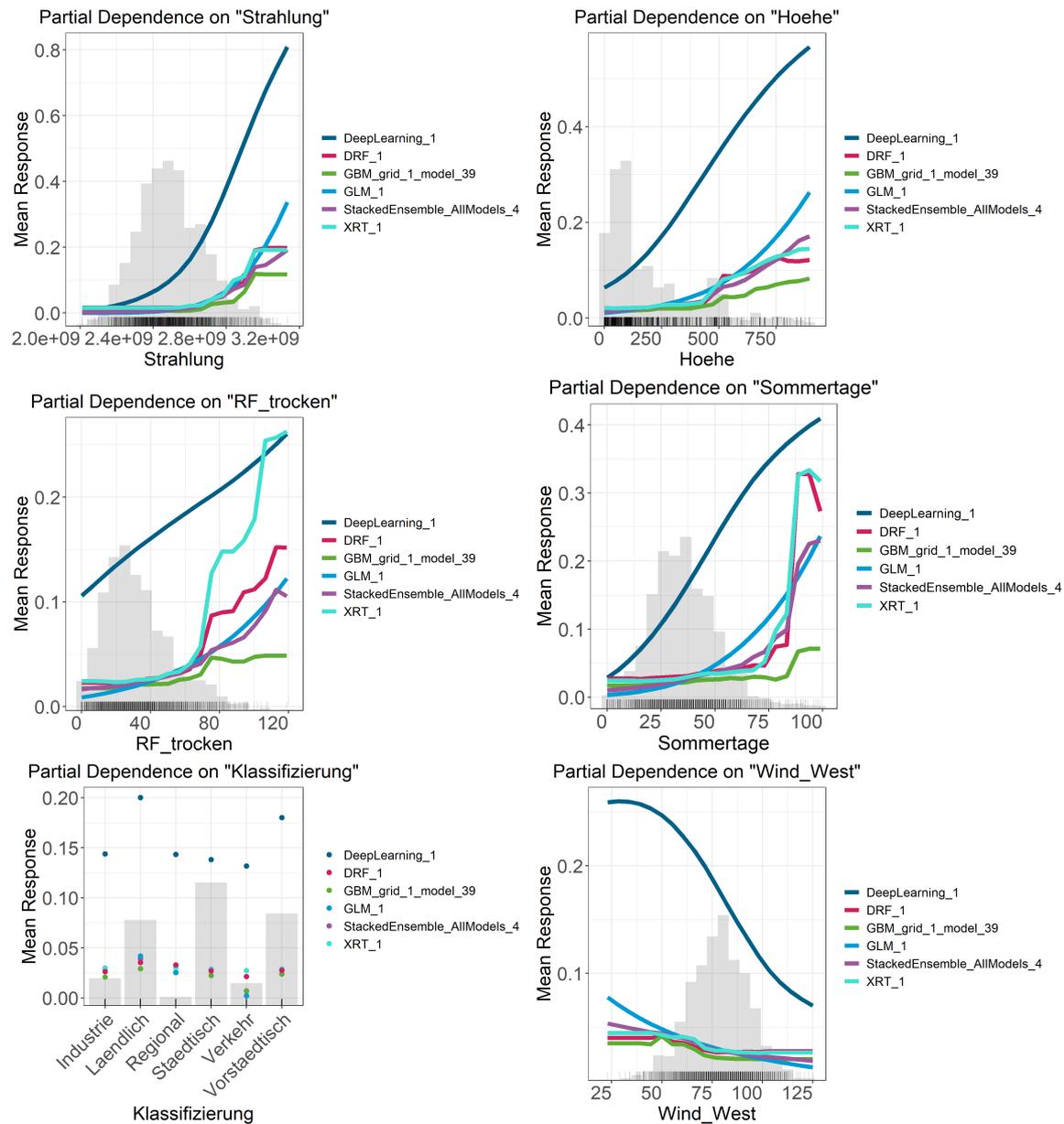
Für das Klassifizierungsmodell (Abbildung 37) wurden die Emissionen bei der Klassifizierung nicht berücksichtigt, da die Jahre 2003 und 2018 in den Daten überrepräsentiert waren. Die Einbeziehung der Emissionen würde dazu führen, dass die Emissionen ein wichtiger Klassifikator wären, was unerwünscht wäre. Die endgültige Klassifizierung hat eine sehr gute Klassifizierungsgenauigkeit von 96 % ($100 \cdot \text{Anzahl der korrekten Klassifizierungen} / \text{Anzahl der Klassifizierungen}$). Das beste Modell war ein Ensemblemodell, was sich schwierig interpretieren lässt. In Abbildung 38 sind die Variable Importance Plots und SHAP-Plots für das GBM-Modell mit der besten Performance dargestellt. Diese Abbildung zeigt, dass die Strahlung der dominierende Diskriminierungsfaktor ist. Es ist jedoch auch ersichtlich, dass die Reihenfolge der Variablen in dem Variable Importance Plot und dem SHAP-Plot unterschiedlich sind. Der Variable Importance Plot wird auf eine grundlegend andere Weise berechnet. Die in einem Variable Importance Plot dargestellten Variablenbedeutungen sind bei Variablen mit nicht linearen Wechselwirkungen weniger stabil als bei einem SHAP-Plot, bei dem die Beiträge der Variablen gründlicher gewichtet werden, indem viele Zufallsziehungen aus dem gesamten Datensatz vorgenommen werden, um den Einfluss einer Variablen zu quantifizieren. Es wurde auch festgestellt, dass die SHAP-Plots besser reproduzierbar (robuster) sind als die Variable Importance Plots. Die Analyse zeigt, dass die Unterscheidung eines „extremen“ Wertes nicht auf der Grundlage einer einzelnen Klimakennzahl, wie der Anzahl der Sommertage, erfolgen kann. Dies ist jedoch auf der Grundlage der Streudiagramme zu erwarten (Abbildung 31), die für Anzahl der Sommertage, Strahlung und Anzahl der RF_trocken-Tage eine Korrelation zwischen den Variablen und der Anzahl der Überschreitungen aufzeigten, aber mit erhebliche Streuung, und nur eine schwache Korrelation für z. B. Anzahl der Schwachwindtage. Dominierende Faktoren sind die Strahlung und die Anzahl der Tage mit niedriger relativer Luftfeuchte (RF_trocken), abgesehen vom Stationstyp und der Stationshöhe. Die Stationsklassifizierung spielt in dem SHAP-Plot eine wichtigere Rolle als in dem Variable Importance Plot, weil der Einfluss in den Beiden anders quantifiziert wird. Es wurde eine hohe Anzahl von Überschreitungen für ländliche Stationen gefunden, nicht aber für Verkehrsstationen. Die Partial Dependence Plots in Abbildung 39 zeigen, wie sich die verschiedenen Modelle des Ensembles und das Modell mit der besten Leistung verhalten. Für die Strahlung zeigen alle Modelle (mit Ausnahme des Deep-Learning-Modells) ein ähnliches Verhalten, für die Anzahl der Sommertage und die Tropennächte ist das Bild uneinheitlicher.

Abbildung 38: Beiträge der Variablen für das Klassifizierungsmodell normale/extreme Anzahl von Ozonüberschreitungen (bestes GBM Modell), Variable Importance Plots (links) und SHAP Plot (rechts)



Quelle: Eigene Darstellung, TNO

Abbildung 39: Partial Dependence Plots für das Klassifizierungsmodell



Quelle: Eigene Darstellung, TNO

6.5 Zusammenfassung der Ergebnisse für Ozonüberschreitungen und Episoden

Die beste Performance wurde für den kürzeren Zeitraum (2014 – 2018, Experiment 2) mit einem RMSE-Wert von 5,67 bei den Testdaten gefunden, was angesichts des Bereichs der Ozonüberschreitungen im Datensatz, der 0 – 100 beträgt, mit den meisten Werten zwischen 0 und 55, gut ist. Die Anwendung dieses Algorithmus auf das Extremjahr 2003 mit vielen Daten über 55 führte jedoch zu einer sehr schlechten Leistung (RMSE von 32,46). Bei Verwendung des gesamten Zeitraums für das Training (siehe auch Anhang B) war der RMSE insgesamt recht gut (RMSE 6,46 für Experiment 3), aber auch hier wurden die Extremwerte zu niedrig vorhergesagt. Die Verwendung mehrerer Subsampling-Methoden mit Oversampling (Überstichproben) von Daten für die Jahre 1995, 2003, 2006 und 2018, um relativ mehr Extreme in den Trainingssatz

aufzunehmen, würde zu einem Modell führen, das für den Großteil der Daten schlechter abschneiden würde als das oben genannte. Für die höchsten Werte würde dieses Modell mit RMSE-Werten der Validierungsdaten zwischen 8,11 und 9,10 zwar besser abschneiden. Diese Werte sind zwar höher als bei den Experimenten 2 und 3, dennoch würde der RMSE für extreme Jahre zwischen 12,06 und 16,07 schwanken. Bei der Erstellung getrennter Modelle für Normal- und Extremwerte waren keine Kompromisse mehr nötig. Die RMSE-Werte für „normale“ Werte (Datenbereich 0 – 54) lagen bei 5,99 und für „extreme“ Werte (Datenbereich 55 – 100) bei 8,68, was deutlich besser ist als die Leistung des allgemeinen Modells. Darüber hinaus wurden Algorithmen für die drei KLENOS-Klimazonen ohne Trennung von „extrem“- und „normal“-Werten trainiert (siehe Anhang B). Die allgemeine Performance war recht gut (RMSE zwischen 5,5 und 7 mit dem niedrigsten Wert für Zone 3 und dem höchsten für Zone 1), aber die Extremwerte werden mit diesem Modell immer noch stärker unterschätzt als mit dem Modell B für Extremwerte.

Eine Zusammenfassung der Performance des endgültigen Modells für „extreme“ und „normale“ Bedingungen, welches Daten aus dem gesamten Zeitraum 1995 – 2018 abdeckt, findet sich in Tabelle 8. Diese Tabelle enthält auch eine Zusammenfassung für die Anzahl der Episoden. Für die Anzahl der Episoden wurde die gleiche Stationsauswahl wie für die Anzahl der Ozonüberschreitungen verwendet. In allen Fällen war ein gestapeltes Ensemble das Modell mit der besten Leistung. Für das Stacked Ensemble war es nicht möglich, Variable Importance Plots und SHAP-Plots zu erstellen. Daher wird in der Analyse der Algorithmen das GBM-Modell mit der besten Performance in den nächsten Abschnitten und in Anhang B vorgestellt.

Tabelle 8: Leistung der Modelle für Überschreitungen und Anzahl der Episoden, unter Verwendung separater Algorithmen für normale und extreme Bedingungen

Bedingung	Beste Leistung	RMSE	Mittlerer absoluter Fehler
Normal, Überschreitungen	Stacked Ensemble, alle Modelle	5.99	4.42
Extrem, Überschreitungen	Stacked Ensemble, best of family	8.68	7.04
Normal, Episoden	Stacked Ensemble, alle Modelle	1.65	1.26
Extrem, Episoden	Stacked Ensemble, best of family	1.80	1.46

6.6 Modell A: normale Bedingungen, Anzahl der Ozonüberschreitungen

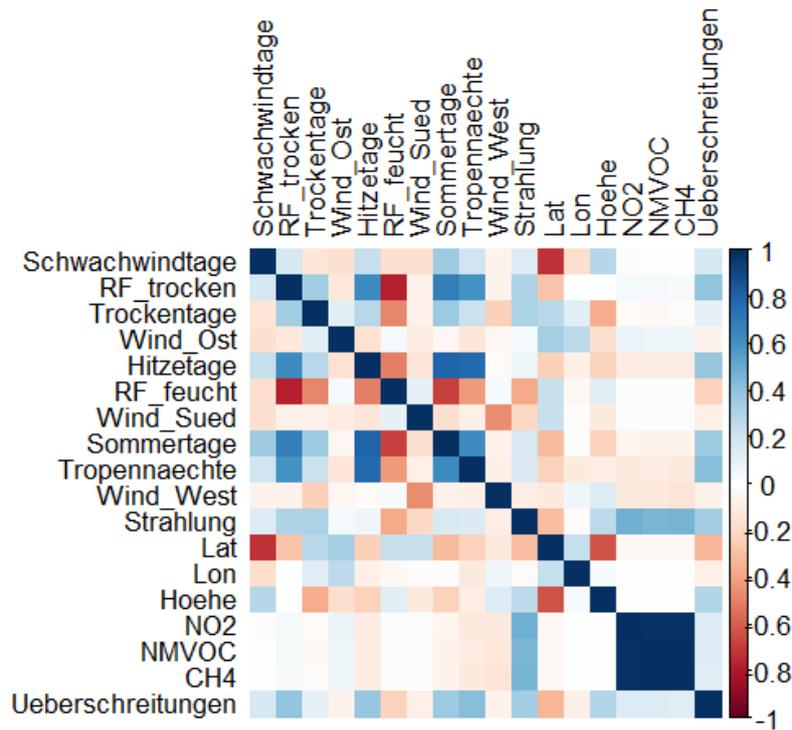
Die zusammenfassende Abbildung der Korrelationen (Abbildung 40) zeigt starke positive Korrelationen zwischen der Anzahl der Ozonüberschreitungen und der Anzahl der Tropennächte, der Anzahl der Sommertage, der Anzahl der Hitzetage und der Strahlung. Die stärkste negative Korrelation besteht mit dem Breitengrad. Die Korrelationen zu den Windrichtungen sind relativ schwach. Auch Korrelationen zwischen den eingehenden Variablen sind erkennbar, wie z. B. eine positive Korrelation zwischen den Temperaturvariablen und eine negative Korrelation zwischen der Anzahl der Schwachwindtage und dem Breitengrad sowie eine negative Korrelation zwischen den Temperaturvariablen und den Feuchtevariablen.

Abbildung 41 und Tabelle 8 zeigen, dass für diesen Fall („normale“ Werte, Anzahl der Ozonüberschreitungen) ein gutes Trainingsergebnis mit einem RMSE von 5,99 erzielt wurde. Der Boxplot zeigt die Verteilung der Anzahl der Überschreitungen vor Verteilung auf Trainings- und Validierungsdaten mit Werten überwiegend im Bereich 10 – 28. Bei den höchsten Werten unterschätzt das Modell die tatsächliche Anzahl der Ozonüberschreitungen meist leicht. Die Residuen zeigen ein zufälliges Muster um 0 (Abbildung 42) mit den künstlichen Streifenmustern, die auf die ganzzahligen Werte zurückzuführen sind, wie in Abschnitt 5.3 erläutert. Das Histogramm in dieser Abbildung zeigt auch, dass die Residuen meist um 0 sind und symmetrisch verteilt sind.

Die trainierten Modelle wiesen eine hohe Korrelation zwischen den Vorhersagen der einzelnen Modelle auf (Abbildung 43) und die Modelle mit der besten Performance stimmten in Bezug auf die Performancestatistiken und die Variablenabhängigkeit gut überein. Die Ähnlichkeit der Performance der unabhängig gestalteten Modelle indiziert, dass die Modelle und ihre Ergebnisse robust sind.

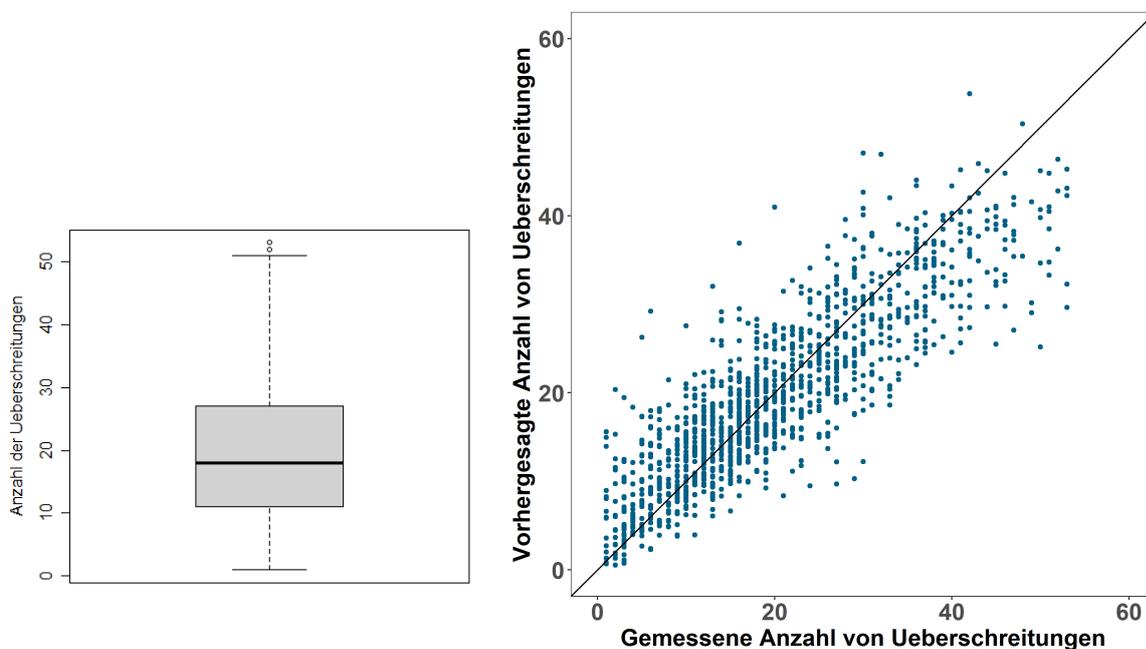
Für „normale“ Werte ist die Variable Anzahl der Tropennächte eine der wichtigsten, mit ebenfalls wichtigen Beiträgen der Variablen Strahlung und Anzahl der Hitzetage (Abbildung 44, Abbildung 46). Die Stationsklasse und die Stationshöhe spielen eine wichtige Rolle. Die Emissionen sind weniger einflussreich, ebenso wie die Variablen bezogen auf die Feuchte und den Wind. Die Partial Dependence Plots (Abbildung 45) zeigen, dass die Abhängigkeiten teilweise Stufen aufweisen, die mit Sprüngen in der Anzahl der Daten in dem jeweiligen Bereich korrelieren. Eine bessere Übereinstimmung zwischen den verschiedenen Modellen wurde dagegen für Bereiche mit guter Datenabdeckung gefunden. Der Einfluss der Stationsklasse ist deutlich, wobei die niedrigsten Werte für Verkehrsstationen und die höchsten Werte für ländliche Stationen ermittelt wurden. Die Korrelation der Tropennächte ist in allen Modellen groß, wie aus der Korrelation der Variablen (Abbildung 40) zu erwarten ist. Obwohl die Anzahl der Tropennächte und die Anzahl der Hitzetage stark korreliert sind, tragen beide zum Endergebnis bei (Abbildung 44, Abbildung 46), während die Anzahl der Sommertage weniger wichtig ist. Die Abhängigkeit vom Breitengrad spiegelt nicht nur die Temperaturbedingungen wider, sondern auch die Anzahl der Schwachwindtage, die in niedrigeren Breitengraden viel höher ist. Die Anzahl der Ozonüberschreitungen nimmt mit den NO_x - und VOC-Emissionen nur sehr geringfügig zu, aber für die niedrigsten und höchsten Emissionswerte zeigen die Linien vor allem für NO_x eine rasche Veränderung mit der Veränderung der Emissionen, was darauf hindeutet, dass die Modelle in diesem Bereich weniger zuverlässig sind und es schwierig ist, die Ergebnisse für die Emissionen zu extrapolieren. Abbildung 45 gibt einen Eindruck von dem variablen Einfluss. Obwohl das GBM-Modell nicht das beste Modell im absoluten Sinne ist, zeigen die Partial Dependence Plots, dass die Abhängigkeiten der Variablen im GBM sehr ähnlich sind zu denen des Ensemble Modells (beste Performance). Die Analyse der Variable Importance und SHAP Plots (Abbildung 46) des GBM Modells unterstützt somit die Analysen von Modell A.

Abbildung 40: Korrelation der Variablen, Anzahl der Ozonüberschreitungen für „normale“ Bedingungen, Modell A, Experiment 5



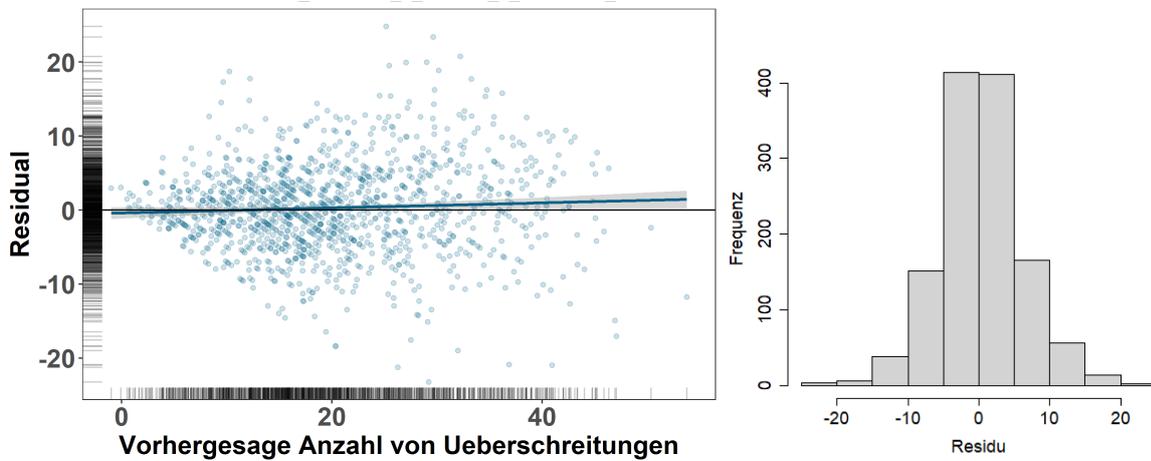
Quelle: Eigene Darstellung, TNO

Abbildung 41: Boxplot der gemessenen Anzahl der Ozonüberschreitungen (links) und vorhergesagte versus gemessene Anzahl der Ozonüberschreitungen (rechts) für „normale“ Bedingungen Modell A, Experiment 5



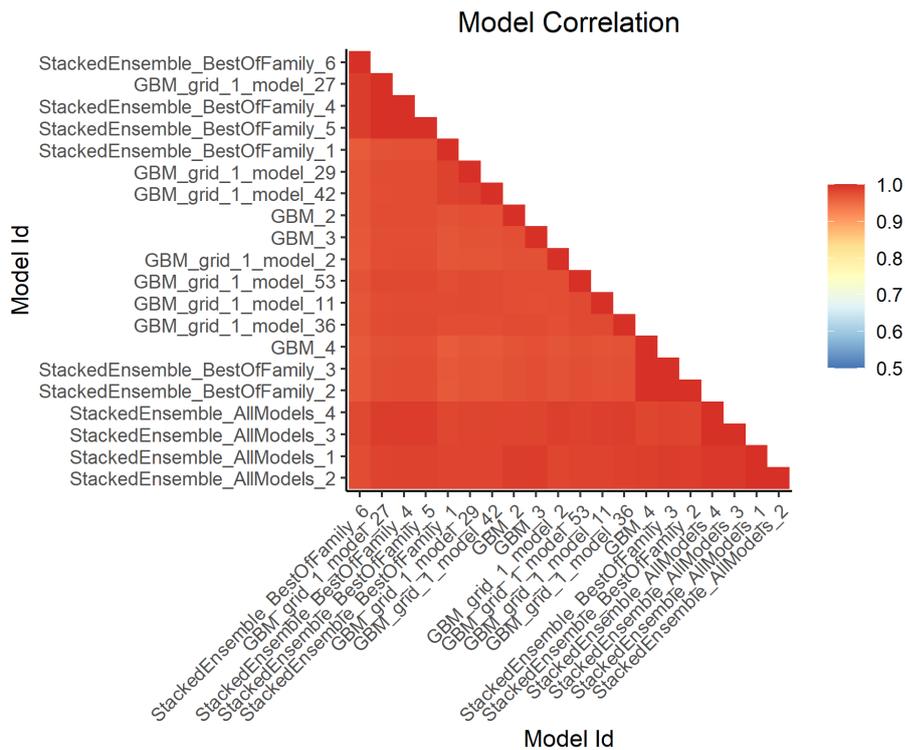
Quelle: Eigene Darstellung, TNO

Abbildung 42: Residuenanalyse für die Anzahl der Ozonüberschreitungen für „normale“ Bedingungen, als Streudiagramm mit vorhergesagten Werten (links) und als Histogramm (rechts)



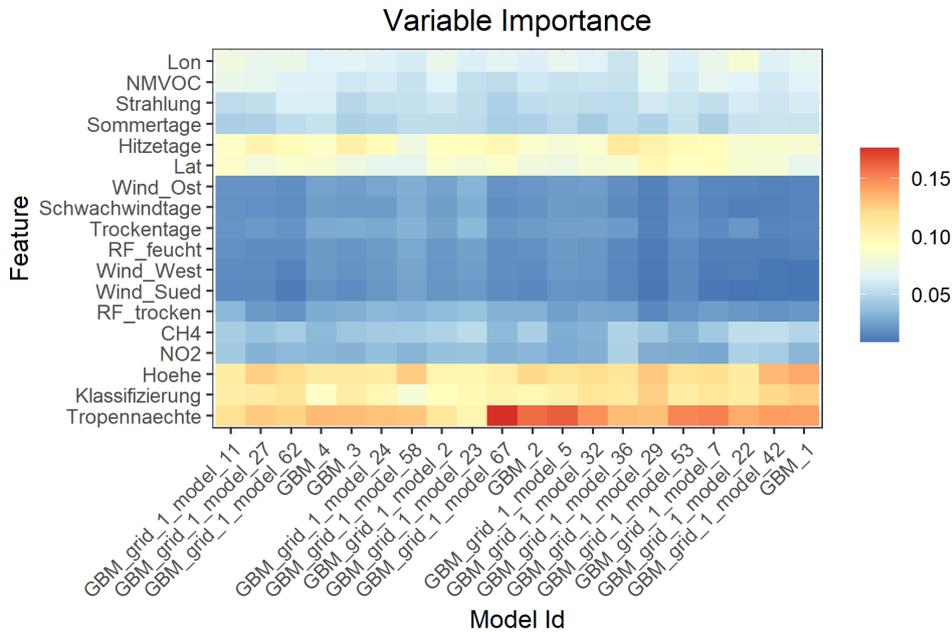
Quelle: Eigene Darstellung, TNO

Abbildung 43: Model Prediction Correlation Heatmap, Anzahl der Ozonüberschreitungen für "normale" Bedingungen, Modell A, Experiment 5



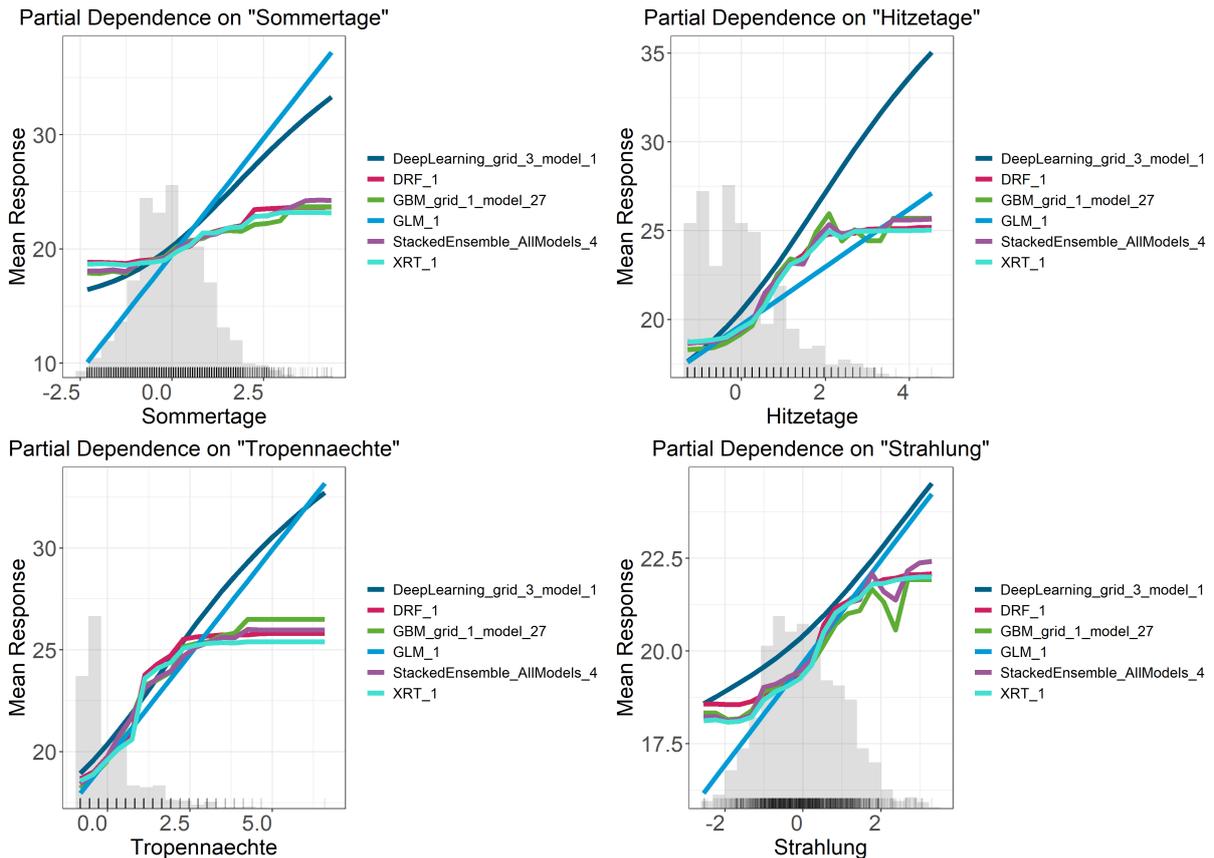
Quelle: Eigene Darstellung, TNO

Abbildung 44: Variable Importance Heatmap, Anzahl der Ozonüberschreitungen für „normale“ Bedingungen, Modell A, Experiment 5

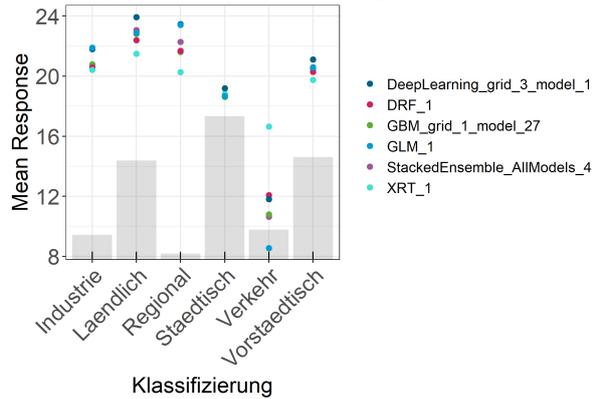


Quelle: Eigene Darstellung, TNO

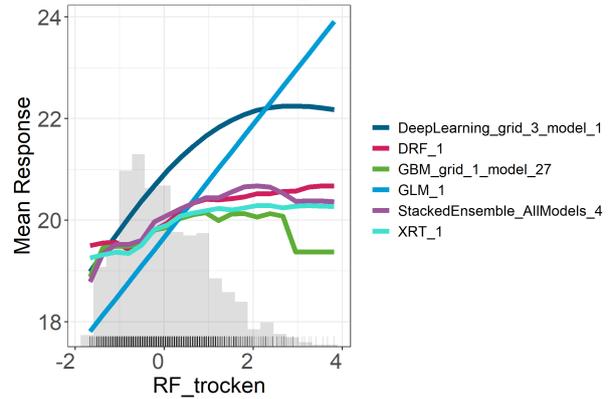
Abbildung 45: Partial Dependence Plots, Anzahl der Ozonüberschreitungen für „normale“ Bedingungen, Modell A, Experiment 5



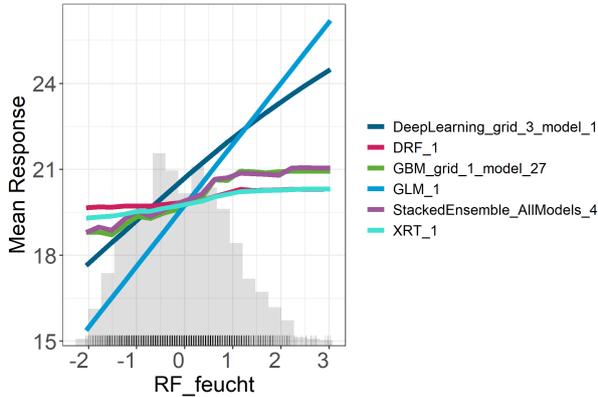
Partial Dependence on "Klassifizierung"



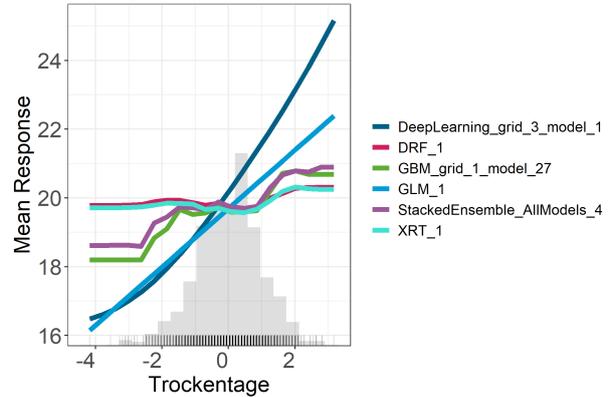
Partial Dependence on "RF_trocken"



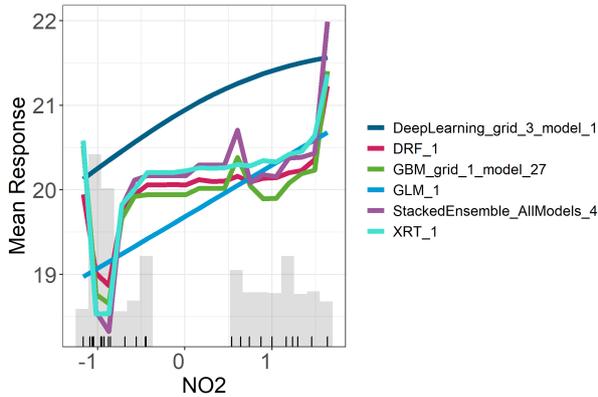
Partial Dependence on "RF_feucht"



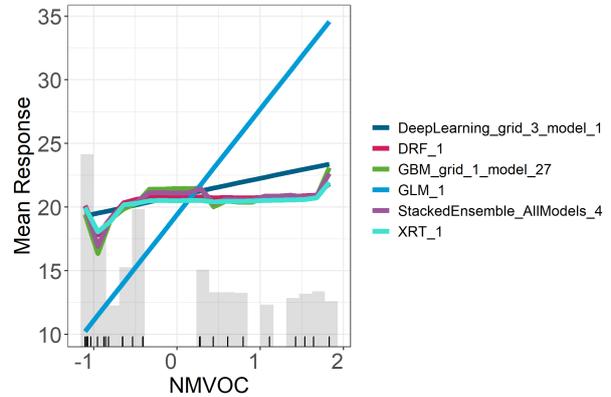
Partial Dependence on "Trockentage"



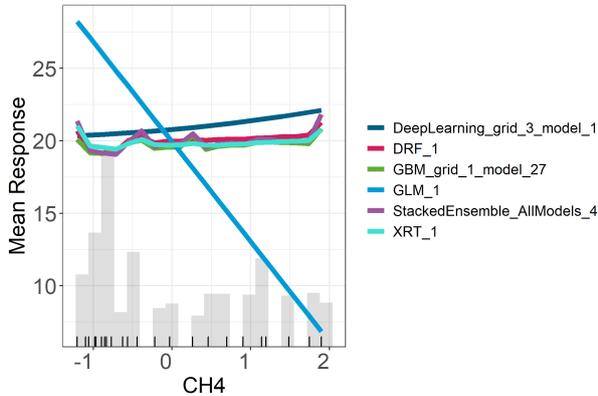
Partial Dependence on "NO2"



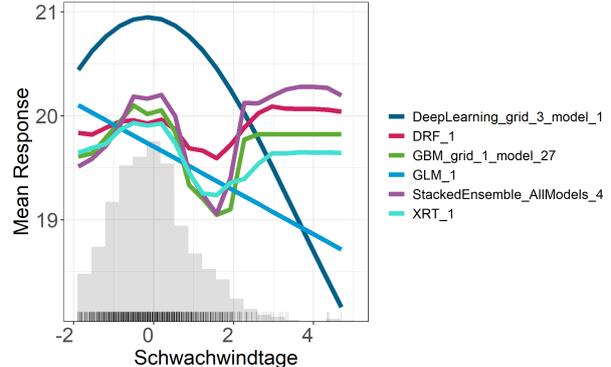
Partial Dependence on "NMVOC"

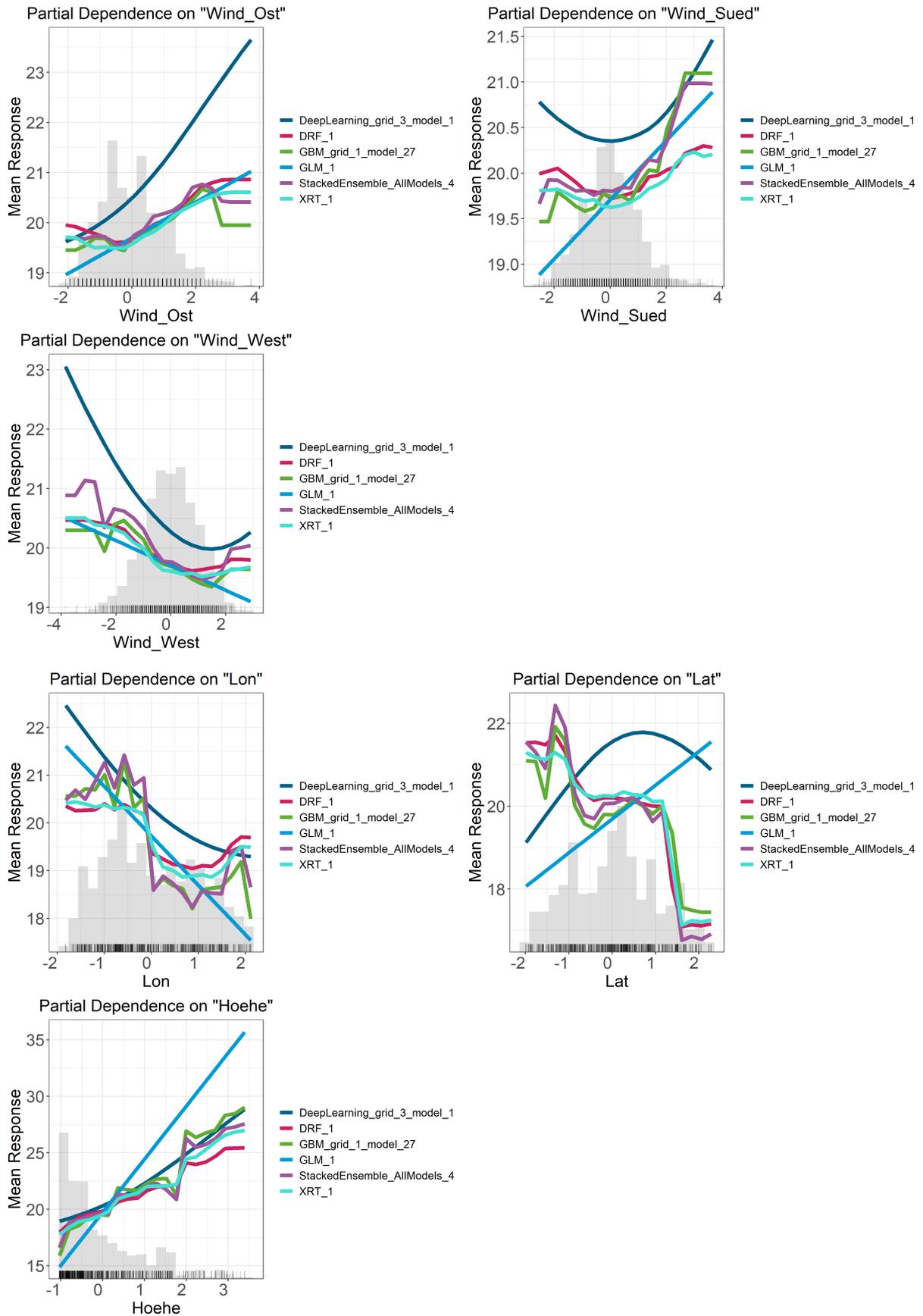


Partial Dependence on "CH4"



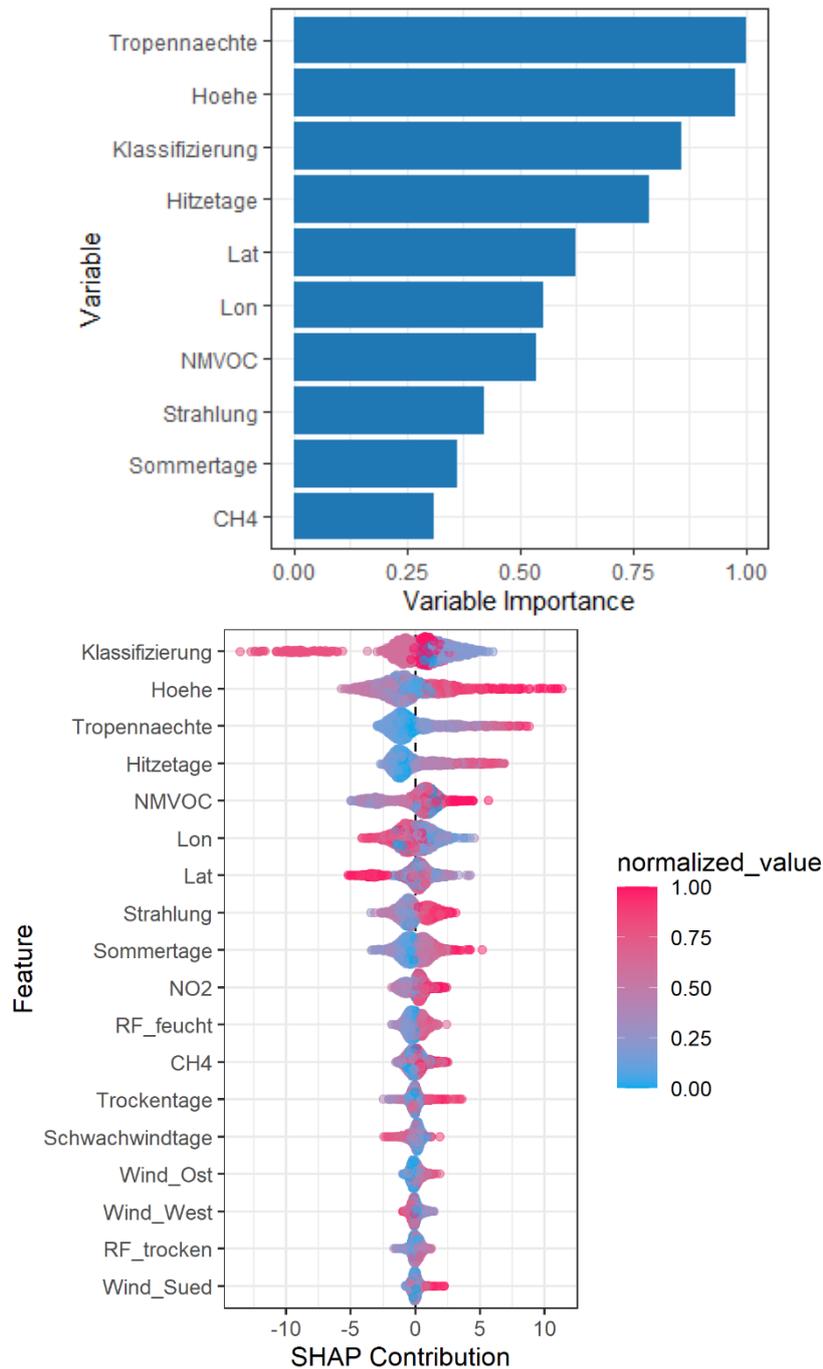
Partial Dependence on "Schwachwindtage"





Quelle: Eigene Darstellung, TNO

Abbildung 46: Variable Importance Plot (oben) und SHAP-Plot (unten) für das GBM-Modell mit der besten Performance, Anzahl der Ozonüberschreitungen für „normale“ Bedingungen, Modell A, Experiment 5



Quelle: Eigene Darstellung, TNO

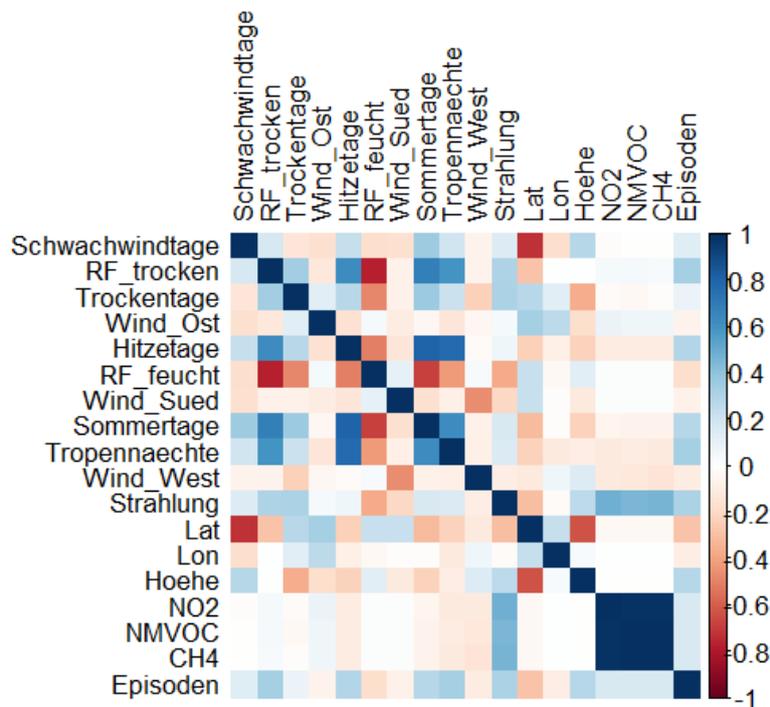
6.7 Modell A: normale Bedingungen, Anzahl der Ozonepisoden

Die Anzahl der Ozonepisoden steht in einem positiven Zusammenhang mit der Anzahl der Tage mit einer relativen Feuchte < 60 % (RF_trocken), den Temperaturvariablen, der Strahlung und der Stationshöhe sowie in einem schwach positivem Zusammenhang mit den Schwachwindtagen, Anzahl der Trockentage und der Emissionsvariablen (Abbildung 47).

Negative Korrelationen gab es für den Breitengrad, Anzahl der Tage mit einer relativen Feuchte

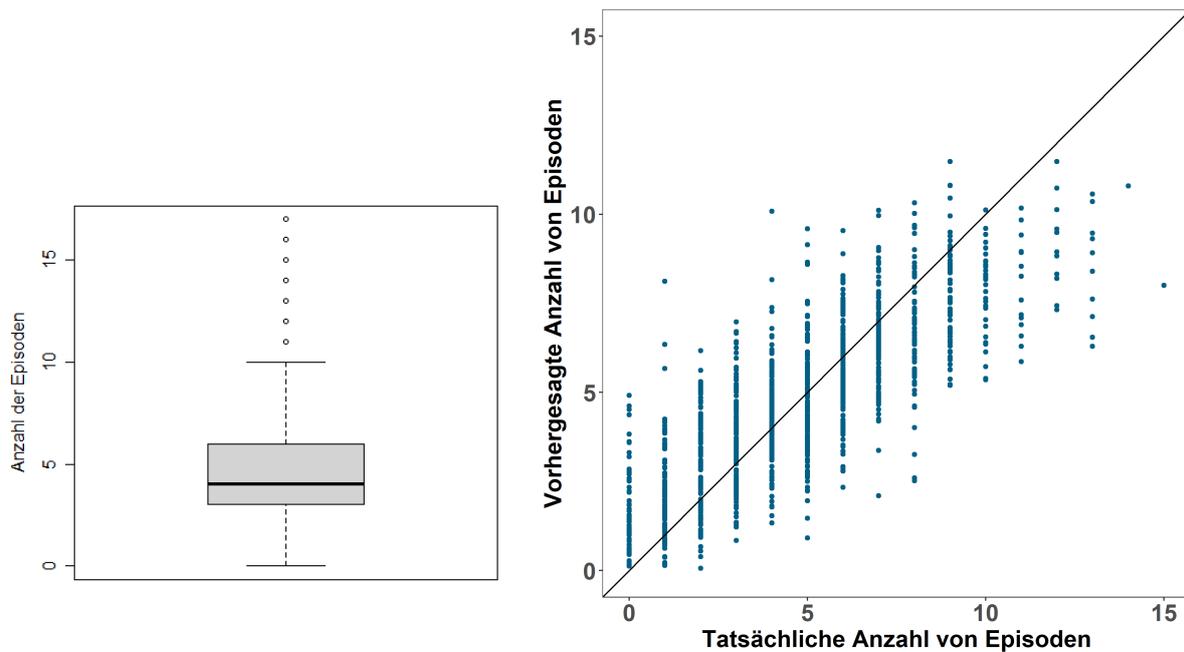
> 80 % (RF_feucht) und dem Längengrad (schwach). Korrelationen für Windrichtung waren sehr schwach. Für die Anzahl der Ozonepisoden wurde ein RMSE von 1,65 ermittelt, was angesichts der Bandbreite der gemessenen Anzahl von Ozonepisoden eine gute Performance darstellt. Abbildung 48 zeigt eine leichte Unterschätzung der höchsten Werte für die Anzahl von Ozonepisoden sowie eine Überschätzung für die niedrigsten. Auch die Residuen zeigen ein angemessenes Verhalten (Abbildung 49). Alle Modelle sind stark korreliert (Abbildung 50). In der Variable Importance Heatmap in Abbildung 51 ist zu sehen, dass alle Modelle die gleichen Variablen als wichtigste kennzeichnen, mit leichten Schwankungen zwischen den Modellen. Die Variable Importance Heatmap, SHAP Plot (Abbildung 53) und Variable Importance plot für das beste GBM-Modell zeigen die wichtigsten Variablen. Für „normale“ Werte sind der Stationsstandort (Breiten- und Längengrad, Stationshöhe) und die Stationsklassifizierung die dominierenden Variablen, gefolgt von den temperaturbezogenen Variablen und der Strahlung. SHAP Plot und Variable Importance Plot zeigen kleine Unterschiede in der Rangfolge. Die Partial Dependence Plots (Abbildung 52) zeigen, dass sich die Abhängigkeit, wie bei der Anzahl der Überschreitungen, für die niedrigsten und höchsten Emissionsbereiche schnell ändert, und die Aussagekraft in diesem Bereich kritisch betrachtet werden soll. Emissionen von CH₄ sind hier aber einflussreicher als für Anzahl der Überschreitungen. Variablen bezogen auf den Wind und die Luftfeuchte spielen auch hier eine untergeordnete Rolle, möglicherweise aufgrund ihrer Korrelation mit der Temperatur und dem Standort der Station.

Abbildung 47: Korrelation der Variablen, Anzahl der Episoden für „normale“ Bedingungen



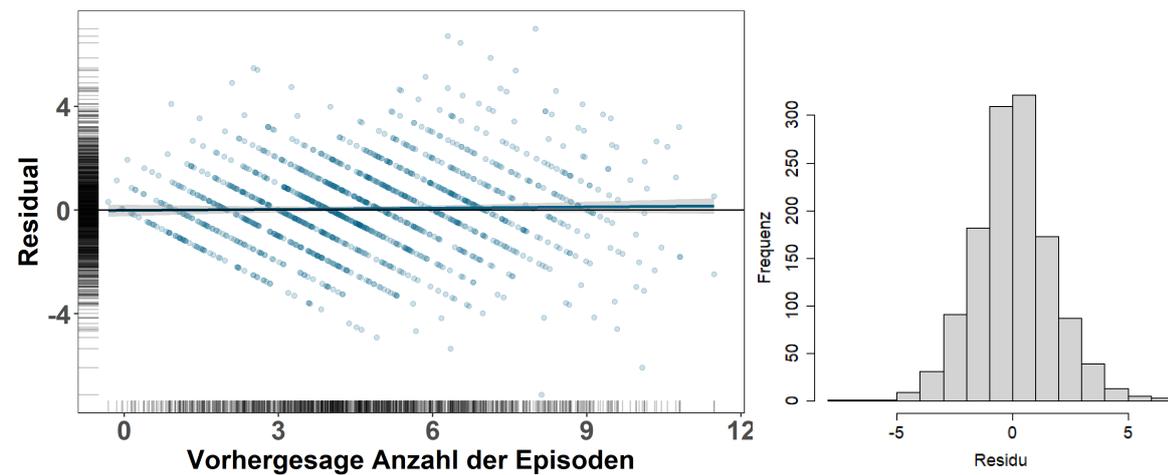
Quelle: Eigene Darstellung, TNO

Abbildung 48: Boxplot (links) und vorhergesagte gegenüber der gemessenen Anzahl der Episoden (rechts) für „normale“ Bedingungen



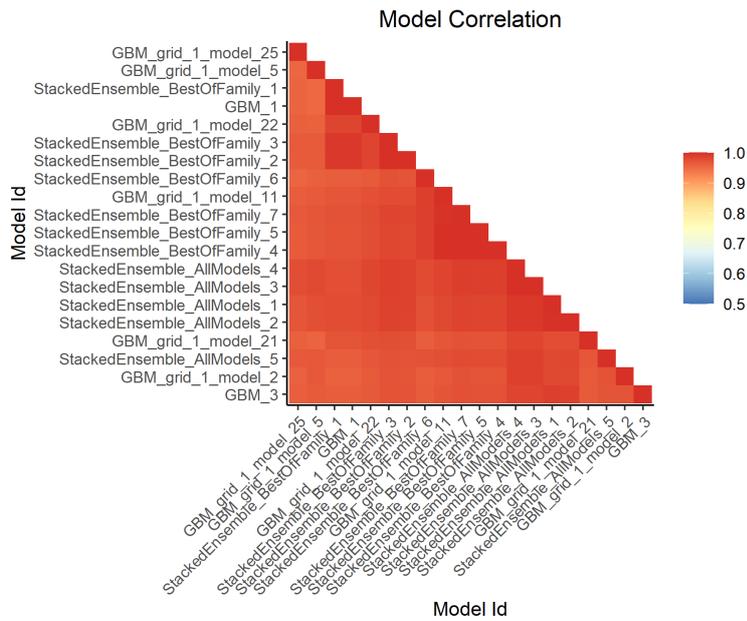
Quelle: Eigene Darstellung, TNO

Abbildung 49: Residuenanalyse für die Anzahl der Episoden für „normale“ Bedingungen, als Streudiagramm mit vorhergesagten Werten (links) und als Histogramm (rechts).



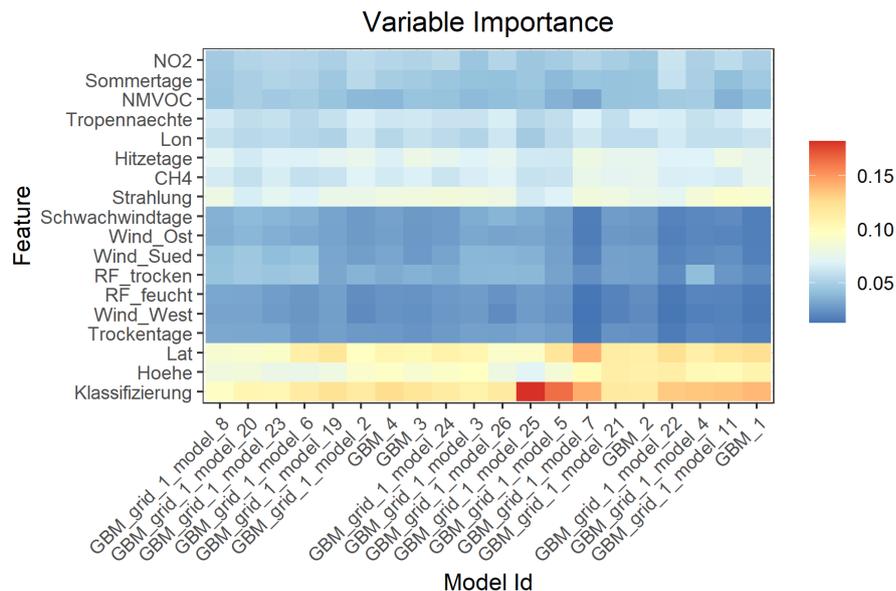
Quelle: Eigene Darstellung, TNO

Abbildung 50: Model Prediction Correlation Heatmap, Anzahl der Episoden für „normale“ Bedingungen



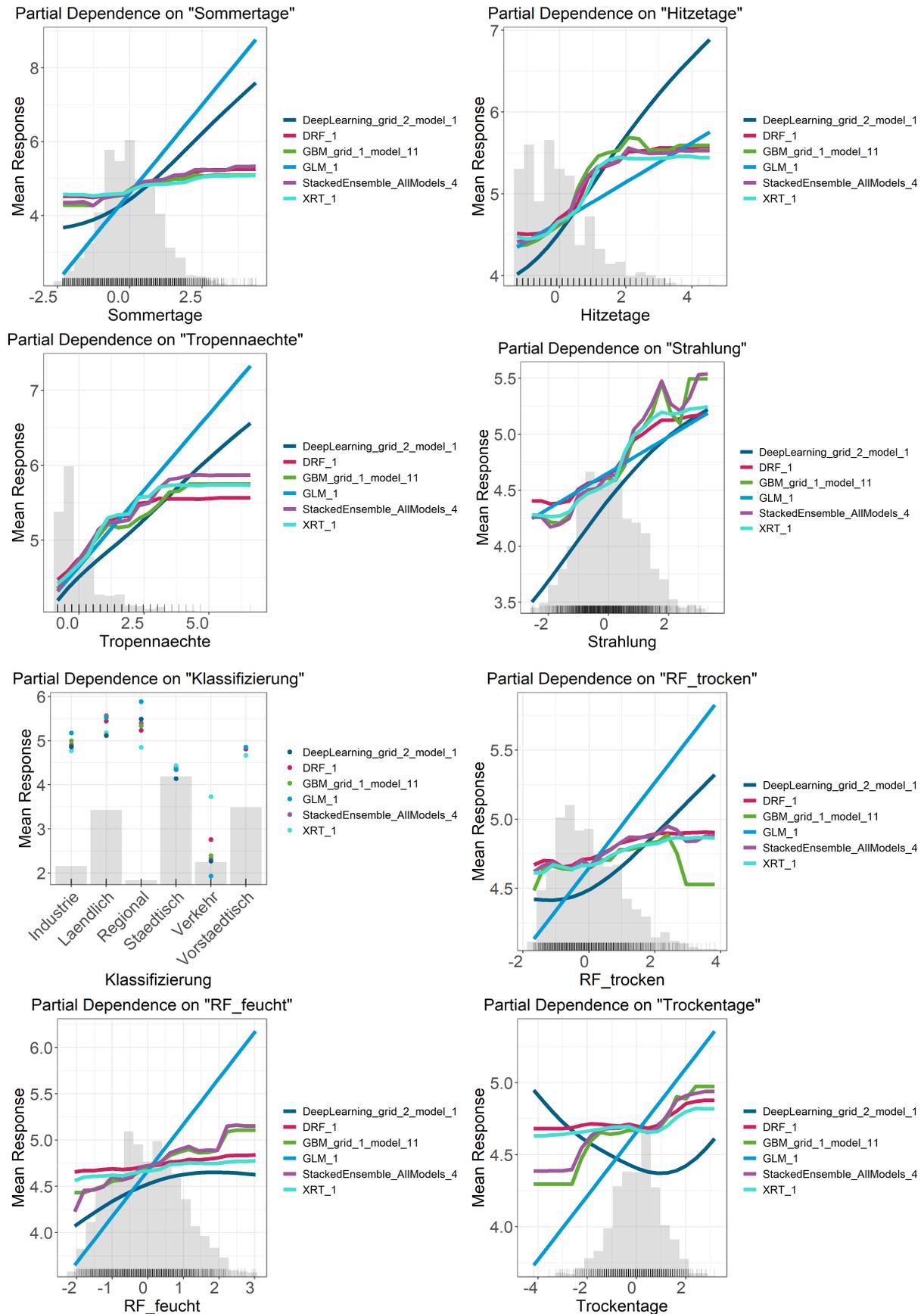
Quelle: Eigene Darstellung, TNO

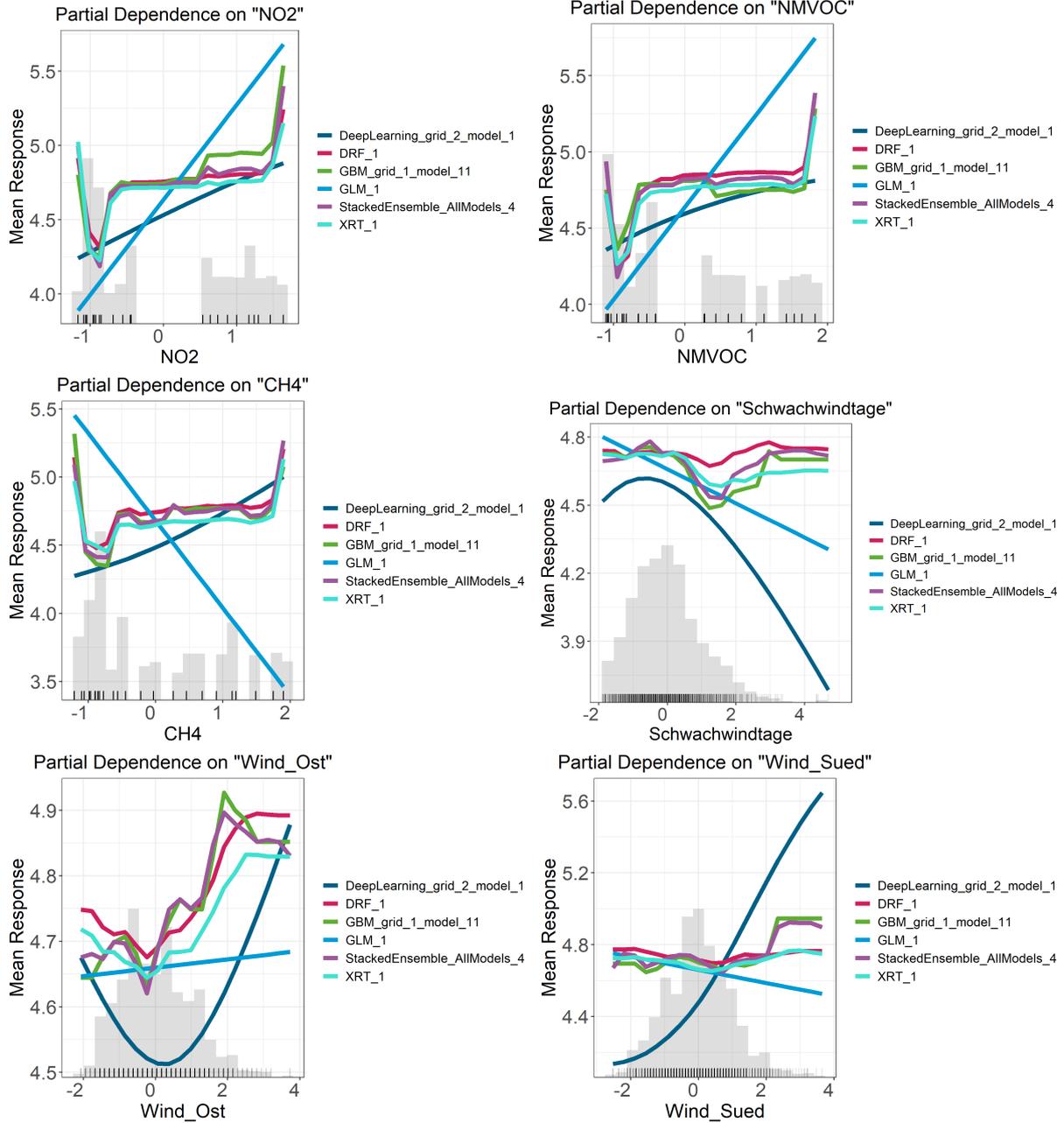
Abbildung 51: Variable Importance Heatmap, Anzahl der Episoden für „normale“ Bedingungen

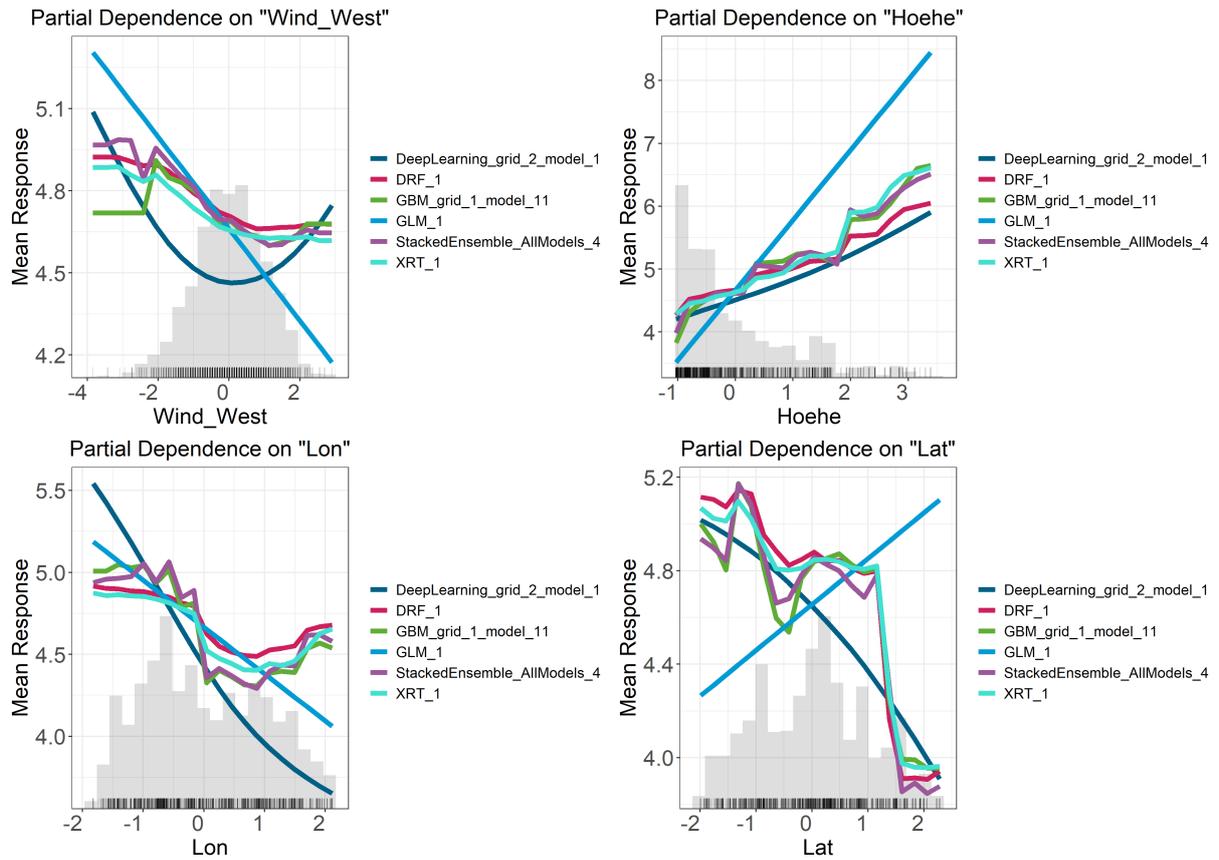


Quelle: Eigene Darstellung, TNO

Abbildung 52: Partial Dependence Plots, Anzahl der Episoden für „normale“ Bedingungen

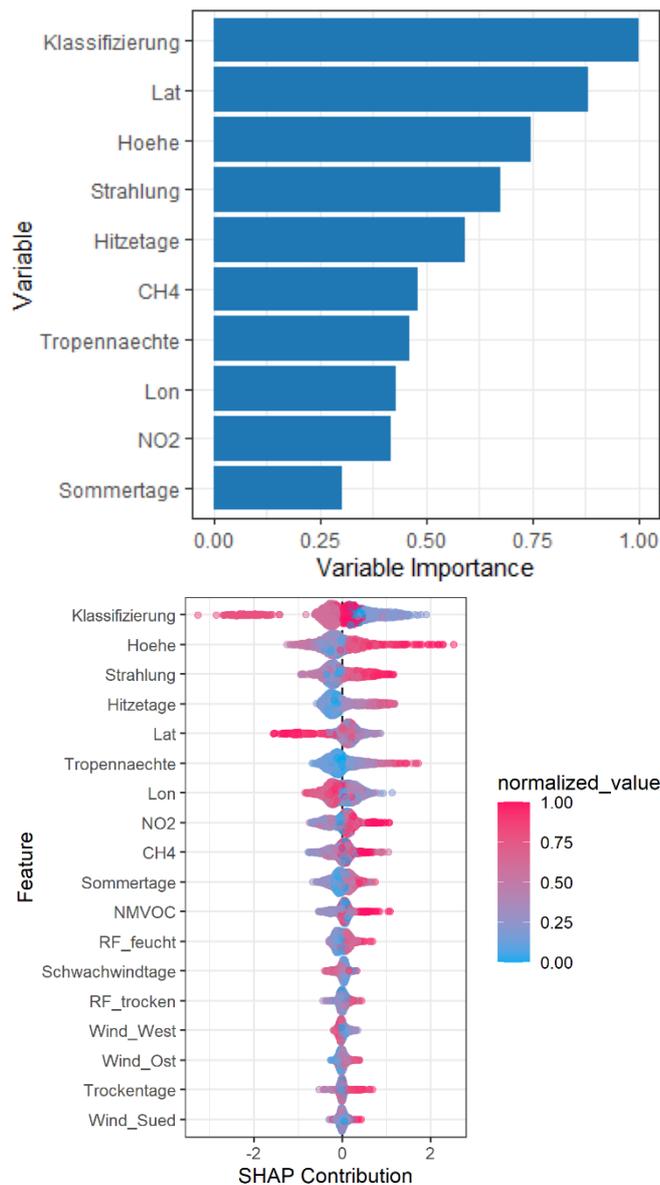






Quelle: Eigene Darstellung, TNO

Abbildung 53: Variable Importance Plot (oben) and SHAP-Plot (unten) für das beste GBM-Modell, Anzahl der Episoden für „normale“ Bedingungen



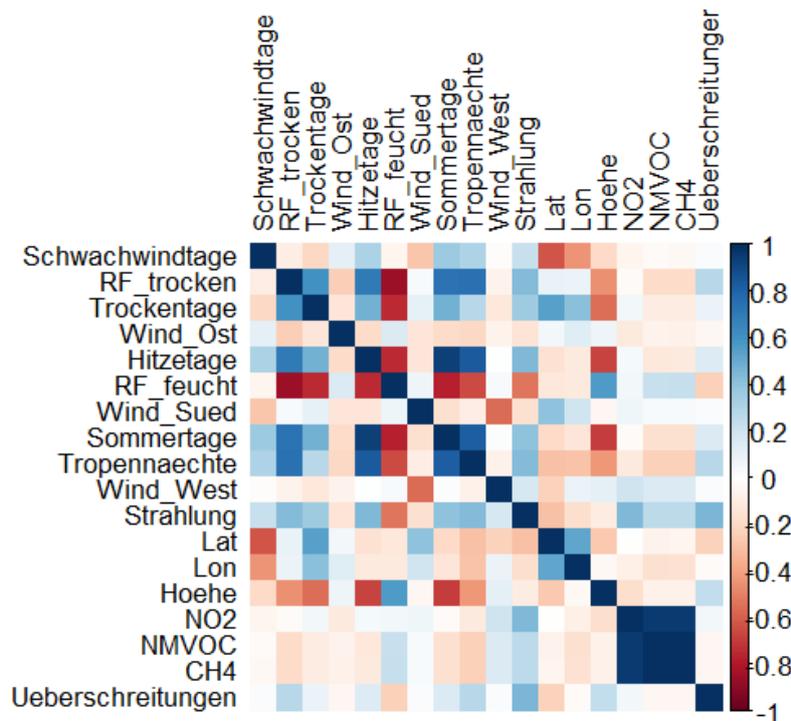
Quelle: Eigene Darstellung, TNO

6.8 Modell B: extreme Bedingungen, Anzahl der Ozonüberschreitungen

Für die „extremen“ Bedingungen stehen weniger Daten für Training und Test zur Verfügung als für die „normalen“ Bedingungen (190 Daten). Abbildung 54 zeigt bereits eine weniger starke negative Korrelation zwischen der Anzahl der Schwachwindtage und der Breitengrade als für die „normalen“ Bedingungen. Dies deutet darauf hin, dass die Unterteilung der meteorologischen Daten entsprechend der Einordnung der „normalen“ und „extremen“ Werte der Ozonüberschreitungen unterschiedliche Regime repräsentiert. Das Modell für „extreme“ Bedingungen neigt dazu, die Anzahl der Ozonüberschreitungen im unteren Bereich zu überschätzen (Abbildung 55), während das Modell für die „normalen“ Bedingungen im oberen Bereich die Anzahl der Ozonüberschreitungen unterschätzt. Das zeigt, dass die beiden Modelle nicht allmählich in einander übergehen. Die geringere Anzahl von Daten führt jedoch zu

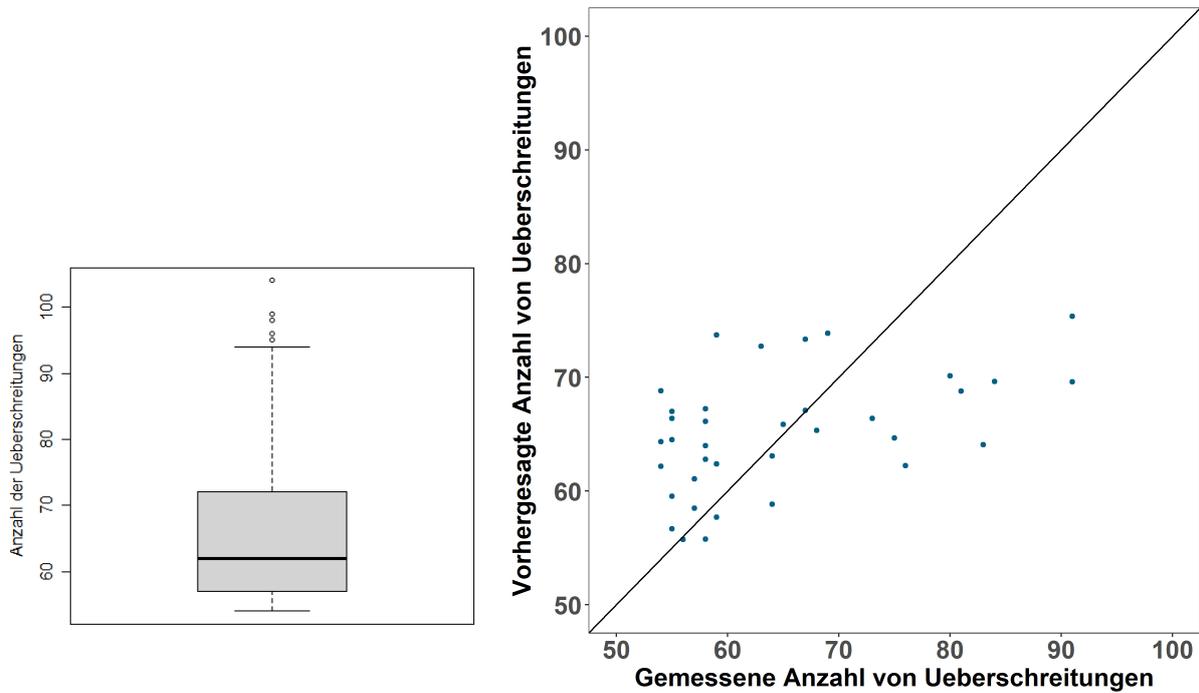
größeren Unsicherheiten, wie aus dem niedrigeren RMSE-Wert des Modells (8,68), den leicht unausgeglichene Abbildungen der Residuenanalyse (Abbildung 56), den geringeren Modellkorrelationen im Ensemble (Abbildung 57, insbesondere eins der Deep Learning Modelle) und den größeren Unterschieden zwischen den Modellen in den Partial Dependence Plots (Abbildung 59) hervorgeht. Die Farbskala der Variable Importance Heatmap (Abbildung 58) ist anders als in Abbildung 44 und Abbildung 51, weil einige Modelle die Stationsklassifizierung als weit einflussreichste Variable auswerten. Für die GBM Modelle sind die Variableneinflüsse der einzelnen Variablen meist weniger stark ausgeprägt als für die Deep Learning Modelle, obwohl auch mit relativ starkem Einfluss der Stationsklassifizierung. Der SHAP-Plot und der Variable Importance Plot (Abbildung 60) zeigen, dass die Stationsklasse und die Anzahl der Tropennächte weiterhin wichtig sind, dass aber jetzt Strahlung und Anzahl der Tage mit einer relativen Feuchte < 60 % (RF_trocken) eine größere Rolle spielen als unter „normalen“ Bedingungen. Die Windkennzahlen haben sehr wenig Einfluss. Die Beziehung zu den Emissionen zeigt eine leichte positive Korrelation zu den NO_x-Emissionen und eine negative Korrelation zu den VOC-Emissionen, aber die Unsicherheit ist aufgrund der begrenzten Datenmenge und der Dominanz der Jahre 2003 und 2018 sehr hoch. Aufgrund der geringen Anzahl von Daten und der Korrelationen zwischen den eingehenden Variablen war das Ergebnis empfindlich gegenüber der genauen Auswahl der Daten für Training und Validierung in der Zufallsauswahl, insbesondere für das Deep-Learning-Modell und weniger stark für die GBM-Modelle. Die Ergebnisse sollten daher nicht überinterpretiert werden. Es sollte jedoch auch bedacht werden, dass die Genauigkeit des Modells besser ist als die des Modells, das alle Daten (also sowohl „normale“ als auch „extreme“ Werte) einschließt.

Abbildung 54: Korrelation der Variablen, Anzahl der Überschreitungen für „extreme“ Bedingungen



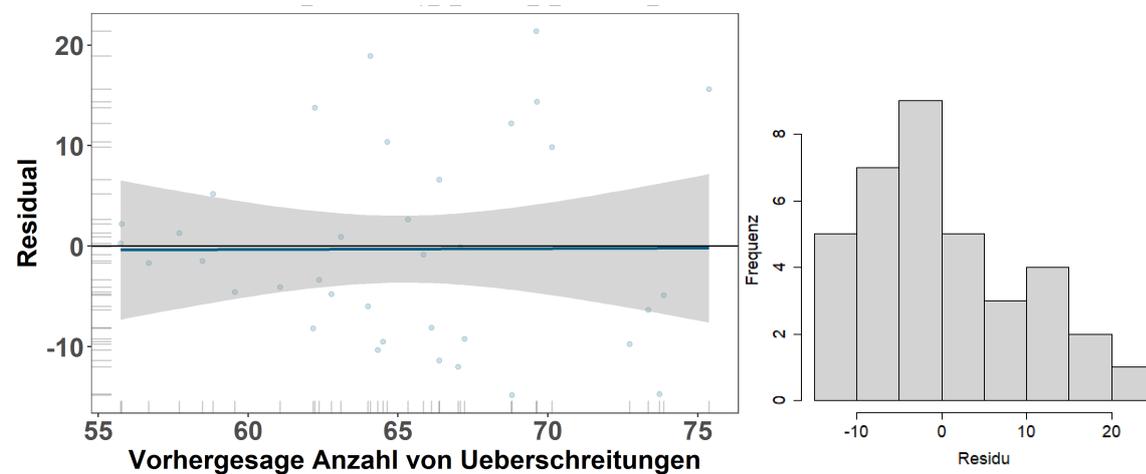
Quelle: Eigene Darstellung, TNO

Abbildung 55: Boxplot der Anzahl der Ozonüberschreitungen (links) und Scatterplot der vorhergesagten gegenüber gemessenen Anzahl der Überschreitungen (rechts) für „extreme“ Bedingungen



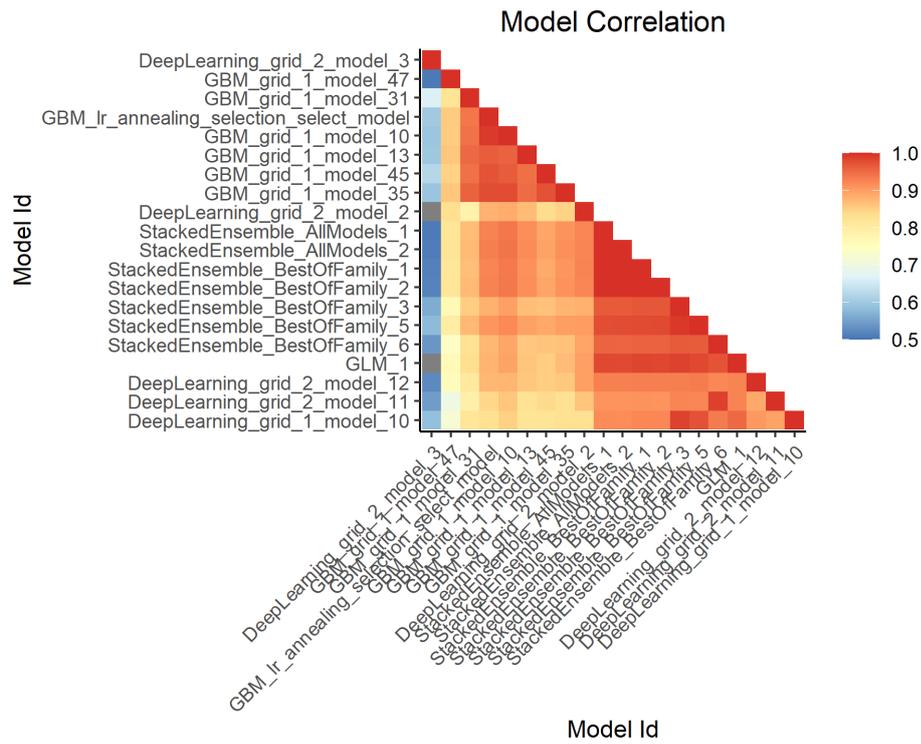
Quelle: Eigene Darstellung, TNO

Abbildung 56: Residuenanalyse für die Anzahl der Ozonüberschreitungen für „extreme“ Bedingungen, als Streudiagramm mit vorhergesagten Werten (links) und als Histogramm (rechts), Modell B, Experiment 6



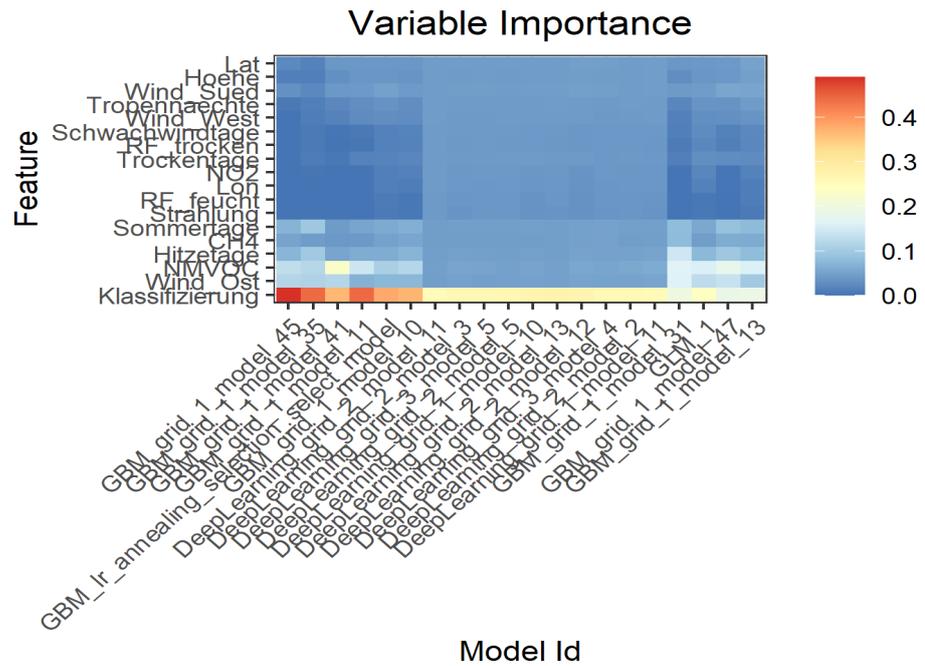
Quelle: Eigene Darstellung, TNO

Abbildung 57: Model Prediction Correlation Heatmap, Anzahl der Überschreitungen für „extreme“ Bedingungen, Modell B, Experiment 6



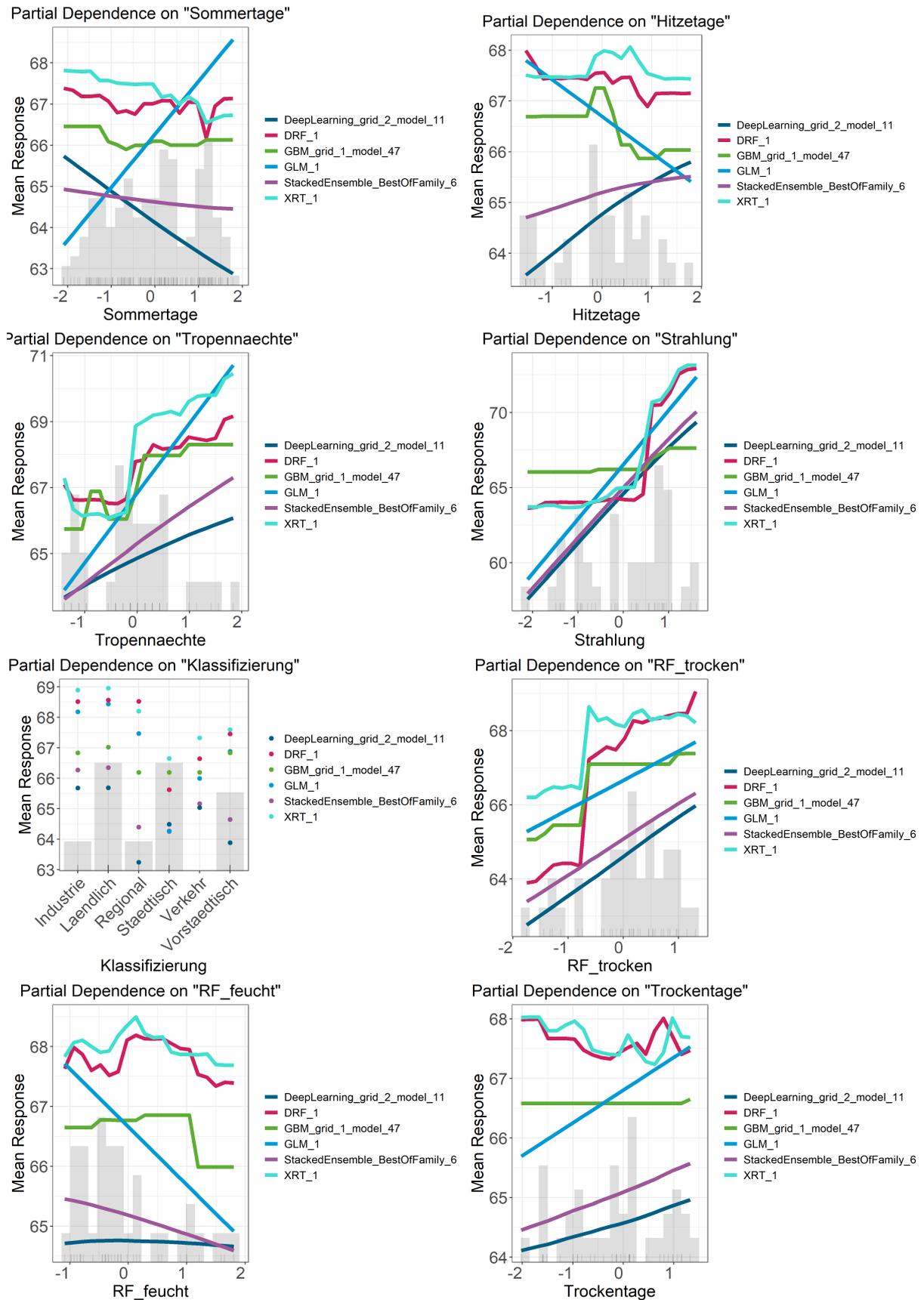
Quelle: Eigene Darstellung, TNO

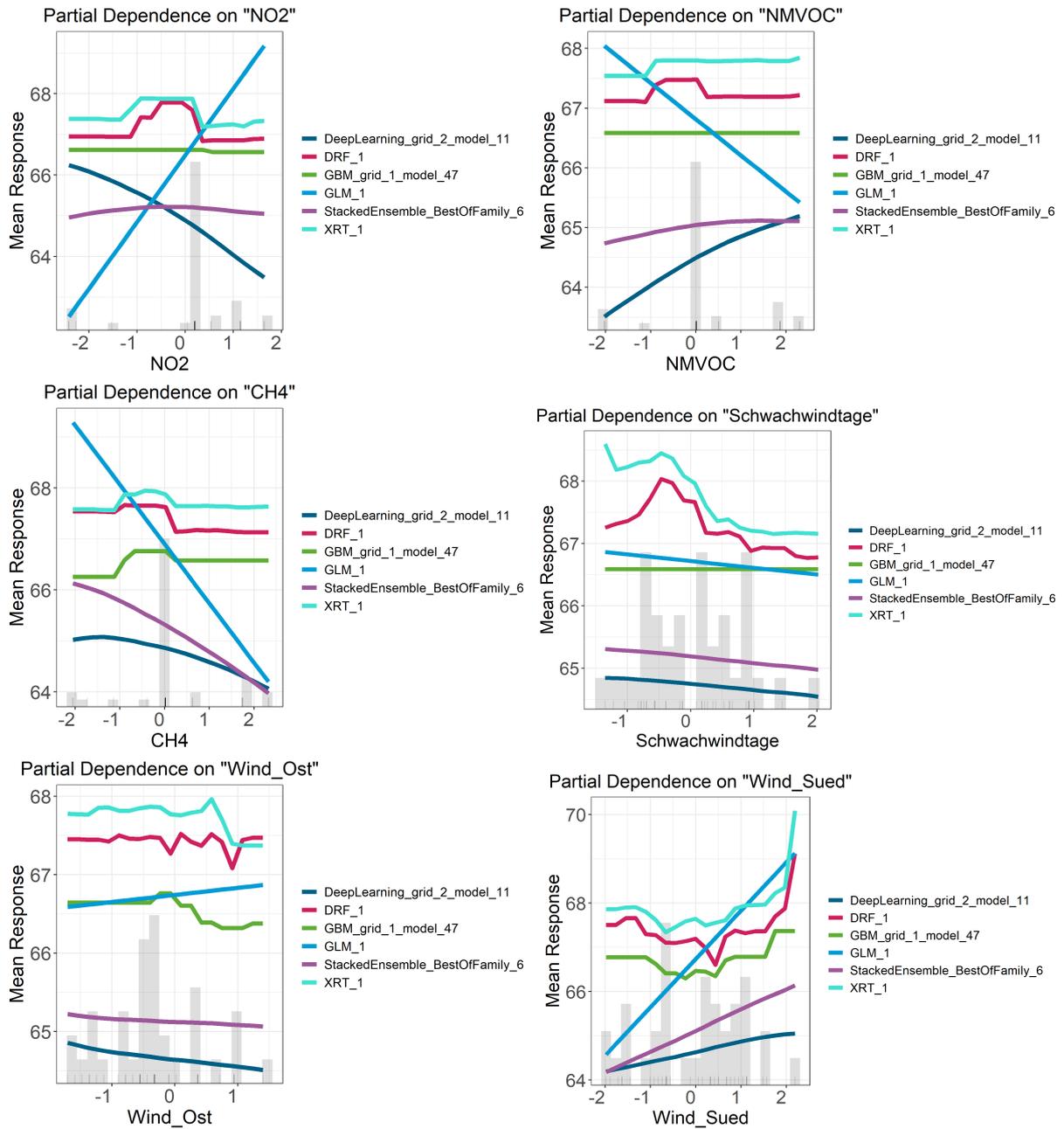
Abbildung 58: Variable Importance Heatmap, Anzahl der Überschreitungen für „extreme“ Bedingungen



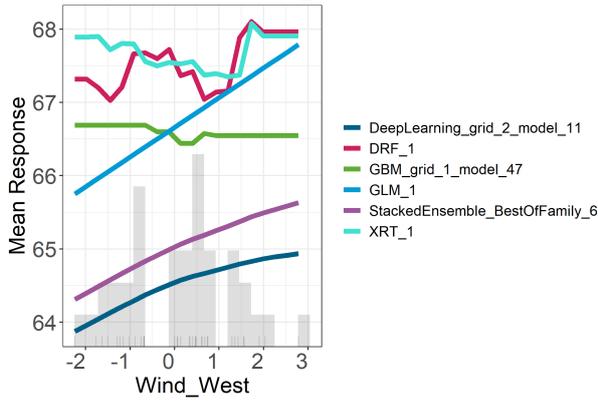
Quelle: Eigene Darstellung, TNO

Abbildung 59: Partial Dependence Plots, Anzahl der Überschreitungen für „extreme“ Bedingungen

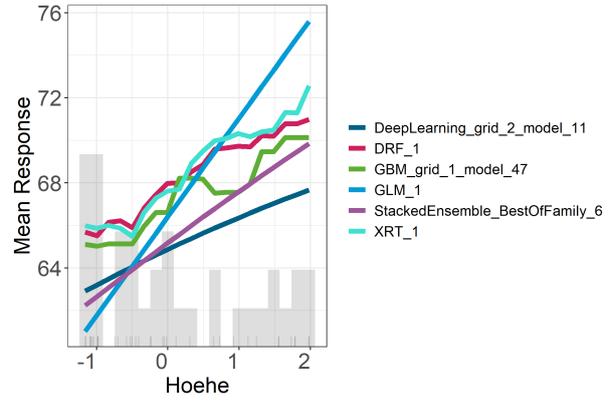




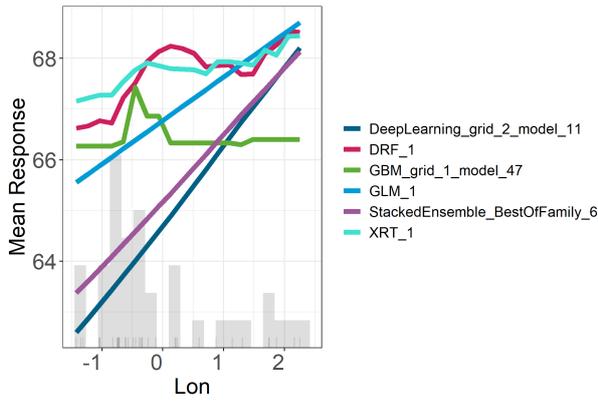
Partial Dependence on "Wind_West"



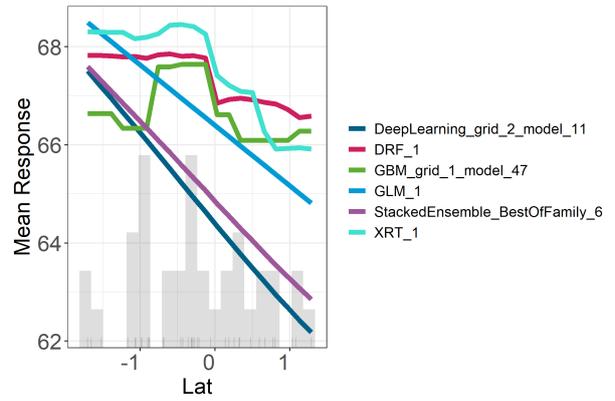
Partial Dependence on "Hoehe"



Partial Dependence on "Lon"

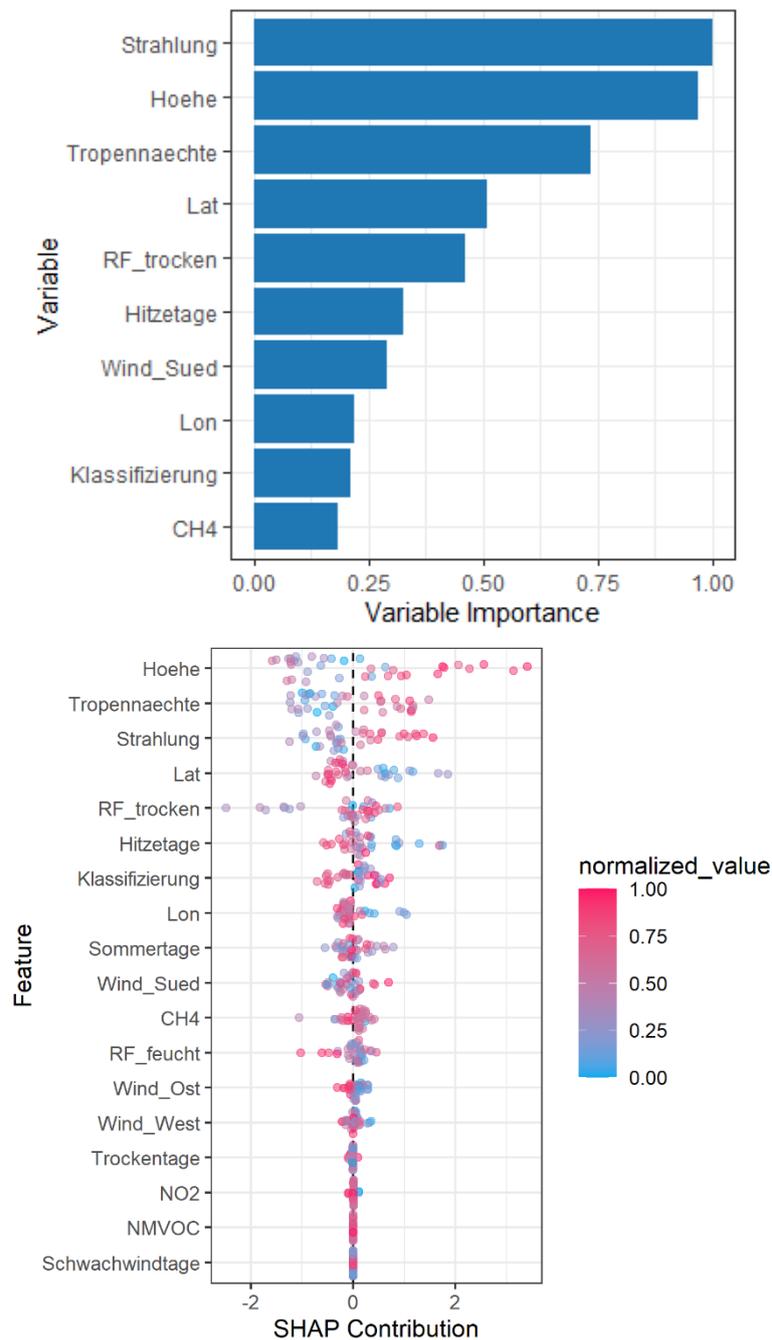


Partial Dependence on "Lat"



Quelle: Eigene Darstellung, TNO

Abbildung 60: Variable Importance Plot (oben) and SHAP-Plot (unten) für das beste GBM-Modell, Anzahl der Überschreitungen für „extreme“ Bedingungen.



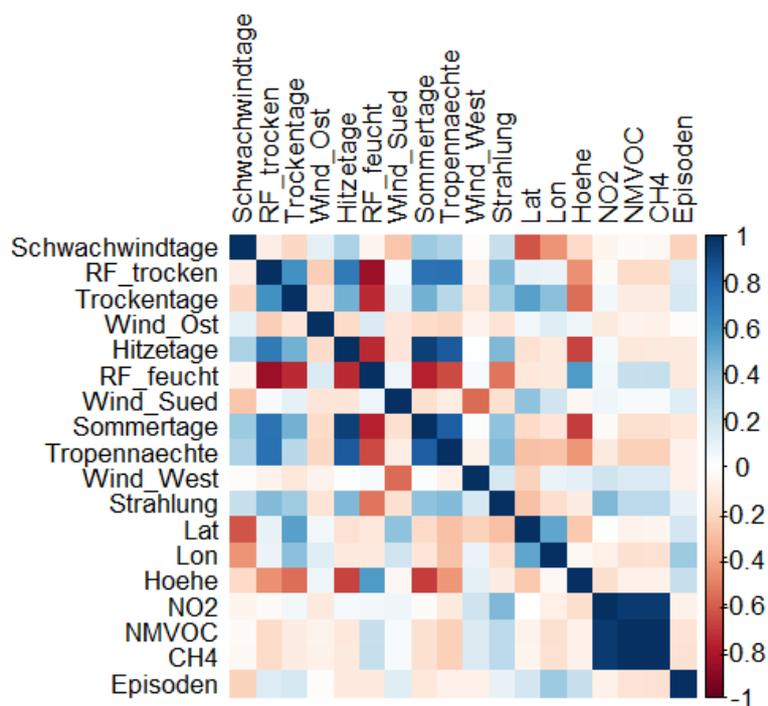
Quelle: Eigene Darstellung, TNO

6.9 Modell B: extreme Bedingungen, Anzahl der Ozonepisoden

Abbildung 61 zeigt andere Korrelationen der eingehenden Variablen als für Anzahl der Überschreitungen. Die Anzahl der Ozonepisoden zeigt eine stärkere negative Korrelation für die Schwachwindtage und negative Korrelationen mit den Temperaturvariablen, aber immer noch eine positive Korrelation für die Strahlung. Es wurden auch schwache negative Korrelationen zu den Emissionsvariablen gefunden. Für die Anzahl der Ozonepisoden gilt die gleiche Einschränkung in Hinblick auf die Anzahl der Daten (190) wie für die Ozonüberschreitungen.

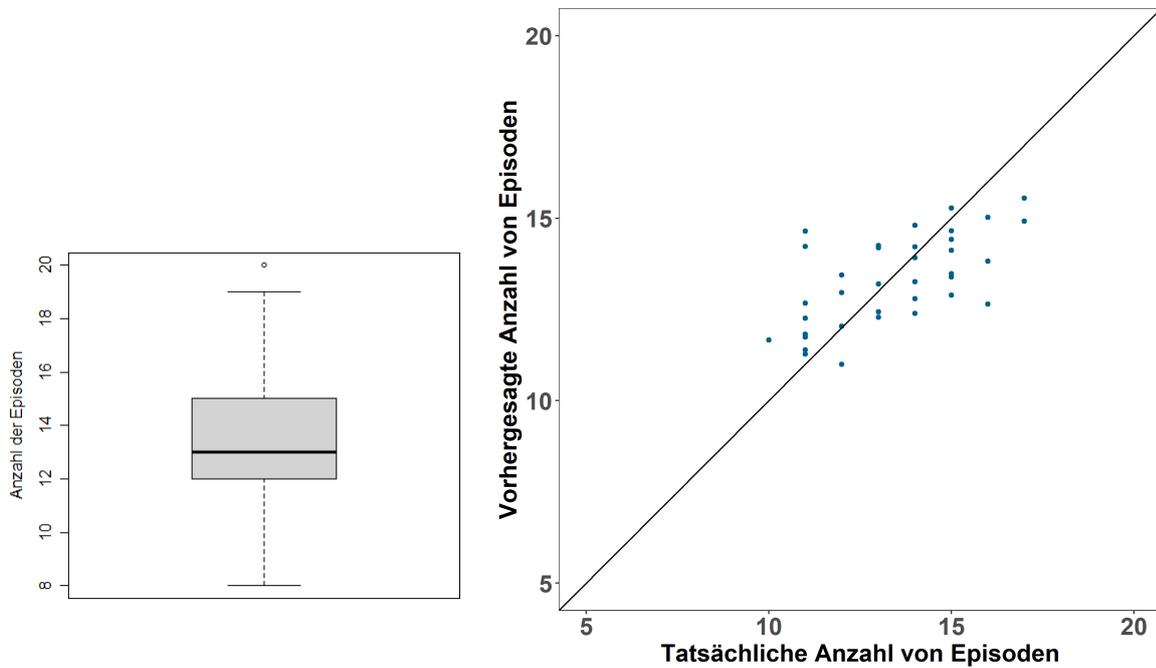
(„extreme“ Werte), was zu weniger robusten Ergebnissen führt. Ergebnisse sind in Bezug auf den RMSE (1,80) in Vergleich zu den typischen Werten (8 – 20 Episoden,) angemessen, wenn auch mit einem kleineren Bereich der vorhergesagten Anzahl (11 – 15) wie gezeigt in Abbildung 62. Die leicht asymmetrischen Residuenplots (Abbildung 63), die niedrigeren Modellkorrelationen im Ensemble (Abbildung 64), der größere Unterschied zwischen den Modellen in der Variable Importance Heatmap (Abbildung 65) und die Partial Dependence Plots (Abbildung 66) ähneln dem, was bei der Anzahl der Überschreitungen beobachtet wurde. Für die „extremen“ Bedingungen ist die Anzahl der Schwachwindtage die dominierende Variable nach der Stationsklasse und der Stationshöhe (Abbildung 66, Abbildung 67). Auch Windrichtung (Wind_West) und Variablen der relativen Feuchte sind wichtiger als Temperaturvariablen. Die Ergebnisse könnten darauf hindeuten, dass bei den „extremen“ Bedingungen eher stagnierende Wetterbedingungen eine Rolle spielen, die nicht direkt durch die Temperatur erfasst werden. Es besteht weiterhin eine schwache negative Korrelation zu den Emissionsvariablen. Die Partial Dependence Plots zeigen erhebliche Unterschiede zwischen den Modellen. Der Einfluss der Emissionsvariablen auf die Endergebnisse ist relativ gering (Abbildung 66, Abbildung 67). Die Ergebnisse sind teilweise kontraintuitiv (negative Korrelationen mit der Temperatur) und sollten angesichts der relativ geringen Anzahl von Daten nicht überinterpretiert werden. Der Scatterplot (Abbildung 62) zeigt nichtsdestotrotz eine Korrelation zwischen vorhergesagter und gemessener Anzahl von Episoden und hat einige Aussagekraft.

Abbildung 61: Korrelation der Variablen, Anzahl der Episoden für „extreme“ Bedingungen



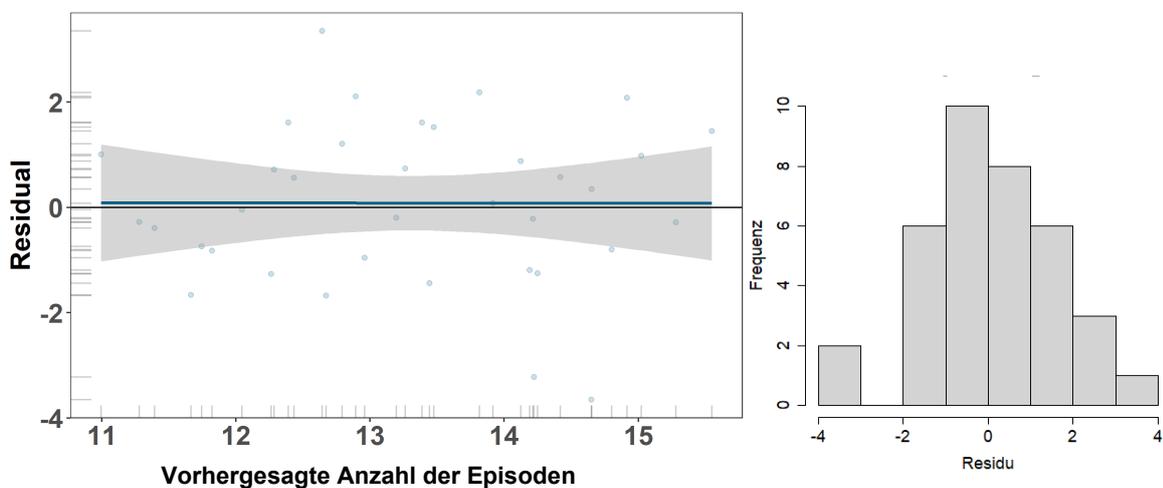
Quelle: Eigene Darstellung, TNO

Abbildung 62: Boxplot (links) der Anzahl der Episoden für „extreme“ Bedingungen und Scatterplot (rechts) der vorhergesagten gegenüber der gemessenen Anzahl der Episoden für „extreme“ Bedingungen



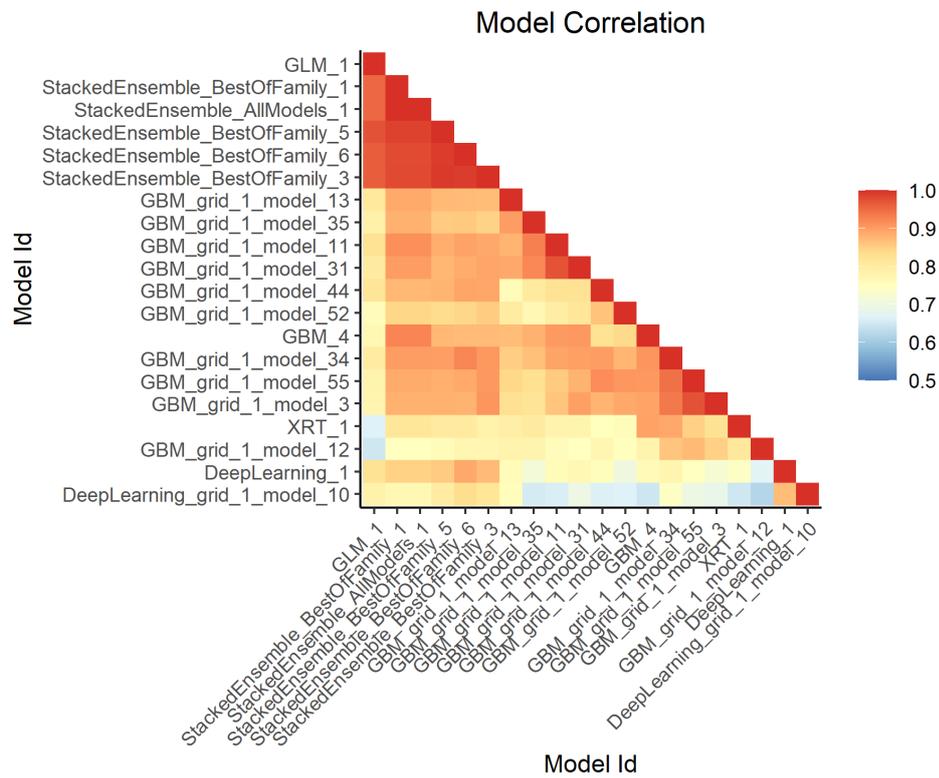
Quelle: Eigene Darstellung, TNO

Abbildung 63: Residuenanalyse für die Anzahl der Episoden für „extreme“ Bedingungen, als Streudiagramm mit vorhergesagten Werten (links) und als Histogramm (rechts)



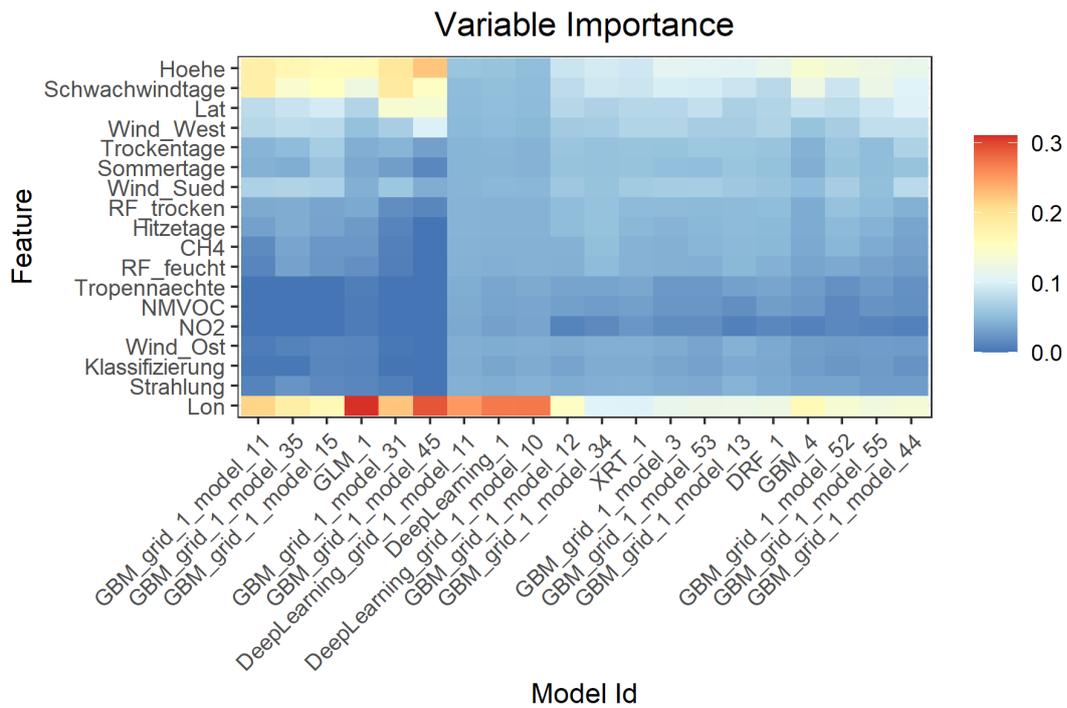
Quelle: Eigene Darstellung, TNO

Abbildung 64: Model Prediction Correlation Heatmap, Anzahl der Episoden für „extreme“ Bedingungen



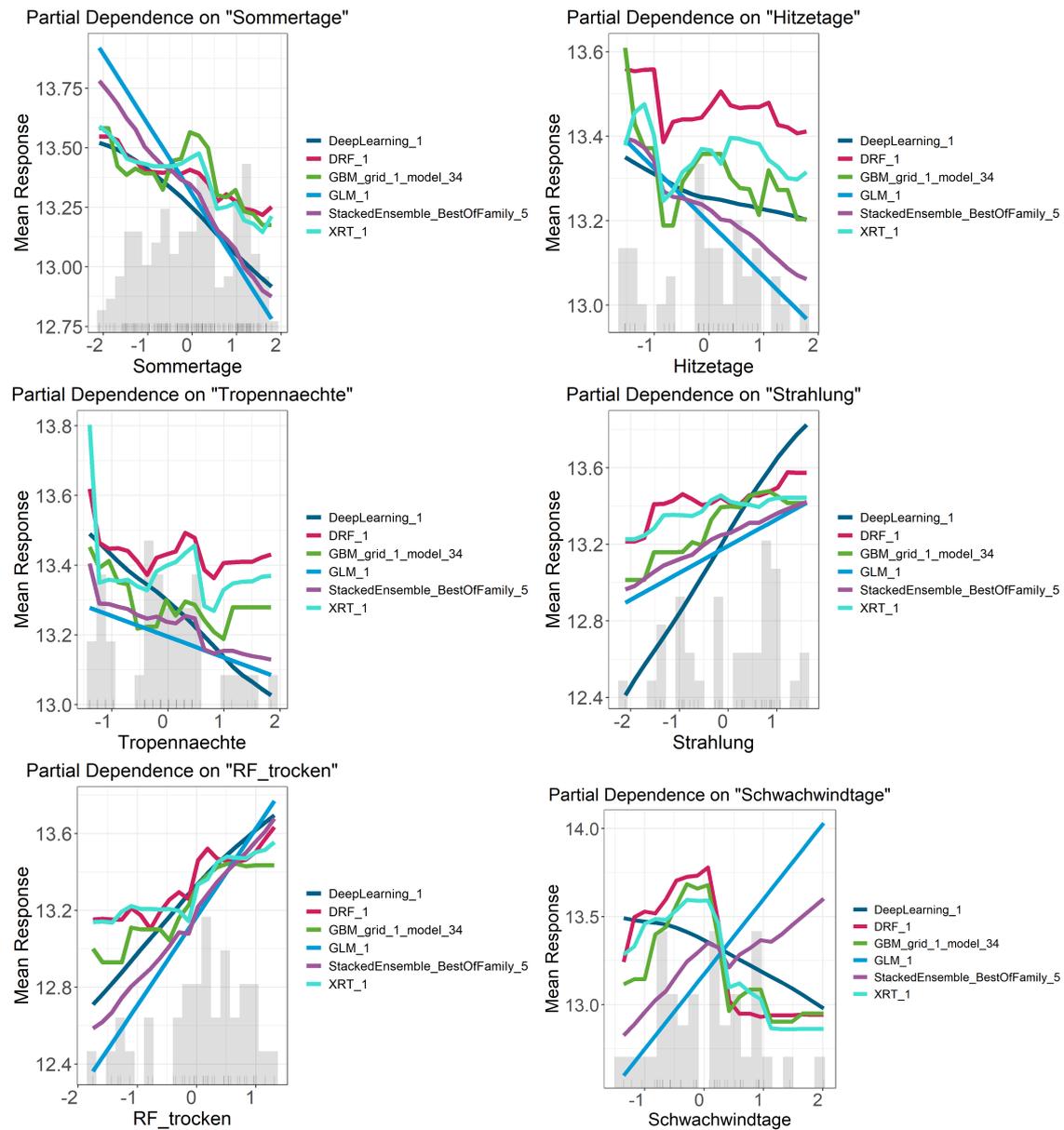
Quelle: Eigene Darstellung, TNO

Abbildung 65: Variable Importance Heatmap, Anzahl der Episoden für „extreme“ Bedingungen



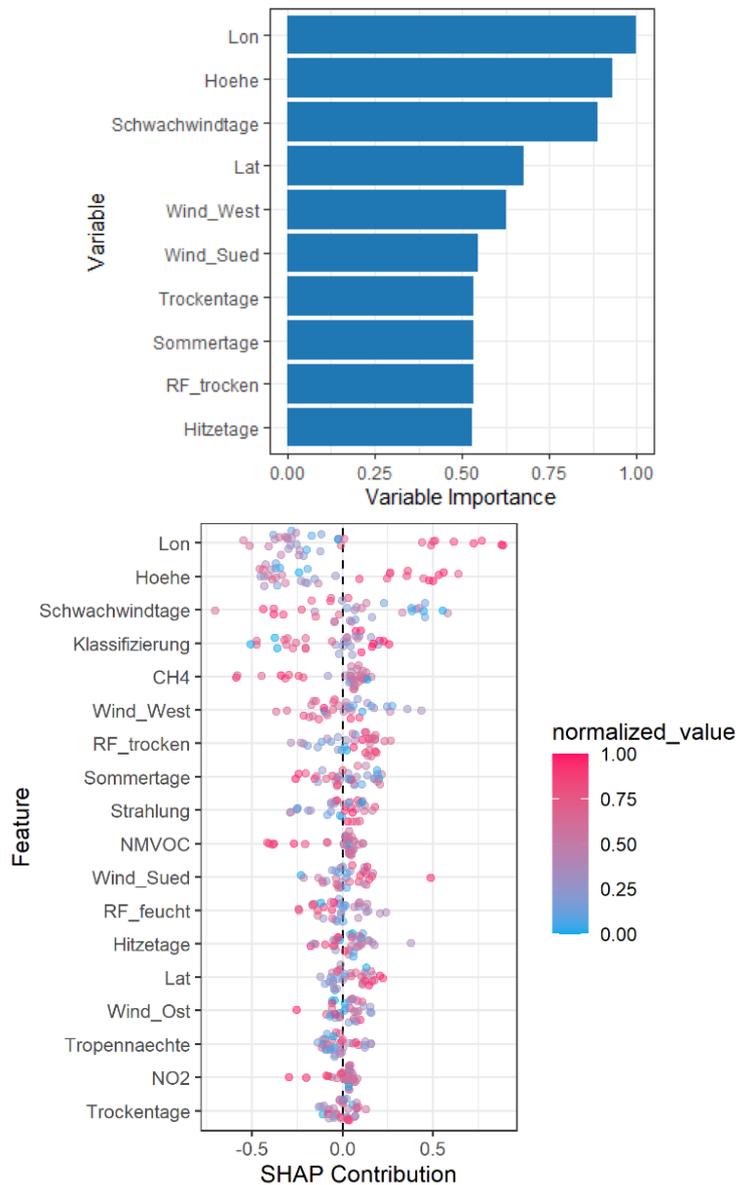
Quelle: Eigene Darstellung, TNO

Abbildung 66: Partial Dependence Plots, Anzahl der Episoden für „extreme“ Bedingungen



Quelle: Eigene Darstellung, TNO

Abbildung 67: Variable Importance Plot (oben) und SHAP-Plot für das GBM-Modell mit der besten Performance (unten), Anzahl der Episoden für „extreme“ Bedingungen



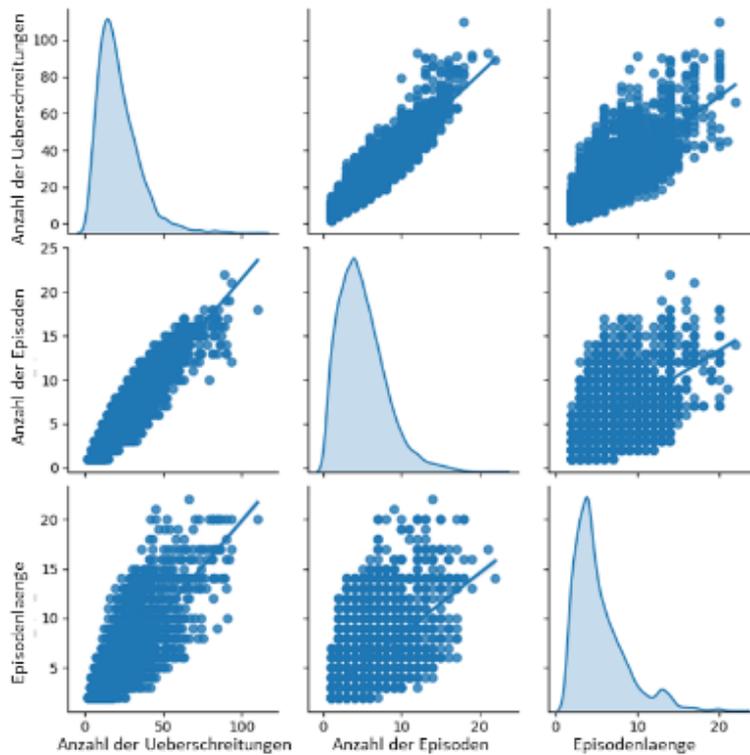
Quelle: Eigene Darstellung, TNO

6.10 Ozonepisodendauer

Die Ozonepisodendauer ist definiert als die maximale Anzahl aufeinanderfolgender Ozonüberschreitungstage (Tage mit $MDA8 > 120 \mu\text{g}/\text{m}^3$) pro Jahr und Station. Die Auswahl dieser Metrik ist von allen in diesem Bericht verwendeten Metriken am empfindlichsten. Abbildung 68 zeigt, dass ein Zusammenhang zwischen der Anzahl der Ozonüberschreitungen und der Ozonepisodendauer besteht, der jedoch schwächer ist als bei der Anzahl der Ozonepisoden. Es wurde untersucht, ob es eine deutliche Korrelation zur Anzahl der aufeinanderfolgenden Sommertage oder Anzahl der aufeinanderfolgenden tropischen Nächte gibt und ob es sinnvoll wäre, eine solche Variable in den Datensatz aufzunehmen. Sommertage und Tropennächte wurden ausgewählt, da sie mit Ozonüberschreitungen korrelieren und relativ einfach zu ermitteln sind. Es konnte eine, wenn auch schwache, Korrelation für die Anzahl der aufeinanderfolgenden Sommertage und Tropennächte gefunden werden (Abbildung 69). Eine

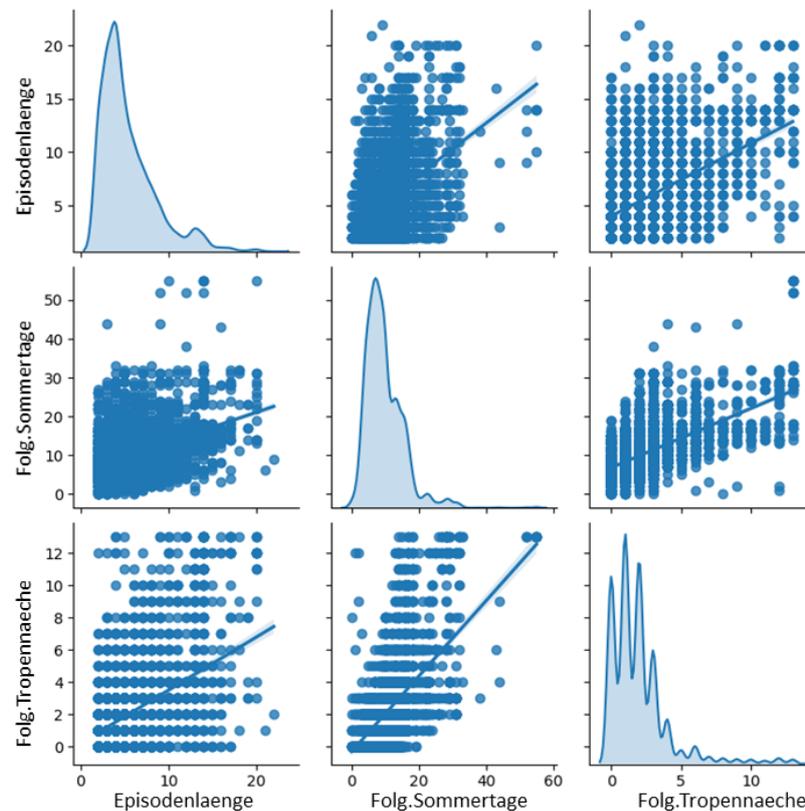
Korrelation zwischen der Ozonepisodenlänge und der Länge dieser hitzewellenbezogenen Variablen war sogar noch schwächer. Deshalb haben wir die Erstellung eines robusten Modells für Episodenlänge mit den heutigen Klimakennzahlen und die zusätzlichen Kennzahlen Anzahl der aufeinanderfolgenden tropischen Nächten und Sommertage nicht für machbar gehalten .

Abbildung 68: Korrelationen zwischen Anzahl der Ozonüberschreitungen, Anzahl der Ozonepisoden und maximale Ozonepisodendauer



Quelle: Eigene Darstellung, TNO

Abbildung 69: Korrelationen zwischen der Ozonepisodenlänge und der Anzahl der aufeinanderfolgenden Sommertage und Tropennächte



Quelle: Eigene Darstellung, TNO

6.11 Reproduzierbarkeit der Ergebnisse

Die Auswahl der Extrem-/Normalfälle war eine Grundfunktion, die bei jeder Verarbeitung des Datensatzes auf den gesamten Datensatz angewendet wurde. Die Extremwerte waren diejenigen über 54 Überschreitungen, aber die Anzahl der Stationen in dieser Kategorie schwankte zwischen 187 und 197 für die verschiedenen Modellläufe von R aufgrund einer Instabilität in R. Es hat Auswirkungen auf die Ergebnisse für die Extremwerte im Hinblick auf die Reproduzierbarkeit der Endergebnisse, da der Datensatz für die Extremwerte relativ klein ist. Es hatte Auswirkungen auf das Modell mit bester Performance und die Variablenbedeutung. Da nur ca. 40 Daten zur Validierung verwendet wurden, wurden visuelle Unterschiede in den Streudiagrammen festgestellt. Es wird daher empfohlen, den Datensatz für „extreme“ und für „normale“ Werte zu speichern und diese Daten für eine vollständige Reproduzierbarkeit im R-Tool einzulesen, wenn z. B. zusätzliche Grafiken erstellt werden müssen.

Beim Maschinellen Lernen werden aus den Datensätzen zufällige Ziehungen vorgenommen, um sie in Trainings- und Validierungsdaten zu trennen. Das Maschinelle Lernen selbst beinhaltet Zufälligkeit bei der Initialisierung (random seed) für die meisten Algorithmen (u. a. GGM, RF). Im Prinzip könnte dieses Problem durch die Verwendung von Zufallsdaten umgangen werden. Wenn jedoch ein anderer Computer oder eine andere H2O-Version verwendet wird, kann dies die Ergebnisse verändern. Im Laufe des Projekts wurde tatsächlich die Version von H2O und der Laptop gewechselt. Die Leistung änderte sich nur geringfügig, aber es hatte Auswirkungen auf das leistungsstärkste Modell und die Variablenbedeutung. Bei den robusteren Ergebnissen für die „normalen“ Werte ist der Effekt relativ gering. Die GBM-Modelle sind im Allgemeinen stabiler

in Bezug auf diesen Effekt, während das Neuronale-Netz-Modell empfindlicher ist. Um eine genaue Reproduzierbarkeit zu erreichen, kann man die Datensätze für „extreme“ und „normale“ Werte getrennt speichern, die Trainings- und Validierungsdaten speichern und das Maschinelle Lernen auf einem einzigen Kern durchführen. Letzteres geht jedoch auf Kosten der Rechenzeit. Die trainierten Modelle können auch für eine spätere Anwendung gespeichert werden. Bei fester Maschine, Reihenfolge des Datensatzes und Anzahl der Daten in den „extremen“ und „normalen“ Teilmengen waren die Ergebnisse des Trainings reproduzierbar.

7 Schlussfolgerungen, Diskussion und Empfehlungen

Die Hauptforschungsfrage lautete:

Lassen sich mit Hilfe statistischer Regressionsverfahren Zusammenhänge zwischen der Überschreitung eines Schwellenwertes für Ozonkonzentrationen und Klimakennzahlen für deutsche Ozonmessstationen finden? Die Antwort auf diese Frage lautet: ja, aber mit Einschränkungen.

Die Studie basiert auf den Ozonmessdaten in Deutschland für den Zeitraum 1995 – 2018, was zu fast 6900 Daten für jährlich aggregierte Daten führt. Als Überschreitungsschwelle wurde der Schwellenwert MDA8 = 120 $\mu\text{g}/\text{m}^3$ gewählt. Diese Metrik ist eine von der EU anerkannte Messgröße für die Gesundheit, sie ist robuster als ein Schwellenwert, der auf einem stündlichen Tagesmaximum basiert und sie repräsentiert Extreme relativ gut in dem Sinne, dass sie nicht zu oft überschritten wird ($< 1/3$ der Sommerperiode) und eine interannuelle und interregionale Variabilität aufweist.

Um sowohl eine außergewöhnlich hohe Anzahl von Ozonüberschreitungen als auch eine eher durchschnittliche Anzahl von Ozonüberschreitungen gut darstellen zu können, wurden zwei getrennte Modelle entwickelt, was in der Datenwissenschaft ein üblicher Ansatz ist. Für das Modell für „normale“ Werte (Anzahl der Überschreitungen ≤ 54) wurde ein RMSE von 5,99 für die Anzahl der Ozonüberschreitungen und für das Modell für „extreme“ Werte (Anzahl der Überschreitungen > 54) ein RMSE von 8,68 ermittelt. Angesichts der Bandbreite der Werte für die Anzahl der Ozonüberschreitungen können diese Ergebnisse als gut angesehen werden. Die Techniken des Maschinellen Lernens waren in der Lage, die Anzahl der Ozonüberschreitungen mit mehreren relevanten Klimakennzahlen in Beziehung zu setzen. Die Ergebnisse sind beziehungsangepasste Modelle, die im Prinzip für Studien zum Klimawandel verwendet werden können. Die Vorteile des Ansatzes bestehen darin, dass die trainierten Algorithmen eine schnelle Alternative zur Verwendung eines Chemie-Transport-Modells darstellen und weitaus geringere Mengen an Eingangsdaten erfordern. Es wurden auch Algorithmen für die Anzahl der Ozonepisoden entwickelt, die eine angemessene Performance aufweisen. Ein Ansatz für die Ozonepisodendauer wurde untersucht, aber nicht weiterverfolgt, da es schwieriger war, sie mit den Klimakennzahlen in Beziehung zu setzen.

Die jährliche Aggregation der Daten bringt jedoch Einschränkungen bei der Darstellung der Beziehungen zwischen Ozonüberschreitungen, meteorologischen Variablen und Emissionen mit sich. Die wichtigste Einschränkung besteht darin, dass die Auswirkungen von Emissionsveränderungen nicht gut quantifiziert werden können, da nur jährliche deutsche Emissionssummen verwendet werden und die meteorologische zwischenjährliche Variabilität starke Auswirkungen hat. Eine weitere Einschränkung ist, dass nicht alle Klimakennzahlen, die hier verwendet wurden, in allen Klimadatensätzen verfügbar sind. Es ist möglich, Modelle zu entwickeln, die diese Variablen ausschließen, aber sie hätten eine geringere Aussagekraft, selbst wenn die Klimakennzahlen mit anderen Variablen korreliert sind. Die letzte Einschränkung besteht darin, dass der gesamte verfügbare Ozonmessdatensatz für das Training und die Validierung verwendet wird. Die Leistung des Modells für unabhängige Jahre (2019 – 2021) wurde nicht überprüft.

Auf die wichtigsten Aspekte wird im Folgenden näher eingegangen.

Ansatz des Maschinellen Lernens

Das Paket H2O wurde für das Maschinelle Lernen ausgewählt, da es robust und frei verfügbar ist, eine große Nutzergemeinschaft hat und mit guten Standardoptionen auch für Nicht-Experten einen guten Einstieg bietet. Dies bedeutet, dass auch ohne tiefgreifende Erfahrung mit begrenztem Zeitaufwand gute Ergebnisse erzielt werden können. Es bietet auch gute Visualisierungen der Ein- und Ausgabedaten, die die Interpretation der Ergebnisse unterstützen. Insbesondere die AutoML-Option ist eine sehr effiziente Möglichkeit, mehrere Algorithmen zu trainieren, einschließlich eines Scans durch den Hyperparameterraum, um die leistungsstärksten Modelle auszuwählen. Dies war besonders nützlich, da nicht immer von vornherein bekannt ist, welcher Algorithmus die besten Ergebnisse liefert und es kann sogar von der Auswahl des jeweiligen Teildatensatzes abhängen, wie es hier der Fall war. Der GBM gehörte zu den leistungsstärksten Algorithmen (mit geeigneten Hyperparametern), das Stacked Ensemble des AutoML (einschließlich des GBM) war das beste Modell, insbesondere für den Fall der „extremen“ Werte. Diese Modelle sind gut in der Lage, mit nichtlinearem Verhalten umzugehen, die Tatsache zu berücksichtigen, dass die Eingabedaten nicht völlig unabhängig sind (z. B. die Anzahl der tropischen Nächte und die Anzahl der Hitzetage stellen beide Temperaturextreme dar) und numerische und nicht-numerische Informationen (z. B. die Stationsklassifizierung) einzubeziehen. Ziel war es, ein Modell zu entwickeln, das die Anzahl der Ozonüberschreitungen oder Ozonepisoden vorhersagt. Modelle für Maschinelles Lernen können nicht gleichzeitig auf die Anzahl der Ozonüberschreitungen und die Anzahl der Ozonepisoden trainiert werden. Für diese Vorhersagewerte wurden daher getrennte Modelle entwickelt.

Die Trennung von Gruppen von Stationen mit ähnlichem Verhalten (Cluster) kann zu genaueren Vorhersagen pro Gruppe führen. Die Clusterbildung könnte auf der Grundlage der Stationsklassifizierung oder der regionalen Verteilung entsprechend den KLENOS-Klimazonen für Ozon erfolgen. Neben dem Clustern von Stationen nach diesen Kriterien kann das Clustern auf der Grundlage anderer Kriterien zu anderen Clustern führen. Es wurde untersucht, ob das Clustering der Stationen mit Hilfe von Maschinellern einen zusätzlichen Nutzen bringen würde. Die Clusterbildung erfolgte anhand der jährlichen Überschreitungen über den gesamten Zeitraum, aber auch anhand kürzerer Zeiträume, wobei einschließlich das zeitliche Zusammentreffen von Überschreitungen an verschiedenen Stationen berücksichtigt wurde. Bergstationen wurden von der Analyse von vornherein ausgeschlossen. Es stellte sich heraus, dass die Clusterbildung robust ist, aber auch sehr empfindlich gegenüber der verwendeten Struktur und Metrik der Eingabedaten ist. So hat die Methode mit kürzeren Zeiträumen die geografische Kohärenz von Ozonepisoden hervor gehoben und die Methode auf Jahresbasis die Ähnlichkeiten zur Stationsklassifizierung gezeigt. Außerdem wird durch das Clustering die Anzahl der Daten pro Datensatz reduziert. Bei den Methoden des Maschinellen Lernens ist es auch möglich, Stationsmerkmale wie Längen- und Breitengrad, Höhe und Stationsklassifizierung direkt zu berücksichtigen. Dies ist eine robustere Vorgehensweise, da mehr Daten gleichzeitig für das Training zur Verfügung stehen. Das Training auf den in KLENOS verwendeten Klimazonen wurde durchgeführt. Dies führte jedoch nicht zu großen Unterschieden, mit dem Nachteil, dass weniger Daten für „extreme“ Werte pro Klimazone zur Verfügung standen. Dadurch war das Ergebnis für die „extremen“ Werte für Anzahl von Ozonüberschreitungen in den Jahren 2003 und 2018 weniger robust.

Es stellte sich heraus, dass ein einziges Modell, das alle Daten für den gesamten Zeitraum berücksichtigt, nicht die beste Performance erbrachte. Bessere Modelle konnten erstellt werden, wenn weniger Jahre berücksichtigt wurde. Für den Zeitraum 2014 – 2018 wurde die beste Performance gefunden. Dieses Modell wäre am besten geeignet, um eine Vorhersage für 2019, 2020 oder 2021 zu machen, da diese Jahre in Bezug auf Klimakennzahlen und Emissionen

ähnlich sind und es kaum Stationswechsel gab. Das für 2014 – 2018 entwickelte Modell zeigt jedoch keine gute Performance (RMSE von 32) für das extreme Jahr 2003, in dem die Emissionen höher waren als für 2014 – 2018, kombiniert mit einem extrem warmen Sommer.

Für den endgültigen Ansatz wurden getrennte Modelle für „extreme“ Werte und „normale“ Werte entwickelt, wobei „normal“ und „extrem“ auf den Interquartilsabständen der Anzahl der Ozonüberschreitungen basieren. Ein separates Klassifizierungsmodell wurde entwickelt, um zu bestimmen, für welchen Bereich von Eingangsdaten das „normale“ oder das „extreme“ Modell für die Vorhersage verwendet werden sollte. Auf diese Weise wurden die RMSE-Werte für „extreme“ Werte im Vergleich zu einem Modell, das sowohl „normale“ als auch „extreme“ Werte berücksichtigt, deutlich verbessert (von etwa 12 – 13 auf etwa 8,6). Es wurden keine separaten Modelle auf Ebene der Stationsklassifizierung gebildet. Modelle für einzelne Klimazonen hatten eine weniger gute Performance für eine hohe Anzahl von Ozonüberschreitungen als das Modell für „extreme“ Werte.

Die Trennung zwischen „normalen“ und „extremen“ Werten konnte mit einer Genauigkeit von etwa 96 % bestimmt werden, was sehr gut ist. Keine einzelne Klimakennzahl wäre als Unterscheidungsfaktor geeignet. Strahlung, Stationsklassifizierung und Variablen bezogen auf die Luftfeuchte und Temperatur waren jedoch zusammen die wichtigsten Prädiktoren, um einen Datenpunkt in die Extrem- oder Normaldatensätze einzuordnen. Die meisten, aber nicht alle Daten im Datensatz der „extremen“ Werte stammten aus den Jahren 2003 oder 2018. Eine beträchtliche Anzahl von Daten aus diesen Jahren gehörte jedoch auch zum Datensatz der „normalen“ Werte. Die Emissionen konnten nicht auf sinnvolle Weise in die Klassifizierung einbezogen werden.

Die Ergebnisse wurden qualitativ mit einer Studie verglichen, in der tägliche Daten verwendet wurden, um eine Beziehung zwischen Ozonwerten und meteorologischen Daten für den Raum Berlin und Brandenburg herzustellen (T. Deng, persönliche Mitteilung, siehe Anhang C). Bei der Verwendung täglicher Daten würde die Vorhersage für 2018 unter Verwendung von Trainingsdaten für 2014 – 2017 und anschließender Aggregation über die Anzahl der Ozonüberschreitungen je nach Station zu Unterschieden von 1 bis 4 Tagen zu den gemessenen Werten führen. Für das Chemie-Transport-Modell LOTOS-EUROS wurden für denselben Zeitraum Unterschiede von 2 bis 11 Tagen gegenüber der gemessenen Anzahl von Überschreitungen festgestellt. Bei diesen Zahlen haben wir nicht berücksichtigt, ob die Überschreitungen für die richtigen Tage modelliert wurden. Das Tagesmodell ist zwar genauer, erfordert aber auch einen gewissen Aufwand, um jede Station einzeln zu trainieren. Außerdem wurde in der Studie von Deng nur eine begrenzte Anzahl von Stationen untersucht, die sich alle in Ostdeutschland befinden. Der Vergleich mit dem CTM und dem auf Tagesdaten basierenden Modell zeigt, dass die Performance des Modells, das jährliche aggregierte Klimadaten verwendet, angesichts des Kompromisses zwischen Datendichte, Allgemeinheit und Genauigkeit angemessen ist.

Der Prädiktand "Anzahl der Ozonepisoden" ist weniger robust als der Prädiktand "Anzahl der Ozonüberschreitungen", da es viele kleine Episoden von nur 2 Tagen oder eine sehr lange Episode geben kann. Die Korrelationen zwischen der Anzahl der Episoden und der Anzahl der Überschreitungen zeigen, dass die Korrelation ziemlich stark ist, und das Training zeigte, dass dieselben dominanten Klimakennzahlen beteiligt waren. Die relative Unsicherheit ist etwas größer als bei der Anzahl der Ozonüberschreitungen, da die Ozoninformationen noch weiter komprimiert sind. Es wurde ein RMSE von 2 Episoden pro Jahr gefunden (in einem Bereich von 0 – 16 Episoden, also etwa 10 %).

Für die maximale Episodenlänge wurde untersucht, ob sie mit einer Metrik in Verbindung gebracht werden kann, die die Episodendauer in Bezug auf die Temperatur angibt. Es wurde die Anzahl der aufeinanderfolgenden tropischen Nächte und die Anzahl der aufeinanderfolgenden Sommertage berechnet. Die direkte Korrelation mit der Anzahl der Ozonepisoden ist jedoch offensichtlicher als mit der Anzahl der aufeinanderfolgenden tropischen Nächte und Sommertage. Wie bei der Anzahl der Ozonepisoden ist die Anzahl der Ozonüberschreitungen robuster, und eine einfache Regression mit der Anzahl der Ozonüberschreitungen könnte informativer sein.

Es ist immer noch möglich, dass eine andere Kombination von Eingangsdaten, das Entfernen einiger stark korrelierter Prädiktoren oder eine Feinabstimmung des Maschinellen Lernens (leicht) bessere Ergebnisse liefert. Außerdem reagieren die Algorithmen in datenarmen Bereichen empfindlich auf die Datenauswahl durch Zufallsziehungen, so dass die Modelle am Rand des Variablenbereichs und die Modelle für „extreme“ Werte weniger robust sind. Wir denken jedoch, dass die aktuelle Arbeit ein gutes Gleichgewicht zwischen den gewählten Variablen, der Komplexität und der Robustheit der Ergebnisse darstellt. Zur Reproduktion der Ergebnisse wird empfohlen, den Trainings- und Validierungsdatensatz sowie das endgültige Modell für eine spätere Verwendung zu speichern. Das leistungsstärkste GBM wäre die beste Wahl, da es die nichtlinearen Abhängigkeiten auf realistische Weise berücksichtigen kann und eine gute Erklärbarkeit aufweist, die sich gut in einem SHAP-Plot zusammenfassen lässt.

Die Modelle des Maschinellen Lernens haben zu wesentlich besseren Ergebnissen geführt, als dies mit linearen Regressionen möglich gewesen wäre, die die starken Korrelationen zwischen meteorologischen Variablen nicht berücksichtigen können. Die Partial Dependence Plots zeigen deutlich das nichtlineare Verhalten. Die Ozonüberschreitungen wurden von den Modellen mit den verschiedenen Klimakennzahlen so in Verbindung gebracht, dass die Modelle aussagekräftig waren. Die Stärke unseres Ansatzes besteht darin, dass wir eine lange Zeitreihe einbezogen haben, die sowohl Emissions- als auch Klimaveränderungen umfasst, was über das hinausgeht, was in der Literatur zu finden ist (z. B. Eskes et al., 2013, die auf Tagesdaten für drei Jahre trainierten, um Klimaszenarien für ausgewählte Stationen zu extrapolieren), trotz der damit verbundenen Herausforderungen und Unsicherheiten. Dadurch war es möglich, die Auswirkungen von Emissionsänderungen einzubeziehen, allerdings mit Unsicherheiten, da die zwischenjährliche Variabilität der Meteorologie größer ist als die zwischenjährliche Variabilität der Emissionen, was es noch schwieriger macht, robuste Ergebnisse für die „extremen“ Werte zu finden.

Anwendung des Instruments für Klimaszenarien

Die Klimakennzahlen wurden aus den stündlichen COSMO-REA6-Daten berechnet, die dem Portal des DWD entnommen wurden. Diese haben eine hohe Auflösung (6 x 6 km²), eine konstante Qualität über einen langen Zeitraum und lassen sich leicht auf die Standorte der Ozonstationen interpolieren. Als Klimakennzahlen wurden die Anzahl der tropischen Nächte, die Anzahl der Sommertage, die Anzahl der Hitzetage, also die Tage, die mit der warmen Jahreszeit zusammenhängen, gewählt. Außerdem wurden die Anzahl der Tage mit Ost-, Süd- und Westwinden, der Tage mit schwachen Winden, der Trockentage, der Tage mit einer relativen Luftfeuchte von < 60 % bzw. > 80 % und die kumulative kurzweilige Strahlung für die Monate April bis September berechnet.

Die Kennzahlen mit Bezug zur Temperatur sind die einflussreichsten Variable für Ozonüberschreitungen im Allgemeinen, und die jährliche Zahl der Sommertage und der tropischen Nächte sind eindeutige Indikatoren, die in Klimadatenbanken leicht verfügbar sind. Allerdings kann auch die Strahlung einen wichtigen Einfluss haben, insbesondere bei „extremen“

Werten. Die Daten für die Strahlung sind in den Klimadatenbanken ECAD und ReKlies nicht direkt verfügbar. Außerdem sind einige relevante Klimakennzahlen nur als Jahreswerte verfügbar, während für Ozonüberschreitungen nur die Werte für das Sommerhalbjahr relevant sind, wie z. B. die Anzahl der Tage mit schwachem Wind. Variablen für die relative Luftfeuchte sind standardmäßig überhaupt nicht verfügbar. Daher wird empfohlen, Klimakennzahlen für die Monate April bis September getrennt zu speichern und auch eine Kennzahl für die relative Luftfeuchte einzubeziehen, wenn eine Klimasimulation erstellt wird, die für Luftqualitätsvorhersagen verwendet werden soll.

Obwohl Emissionsänderungen eine wichtige Rolle für das Ozonverhalten spielen, wurden die Auswirkungen von Emissionsänderungen im Rahmen dieser Studie nur begrenzt quantifiziert (nur deutsche Gesamtsumme, keine räumliche Disaggregation, keine Informationen aus Nachbarländern, nur Untersuchung der Überschreitungsanzahl). Für „normale“ Werte zeigte sich eine Tendenz zu einer geringeren Anzahl von Ozonüberschreitungen und Ozonepisoden mit abnehmenden Emissionen. Für „extreme“ Werte wurde eine negative Korrelation festgestellt, allerdings mit großer Unsicherheit (weniger Trainingsdaten).

Das derzeitige Instrument ist ein „relationship fitting“-Modell, das im Prinzip auch für Klimaszenarien verwendet werden kann. Dies würde das Tool zu einer sehr effizienten Möglichkeit machen, die Bandbreite der Auswirkungen des Klimawandels auf Ozonüberschreitungen zu erforschen. Der trainierte Algorithmus ist sehr viel schneller als die Ausführung einer CTM-Simulation. Der Nachteil ist, dass eine Extrapolation möglicherweise nicht sinnvoll ist, wenn die Klimakennzahlen für das künftige Klima außerhalb der derzeit verwendeten Werte liegen. Außerdem können sich die Beziehungen zwischen den Klimakennzahlen ändern. Dies könnte durch einen Vergleich mit CTM-Modellen oder durch die Verwendung von CTMs zur Erzeugung zusätzlicher Trainingsdaten für die Extrapolation untersucht werden. Auch CTMs sind nicht perfekt, da sie die Tendenz haben, Extreme zu unterschätzen. Andererseits sind CTMs besser geeignet, um die Auswirkungen von Emissionsszenarien zu untersuchen. Außerdem verwenden sie in ihren Parametrisierungen ausdrücklich meteorologische Variablen, so dass eine mögliche Änderung ihrer Interdependenz aufgrund des Klimawandels nicht gesondert untersucht werden muss. Obwohl es sich bei den hier entwickelten Maschinellen Lernmodellen um „relationship fitting“-Modelle handelt, können die nichtlinearen Abhängigkeiten der Klimakennzahlen die Modellperformance beeinträchtigen, wenn sich diese Beziehungen erheblich ändern. Die aktuelle Studie hat die Stärken und Grenzen des Ansatzes aufgezeigt. Bevor das Instrument sinnvoll für Klimaszenarien eingesetzt werden kann, ist weitere Forschung erforderlich.

Empfehlungen

- ▶ Die Performance für die Jahre 2019, 2020 und 2021 als unabhängige Validierung des Verfahrens und der Modelle sollten überprüft werden, bevor sie als Instrument zum Einsatz gebracht wird, um den Effekt von Klimawandel auf Ozon zu quantifizieren. Die Jahre 2019 und 2020 sind durch Sommer mit hohen Temperaturen charakterisiert und könnten als Test für die Anwendung von Klimawandelszenarios dienen.
- ▶ Es wird derzeit *nicht* empfohlen, das hier entwickelte Tool für Klimaszenarien als einzige Informationsquelle zu verwenden. Die Emissionen von NO_x, VOC und CH₄ werden in den kommenden Jahrzehnten voraussichtlich weiter gesenkt und unter die derzeit untersuchten Emissionsniveaus hinausgehen. Eine Extrapolation in Bezug auf die Emissionen ist schwer zu rechtfertigen. Das Tool könnte jedoch sehr effektiv sein, wenn es darum geht, eine Bandbreite für ein Ensemble von Szenarien festzulegen, um aufwändige Simulationen mit einem CTM zu ergänzen, die nur für ein oder zwei Szenarien durchgeführt werden können.

CTM Simulationen könnten Daten für das Training neuer Emissionsszenarien liefern, und das hier entwickelte ML-Modell könnte verwendet werden, um die Auswirkungen der Meteorologie zu untersuchen. Diese Richtung könnte weiter untersucht werden.

- ▶ Klimakennzahlen für neue Klimasimulationen sollten auf monatlicher Basis gespeichert werden anstelle von jährlichen Aggregationen, wie dies z. B. in ReKliES bereits geschieht. Dies ermöglicht eine bessere Analyse von saisonalen Effekten (z. B. Ozonüberschreitung, würde aber auch für Wintersmog gelten) bei gleichzeitiger Reduzierung der Datenlast im Vergleich zur Speicherung von stündlichen bis täglichen Daten.
- ▶ Eine Klimakennzahl für die Luftfeuchte sollte definiert und gespeichert werden. Die relative Luftfeuchte spielt bei vielen atmosphärischen Prozessen, einschließlich der Ozonbildung, eine wichtige Rolle und trägt nachweislich positiv zur Qualität des Algorithmus zur Vorhersage von Ozonüberschreitungen bei. Die hier entwickelten Kennzahlen haben sich als nützlich erwiesen.

8 Quellenverzeichnis

- Andersson, C. and Engardt, M. (2010): European ozone in a future climate: Importance of changes in dry deposition and isoprene emissions. *Journal of Geophysical Research: Atmospheres*, 115, D02303, doi:10.1029/2008JD011690
- Anttila, P. and Tuovinen, J.P. (2010): Trends of primary and secondary pollutant concentrations in Finland in 1994–2007. *Atmospheric Environment*, 44(1), S 30-41
- Barnpadimos, I., Hueglin, C., Keller, J., Henne, S., and Prévôt, A. S. H. (2011): Influence of meteorology on PM10 trends and variability in Switzerland from 1991 to 2008. *Atmospheric Chemistry and Physics*, 11,4, S1813–1835
- Bollmeyer, C., J. D. Keller, C. Ohlwein, S. Bentzien, S. Crewell, P. Friederichs, A. Hense, J. Keune, S. Kneifel, I. Pscheidt, S. Redl, S. Steinke (2015): Towards a high-resolution regional reanalysis for the European CORDEX domain, *Q. J. Royal Met. Soc.*, 141, Doi: 10.1002/qj.2486
- Boleti, E., Hueglin, C. and Takahama, S. (2018): Ozone time scale decomposition and trend assessment from surface observations in Switzerland. *Atmospheric Environment*. 191, Doi 10.1016/j.atmosenv.2018.07.039
- Boleti, E., Hueglin, C. and Takahama, S. (2019a): Trends of surface maximum ozone concentrations in Switzerland based on meteorological adjustment for the period 1990–2014. *Atmospheric Environment*, 213, Doi 10.1016/j.atmosenv.2019.05.018
- Boleti, E., Hueglin, C., Grange, S. K., Prévôt, A. S. H., and Takahama, S. (2019b): Temporal and spatial analysis of ozone concentrations in Europe based on time scale decomposition and a multi-clustering approach, *Atmos. Chem. Phys. Discuss.*, Doi 10.5194/acp-2019-909
- Brasseur, G. P., Jacob, D., and Schuck-Zöller, S. (2017): *Klimawandel in Deutschland: Entwicklung, Folgen, Risiken und Perspektiven*. Springer Verlag.
- Breiman, L. (2001): Statistical Modeling: The Two Cultures, *Statistical Science*, Vol. 16, No. 3, S. 199–231
- Carslaw, D. C., S. D. Beevers, S. D. and Tate, J. E. (2007): Modelling and assessing trends in traffic-related emissions using a generalised additive modelling approach. *Atmospheric Environment*, 41, S. 5289–5299
- Colette, A., Granier, C., Hodnebrog, O., Jakobs, H., Maurizi, A., Nyiri, A., Bessagnet, B., D'Angiola, A., D'Isidoro, M., Gauss, M., Meleux, F., Memmesheimer, M., Mieville, A., Roüil, L., Russo, F., Solberg, S., Stordal, F., Tampieri, F. (2011): Air quality trends in Europe over the past decade: a first multi-model assessment. *Atmos. Chem. Phys.* 11, Doi:10.5194/acp-11-11657-201
- Colette, A., Andersson, C., Baklanov, A., Bessagnet, B., Brandt, J., Christensen, J.H., Doherty, R., Engardt, M., Geels, C., Giannakopoulos, C. and Hedegaard, G.B., (2015): Is the ozone climate penalty robust in Europe? *Environmental Research Letters*, 10,8, S.084015.
- Colette, A., Solberg, S., Beauchamp, M., Bessagnet, B., Malherbe, L., Guerreiro, C., (2017): Long-term air quality trends in Europe, ETC/ACM Technical paper, Bilthoven, Netherlands
- De Hoogh, K., Saucy, A., Shtein, A., Schwartz, J., West, E. A., Strassmann, A., and Kloog, I. (2019): Predicting fine-scale daily NO₂ for 2005–2016 incorporating OMI satellite data across Switzerland. *Environmental science & technology*, 53, 17, S. 10279-10287
- Dentener F., Keating T., and Akimoto H. (2010): Hemispheric transport of air pollution, part a: Ozone and particulate matter. Technical Report 11.II.E.7, UNECE, New York and Geneva
- Doherty, R.M., Wild, O., Shindell, D.T., Zeng, G., Mackenzie, I.A., Collins, W.J., Fiore, A.M., Stevenson, D. S., Dentener, F.J., Schultz, M.G., Hess, P., Derwent, R.G., and Keating, T. J. (2013): Impacts of climate change on

- surface ozone and intercontinental ozone pollution : A multi-model study. *Journal of Geophysical Research: Atmospheres*, 118:3744–3763, Doi: 10.1002/jgrd.50266.
- Eskes, H., Van Weele, M., Van Noije, T. (2013): Meteorological scenarios translated to change in air pollution levels, KNMI'14 report, Version 01, 22 January 2013, KNMI, De Bilt, Netherlands
- Eslami, E., Choi, Y., Lops, Y. and Sayeed, A. (2020): A real-time hourly ozone prediction system using deep convolutional neural network. *Neural Computing and Applications*, 32,13, S.8783-8797
- Feng, R., Zheng, H., Zhang, A., Huang, C., Gao, H. and Ma, Y. C. (2019): Unveiling tropospheric ozone by the traditional atmospheric model and machine learning, and their comparison: A case study in Hangzhou, China. *Environmental pollution*, 252: 366-378.
- EC, (2008): Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe, OJ L 152, 11.06.08, 1–44. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF> [Accessed 14 May 2009]
- Fleming, Z.L., Doherty, R.M., Schneidmeyer, E.V., Malley, C.S., Cooper, O.R., Pinto, J.P., Colette, A., Xu, X., Simpson, D., Schultz, M.G., Lefohn, A.S., Hamad, S., Moolala, R., Solberg, S. (2018): Tropospheric Ozone Assessment Report: present-day ozone distribution and trends relevant to human health. *Elementa: Sci. Anthropocene* 6, Doi10.1525/elementa.273.
- Friedman, J.H. (2006): Recent Advances in Predictive (Machine) Learning, *J. Classif.*, 23, 175–197, <https://doi.org/10.1007/s00357-006-0012-4>
- Giordano, L., Brunner, D., Flemming, J., Hogrefe, C., Im, U., Bianconi, R., Badia, A., Balzarini, A., Baró, R., Chemel, C., Curci, G., Forkel, R., Jiménez-Guerrero, P., Hirtl, M., Hodzic, A., Hozak, L., Jorba, O., Knöbe, C., Kuenen, J.J.P., Makar, P.A., Manders-Groot, A., Neal, L., Pérez, J.L., Pirovano, G., Pouliot, G., San José, R., Savage, N., Schröder, W., Sokhi, R.D., Syrakov, D., Törán, A., Tuccella, P., Werhahn, J., Wolke, R., Yahya, K., Žabkar, R., Zhang, Y., Galmarini, S. (2015): Assessment of the MACC reanalysis and its influence as chemical boundary conditions for regional air quality modeling in AQMEII-2, *Atmospheric Environment*, Volume 115, Pages 371-388, Doi 10.1016/j.atmosenv.2015.02.034
- Hastie T., Tibshirani, R. Friedman, J. (2001): *The Elements of Statistical Learning*, 2nd edition, Springer
- Hastie, T. and Tibshirani, R. (1990): Generalized Additive Models. *Statistical Science*, 1, 3, S. 297–310
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., .. & Thépaut, J. N. (2020): . The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 730, S. 1999-2049.
- Hertig, E. (2020): Health-relevant ground-level ozone and temperature events under future climate change using the example of Bavaria, Southern Germany, *Air Quality, Atmosphere & Health*, 13, Doi 10.1007/s11869-020-00811-z
- Hendriks, C., Forsell, N., Kiesewetter, G., Schaap, M. and Schoepp, W. (2016): Ozone concentrations and damage for realistic future European climate and air quality scenarios. *Atmospheric Environment*, 144, S.208-219.
- Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., Baró, R., Bellasio, R., Brunner, D., Chemel, C. and Curci, G. (2015): Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone. *Atmospheric environment*, 115, S. 404-420
- Immitzer, M., Atzberger, C., and Koukal, T. (2012): Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data, *Remote Sens.*, 4, Doi 10.3390/rs4092661
- Jackson, L. S., Carslaw, N., Carslaw, D. C., Emmerson, K. M. (2009): Modelling trends in OH radical concentrations using generalized additive models. *Atmospheric Chemistry and Physics*, 9,6, S.2021–2033

- Jacob, D. J., and Winner, D. A. (2009): Effect of Climate Change on Air Quality. *Atmospheric Environment*, 43, Doi: 10.1016/j.atmosenv.2008.09.051
- Kotsiantis, S. B. (2013): Decision trees: a recent overview, *Artif. Intell. Rev.*, 39, Doi 10.1007/s10462-011-9272-4
- Kuhn, M. (2013): Predictive Modeling with R and the caret Package, available at: <https://www.r-project.org/conferences/useR-2013/Tutorials/Kuhn.html> (last access: 30 April 2018), useR! conference, 10–12 July 2013, University of Castilla-La Mancha, Al-bacete, Spain.
- Lefohn, A.S., Malley, C.S., Smith, L., Wells, B., Hazucha, M., Simon, H., Naik, V., Mills, G., Schultz, M.G., Paoletti, E., De Marco, A. (2018): Tropospheric ozone assessment report: Global ozone metrics for climate change, human health, and crop/ecosystem research. *Elementa, Science of the Anthropocene*, 6, Doi /doi.org/10.1525/elementa.279
- Liu, Z., Sullivan, C. (2019): Prediction of weather induced background radiation fluctuation with recurrent neural networks. *Radiat. Phys. Chem.* 155, Doi 10.1016/j.radphyschem.2018.03.005.
- Lyapina, O., Schultz, M.G., Hense, A. (2016): Cluster analysis of European surface ozone observations for evaluation of MACC reanalysis data, *Atmos. Chem. Phys.*, 16, 6863–6881, doi: 10.5194/acp-16-6863-2016
- Manders, A. M. M., van Meijgaard, E., Mues, A. C., Kranenburg, R., van Ulft, L. H., and Schaap, M. (2012): The impact of differences in large-scale circulation output from climate models on the regional modeling of ozone and PM, *Atmos. Chem. Phys.*, 12, 9441-9458, Doi10.5194/acp-12-9441-2012,
- Munir, S., Chen, H. and Ropkins, K., 2013. Quantifying temporal trends in ground level ozone concentration in the UK. *Science of the total environment*, 458, S..217-227
- Noordijk, H.(2003): PROZON EN PROPART; statistische modellen voor smogprognose. RIVM Report, Bilthoven, The Netherlands
- Otero N., Sillmann, J., Schnell, J. L., Rust, H.W, and Butler, T. (2016):Synoptic and meteorological drivers of extreme ozone concentrations over Europe. *Environmental Research Letters*, 11,2, S.24005
- Otero, N., Sillmann, J., Mar, K. A., Rust, H. W., Solberg, S., Andersson, C., Engardt, M., Bergström, R., Bessagnet, B., Colette, A., Couvidat, F., Cuvelier, C., Tsyro, S., Fagerli, H., Schaap, M., Manders, A., Mircea, M., Briganti, G., Cappelletti, A., Adani, M., D'Isidoro, M., Pay, M.-T., Theobald, M., Vivanco, M. G., Wind, P., Ojha, N., Raffort, V., and Butler, T. (2018): A multi-model comparison of meteorological drivers of surface ozone over Europe, *Atmos. Chem. Phys.*, 18, S. 12269-12288,
- Pavlik, D., Heidenreich, M., Wolke, R., Stern R. (2016): KLENOS – Einfluss einer Änderung der Energiepolitik und des Klimas auf die Luftqualität sowie Konsequenzen für die Einhaltung von Immissionsgrenzwerten und Prüfung weitergehender emissionsmindernder Maßnahmen, Endbericht, 2016, Umweltbundesamt, Texte 84/2016
- Paoletti E., De Marco A., Beddows D.C S, Harrison R. M., and Manning W.J. (2014): Ozone levels in European and USA cities are increasing more than at rural sites, while peak values are decreasing. *Environmental Pollution*, 192, doi: 10.1016/j.envpol.2014.04.040
- Pearce, J. L., Beringer, J., Nicholls, N., Hyndman, R. J. and Tapper, N. J. (2011): Quantifying the influence of local meteorology on air quality using generalized additive models. *Atmospheric Environment*, 45,6, S. 1328-1336
- Querol, X., Alastuey, A., Reche, C., Orto, A., Pallares, M., Reina, F., Dieguez, J.J., Mantilla, E., Escudero, M., Alonso, L. and Gangoiiti, G., (2016): On the origin of the highest ozone episodes in Spain. *Science of the Total Environment*, 572, S. 379-389
- Reid, C. E., Jerrett, M., Petersen, M. L., Pfister, G. G., Morefield, P. E., Tager, I. B., Raffuse, S. M. and Balmes, J. R., (2015): Spatiotemporal prediction of fine particulate matter during the 2008 northern California wildfires using machine learning. *Environmental science & technology*, 49, 6, S. 3887-3896

- ReKliEs-DE (2017): Regionale Klimaprojektionen Ensemble für Deutschland, Ergebnisbericht, Word Data Center for Climate, doi: 10.2312/WDCC/ReKliEsDe_Ergebnisbericht
- Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986): Learning Internal Representations by Error Propagation, Institute for cognitive Science, University of California, San Diego
- Rumelhart, D.E., James L. McClelland, and the PDP research group (1987): Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundation. ISBN 9780262680530, MIT Press
- Schmidtke, J., Schmidt, K. (2019): Ableitung sicherer Trendaussagen zur Entwicklung der Luftqualität in Deutschland Abschlussbericht UBA-Texte, ISSN 1862-4804
- Schratz, P., Muenchow, J., Iturritya, E., Richter, J., Brenning, A. (2018): Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data. arXiv preprint arXiv:1803.11266
- Staehelin, J., Thudium, J., Buehler, R., Volz-Thomas, A., Graber, W. (1994): Trends in surface ozone concentrations at Arosa (Switzerland). *Atmos. Environ*, 28, Doi 10.1016/1352-2310(94)90024-8
- Sillman, S., Samson, P.J., (1995): Impact of temperature on oxidant photochemistry in urban, polluted rural and remote environments. *J. Geophys. Res.-Atmos.* 100, Doi10.1029/94JD02146
- Smola, A. and Vishwanathan, S. V. N. (2008): Introduction to Machine Learning, Cambridge University Press, UK
- Thompson, M.L., Reynolds, J., H. Cox, L.H., Guttorp, P., Sampson, P.D. (2001): A review of statistical methods for the meteorological adjustment of tropospheric ozone, *Atmospheric Environment*, 35, S. 617-630
- UBA (2020): Luftqualität 2019 vorläufige Auswertung, ISSN 2363-829X, www.umweltbundesamt.de/publikationen/luftqualitaet-2019, Umweltbundsamt, Dessau
- Varotsos, K.V, Giannakopoulos, C., Tombrou, M. (2019): Ozone-temperature relationship during the 2003 and 2014 heatwaves in Europe, *Regional Environmental Change*, 19, S. 1653-1665
- Watson, G.L., Telesca, D., Reid, C.E., Pfister, G.G. and Jerrett, M., 2019. Machine learning models accurately predict ozone exposure during wildfire events. *Environmental Pollution*, 254, p.112792.
- Van Dingenen, R., Crippa, M., Maenhout, G., Guizzardi, D., Dentener, F. (2018): Global trends of methane emissions and their impacts on ozone concentrations, EUR 29394 EN, Publications Office of the European Union, Luxembourg, JRC113210, ISBN 978-92-79-96550-0, Doi:10.2760/820175
- WHO (2021): Global air quality guidelines. Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. ISBN 9789240034211 <https://apps.who.int/iris/handle/10665/345329>, World Health Organization, Geneva
- Wilson, R.C., Fleming, Z.L., Monks, P.S., Clain, G., Henne, S., Konovalov, I.B., Szopa, S., Menut, L. (2012): Have primary emission reduction measures reduced ozone across Europe? An analysis of European rural background ozone trends 1996-2005. *Atmos. Chem. Phys.* 12, Doi:10.5194/acp-12-437-2012
- Wood, S. (2006): Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC press.
- Yan Y., Pozzer A., Ojha, N., Lin, J., and Lelieveld J. (2018): *Analysis of European ozone trends in the period 1995-2014*. *Atmospheric Chemistry and Physics*, 18, Doi: 10.5194/acp-18-5589-2018
- Zhan, Y., Luo, Y., Deng, X., Chen, H., Grieneisen, M.L., Shen, X., Zhu, L. and Zhang, M. (2017): Spatiotemporal prediction of continuous daily PM_{2.5} concentrations across China using a spatially explicit machine learning algorithm. *Atmospheric Environment*, 155, S.129-139
- Zhan, Y., Luo, Y., Deng, X., Grieneisen, M.L., Zhang, M. and Di, B., 2018. Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment. *Environmental Pollution*, 233, S. 464-473

A Benutzerhandbuch R-Skripte

Systemanforderungen:

- R + R-Studio. R Version 4.0.4 und 4.0. R-Studio Version ist: 1.4.1103 für Entwicklung verwendet, Version 3.6.1 und 1.2.1335 beim UBA wäre ausreichend.
- aktuelle Java-Version, um den Zugang zu H2O zu erleichtern
- Zugriff auf H2O (library(h2o), Meldungen über die Notwendigkeit der Installation neuer Versionen von H2O können ignoriert werden)

Erforderliche Eingabedateien sind:

- Jährliche Ozondateien mit höchstem täglichen 8-Stunden-Mittelwert (MDA8), die der UBA-Datenbank entnommen wurden
- COSMO-Reanalysedateien, wie vom DWD heruntergeladen. Dies sind monatliche Dateien von T2M, U, V, RH, RAIN_GSP, RAIN_CON, ASOB_S. Die Dateien werden im grib-Format auf dem COSMO-REA6 Gitter geliefert und können mit Hilfe der von TNO bereitgestellten Python-Skripte in geeignete netcdf-Dateien umgewandelt werden.
- Datei mit Stationsmetadaten (Name, Code, Koordinaten, Stationsklassifizierung) stations_metadata.csv

Es ist möglich, den Arbeitsbereich (z. B. das trainierte Modell) zu speichern und für eine neue Sitzung wieder zu laden.

Hauptebene: Library_Parameter.R

Das Skript Libraries_Parameters.R ist die höchste Ebene, in der die Pfade zu den Eingabe- und Ausgabedaten definiert werden. In diesem Skript können der Schwellenwert für Ozonüberschreitungen (derzeit 120,5 $\mu\text{g}/\text{m}^3$) sowie die Jahre, für die die Meteorologiedaten extrahiert werden sollen, festgelegt werden. Da die Extraktion der Meteorologiedaten ein zeitaufwändiger Prozess ist, empfiehlt es sich, das Skript in Abschnitten von wenigen Jahren auszuführen.

Es wird empfohlen, neue Pfade zu den Daten (Results_Path, Plot_Path) zu erstellen, wenn ein neuer Datensatz für das Training verwendet wird. Alternativ müssen die Namen für die automatisch erstellten Abbildungen und die Modellstatistiken, die die Performance des Modells darstellen, in den einzelnen R-Skripten, die sie erzeugen, angepasst werden.

Das Skript enthält Aufrufe zur Ausführung anderer Skripte, die Teile der Datenaufbereitung oder des Maschinellen Lernens durchführen. Diese Aufrufe können kommentiert/auskommentiert werden, um flexibel zu sein, welcher Teil des Codes ausgeführt werden muss. Die Datenaufbereitung muss nur einmal durchgeführt werden, kann jedoch für neue Jahre ausgeführt werden, um den Datensatz zu erweitern. Das Maschinelle Lernen ist jetzt optimiert, aber für einen neuen Datensatz könnte man das Training eines neuen Algorithmus oder die Auswahl von Stationen untersuchen wollen. Bei der Verwendung von Aufrufen zu anderen Skripten werden nicht alle Ausgaben des Skripts angezeigt. Wenn alle Zwischenergebnisse angezeigt werden sollen (z. B. Statistiken), empfiehlt es sich, diese Skripte Zeile für Zeile

auszuführen. Dies gilt insbesondere für die Skripte, mit denen die Modelle für Maschinelles Lernen erstellt werden.

Vor der Durchführung des Regressionsmodells muss angegeben werden, welcher Datensatz verwendet werden soll (Extrem-/Nicht-Extremfälle, Einbeziehung/Ausschluss von Variablen). Dies kann in dem Abschnitt vor Task 6 geschehen. Die Abschnittsnummern in diesem Anhang entsprechen den Tasks, wie sie im Hauptskript angegeben sind. Die Ergebnisse in Kapitel 6 basieren auf der Auswahl der Prädiktoren wie in Tabelle 6, aber es werden auch sinnvolle alternative Optionen angegeben, so dass man leicht herausfinden kann, worin die Unterschiede bestehen.

A.1 Verarbeitung von Ozondaten, Ozone_Exceedances.R

Das Skript Ozone_Exceedances.R liest alle verfügbaren Ozondateien ein, die tägliche maximale 8-Stunden-Ozonkonzentrationen enthalten. Sie sind nach Jahren geordnet und umfassen alle für das jeweilige Jahr verfügbaren Stationen.

Ermittelt werden die Anzahl der Überschreitungen, die Anzahl der Episoden und die maximale Episodenlänge.

Es wird eine Ausgabedatei erzeugt (Processed_Station_Data.csv), die für jede Station und jedes Jahr mit ausreichender Datenabdeckung die Anzahl der Überschreitungen, die Anzahl der Episoden und die maximale Episodenlänge enthält.

Das Skript läuft schnell, so dass es nicht notwendig ist, Daten für bestimmte Jahre auszuwählen. Es werden alle Dateien im angegebenen Verzeichnis berücksichtigt. Die Auswahl der zu berücksichtigenden Jahre kann bei der Erstellung des Datensatzes für das Training getroffen werden. Der Vorteil ist, dass auf diese Weise die Namen der Dateien der Eingangsdaten beliebig sind. Das Skript erzeugt eine Datei, die nur die Stationen mit gültigen Daten für den gesamten Zeitraum enthält.

A.2 Stationsrasterindizes für COSMO Meteo_CoLocation.R

Für die Stationsstandorte müssen Meteorologiedaten extrahiert werden. Die Meteorologiedaten werden aus den COSMO-REA6-Daten eingelesen, die auf einem gedrehten Polgitter liegen. Das macht es nicht einfach, die Indizes der Stationen auf dem COSMO-Gitter zu finden. Das Skript Meteo_CoLocation.R geht die Liste der aktiven Stationen durch und berechnet für jede Station die Entfernung zu allen Gitterpunkten im COSMO-Gitter.

Eingabedateien:

- Aktive_Stationen.csv
- T_2M.2D.201808_2dlonlat.nc (COSMO-Datei zum Einlesen von Gitterinformationen)

Dann wählt es die Gitterindizes mit der kürzesten Entfernung zur Station als Gitterzellenindizes aus.

Die Ausgabe des Skripts ist eine Datei: Colocation_grid.csv.

Die Definition der Rasterpunkte ist ein zeitaufwändiger Prozess für eine lange Liste von Stationen. Im Prinzip muss er nur einmal durchgeführt werden. Für neue Stationen könnte das Skript ausgeführt werden, um die Datenindizes für diese Punkte zu generieren, und diese

können mit der vorhandenen Rasterdatei kombiniert werden. Die Dateinamen für Eingabe und Ausgabe können im Skript geändert werden.

A.3 Extraktion von Klimakennzahlen: Meteo_Extraction.R

Eingabe:

- Colocatoin_Grid.csv
- COSMO REA6 Meteodaten (stündliche Daten pro Monat)

Dieses Skript extrahiert die Meteorologiedaten aus den monatlichen COSMO-Dateien an den Ozonstationen und berechnet und speichert die Klimakennzahlen für April-September.

Ausgabe: Jährliche Dateien pro Klimakennzahlen mit Daten pro Station für die 6 Monate.

Der Prozess ist zeitaufwändig und es wird empfohlen, ihn für mehrere Jahre zu starten, anstatt alle Jahre auf einmal.

Es wurde auch festgestellt, dass für das Jahr 1996 für den 4. und 5. Mai Daten in den heruntergeladenen COSMO-Dateien für die Variablen Strahlung, relative Feuchte und Windvektoren fehlten. Bei der relativen Luftfeuchte und den Windvektoren ist der Verlust dieser beiden Tage unbedeutend, bei der Strahlung jedoch dominieren die Füllwerte die Jahreswerte, da sie kumulativ sind. Daher werden die Strahlungsdaten zunächst pro Tag gespeichert. Bei der Verarbeitung wurden die Daten des 4. Mai durch die des 3. Mai, und die des 5. Mai durch die des 6. Mai ersetzt.

A.4 Konstruktion eines Datensatzes: Construct_Dataframe.R

Eingabe

- verarbeitete Ozondaten
- verarbeitete Klimadaten
- jährliche Emissionsdaten (Emissionen_NOX-VOC.csv)

Dieses Skript kombiniert alle Klimadaten und Daten zu Ozonüberschreitungen zu Datensätzen, die für das Training eines Algorithmus verwendet werden können. Eine Unterauswahl wird durch den Ausschluss von Bergstationen (Höhe > 900m) getroffen. Auch andere Teildatensätze können manuell definiert werden (z. B. nur bestimmte Jahre).

Es werden standardmäßig mehrere Datensätze erstellt:

- Full_Data_Meteo_Exceedance_Episodes_EpisLength_Emissions_No_Mounts: vollständiger Datensatz, keine Bergstationen, enthält vollständige Stationsklassifizierungen
- Data_Meteo_Exceedance_Episodes_EpisLength_Emissions_No_Mounts: dasselbe, aber mit nur 6 Klassifizierungstypen für Stationen
- Teilmenge dieser Daten für 2003 zur Prüfung und Bewertung des Extremfalls (Samplefrom2003, TestData_from2003)
- einfache Erweiterung durch den Benutzer zur Definition weiterer Datensätze

Die Konstruktion des Datensatzes ist schnell und der resultierende Datensatz wird direkt in den nachfolgenden Schritten für die Datenauswahl und das Training verwendet.

Die Erstellung des Datensatzes kann einige Minuten dauern. Die Datensätze können für weitere Schritte gespeichert und importiert werden.

A.5 Einstufung extreme und normale Jahre: Anomaly_Detection.R

Eingabe: Datensatz aus dem vorherigen Schritt.

Das Skript ermittelt, welche Jahre und Stationen in Bezug auf die Anzahl der Überschreitungen als extrem zu betrachten sind. Basiert auf der R-Library rstatix.

Ausgabe: Datensatz mit Extremwerten und Datensatz mit normalen Bedingungen

Die Auswahl ist schnell und die resultierenden Datensätze werden direkt in den nachfolgenden Schritten für die Datenauswahl und das Training verwendet.

A.6 Klassifizierung von Klimakennzahlen für extreme Jahre und normale Jahre Classification_Model.R

Dieses Skript bestimmt, welche Klimabedingungen einem extremen Jahr entsprechen und welche Bedingungen zu einem normalen Jahr gehören. Es stellt eine Verbindung zu H2O für Maschinelles Lernen her.

Eingabe der Daten aus dem vorangegangenen Schritt: pro Station die Klassifizierung auf der Grundlage der Anzahl der Ozonüberschreitungen, Klimakennzahlen und Stationsklassifizierung.

In Classification_Model.R ist es möglich, die Parameter des Maschinellen Lernens zu ändern, insbesondere die Laufzeit (max_run_time, derzeit 300 s) kann an längere Trainingszeiten angepasst werden, was in einigen Fällen zu Verbesserungen führt. Außerdem kann die explizite Hinzunahme oder der Ausschluss von Algorithmen erzwungen werden.

Ausgabe ist eine Reihe von trainierten Klassifizierungsmodellen mit einer Rangfolge ihrer Performance (exm_class).

Die Ausführung dieses Skripts kann einige Minuten dauern, ein Fortschrittsbalken wird angezeigt.

A.7 Training des Modells Regression_Model_Exceedances.R oder Regression_Model_Episodes.R

Dieses Skript trainiert das Modell für Ozonüberschreitungen.

Es stellt eine Verbindung zu H2O für Maschinelles Lernen her.

Eingabe: Datensatz der Ozonüberschreitung für Extremwerte oder für normale Bedingungen (festgelegt in Libraries_Parameters.R)

Standardmäßig werden 80 % der Daten zum Training und 20 % zum Testen (Validierung) verwendet. In Regression_Model_Exceedances.R ist es möglich, die Parameter für Maschinelles Lernen zu ändern, insbesondere die Laufzeit (max_runtime_secs, derzeit 500 s) wurde vom H2O-Standard (300 s) an längere Trainingszeiten angepasst, was zu einer Verbesserung führte. Außerdem kann die/der explizite Hinzunahme/Ausschluss von Algorithmen erzwungen werden, und es können unterschiedliche relative Verteilungen auf Test- und Validierungsdaten eingestellt werden.

Die Ausgabe ist eine Reihe von trainierten Modellen `Ozone_exceedance_model` oder `Ozone_episode_model` entweder für extreme oder normale Bedingungen, mit einer Reihenfolge ihrer Performance (`exm_class`). Eine Zusammenfassung der Eingabedaten wird auf dem Bildschirm angezeigt, ebenso wie ein Boxplot der Überschreitungen.

Die Leistungen (RMSE, MSE, MAE) der verschiedenen Teilmodelle werden auf dem Bildschirm angezeigt und grafisch aufbereitet.

Die Wichtigkeit von Variablen kann für einige ML-Modelle, wie GBM, dargestellt werden, nicht aber für das Stacked Ensemble-Modell, das in vielen Trainingsfällen die beste Option darstellt.

Die Ausführung dieses Skripts kann etwa 10 Minuten dauern (Trainingszeit 500 s), ein Fortschrittsbalken wird angezeigt.

A.8 Visualisierung und Interpretation der Ergebnisse

Die Performance des trainierten Modells kann auf verschiedene Weise visualisiert werden. Die Funktion „H2Oexplain“ bietet Visualisierungen, die im R-Studio-Darstellungsfenster angezeigt werden. Diese sind:

- Residuenanalyse-Plot
- Variable Importance Plots (für die Modelle, für die sie verfügbar ist, insbesondere GBM)
- Variable Importance Heatmap (Wichtigkeit der Variablen)
- Korrelation der Modelle
- SHAP-Plots
- Partial Dependence Plots für die 5 am meisten beitragenden Variablen
- Conditional expectations (nicht erläutert in diesem Bericht)

Für jede dieser Darstellungen wird die Art der in der Darstellung enthaltenen Informationen im Konsolenfenster erläutert. Im Fenster "Diagramme" können die Diagramme durchgeblättert werden, und der Benutzer kann die Abbildung manuell über die Schaltfläche "Exportieren" im Fenster "Diagramme" speichern.

Darüber hinaus kann die Modellkorrelations-Heatmap aufgezeichnet werden, um zu sehen, wie unabhängig die verschiedenen Modelle sind.

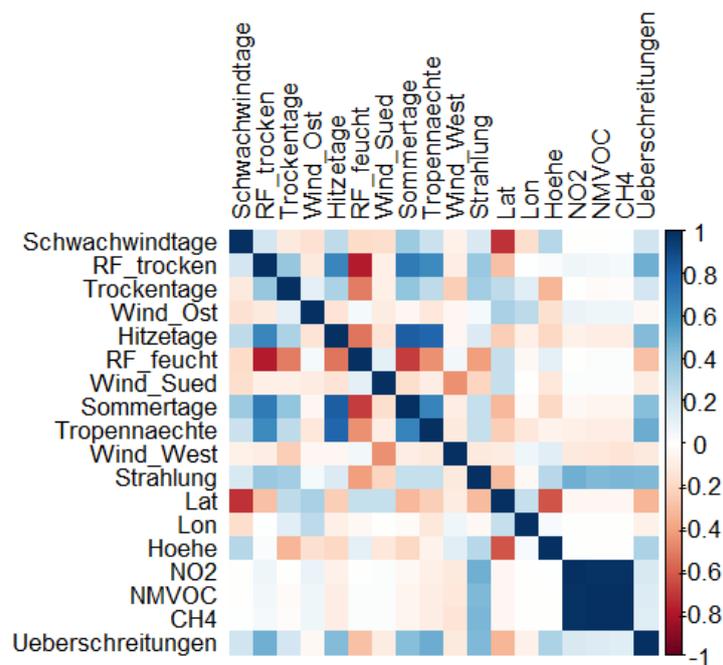
Im Idealfall sollten die Residuen zufällig verteilt sein. Muster in diesem Diagramm können auf potenzielle Probleme bei der Modellauswahl hinweisen, z. B. die Verwendung eines einfacheren Modells als erforderlich, die Nichtberücksichtigung von Heteroskedastizität, Autokorrelation usw. Wenn Sie "gestreifte" Linien von Residuen sehen, ist dies ein Artefakt einer ganzzahligen (im Gegensatz zu einer reellen) Antwortvariablen.

B Zusätzliche Experimente

B.1 Berücksichtigung aller Daten

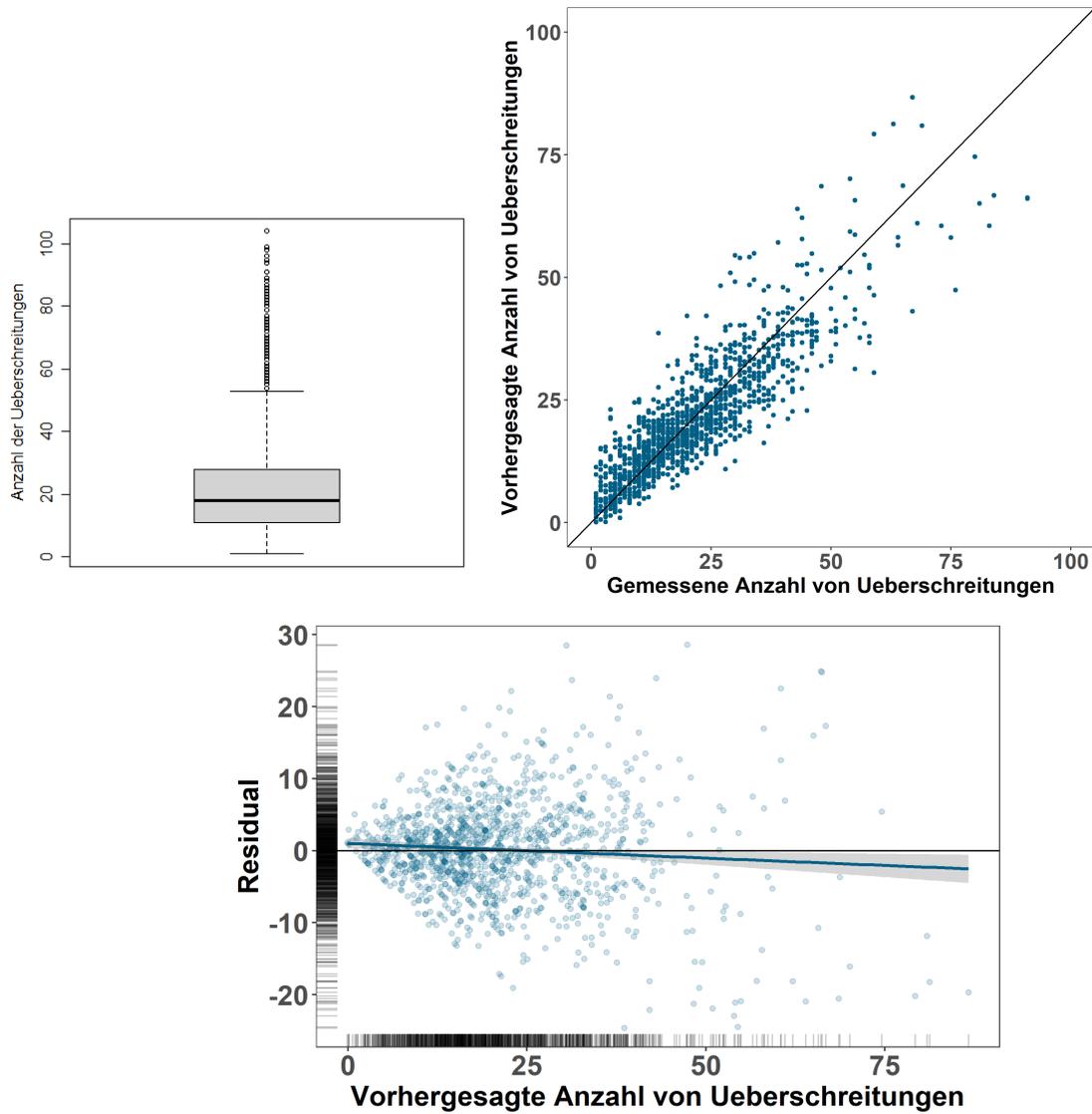
Wenn alle Daten (ohne Bergstationen) über alle Jahre und Stationen einbezogen werden, wird ein recht gutes Ergebnis erzielt. Allerdings sind die Residuen etwas höher als im Fall der Modelle für „normale“ bzw. „extreme“ Werte, für die die Residuen kaum einen absoluten Wert von 20 überschreiten. Dies zeigt, dass der Ansatz für „normale“ und „extreme“ Werte separate Modelle zu erstellen, der bessere Ansatz ist.

Abbildung 70: Korrelation von Eingangsvariablen/allen Klimakennzahlen, Emissionen und Anzahl der Ozonüberschreitungen, alle Daten



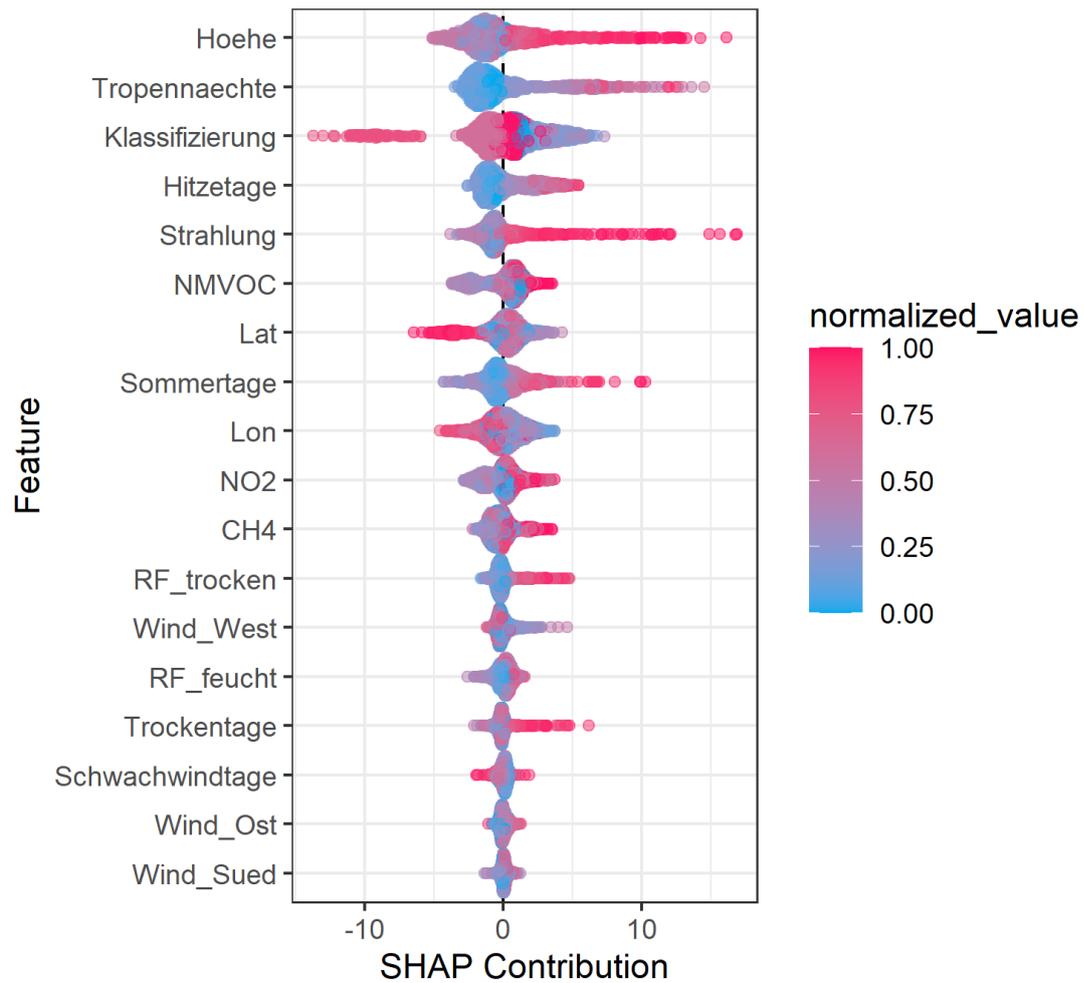
Quelle: Eigene Darstellung, TNO

Abbildung 71: Boxplot der Anzahl der Ozonüberschreitungen (links), vorhergesagte versus tatsächliche Anzahl der Überschreitungen (rechts) und Residuenanalyse (unten), alle Daten



Quelle: Eigene Darstellung, TNO

Abbildung 72: SHAP-Plot, Anzahl der Ozonüberschreitungen, alle Daten



Quelle: Eigene Darstellung, TNO

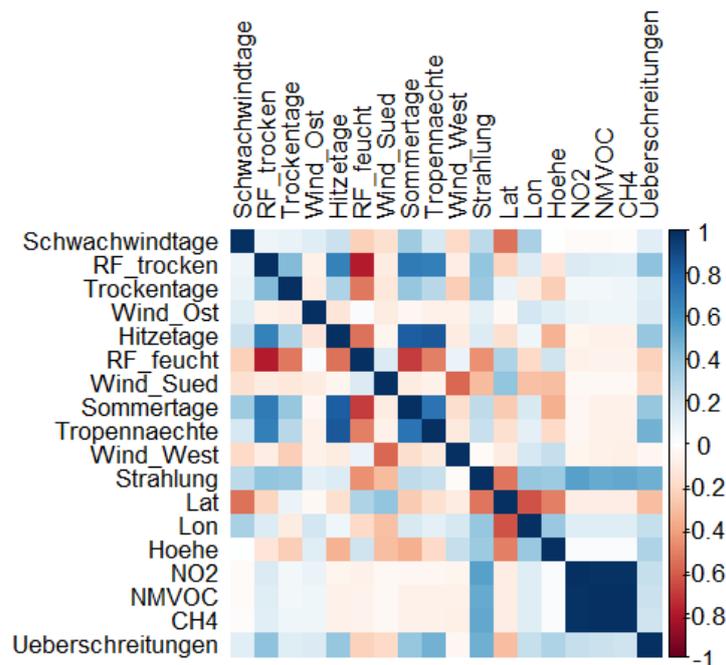
B.2 Verwendung von Klimazonen

Durch die Verwendung von Klimazonen für Ozon, wie sie im KLENOS-Projekt definiert sind, wären die Daten pro Zone homogener, was die Performance verbessern könnte. Der Nachteil ist, dass weniger Daten verfügbar sind, insbesondere für die „extremen“ Werte, was die Ergebnisse weniger robust macht. Für Zone 1 sind 2187 Daten verfügbar, für Zone 2 2828 Daten und für Zone 3 1189. Die Performance für Zone 2 (Mitte Deutschlands von Nord nach Süd) ist besser als die für Zone 1 (Südwesten Deutschlands), für die die niedrigen Werte tendenziell überschätzt und die hohen Werte tendenziell unterschätzt werden. Für Zone 2 sind die Kennzahlen für die Modellperformance ähnlich wie beim Modell A für „normale“ Werte.

B.2.1 Zone 1

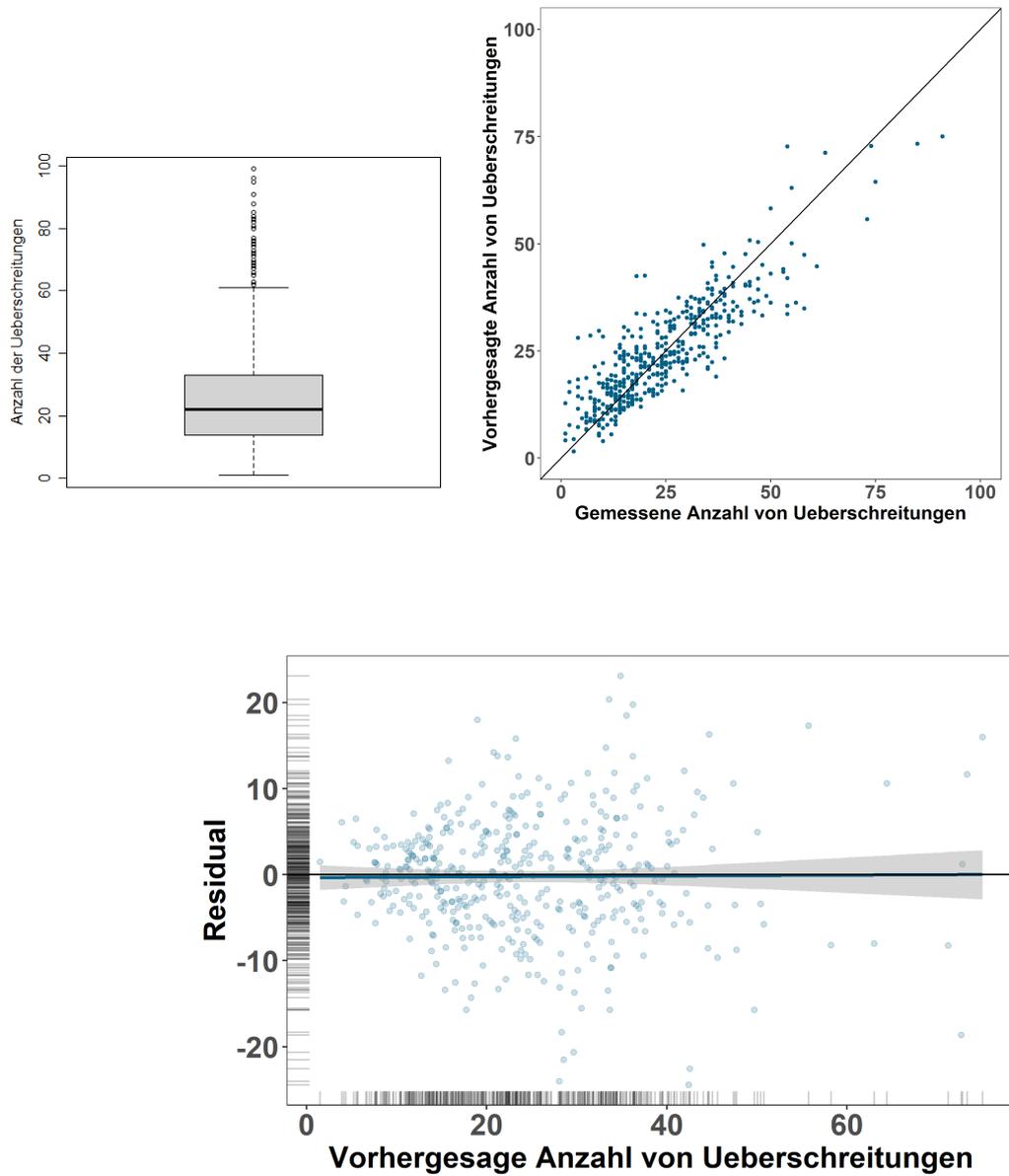
Zone 1 enthält 2187 Daten und umfasst das Ruhrgebiet und die Mainregion. Die Anzahl der tropischen Nächte zeigt die stärkste Korrelation, dicht gefolgt von der Strahlung. Für die Anzahl der Tage mit geringer Luftfeuchte (RF_trocken) und für Sommertage sind die Korrelationen etwa gleich hoch.

Abbildung 73: Korrelationsdiagramm Anzahl der Ozonüberschreitungen für Zone 1



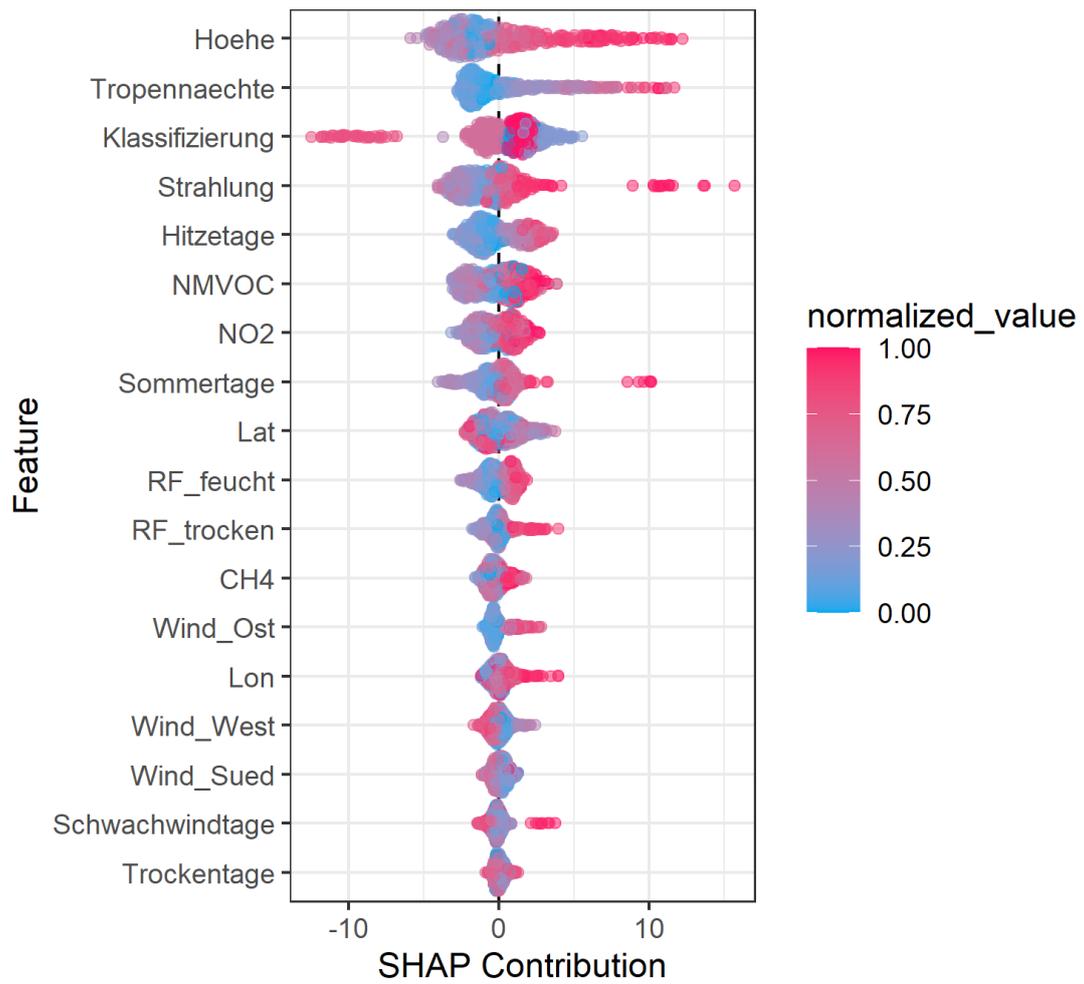
Quelle: Eigene Darstellung, TNO

Abbildung 74: Boxplot der gemessenen Anzahl der Ozonüberschreitungen (links), vorhergesagte versus gemessene Anzahl der Ozonüberschreitungen (rechts) und Residuenanalyse (unten), Zone 1



Quelle: Eigene Darstellung, TNO

Abbildung 75: SHAP-Plot, Anzahl der Ozonüberschreitungen, Zone 1

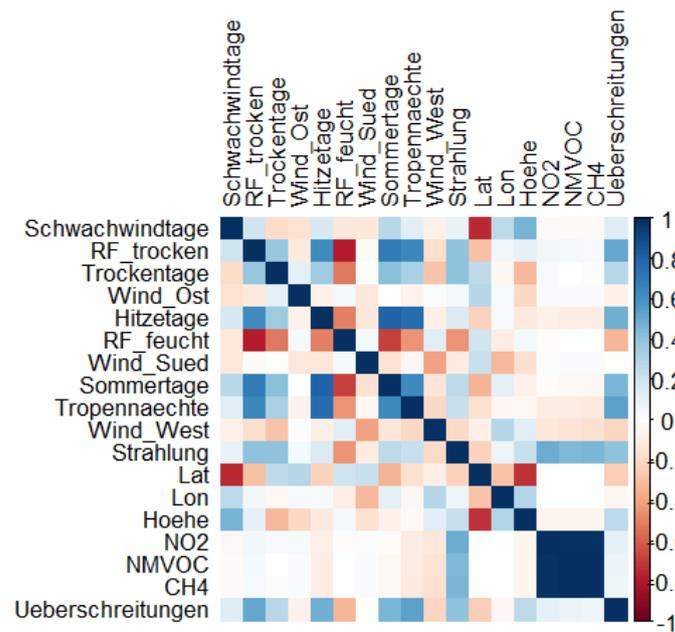


Quelle: Eigene Darstellung, TNO

B.2.2 Zone 2

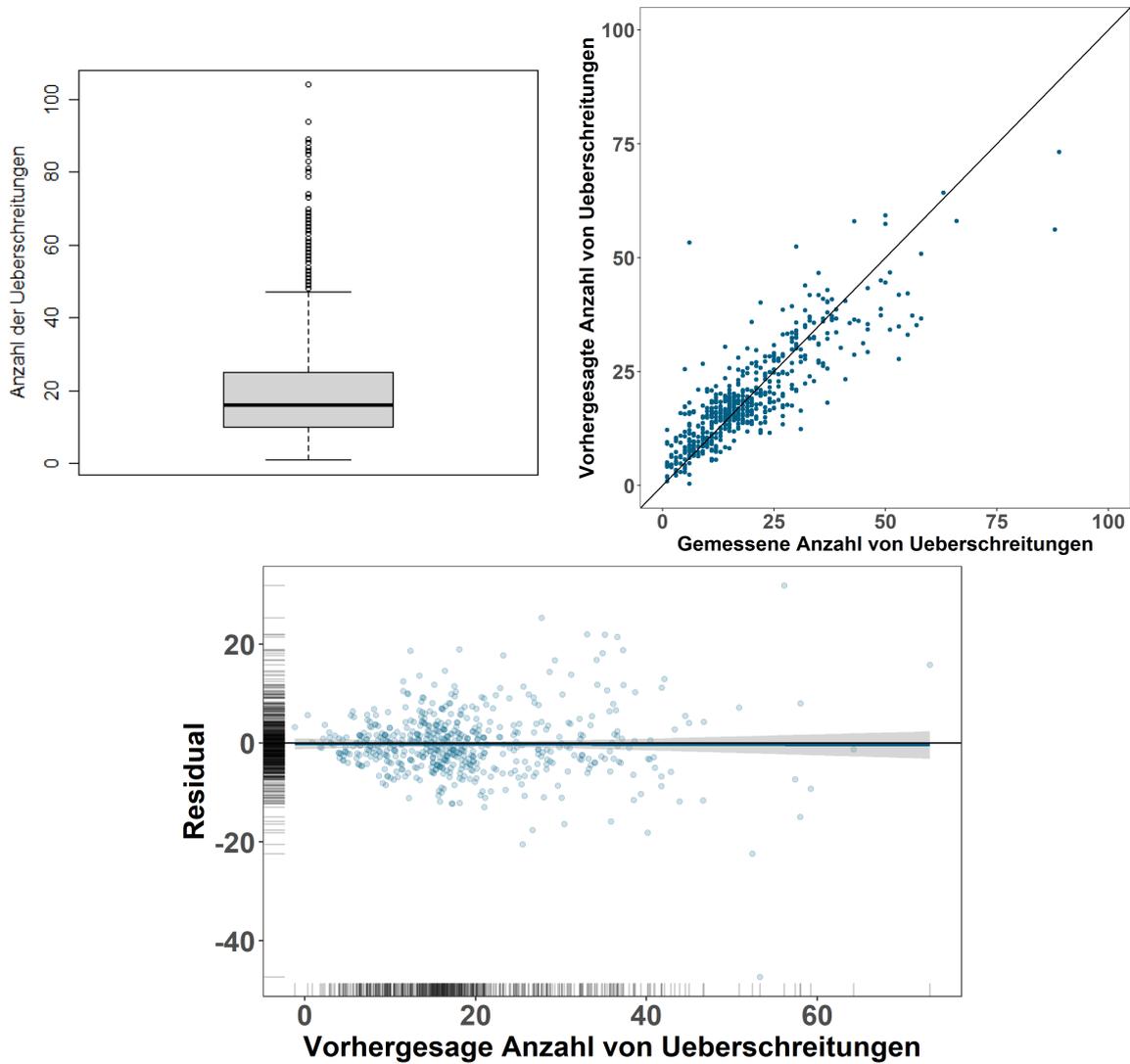
Dies ist die größte Zone mit 2828 Daten. Sie deckt die größte Fläche und den größten Höhenbereich ab, was sich in dem größeren Einfluss der Stationsstandortparameter im SHAP-Plot widerspiegelt. Tropische Nächte und Hitzetage weisen eine etwas stärkere Korrelation auf als Sommertage und die Strahlung. Für das endgültige Modell werden die größten Werte leicht unterschätzt.

Abbildung 76: Korrelationsdiagramm Anzahl der Ozonüberschreitungen für Zone 2



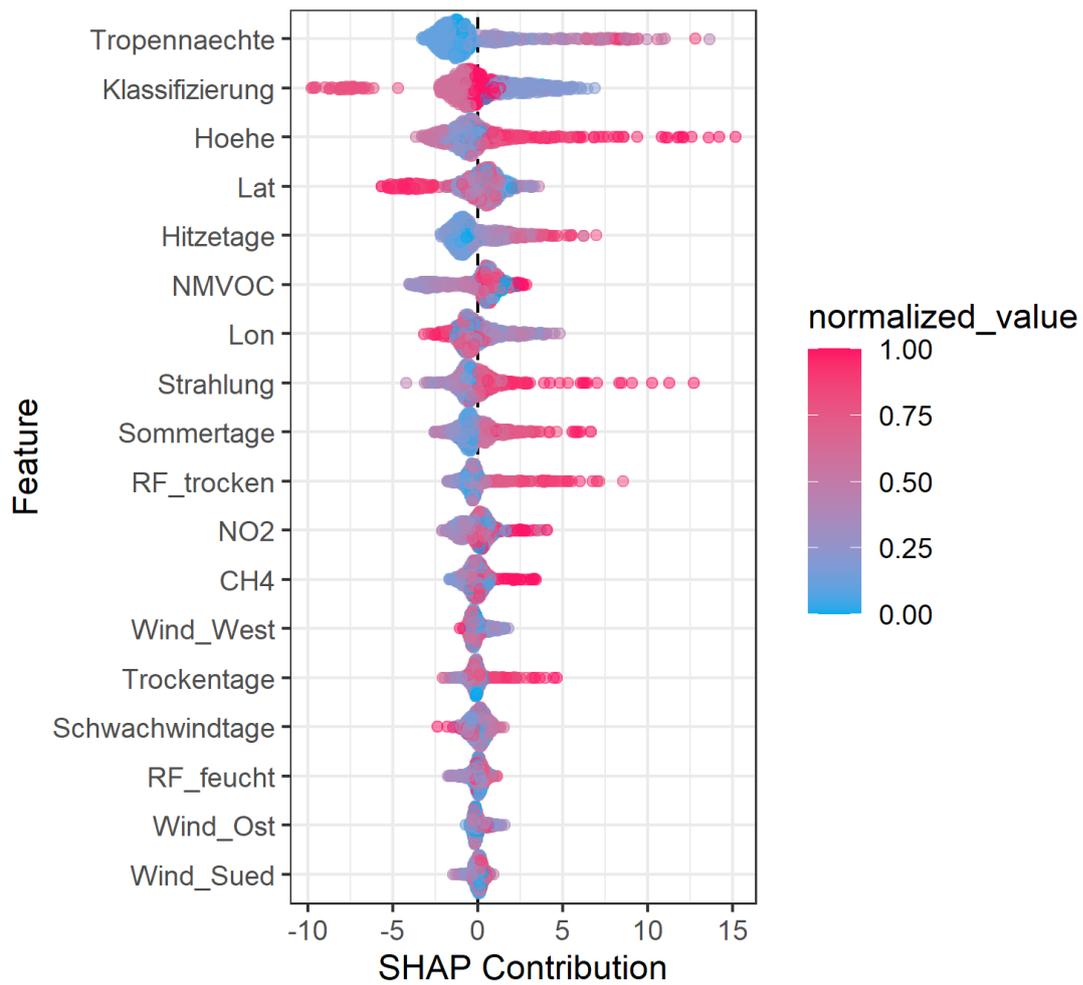
Quelle: Eigene Darstellung, TNO

Abbildung 77: Boxplot der Anzahl der Ozonüberschreitungen (links), vorhergesagte versus tatsächliche Anzahl der Ozonüberschreitungen (rechts) und Residuenanalyse (unten), Zone 2



Quelle: Eigene Darstellung, TNO

Abbildung 78: SHAP-Plot, Anzahl der Ozonüberschreitungen, Zone 2

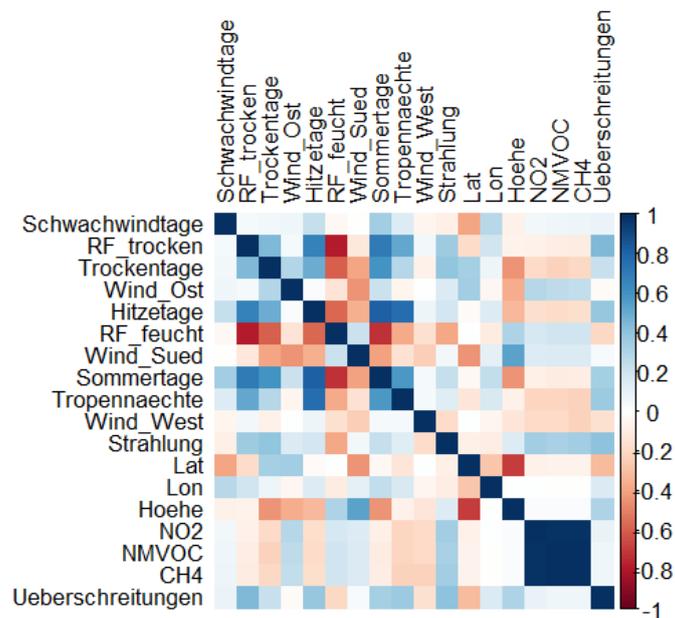


Quelle: Eigene Darstellung, TNO

B.2.3 Zone 3

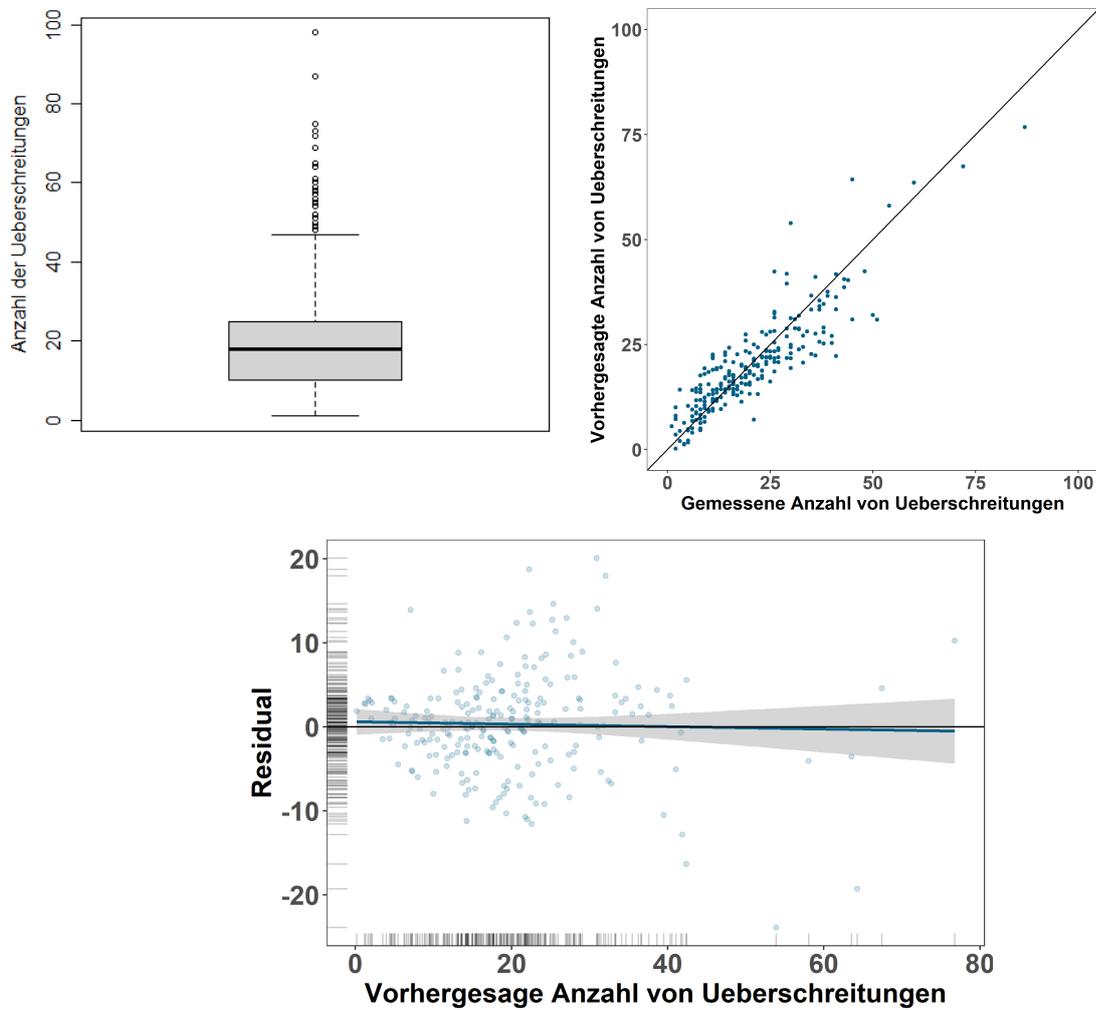
Zone 3 ist die klimatisch homogenste Zone mit der geringsten Anzahl von Daten (1189). Die Region umfasst Berlin als Großstadt mit einem relativ dichten Messnetz. In dieser Region sind die positiven Korrelationen mit temperaturbezogenen Prädiktoren und der Strahlung offensichtlich, aber weniger stark als in den anderen Regionen. Dadurch ist die positive Korrelation mit der Anzahl der Tage mit geringer Luftfeuchte (RF_trocken) stärker ausgeprägt als in den anderen Regionen. Die Residuenanalyse zeigt Unterschätzungen insbesondere für die höchsten Werte, die jedoch innerhalb von 20 Ozonüberschreitungen bleiben. Strahlung und Anzahl der Hitzetage sind die einflussreichsten Prädiktoren im besten GBM-Modell direkt nach der Rolle des Stationsstandorts. Die Unterschätzung der höchsten Werte ist stärker als in Zone 2, aber die Anzahl der Daten ist für eine zuverlässige Bewertung zu gering.

Abbildung 79: Korrelationsdiagramm Anzahl der Ozonüberschreitungen Zone 3



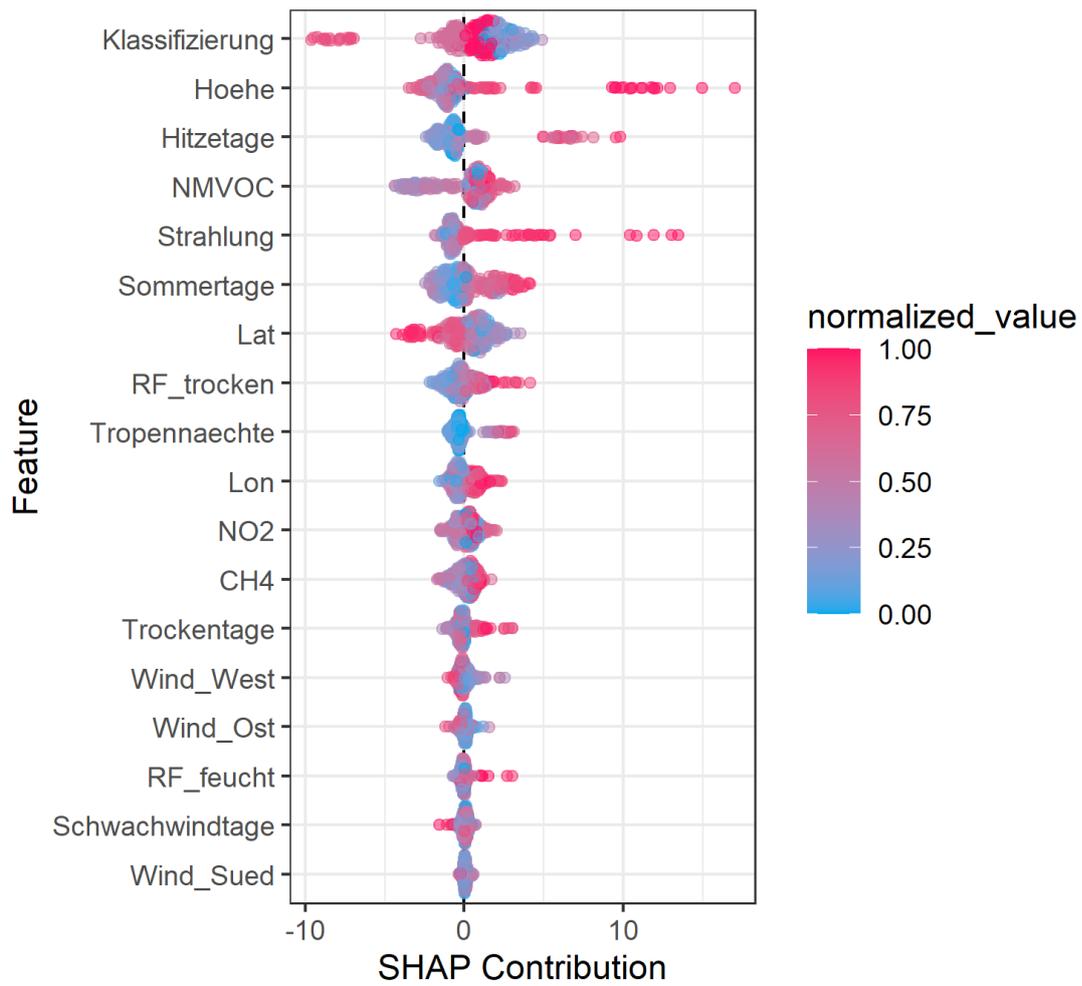
Quelle: Eigene Darstellung, TNO

Abbildung 80: Boxplot der Anzahl der Ozonüberschreitungen (links), vorhergesagte versus tatsächliche Anzahl der Überschreitungen (rechts) und Residuenanalyse (unten), Zone 3



Quelle: Eigene Darstellung, TNO

Abbildung 81: SHAP-Plot, Anzahl der Ozonüberschreitungen, Zone 3



Quelle: Eigene Darstellung, TNO

C Alternative Ansätze: CTM und Training auf tägliche Daten

Die Verwendung aggregierter Daten schränkt die Verfügbarkeit von Daten für das Training und die Wiedergabe der Beziehung zwischen meteorologischen Prädiktoren und dem Prädiktand erheblich ein. Im Rahmen seiner Doktorarbeit hat Tuo Deng (TU Delft, Niederlande) ein Modell zur Anpassung der Beziehung pro Station erstellt, das auf täglichen meteorologischen Werten und der Ozon-MDA8 basiert und einen Multilayer Perceptron-Algorithmus (MLP, Neuronales Netz) verwendet. Es wurde nur der Zeitraum 2014 – 2018 verwendet, wobei 2014 – 2017 als Trainingsdaten und 2018 als Validierungsdaten dienten.

Aufbau des MLP-Modells:

- ▶ Eingabeschicht („input layer“): Tägliche maximale 2m-Temperatur, tägliche mittlere Windgeschwindigkeit (10 m), tägliche mittlere Windrichtung (10 m), tägliche mittlere relative Luftfeuchte, tägliche akkumulierte Sonneneinstrahlung, täglich akkumulierter Niederschlag.
- ▶ Es wurden zwei „hidden layers“ mit je 100 Neuronen verwendet
- ▶ Die Ausgangsschicht („output layer“) waren Ozon MDA8-Werte.
- ▶ Es wurde eine e-gewichtete Verlustfunktion verwendet $L(y) = e^{\alpha \cdot y} * (y - \hat{y})^2$ mit $\alpha > 0$, um hohe MDA8-Werte stärker zu gewichten und Extremwerte besser darzustellen. Das Optimum für α reichte von 0 für DEBB007 bis 9 für DEBB066.

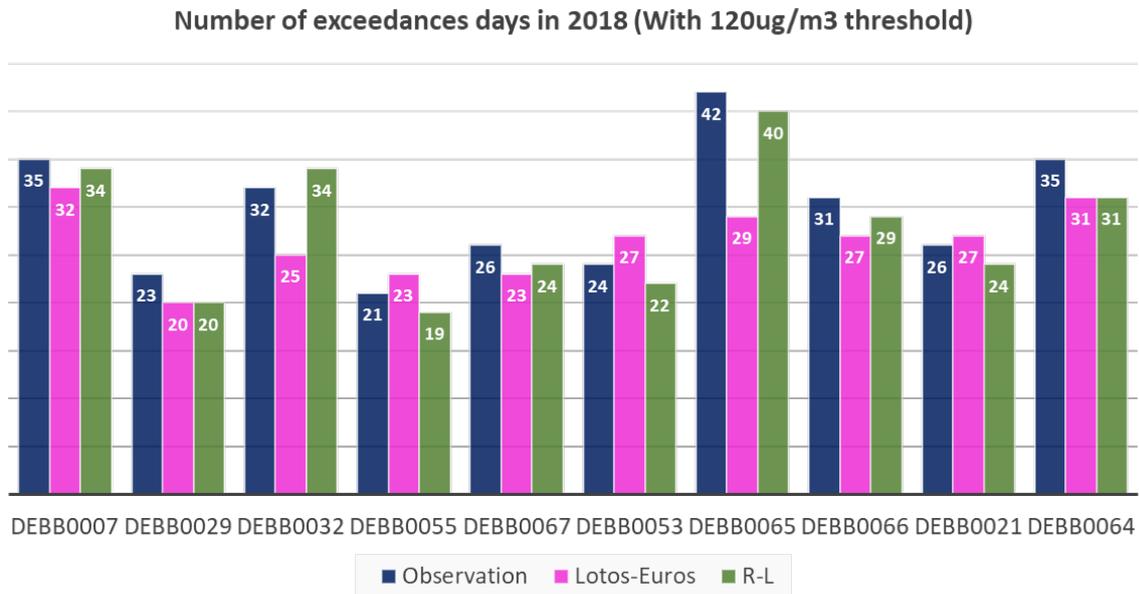
In dieser Studie konzentrierte sich die Validierung auf die Analyse der korrekt vorhergesagten Überschreitungen der Grenze von 100 oder 120 $\mu\text{g}/\text{m}^3$. Hier stellen wir die Gesamtzahl der vorhergesagten Überschreitungen und die Gesamtzahl der beobachteten Überschreitungen dar, unabhängig davon, ob sie für den richtigen Tag vorhergesagt wurden. Wir haben auch MDA8 aus einer LOTOS-EUROS-Simulation, wie sie in der Konfiguration für den Copernicus Atmospheric Monitoring Services verwendet wird, als Referenz für eine CTM-Simulation, einbezogen.

Für die einzelnen Stationen ergaben die MLP RMSE-Werte zwischen 12,6 und 13,6, und die Korrelationen (R^2) lagen zwischen 0,8 und 0,84. Die Variablen, die bei weitem den größten Beitrag leisteten, waren die tägliche Höchsttemperatur und die einfallende Strahlung, wobei die relativen Beiträge je nach Station variierten. Die relative Luftfeuchte war ebenfalls von Bedeutung, wenn auch um einen Faktor 3 - 4 niedriger als die Variable mit dem höchsten Beitrag. Windgeschwindigkeit und -richtung leisteten einen sehr geringen Beitrag, und der Beitrag des Niederschlags war bei den meisten Stationen vernachlässigbar.

Abbildung 82 zeigt, dass das MLP die Anzahl der Überschreitungen mit einer maximalen Abweichung von 4 Tagen erfasst, wobei eine Tendenz zur Unterschätzung besteht. LOTOS-EUROS liegt größtenteils im gleichen Bereich, mit Ausnahme von DEBB065 und DEBB032, für die starke Unterschätzungen festgestellt wurden.

Das Validierungsjahr war 2018, in dem extreme Ozonwerte in weiten Teilen Deutschlands gemessen wurden. Für die in Anhang C untersuchten Stationen war die Anzahl der Überschreitungen kleiner als 55.

Abbildung 82: Gemessene und vorhergesagte Anzahl von Überschreitungen für 2018 unter Verwendung eines „Relationship fitting“-Modells auf der Grundlage täglicher Daten oder der CTM LOTOS-EUROS



Quelle: T. Deng, persönliche Mitteilung

