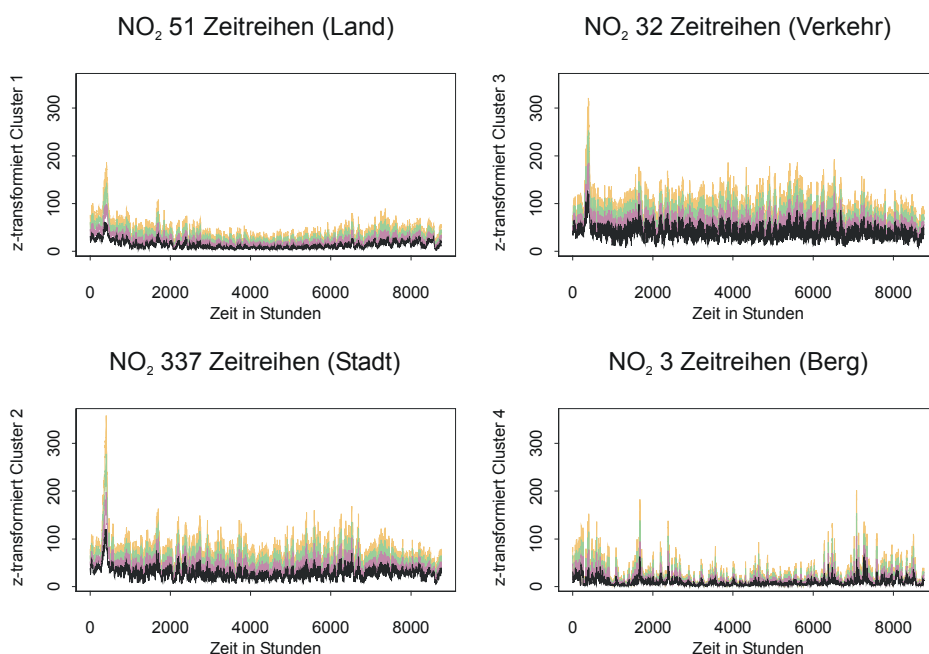


Umweltforschungsplan des Bundesministeriums für Umwelt,  
Naturschutz und Reaktorsicherheit

Förderkennzeichen (UFOPLAN) 200 42 265

Automatische Klassifizierung der Luftschadstoff-  
Immissionsmessungen aus dem LIMBA-Meßnetz  
Klassifizierung · 2. Teilbericht



für  
Umweltbundesamt II 6.4  
Postfach 33 00 22  
D-14191 Berlin  
von  
**IVU** Umwelt GmbH  
Burgweg 10 · D-79350 Sexau  
Telefon: 07641 53046  
Telefax: 07641 53047  
Dr.-Ing. Götz Wiegand  
eMail: gw@ivu-umwelt.de  
URL: <http://www.ivu-umwelt.de>  
Juli 2002

Die Abbildung zeigt Ergebnisse der Klassifizierung für die Zeitreihen von NO<sub>2</sub> in µg/m<sup>3</sup>. Schwarz sind die Mittelwerte der Zeitreihen des jeweiligen Clusters gezeichnet, farbig sind die einfachen- bis dreifachen Standardabweichungsintervalle eingetragen. Die Skalierung entspricht den Originalzeitreihen.



## Inhaltsverzeichnis

1	ALLGEMEINES.....	4
1.1	Aufgabenstellung.....	4
2	DIE DATEN UND DAS VERFAHREN.....	5
2.1	LIMBA-Dateisystem.....	5
2.2	Das multivariate Verfahren: Clusteranalyse.....	5
2.2.1	Allgemeines.....	5
2.2.2	Clusteranalyse und Statistik.....	5
2.2.3	Anwendungsbereiche der Clusteranalyse.....	5
2.3	Clusteranalyse zur Klassifikation von Meßstationen.....	6
2.4	Durchführung der Clusteranalyse.....	10
2.4.1	Auffälligkeiten in den Daten bei der Clusteranalyse.....	10
2.4.1.1	CO-Daten.....	10
2.4.1.2	PM10-Daten.....	10
2.4.2	Softwaresystem.....	11
2.4.3	Ergebnis.....	11
2.5	Interpretation der Cluster.....	12
2.5.1	Allgemeines.....	12
2.5.2	Geografische und soziodemografische Merkmale.....	12
2.5.3	Mittelwerte der Schadstoffe.....	14
3	ZUSAMMENFASSUNG.....	15
4	LITERATUR.....	16

## Abbildungsverzeichnis

Abbildung 1	Ablaufschema zur Berechnung der Distanzmatrix D (423 · 423) aus der Datenmatrix M (8'760 · 423) im Beispielfall.....	7
Abbildung 2	Dendrogramm der NO <sub>2</sub> -Konzentrationen von 1997..	8
Abbildung 3	Das Cluster 3 aus dem Dendrogramm der Abbildung 2 bei Distanz 130 vergrößert. ....	9
Abbildung 4	Ergebnisse der Klassifizierung für die Zeitreihen von NO <sub>2</sub> in µg/m <sup>3</sup> .....	11

## Tabellenverzeichnis

Tabelle 1	32 Messstellen aus dem Cluster 3 in Abbildung 3..	9
Tabelle 2	Name der Schadstoffe und Anzahl der Zeitreihen im Jahre 1997.....	10
Tabelle 3	UBA-Klassifikation mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km <sup>2</sup> der Gemeinden, in denen die Messstationen liegen.....	12
Tabelle 4	CO-Clusterzuordnung für 293 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km <sup>2</sup> der Gemeinden, in denen die Messstationen liegen. ....	12
Tabelle 5	NO <sub>2</sub> -Clusterzuordnung für 423 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km <sup>2</sup> der Gemeinden, in denen die Messstationen liegen. ....	13
Tabelle 6	NO-Clusterzuordnung für 396 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km <sup>2</sup> der Gemeinden, in denen die Messstationen liegen. ....	13
Tabelle 7	NO <sub>x</sub> -Clusterzuordnung für 390 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km <sup>2</sup> der Gemeinden, in denen die Messstationen liegen. ....	13
Tabelle 8	O <sub>3</sub> -Clusterzuordnung für 362 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km <sup>2</sup> der Gemeinden, in denen die Messstationen liegen. ....	13
Tabelle 9	PM10-Clusterzuordnung für 407 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km <sup>2</sup> der Gemeinden, in denen die Messstationen liegen. ....	13
Tabelle 10	SO <sub>2</sub> -Clusterzuordnung für 432 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km <sup>2</sup> der Gemeinden, in denen die Messstationen liegen. ....	13
Tabelle 11	Inhaltliche Zuordnung der 4 Cluster nach dem Kriterium geografische Merkmale. ....	14
Tabelle 12	Aufteilung und Mittelwert für die Clusteranalyse und die UBA-Klassifikation für alle verfügbaren Jahre und 6 Schadstoffe. ....	15

## 1 Allgemeines

### 1.1 Aufgabenstellung

Die EU-Luftqualitätsrahmenrichtlinie (RRL 96/62/EU) und die Tochterrichtlinien (TRL 1999/30/EG) und (TRL 2000/69/EG) setzen zur Bewertung der Luftqualität die Kenntnis von statistischen Kenngrößen von Zeitreihen voraus, die man aus gemessenen Zeitreihen direkt ableiten kann.

Modellrechnungen werden in der RRL 96/62/EU ausdrücklich erwähnt. Allerdings wird nicht darauf eingegangen, wie man die Kenngrößen der Zeitreihen durch Modellrechnungen erhält. Es ist nicht trivial, gewisse Kenngrößen der Zeitreihen mit Modellen zu berechnen. So ist z. B. die 18-malige Überschreitung von  $200 \mu\text{g}/\text{m}^3$  des  $\text{NO}_2$ -Wertes mit Modellen kaum zu ermitteln. Hier ist die RRL 96/62/EU unseres Erachtens nicht vollständig.

Das heute übliche Vorgehen, den statistischen Zusammenhang zwischen dem z. B. Jahresmittelwert und den interessierenden Kenngrößen aus *allen* verfügbaren Zeitreihen abzuleiten, ist ein suboptimales Vorgehen, denn es werden Zeitreihen von z. B. Bergstationen mit Zeitreihen von ausgeprägten Verkehrsstationen zusammen ausgewertet, obwohl sie unterschiedliche Zeitreihencharakteristika und statistische Kennwerte haben.

Diese unterschiedlichen Zeitreihencharakteristika auszunutzen, um ein automatisches Klassifikationsverfahren zu finden, ist die Grundidee des hier beschriebenen Vorhabens.

In dem Dateisystem des UBA liegen Zeitreihen für über 1'000 Stationen vor. Die UBA-Stationssklassifikation gibt eine Einordnung für 377 Stationen: 228 Stadt, 72 Land, 69 Verkehr und 8 Berg.

Die UBA-Stationssklassifikation soll mit nachvollziehbaren Verfahren fundiert werden. Dabei sollen möglichst viele Stationen klassifiziert werden.

In diesem Bericht wird die Methodik und Durchführung der Klassifikation behandelt.

## 2 Die Daten und das Verfahren

### 2.1 LIMBA-Dateisystem

Die Daten stammen aus dem LIMBA-Dateisystem des UBA. Die Vorprozessierung der Daten für die hier beschriebene multivariate Analyse ist in einem getrennten Bericht beschrieben (IVU Umwelt 2002.1).

### 2.2 Das multivariate Verfahren: Clusteranalyse

#### 2.2.1 Allgemeines

Die Bezeichnung Clusteranalyse<sup>1</sup> bezeichnet eine Anzahl von unterschiedlichen Klassifikationalgorithmen. Die Frage nach der Organisation der Beobachtungen in aussagefähige Strukturen ist eine allgemeine Frage, die sich Forschern in vielen Bereichen der Wissenschaft stellt.

Biologen müssen z. B. die unterschiedlichen Tierarten organisieren, bevor eine aussagefähige Beschreibung der Unterschiede zwischen Tieren möglich ist. Entsprechend dem modernen Klassifikationssystem, das in der Biologie eingesetzt wird, gehört der Mensch den Primaten, den Säugetieren, den Amnioten, den Wirbeltieren und den Tieren an. Je höher die Stufe der Gruppierung, desto weniger ähnlich sind die Mitglieder in der jeweiligen Kategorie. Der Mensch hat mehr Ähnlichkeiten mit weiteren Primaten als mit den mehr "entfernten" Mitgliedern der Säugetiere z. B. den Hunden.

#### 2.2.2 Clusteranalyse und Statistik

Die Clusteranalyse ist weniger eine typische statistische Prozedur, als eine "Ansammlung" unterschiedlicher Algorithmen, deren Aufgabe es ist, Objekte verschiedenen Gruppen zuzuordnen. Wichtig ist hier, dass Clusteranalyse – anders als viele andere statistische Prozeduren - meistens eingesetzt wird, wenn noch keine Hypothesen vorliegen, sondern man sich noch in der suchenden Phase befindet. Folglich ist die statistische Signifikanzprüfung hier nicht angebracht.

#### 2.2.3 Anwendungsbereiche der Clusteranalyse

Clusteranalyse-Techniken sind auf eine große Anzahl von Forschungsprobleme angewendet worden. Hartigan (Hartigan, 1975) stellt eine ausgezeichnete Zusammenfassung der vielen erschienenen Studien, welche über die Ergebnisse der Clusteranalysen berichten, zur Verfügung.

Auf dem Gebiet der Medizin, können z. B. die Zusammenfassung von Krankheiten, Heilungen für Krankheiten oder Symptome von Krankheiten zu sehr nützlichen Kategorisierungen führen.

Auf dem Gebiet der Psychiatrie ist die korrekte Diagnose der Gruppen von Symptomen wie Paranoia, Schizophrenie, usw. für erfolgreiche Therapie wesentlich.

---

<sup>1</sup> Der Ausdruck Clusteranalyse wurde zuerst verwendet von Tryon (Tryon R. C. (1939)), siehe auch z. B. Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979)



In der Archäologie haben Forscher versucht, systematische Einteilungen der Steinwerkzeuge, der Bestattungsobjekte, etc. herzustellen, indem sie Techniken der Clusteranalyse anwendeten.

Wann immer man einen "Berg" von Informationen in handliche aussagefähige Gruppen einstufen muss, ist die Clusteranalyse vom großen Nutzen.

Hier steht man vor der Aufgabe, verschiedene Zeitreihen bestimmten Gruppen zuzuordnen.

### 2.3 Clusteranalyse zur Klassifikation von Meßstationen

Ausgegangen wird von einer Menge vollständiger<sup>2</sup> Zeitreihen mit Messungen eines Stoffes für einen bestimmten Zeitraum. Die einzelnen Zeitreihen müssen z-transformiert<sup>3</sup> sein, damit nicht das Niveau einer Zeitreihe das Ergebnis verfälscht, denn die Zeitreihencharakteristik soll das Ergebnis prägen, nicht das Niveau einer Zeitreihe.

Das Verfahren soll am Beispiel der NO<sub>2</sub>-Messungen des Jahres 1997 demonstriert werden. Im Beispielsfall ist die Datenmatrix (M) eine 8'760 · 423-Matrix, das ist die Anzahl der Stunden des Jahres 1997 mal der Anzahl der Zeitreihen. Die Distanzmatrix (D) ist eine 423 · 423-Matrix (siehe Abbildung 1).

Die Elemente von D sind auf der Hauptdiagonalen Null, denn sie geben den Abstand einer Zeitreihe von sich selbst wieder. Die Matrix D ist symmetrisch und positiv definit. Sie enthält in den einzelnen Nicht-Diagonal-Elementen die euklidische Entfernung der z-transformierten Zeitreihen, wobei die Entfernung als Summe der stündlichen Entfernungen berechnet wird.

---

<sup>2</sup> Die Zeitreihen wurden fehlwertersetzt, siehe IVU Umwelt, 2002.1

<sup>3</sup> z-Transformation ist Überführung einer beliebigen Reihe in eine Reihe mit Mittelwert 0 und Varianz 1.

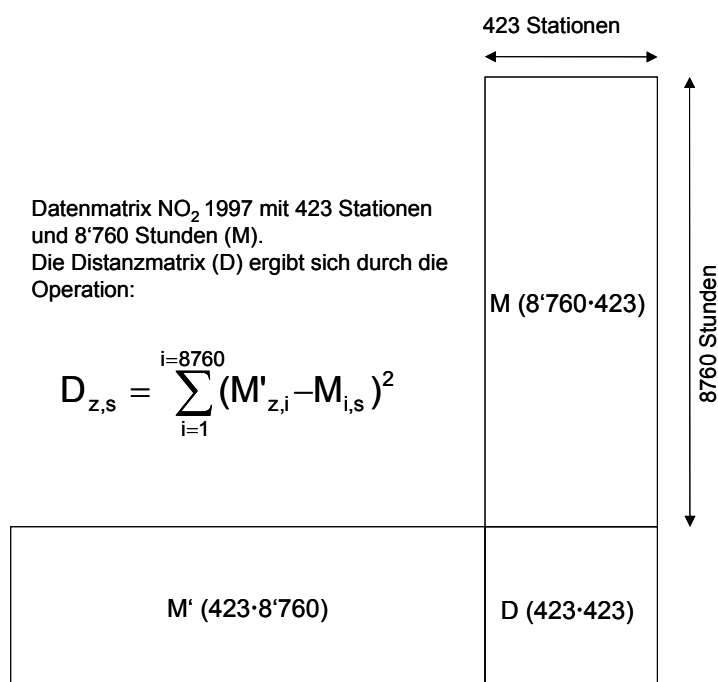


Abbildung 1 Ablaufschema zur Berechnung der Distanzmatrix D (423 · 423) aus der Datenmatrix M (8'760 · 423) im Beispielfall, Näheres siehe Text.

Von der Distanzmatrix ausgehend wird jetzt die hierarchische Clusteranalyse durchgeführt. Die Clusteranalyse wird ohne jedes statistische Modell durchgeführt, das heißt es sind keine Schlüsse auf Signifikanz oder ähnliches möglich.

Geht man von der Distanz Null aus, auf der die einzelnen Zeitreihen isoliert sind, hat man im Beispiel 423 Elemente. Die erste Verschmelzung von zwei Zeitreihen geschieht bei der minimalen endlichen Distanz. Das dadurch neu entstandene Element hat als Koordinaten den Mittelwert der einzelnen Koordinaten. Der Prozess wird fortgesetzt bis zur maximalen Distanz, bei der es nur noch ein Cluster gibt.

Abbildung 2 gibt Darstellung einer hierarchischen Clusteranalyse am Beispielfall wieder.



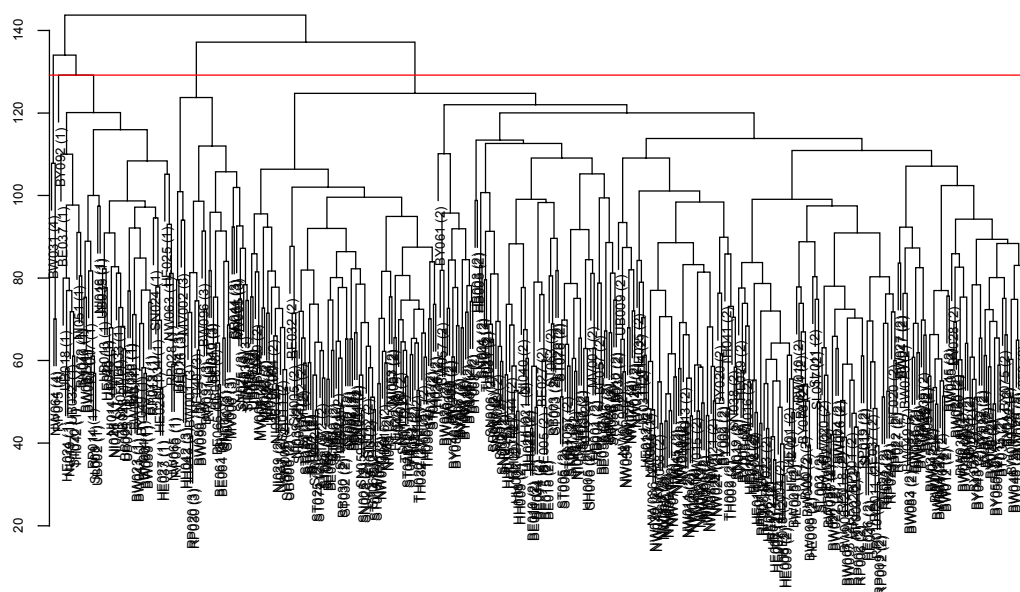


Abbildung 2 Dendrogramm der  $\text{NO}_2$ -Konzentrationen von 1997. Ausgangspunkt sind 423 z - transformierte Zeitreihen auf der Abszisse. Mit zunehmender Distanz (Ordinate) verschmelzen die einzelnen Zeitreihen zu Clustern. Näheres siehe Text und z. B. Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979).

An welcher Stelle man den Baum schneidet, also die Anzahl der Cluster festlegt, ist dem Anwender überlassen. Die UBA-Klassifikation enthält vier Klassen: Stadt, Land, Verkehr und Berg. In Anlehnung daran wird der Clusterbaum bei vier Clustern geschnitten, der Schnitt ist durch eine horizontale Linie bei Distanz 130 angegeben.

Da die Abbildung 2 ein wenig unübersichtlich ist, wurde das Cluster 3 aus dem Schnitt bei 130 vergrößert, es ist in Abbildung 3 dargestellt.

In Tabelle 3 sind die Namen der Messstationen, die dem Cluster 3 zugeordnet wurden, aufgeführt. Neben den Namen ist noch die UBA-Klassifikation, die Höhe in Metern und die Einwohnerdichte in Einwohner/ $\text{km}^2$  dargestellt.

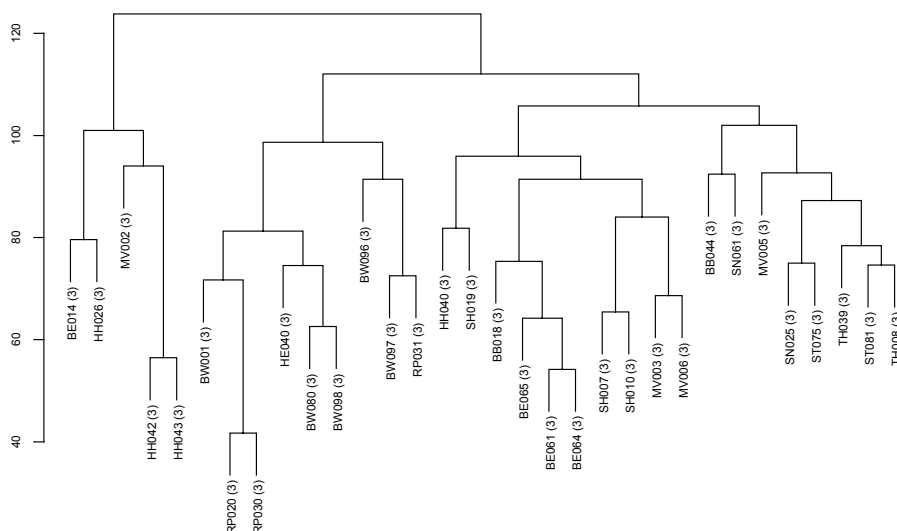


Abbildung 3 Das Cluster 3 aus dem Dendrogramm der Abbildung 2 bei Distanz 130 vergrößert. Näheres siehe Text.

Messstellen-Name	Code	UBA	Höhe	Einwohner
Oranienburg	BB018	Stadt	35	652.81
Cottbus, Bahnhofstr.	BB044	-	70	737.52
B Charlottenburg-Stadtautobahn	BE014	Verkehr	50	4359.07
B Steglitz-Schildhornstr.	BE061	-	45	4359.07
B Neukölln-Karl-Marx-Str. 76	BE064	-	40	4359.07
B Friedrichshain-Frankfurter Al	BE065	-	40	3128.88
Karlsruhe-Mitte	BW001	Verkehr	115	1597.81
Karlsruhe-Straße	BW080	-	115	1597.81
Ulm-Straße	BW096	-	490	978.2
Freiburg-Straße	BW097	-	260	1322.72
Mannheim-Straße	BW098	-	95	2122.86
Darmstadt-Hügelstraße	HE040	-	140	1127.37
HH Stresemannstraße (Straße)	HH026	-	20	2256.94
HH Theodor-Heuss-Platz (Straße)	HH040	-	9	2256.94
HH Tarpenbekstraße (Straße)	HH042	-	8	2256.94
HH Bremer Straße (Straße)	HH043	-	14	2256.94
Rostock	MV002	Stadt	10	1125.14
Neubrandenburg	MV003	Stadt	15	869.93
Schwerin	MV005	Stadt	42	789.36
Stralsund	MV006	Stadt	15	1577.7
Trier-Ostallee	RP020	Verkehr	140	852.75
Trier Th.-Heuss-Allee	RP030	-	140	852.75
Pirmasens Park-Brauerei	RP031	-	355	745.85
Kiel-Schützenwall Verk.	SH007	Stadt	17	1974.79
Lübeck-Lindenplatz Verk.	SH010	Verkehr	10	996.2
Kiel-Westring Verk.	SH019	-	25	1974.79
Leipzig-Mitte	SN025	Verkehr	110	1685.43
Dresden-Nord	SN061	Verkehr	112	1451.93
Halle/Verkehr	ST075	-	100	1884.15
Weißenfels/Verkehr	ST081	-	120	1311.92
Pößneck	TH008	-	215	595.71

Tabelle 1 32 Messstellen aus dem Cluster 3 in Abbildung 3. Hinter dem Messstellen-Name steht der Messstellen-Code, der eine Referenz zu der Abbildung 3 erlaubt. Es folgt die UBA-Klassifikation, die Höhe der Messstelle über NN in m und die Einwohnerdichte der Gemeinde, in der die Messstelle liegt, in Einwohner/km<sup>2</sup>.

Die Clusteranalyse wurde für die unterschiedlichen Schadstoffe für das Jahr 1997 durchgeführt. Zu der Anzahl der Zeitreihen im Jahr 1997 siehe Tabelle 2.

Tabelle 2 Name der Schadstoffe und Anzahl der Zeitreihen im Jahre 1997

Name	Anzahl
Kohlenmonoxid	294
Stickstoffdioxid	423
Stickstoffdioxid	396
Stickstoffdioxid	390
Ozon	362
Partikel	407
Schwefeldioxid	432

## 2.4 Durchführung der Clusteranalyse

### 2.4.1 Auffälligkeiten in den Daten bei der Clusteranalyse

#### 2.4.1.1 CO-Daten

Bei den CO-Clustern aus den z-transformierten Daten fielen die Stationen mit den Stationsnamen HB001, HB002 und HB005 aus dem Muster der sonstigen vier Cluster heraus. Eine Inspektion ergab, dass die drei Zeitreihen nur aus den Werten 1, 5, 10, 15, 20 und dem Fehlwertcode bestanden. Sie wurden aus der Clusteranalyse entfernt.

Bei den CO-Clustern aus den z-transformierten Daten fiel die Station mit dem Stationsnamen ST070 (Herzgerode) als eigenes Cluster auf. Da ein offensichtlicher Grund hierfür nicht zu erkennen war, wurde die Anzahl Cluster bei CO auf fünf erhöht.

Zwei rheinland-pfälzische Stationen wurden bei der Distanz Null verschmolzen, d.h. die Messzeitreihen sind identisch! Es handelt sich um die Stationen:

CODE	NAME	Höhe	Rechtswert	Hochwert	O <sub>3</sub> -Klassif.
DERP006	Ludwigshafen-Pfalzgrafenpl.	94.00	460241.11	5480487.63	none
DERP007	Mainz-Mombach	120.00	443868.12	5540843.64	ST

Weil nicht entschieden werden kann, welche der Stationen richtig ist, wurden beide aus der Clusteranalyse ausgeschlossen.

#### 2.4.1.2 PM10-Daten

Bei den PM10-Clustern aus den z-transformierten Daten fiel die Station mit dem Stationsnamen SL008 (Nonnweiler) als eigenes Cluster auf. Die PM10-Messwerte aus der Station Nonnweiler waren in dem Fehlwertersetzungsschritt mit den umgebenden Stationen kaum korreliert, darum wurde die Ausfallperiode zwischen der 7'377. und der 7'932. Stunde von dem Fehlwertalgorithmus gleich dem letzten beobachteten Wert gesetzt. Damit wird diese Reihe unplausibel und sie wurde bei der Clusteranalyse ausgeschlossen.

## 2.4.2 Softwaresystem

Das Softwaresystem mit dem die Clusteranalyse durchgeführt wurde ist S+4.52 auf Windows 2000<sup>4</sup>.

Das verwendete Verfahren ist das hierarchische Clusteranalyseverfahren (complete linkage), das in der Standard-Prozedur "hclust" implementiert ist.

Die Verwendung von Clusteranalyseverfahren setzt Informationen über die Distanzstruktur bzw. Ähnlichkeitsstruktur voraus (siehe Kapitel 2.3). Die euklidische Distanzmatrix wurde mit der standardmäßigen Prozedur "dist" ermittelt.

## 2.4.3 Ergebnis

Als Ergebnis liefert die Analyse je eine – verständlicherweise nicht identische - Clusteraufteilung für die Schadstoffe CO, NO<sub>2</sub>, NO, NO<sub>x</sub>, O<sub>3</sub>, PM<sub>10</sub> und SO<sub>2</sub>.

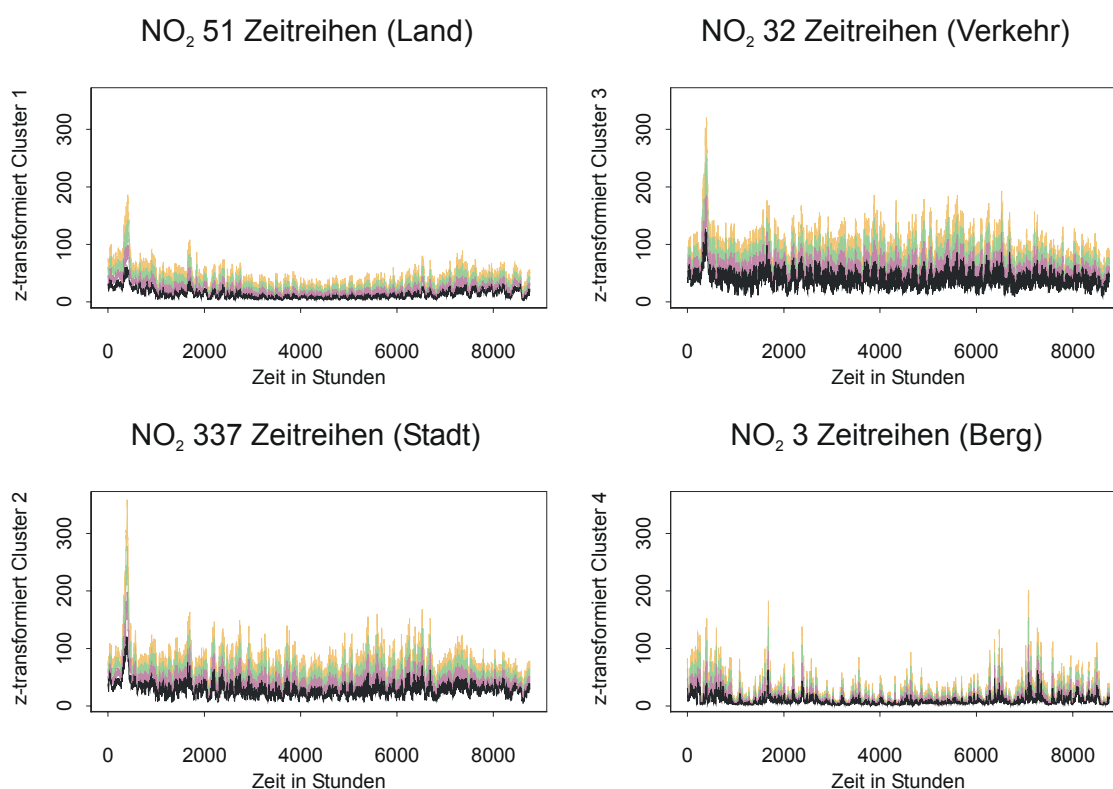


Abbildung 4 Die Abbildung zeigt Ergebnisse der Klassifizierung für die Zeitreihen von NO<sub>2</sub> in µg/m<sup>3</sup>. Schwarz sind die Mittelwerte der Zeitreihen des jeweiligen Clusters gezeichnet, farbig sind die einfachen- bis dreifachen Standardabweichungsintervalle eingetragen. Die Skalierung entspricht den Originalzeitreihen.

Die Abbildung 4 zeigt Ergebnisse der Klassifizierung für die Zeitreihen von NO<sub>2</sub> in µg/m<sup>3</sup>. Die Bezeichnung der Cluster erfolgt im Vorgriff auf Kapitel 2.5.

<sup>4</sup> Siehe z. B. <http://www.insightful.com>

Die Clusteranalyse liefert nur homogene Cluster, keine inhaltliche Interpretation der Cluster. Wie die einzelnen Cluster inhaltlich zu interpretieren sind, ist nicht Gegenstand der Clusteranalyse sondern Gegenstand substanzwissenschaftlicher Erörterungen. Diese Frage wird im Kapitel 2.5 behandelt.

## 2.5 Interpretation der Cluster

### 2.5.1 Allgemeines

Die sich an die Clusteranalyse notwendig anschließende Phase ist die inhaltliche Interpretation der Cluster. In dem hier vorliegenden Fall geht es um Zuordnung von Messstationen zu Clustern. Eine Messstation liegt in einem bestimmten Umfeld, das sich durch soziodemografische und durch geografische Merkmale beschreiben lässt. Ein weiteres externes Unterscheidungskriterium sind die Mittelwerte der Schadstoffe in den unterschiedlichen Clustern.

### 2.5.2 Geografische und soziodemografische Merkmale

Die UBA-Klassifikation lässt sich nach leicht verfügbaren geografischen und soziodemografischen Merkmalen eindeutig sortieren. Die Merkmale sind Höhe der Station über NN in m und Einwohnerdichte der Gemeinde, in der die Station liegt, in Einwohner/km<sup>2</sup>.

Tabelle 3 UBA-Klassifikation mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km<sup>2</sup> der Gemeinden, in denen die Messstationen liegen. Näheres siehe Text.

Verkehr	[n= 68]	150 m	[n= 68]	1'569 Einwohner pro km <sup>2</sup>
Stadt	[n=215]	196 m	[n=215]	965 Einwohner pro km <sup>2</sup>
Land	[n= 69]	323 m	[n= 69]	252 Einwohner pro km <sup>2</sup>
Berg	[n= 8]	1'146 m	[n= 8]	93 Einwohner pro km <sup>2</sup>

Man erkennt, dass die 68 Verkehrsstationen in dicht besiedelten Gemeinden (1'569 Einwohner/km<sup>2</sup>) und relativ tief über NN (150 m) liegen. Die 8 Bergstationen liegen relativ hoch (1'146 m), wie der Name schon sagt. Gleichzeitig liegen sie auf Gemeindegebieten mit relativ geringer Einwohnerdichte (93 Einwohner/km<sup>2</sup>).

Die Stadt- und Land-Messstationen liegen dazwischen, insgesamt ist die Ausprägung dieser zusätzlichen Merkmale als plausibel zu bewerten.

Man kann nun für die gefundenen Cluster (Kapitel 2.4.3) die gleiche Einteilung vornehmen (Tabelle 4 bis Tabelle 10):

Tabelle 4 CO-Clusterzuordnung für 293 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km<sup>2</sup> der Gemeinden, in denen die Messstationen liegen. Näheres siehe Text.

Cluster 3	[n= 16]	34 m	2313 Einwohner pro km <sup>2</sup>
Cluster 2	[n= 11]	85 m	1955 Einwohner pro km <sup>2</sup>
Cluster 1	[n=263]	199 m	1336 Einwohner pro km <sup>2</sup>
Cluster 4	[n= 3]	823 m	226 Einwohner pro km <sup>2</sup>

Tabelle 5 NO<sub>2</sub>-Clusterzuordnung für 423 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km<sup>2</sup> der Gemeinden, in denen die Messstationen liegen. Näheres siehe Text.

Cluster 3	[n= 32]	100 m	1713 Einwohner pro km <sup>2</sup>
Cluster 2	[n=337]	172 m	1282 Einwohner pro km <sup>2</sup>
Cluster 1	[n= 51]	410 m	315 Einwohner pro km <sup>2</sup>
Cluster 4	[n= 3]	724 m	153 Einwohner pro km <sup>2</sup>

Tabelle 6 NO-Clusterzuordnung für 396 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km<sup>2</sup> der Gemeinden, in denen die Messstationen liegen. Näheres siehe Text.

Cluster 3	[n= 15]	95 m	1364 Einwohner pro km <sup>2</sup>
Cluster 2	[n=227]	115 m	1196 Einwohner pro km <sup>2</sup>
Cluster 1	[n=150]	288 m	1056 Einwohner pro km <sup>2</sup>
Cluster 4	[n= 4]	724 m	242 Einwohner pro km <sup>2</sup>

Tabelle 7 NO<sub>x</sub>-Clusterzuordnung für 390 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km<sup>2</sup> der Gemeinden, in denen die Messstationen liegen. Näheres siehe Text.

Cluster 3	[n=172]	119 m	1460 Einwohner pro km <sup>2</sup>
Cluster 2	[n=185]	199 m	942 Einwohner pro km <sup>2</sup>
Cluster 1	[n= 8]	297 m	967 Einwohner pro km <sup>2</sup>
Cluster 4	[n= 25]	536 m	368 Einwohner pro km <sup>2</sup>

Tabelle 8 O<sub>3</sub>-Clusterzuordnung für 362 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km<sup>2</sup> der Gemeinden, in denen die Messstationen liegen. Näheres siehe Text.

Cluster 3	[n= 48]	25 m	789 Einwohner pro km <sup>2</sup>
Cluster 2	[n=239]	219 m	1028 Einwohner pro km <sup>2</sup>
Cluster 1	[n= 62]	275 m	769 Einwohner pro km <sup>2</sup>
Cluster 4	[n= 13]	1036 m	133 Einwohner pro km <sup>2</sup>

Tabelle 9 PM10-Clusterzuordnung für 407 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km<sup>2</sup> der Gemeinden, in denen die Messstationen liegen. Näheres siehe Text.

Cluster 3	[n= 71]	27 m	1712 Einwohner pro km <sup>2</sup>
Cluster 2	[n=287]	221 m	1086 Einwohner pro km <sup>2</sup>
Cluster 1	[n= 43]	249 m	785 Einwohner pro km <sup>2</sup>
Cluster 4	[n= 6]	1020 m	127 Einwohner pro km <sup>2</sup>

Tabelle 10 SO<sub>2</sub>-Clusterzuordnung für 432 Stationen 1997 mit der mittleren Höhe der jeweiligen Messstationen und der mittleren Einwohnerzahl pro km<sup>2</sup> der Gemeinden, in denen die Messstationen liegen. Näheres siehe Text.

Cluster 3	[n= 10]	10 m	1786 Einwohner pro km <sup>2</sup>
Cluster 2	[n=136]	82 m	1263 Einwohner pro km <sup>2</sup>
Cluster 1	[n=250]	267 m	977 Einwohner pro km <sup>2</sup>
Cluster 4	[n= 36]	490 m	1589 Einwohner pro km <sup>2</sup>

Die Tabelle 4 bis Tabelle 10 sind zeilenweise nach aufsteigender Höhe geordnet. Man kann erkennen, dass sich bei allen Schadstoffen die gleiche Reihenfolge ergibt: Cluster 3, 2, 1 und 4.

Die selbe Einteilung nach aufsteigender Höhe zeigt bei der UBA-Klassifikation die Reihenfolge Verkehr, Stadt, Land und Berg. Nach dem Kriterium geografische Merkmale könnte man folgende Einteilung treffen:

Tabelle 11 Inhaltliche Zuordnung der 4 Cluster nach dem Kriterium geografische Merkmale. Näheres siehe Text.

Verkehr	Cluster 3
Stadt	Cluster 2
Land	Cluster 1
Berg	Cluster 4

Auch nach der mittleren Einwohnerdichte der Gemeinde, in der die jeweilige Messstation liegt, ergibt sich bei den anderen Schadstoffen die gleiche Reihenfolge.

Die Anzahlen der Stationen pro Cluster sind allerdings unterschiedlich, das liegt daran, dass die Gesamtanzahl der Stationen pro Schadstoff unterschiedlich ist und daran, dass es Grenzfälle gibt, die je nach Schadstoff zu einer unterschiedlichen Einordnung führen. Das ist verständlich, denn die Schadstoffe stammen aus sehr unterschiedlichen Quellen.

### 2.5.3 Mittelwerte der Schadstoffe

Ein weiteres inhaltliches Unterscheidungskriterium ist der Mittelwert der Schadstoffe je Cluster. Ein Vergleich mit den Mittelwerten der UBA-Klassifikation ist ein weiteres Indiz für die substanzwissenschaftliche Interpretation der Cluster. Die Auswertung in Tabelle 12 erfolgt nur für die 6 Schadstoffe NO, NO<sub>2</sub>, NO<sub>x</sub>, O<sub>3</sub>, PM10 und SO<sub>2</sub>, weil die Dimension der Aufzeichnung von CO über die Jahre wechselt (hierzu IVU Umwelt 2002.1).

Die Sortierung in Tabelle 12 ist für die Clusteranalyse (aufsteigende Clusternummern) und die UBA-Klassifikation (LA=Land, ST=Stadt, VK=Verkehr, BG=Berg) für alle Schadstoffe die selbe:

Die letzten beiden Spalten der Tabelle 12 enthalten den Rang der Mittelwerte für die IVU-Klassifikation und die UBA-Klassifikation. Man erkennt, dass sich für nahezu alle Schadstoffe die gleiche Reihenfolge ergibt. Lediglich bei SO<sub>2</sub> sind die Cluster 2 (ST) und 3 (VK) vertauscht. Aber SO<sub>2</sub> ist ein Schadstoff, der nicht aus verkehrsbezogenen Quellen stammt und dessen Messniveau im Zeitverlauf stark abnehmen.

	IVU	$\mu$ (IVU)	n(IVU)	UBA	$\mu$ (UBA)	n(UBA)	r(IVU)	r(UBA)
NO	1	3.93	327	LA	3.02	331	3	3
NO	2	25.77	2327	ST	15.67	1192	2	2
NO	3	57.48	141	VK	39.22	660	1	1
NO	4	2.42	25	BG	1.78	16	4	4
NO <sub>2</sub>	1	17.73	350	LA	16.37	375	3	3
NO <sub>2</sub>	2	36.97	2580	ST	31.22	1296	2	2
NO <sub>2</sub>	3	50.48	148	VK	45.45	722	1	1
NO <sub>2</sub>	4	11.83	33	BG	7.12	16	4	4
NO <sub>x</sub>	1	25.06	319	LA	21.86	322	3	3
NO <sub>x</sub>	2	83.82	2244	ST	59.95	1172	2	2
NO <sub>x</sub>	3	156.16	141	VK	118.21	633	1	1
NO <sub>x</sub>	4	15.87	25	BG	10.40	16	4	4
O <sub>3</sub>	1	43.52	499	LA	44.51	522	2	2
O <sub>3</sub>	2	30.78	2125	ST	32.94	1513	3	3
O <sub>3</sub>	3	25.04	99	VK	24.13	671	4	4
O <sub>3</sub>	4	58.07	40	BG	68.79	23	1	1
PM10	1	27.39	221	LA	35.33	264	3	3
PM10	2	39.67	2110	ST	35.87	1228	2	2
PM10	3	47.45	109	VK	41.27	521	1	1
PM10	4	22.48	7	BG	19.37	5	4	4
SO <sub>2</sub>	1	13.98	335	LA	19.56	375	3	3
SO <sub>2</sub>	2	24.28	2689	ST	22.39	1356	1	2
SO <sub>2</sub>	3	23.48	112	VK	25.17	728	2	1
SO <sub>2</sub>	4	11.39	32	BG	5.90	16	4	4

Tabelle 12 Aufteilung und Mittelwert für die Clusteranalyse und die UBA-Klassifikation für alle verfügbaren Jahre und 6 Schadstoffe. Spalte IVU enthält die Clustereinteilung,  $\mu$ (IVU) den Mittelwert und n(IVU) die Anzahl der Zeitreihen. Die gleichen Spalten schließen sich für die UBA-Klassifikation an. Die letzten beiden Spalten geben den Rang der Mittelwerte an. Näheres siehe Text.

Es ergibt sich nach dem Kriterium "Mittelwerte der Schadstoffe" die gleiche Interpretation der Cluster wie in Tabelle 11 in Kapitel 2.5.2.

### 3 Zusammenfassung

Die Clusteranalyse der einzelnen Schadstoffe liefert verschiedene Einteilungen der Messstationen in die Gruppen Verkehr, Stadt, Land und Berg. Diese Einteilungen unterscheiden sich aber nicht gravierend.

Die Clustereinteilung von NO<sub>2</sub> wird im Folgenden zugrundegelegt,

- × weil für NO<sub>2</sub> die meisten verwertbaren Zeitreihen vorlagen<sup>5</sup> und
- × weil NO<sub>2</sub> der Schadstoff ist, auf den in der EU-Luftqualitätsrichtlinie verstärkt Bezug genommen wird.

Die Anwendung dieser Clustereinteilung wird im 3. Teilbericht beschrieben (IVU Umwelt 2002.3).

<sup>5</sup> Für SO<sub>2</sub> ist die Zahl der Zeitreihen absolut größer. Aber SO<sub>2</sub> hat die Bedeutung als Leitkomponente der Luftverunreinigungen verloren.





## 4 Literatur

- Hartigan, J. A. (1975)  
Clustering algorithms  
New York: Wiley.
- Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979)  
Multivariate Analysis. Academic Press, London
- Tryon, R. C. (1939)  
Cluster Analysis  
Ann Arbor, MI: Edwards Brothers.
- IVU Umwelt 2002.1  
Automatische Klassifizierung der Luftschadstoff-Immissionsmessungen aus dem LIMBA-Meßnetz, Förderkennzeichen (UFOPLAN) 200 42 265, Datenaufbereitung · 1. Teilbericht
- IVU Umwelt 2002.3  
Automatische Klassifizierung der Luftschadstoff-Immissionsmessungen aus dem LIMBA-Meßnetz, Förderkennzeichen (UFOPLAN) 200 42 265, Anwendung · 3. Teilbericht
- RRL 96/62/EG  
Richtlinie des Rates über die Beurteilung und die Kontrolle der Luftqualität vom 27.9.1996 (Rahmenrichtlinie)
- TRL 1999/30/EG  
Richtlinie des Rates vom 22.4.1999 über Grenzwerte für Schwefeldioxid, Stickstoffdioxid und Stickstoffoxide, Partikel und Blei in der Luft (1. Tochterrichtlinie)
- TRL 2000/69/EG  
Richtlinie des Europäischen Parlaments und des Rates vom 16.11.2000 über Grenzwerte für Benzol und Kohlenmonoxid in der Luft (2. Tochterrichtlinie)