

Praktikumsbericht

Robin Mayer, Nils Nover, Mariella Zunker

June 23, 2020

Contents

1 Fragestellung und Auswahl der Datensätze	2
2 Datenaufbereitung	2
3 Datenauswertung	3
3.1 Vergleich der Stickoxidwerte zwischen den Bundesländern	5
3.2 Vergleich der Stickoxidwerte zwischen ausgewählten ländlichen und städtischen Gebieten	5
4 Interpretation der Ergebnisse	5
5 Diskussion	5

1 Fragestellung und Auswahl der Datensätze

Die Frage der Schädlichkeit von Stickoxiden und deren Zusammenhang mit Dieselfahrzeugen hat in der Vergangenheit die Debatte um die Verkehrswende dominiert. Um einen Überblick über die Rolle von verschiedenen Antriebsarten zu bekommen, sollten in diesem Projekt Daten zu Antriebsdaten mit Stickoxidwerten in Deutschland verglichen werden. Hierbei sind einerseits die Unterschiede in verschiedenen Stadt- und Landkreisen zu berücksichtigen, auf der anderen Seite sollte auch die zeitliche Entwicklung und der lokale Einfluss von sogenannten Umweltzonen betrachtet werden. In Umweltzonen sind als nicht-schadstoffarm gekennzeichnete Fahrzeuge verboten, um die Luftqualität zu verbessern.

Zu Analyse der Fragestellung wurden Datensätze des Umweltbundesamtes [QUELLE] sowie des Kraftfahrtbundesamtes [QUELLE] genutzt. Die Datensätze des Umweltbundesamtes lagen als .xlsx-Dateien vor und beinhalteten Informationen zu jahresgemittelten Stickoxidwerten an verschiedenen deutschen Messstationen(1), während das Kraftfahrtbundesamt hauptsächlich .pdf-Dateien veröffentlichte, in denen die zu einem Stichtag zugelassenen Anzahlen von Autos nach Antriebsart in den jeweiligen Orten aufgelistet waren (2). Beide Quellen stellen Datensätze nach Jahr zu Verfügung.

bundesland	jahr	station	name	umgebungstyp	emissionstyp	jahresmittel	maxstundenwert
Brandenburg	2002	'DEBB001'	Burg (Spreewald)	vorstädtisches Gebiet	Hintergrund	10	69
Brandenburg	2002	'DEBB006'	Cottbus-Süd	städtisches Gebiet	Hintergrund	19	107
Brandenburg	2002	'DEBB009'	Forst	vorstädtisches Gebiet	Hintergrund	16	83
Brandenburg	2002	'DEBB021'	Potsdam-Zentrum	städtisches Gebiet	Hintergrund	21	111

Figure 1: Auszug aus dem NO2-Datensatz.

2 Datenaufbereitung

Zunächst mussten die Datensätze in ein Format konvertiert werden, in dem sie sinnvoll verarbeitet werden konnten. Dazu wurden die .pdf-Dateien ins .xlsx-Format konvertiert. Aufgrund der großen Unterschiede der beiden Formate musste hierbei jedoch manuell noch viel nachgebessert werden (WAS?), weshalb nicht alle zur

land	rb	Stadt	Insgesamt	Benzin	Diesel	Gas (einschl. bivalent)	Hybrid	Elektro
BADEN- WUERTTEMBERG	STUTTGART	08111 STUTTGART,STADT	298.172	182.451	111.148	1.893	1.788	814
BADEN- WUERTTEMBERG	STUTTGART	08115 BOEBLINGEN	244.396	155.578	85.499	1.452	1.336	474
BADEN- WUERTTEMBERG	STUTTGART	08116 ESSLINGEN	319.920	208.669	107.640	1.907	1.301	325

Figure 2: Auszug aus dem KFZ-Datensatz.

Verfügung stehenden Jahrgänge ausgewertet werden konnten. Die .xlsx-Dateien konnten im Anschluss in ein .csv-Format konvertiert und als solches eingelesen werden. Aufgrund der guten Verfügbarkeit bestehender Tools und Libraries wurde für die Auswertung Python gewählt. Hierbei konnte vor allem auf das Datenanalysetool Pandas sowie Numpy und Matplotlib für die Auswertung zugegriffen werden. Das Jupyter Notebook stellt zudem eine übersichtliche und gut nachvollziehbare Programmierungsumgebung dar, in der Code gut im Team erarbeitet werden kann. Nach dem Einlesen wurden die Datensätze auf Vollständigkeit und Fehler durchsucht. Dabei wurde auf nicht erfasste Datenpunkte sowie offensichtliche Abweichungen wie negative Zahlen geachtet. Zudem mussten die Städtenamen und alle als String codierten Variablen überprüft und vereinheitlicht werden. So wurden alle Namen in Großbuchstaben und ohne Umlaute dargestellt, sowie Rechtschreibfehler und Unterschiede in der Darstellung von Doppelnamen korrigiert. Trennzeichen wie Komma oder Slash wurden vereinheitlicht und Zahlenwerte als Float gecastet. Zudem wurden Messwerte, die keinem eindeutigen Ort zugewiesen werden konnten (für einige Messwerte wurde als Bundesland "Umweltbundesamt" angegeben), aus dem Datensatz gelöscht, da eine nicht-automatisierte Zuordnung in der verfügbaren Zeit nicht möglich war.

Insgesamt nahm die Aufbereitung der Daten viel Zeit in Anspruch. !?!?!?

3 Datenauswertung

Die Datensätze wurden zuerst isoliert betrachtet. Zur Messung der Luftwerte existierten Daten aus den Jahren 2002-2019, KfZ-Daten konnten von 2012-2019 genutzt werden. Die Datensätze enthielten auch Zuordnungen der einzelnen Orte zu verschiedenen Abstufungen der Urbanität wie "vorstädtisches Gebiet". Die bestehende Einteilung wurde zur Vereinfachung zu den drei Kategorien "städtisch", "vorstädtisch" sowie "ländlich" zusammengefasst. Für einen ersten Überblick wurden die Stickoxidwerte über die Zeit geplottet (3). In der Tendenz stimmt das Ergebnis hierbei mit der Auswertung des Umweltbundesamtes (QUELLE) überein, die absoluten Werte weisen jedoch leichte Abweichungen auf. Dies ist wohl auf die leicht veränderte Einteilung der Ort in die drei Kategorien zurückzuführen.

Zur weiteren Analyse wurden die Datensätze im Anschluss zusammengefügt. Dabei sollten die Stickstoffdaten über die Orte den Kraftfahrzeugdaten zugeordnet werden. Um Dopplungen zu vermeiden, wurde dazu pro Jahr und Bundesland überprüft, ob eine Stadt im Stickstoff-Datensatz auch im KfZ-Datensatz existiert und bei Übereinstimmung wurden die Daten in neuen Spalten hinzugefügt. Dieses Merging stellte eine größere Herausforderung dar als zunächst angenommen,

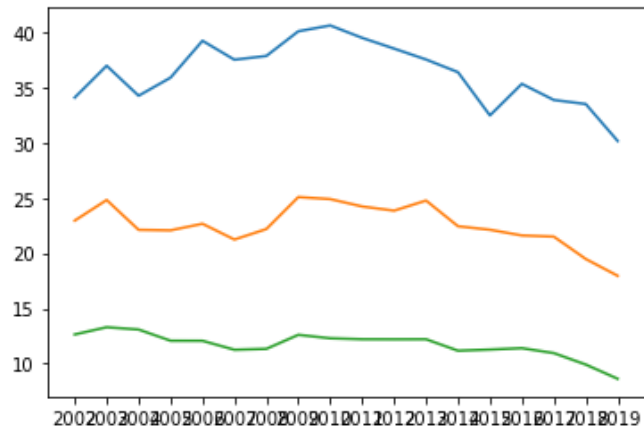


Figure 3: Entwicklung der Stickoxide über die Zeit. Blau: Städtisch. Gelb: Vorstädtisch. Grün: Ländlich.

da die Datensätze nicht die gleiche Zuordnung der Werte zu den Orten enthielten, wie zunächst angenommen. So bezogen sich die Daten des Kraftfahrtbundesamtes auf die regulären Landkreise, die Messungen des Umweltbundesamtes waren jedoch lediglich dem nächsten Ort zugewiesen und nicht der offiziellen Kreisstadt. Diese Zuordnung führte dazu, dass beim Merging ländliche Gebiete weniger gut zugeordnet werden konnten als größere Städte (4).

umgebungstyp	
ländlich Gebiet	2
ländlich regional	6
ländlich stadtnah	4
städtisches Gebiet	264
vorstädtisches Gebiet	37

Figure 4: Anzahl an Orten in den Kategorien, die zugeordnet werden konnten, beispielhaft am Jahr 2017.

Zudem ist zu beachten, dass einige Städte in den Datensätzen mehrfach vorkommen (z.B. Weimar Schwansestr. und Weimar Steubenstr.) (WIE wurde das noch mal gehandhabt)

Im Anschluss an die Zuordnung wurden die Stickstoffwerte gegen die Fahrzeugzahlen aufgetragen (5). Dabei konnte bereits ein grober linearer Trend erkannt werden, der später noch mithilfe linearer Regression genauer aufgezeigt werden sollte. Außerdem zeigte sich, dass zwei Städte eine extrem hohe Anzahl an gemeldeten Fahrzeugen aufwiesen. (NUR Das Jahr)

Als nächstes wurde der Anteil an Dieselfahrzeugen betrachtet und die Stickstoffwerte gegen diesen aufgetragen (6).

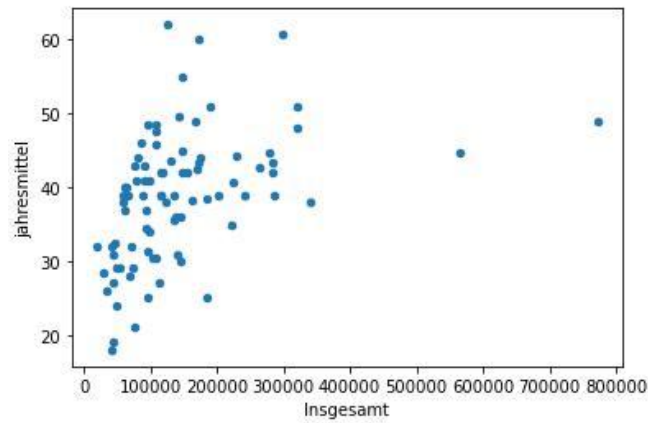


Figure 5: Stickstoffwerte in Abhängigkeit der Fahrzeugzahlen, beispielhaft am Jahr 2017.

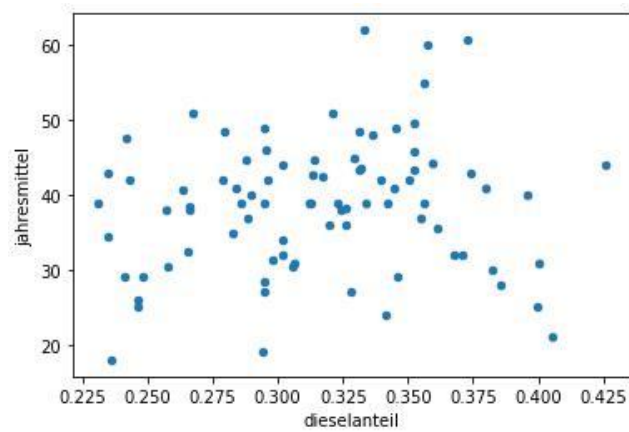


Figure 6: Stickstoffwerte in Abhängigkeit des Dieselanteil, beispielhaft am Jahr 2017.

3.1 Vergleich der Stickoxidwerte zwischen den Bundesländern

3.2 Vergleich der Stickoxidwerte zwischen ausgewählten ländlichen und städtischen Gebieten

4 Interpretation der Ergebnisse

5 Diskussion

Frage: dort wo autos gemeldet werden nicht unbedingt gefahren (Hamburg+Hannover), Aufwand pdf-Parsen –i nur neuere Daten, Matching der Orte in beiden Datensätzen, Städte mehrmals aufgezählt,