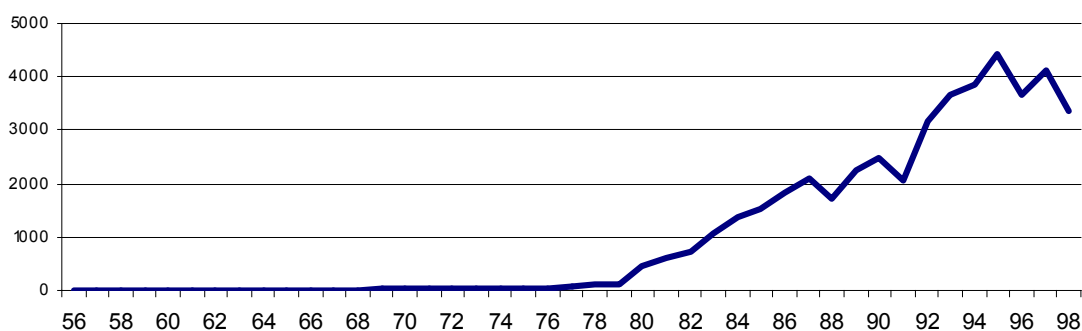


Umweltforschungsplan des Bundesministeriums für Umwelt,
Naturschutz und Reaktorsicherheit

Förderkennzeichen (UFOPLAN) 200 42 265

Automatische Klassifizierung der Luftschadstoff-
Immissionsmessungen aus dem LIMBA-Meßnetz
Datenaufbereitung · 1. Teilbericht



Anzahl der Zeitreihen in der LIMBA-Datenbestand pro Jahr

für
Umweltbundesamt II 6.4
Postfach 33 00 22
D-14191 Berlin
von
erarbeitet 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



Inhaltsverzeichnis

1	ALLGEMEINES.....	5
1.1	Aufgabenstellung des Projektes	5
2	DATENAUFBEREITUNG	6
2.1	Allgemeines	6
2.2	Vorbereitung	6
2.2.1	Allgemeine Änderungen	6
2.2.2	Windgeschwindigkeit-Zeitreihen	7
2.2.3	Ozon-Zeitreihen.....	7
2.2.4	CO-Zeitreihen.....	7
2.3	Ergebnis	8
3	EU-MINDESTDATENBELEGUNG	9
4	BEREINIGUNG VON OFFENSICHTLICHEN FEHLCODIERUNGEN.....	10
4.1	Allgemeines	10
4.2	NO-Daten	10
4.3	NO ₂ -Daten	11
4.4	CO-Daten	13
4.5	SO ₂ -Daten	13
4.6	TSP/PM10-Daten	15
5	ERZEUGUNG VON ZEITREIHEN	17
5.1	NO _x -Zeitreihen.....	17
5.2	Erzeugung von PM10-Zeitreihen	17

6	FEHLWERTERSETZUNG	18
6.1	Allgemeines	18
6.2	Fehlwertersetzungsverfahren	19
6.2.1	Allgemeines	19
6.2.2	Fehlwerte als Elemente des raum-/zeitlichen Prozesses	19
6.2.2.1	Einführung	19
6.2.2.2	Zeitliches Aggregationsniveau	20
6.2.2.3	Räumliche Korrelation.....	20
6.2.2.4	Zeitliche Korrelation	21
6.2.2.5	Synthese	22
6.2.2.6	Anwendbarkeit	22
6.3	Anwendung des Fehlwertersetzungsverfahrens.....	23
7	LITERATUR	24



Abbildungsverzeichnis

Abbildung 1	Unterschiedliche Kodierung der CO-Zeitreihen.....	7
Abbildung 2	Anzahl der Zeitreihen pro Jahr in dem LIMBA-Dateisystem	8
Abbildung 3	Anzahl der vollständigen Beobachtungen gegen Anzahl der Zeitreihen pro Jahr am Beispiel der NO ₂ -Zeitreihen.....	18
Abbildung 4	Fehlwertersetzung mit den Stundenmitteln Ozon und den auf Stundenmittelwerte aggregierten meteorologischen Daten aus den Ländermeßnetzen	21



Tabellenverzeichnis

Tabelle 1	Anzahl und Gründe für das Ausschließen von Zeitreihen aus der weiteren Untersuchung	9
Tabelle 2	Auffällige Codierungen in 24 sächsischen Stationen, näheres siehe Text.....	10



1 Allgemeines

1.1 Aufgabenstellung des Projektes

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 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-Stationsklassifikation gibt eine Einordnung für 377 Stationen: 228 Stadt, 72 Land, 69 Verkehr und 8 Berg.

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

In diesem Bericht wird die Datenaufbereitung für die Klassifikation beschrieben.

2 Datenaufbereitung

2.1 Allgemeines

Im März dieses Jahres erhielt IVU Umwelt vom UBA II 4.6 eine CD mit dem LIMBA-Dateisystem. Dieses Kapitel enthält eine Beschreibung der Arbeitsschritte, die zur Vorbereitung der multivariaten Analyse zur Klassifizierung der Stationen erforderlich war.

Diese vorbereitenden Arbeitsschritte waren notwendig, um aus dem Datenbestand diejenigen Dateien herauszufiltern,

- × die frei von offensichtlichen Codierungsfehlern sind,
- × die formal die Ansprüche an eine automatisierte Datenverarbeitung erfüllen, und
- × bei denen die gleiche Systematik bei der Dateinamensgebung eingehalten wurde.

2.2 Vorbereitung

2.2.1 Allgemeine Änderungen

Das LIMBA-Dateisystem enthielt redundante Informationen, die beseitigt wurden. Es handelte sich um Dateien mit den Endungen *.prn, *.doc, usw., die gelöscht wurden.

Die gleiche Dateinamenssystematik für alle Zeitreihen ist erforderlich, da die Dateien mit Algorithmen bearbeitet werden, die Informationen darüber brauchen, was welche Datei enthält und von wo und wann sie stammt. Dateinamen sollten folgende Systematik haben:

LLNNNSJJ.KKK mit	"LL"	Kennung für Datenquelle
	"NNN"	Messstellennummer
	"S"	Kennung für Stundenwerte
	"JJ"	zwei Endziffern des Jahres
	"KKK"	Kennung für Komponente

Die Dateinamen der kommunalen Messstationen, bei denen festgestellt werden konnte, von wann sie stammen, wurden auf 8 Zeichen gebracht. Dateien, die zeitlich nicht zugeordnet werden konnten, wurden gelöscht. Bei kommunalen Dateien, die zeitlich zugeordnet werden konnten, wurden die Dateinamen aus eindeutigen Nummern konstruiert und das Jahr angefügt.

Dateien mit Dateiname der Form X*.* und xx*.* wurden gelöscht, weil nicht festzustellen war, woher sie sind.

Circa 1'700 Dateien waren inhaltlich Stundendateien, aber mit der falschen Kennung, sie wurden von ?????H??.* in ?????S??.* umbenannt. 707 Dateien davon existierten schon als ?????S??.* und die Inhalte waren identisch, so dass diese 707 Dateien (????S??.*) gelöscht wurden.

Zwei gleichlautende Stationsbezeichnungen wurden entdeckt:

- × DESL006 Lauterbach Köhlerstr.
- × DESL006 Lauterbach Alte Schule, umbenannt in DESL018

2.2.2 Windgeschwindigkeit-Zeitreihen

Die Datei wd501s87.wig wurde gelöscht, weil die Länge Null war.

2.2.3 Ozon-Zeitreihen

BW005OUT.O3, HE019x1.O3, nw076out.O3, ub005out.O3 wurden gelöscht, da sie ein falsches Format hatten, die Werte standen untereinander.

Datei "ub029s90.O3" enthält Dezimalkommas statt Dezimalpunkte, was korrigiert wurde. Manche der O3-Dateien enthalten verschiedene Längen der Sätze. Da solche Datensätze kaum einzulesen sind, wurden diese Zeitreihen verworfen.

2.2.4 CO-Zeitreihen

Von den 3'860 CO-Zeitreihen enthalten 408 Halbstundenwerte und werden nicht weiter betrachtet, die Datei hh020s92.CO wurde gelöscht, weil die Datei 8764 · Null enthält.

Die 2'742 CO-Daten sind offensichtlich in unterschiedlichen Dimensionen abgelegt, der Bereich des Mittelwertes der Zeitreihen überdeckt 5 Größenordnungen: 0.01 (MV012S94.CO) bis 1'280 (HH009S88.CO) in der Zeit zwischen 1980 und 1998.

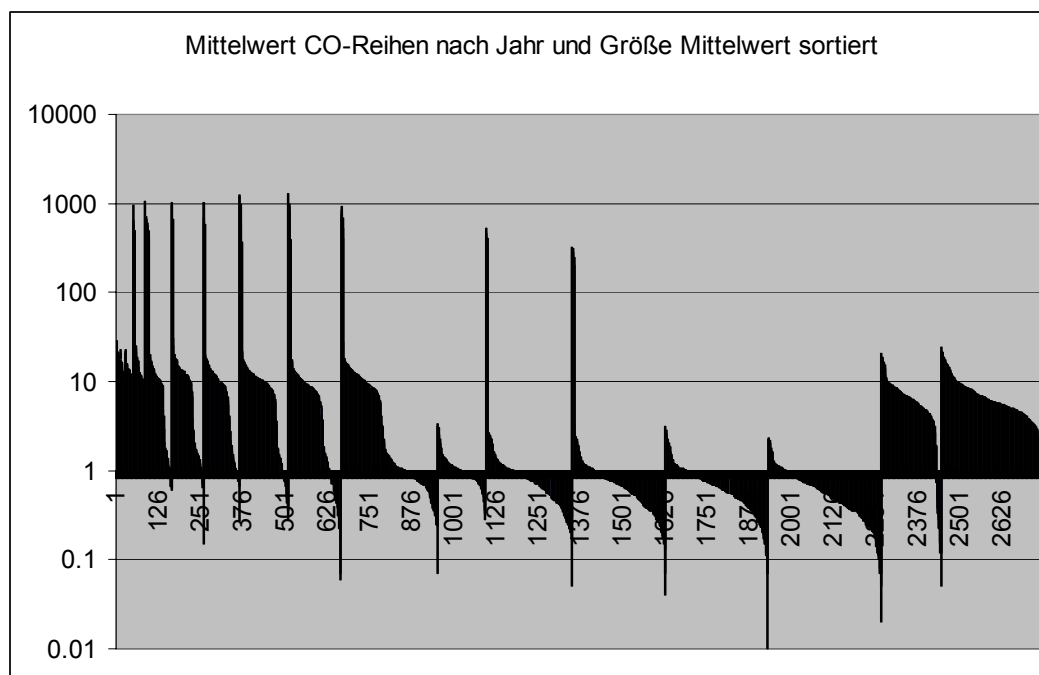


Abbildung 1 Unterschiedliche Kodierung der CO-Zeitreihen. Aufgetragen ist hier der CO-Jahresmittelwert der unterschiedlichen Stationen. Sortiert ist von links nach rechts nach dem Jahr der Beobachtung und innerhalb eines Jahres nach der Größe der Mittelwerte. Man kann erkennen, dass Größenordnungen zwischen den einzelnen Jahresmittelwerten liegen. Das ist sicherlich auf unterschiedliche Dimensionierung zurückzuführen.

2.3 Ergebnis

Es bleiben circa 45'000 Zeitreihen aus den Jahren 1956 bis 1998.

Die Anzahl der Zeitreihen pro Jahr ist aus der Abbildung 2 zu erkennen.

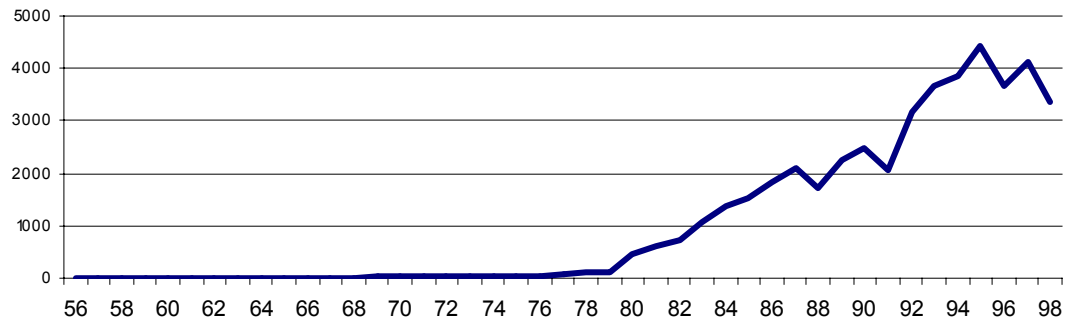


Abbildung 2 Anzahl der Zeitreihen pro Jahr in dem LIMBA-Dateisystem

3 EU-Mindestdatenbelegung

Die EU-Rahmenrichtlinie 96/62 (RRL96/62/EG) bzw. die dazu gehörenden Tochterrichtlinien (TRL 1999/30/EG und TRL 2000/69/EG) definieren Vorschriften für die Mindestdatenbelegung einer Zeitreihe, die allerdings einigen Definitionsspielraum lässt. Hier wird die Vorschrift für die Mindestdatenbelegung folgendermaßen verstanden:

Mindestbelegungshäufigkeit (laut Text der RR96/62)	90.00%
minus Wartung (2 Tage)	00.55%
minus Kalibrierung (1h pro Tag)	04.14%
Mindestgesamtbelegung	85.31%

Diese Vorschrift wird abgeprüft und es werden lediglich diejenigen Zeitreihen weiterverwendet, bei denen diese Vorschrift eingehalten ist.

Am Ende dieses Arbeitsschrittes ergeben sich 33'772 Zeitreihen, die in Ordnung zu sein scheinen. Die Anzahl und Gründe der Nichtberücksichtigung der Zeitreihen ergeben sich aus Tabelle 1.

Tabelle 1 Anzahl und Gründe für das Ausschließen von Zeitreihen aus der weiteren Untersuchung

Bezeichnung	Anzahl	Prozent
Zeitreihen in Ordnung	33'772	74.94%
unter dem EU96/62 Belegungsgrad (minimum data capture)	10'092	22.39%
Nur Fehlwerte	1'005	2.23%
Anzahl der Felder (durch "white space" getrennt) stimmt nicht	103	0.23%
keine Variation	92	0.20%
Gesamt	45'064	100.00%

4 Bereinigung von offensichtlichen Fehlcodierungen

4.1 Allgemeines

Dieser Zeitschritt ist sehr zeitaufwendig aber notwendig, wenn man sinnvolle Ergebnisse bei der sich anschließenden multivariaten Analyse erhalten will. Denn die multivariate Analyse (IVU Umwelt 2002.2) zur Klassifizierung der Stationen stellt auf die Ähnlichkeit des zeitlichen Verlaufs der verschiedenen Zeitreihen ab, hierbei können Kodierungsfehler zu Fehlzusammenhängungen führen.

Wegen unterschiedlicher Dimensionierung bei bestimmten Stoffen in den unterschiedlichen Jahren und wegen des enormen Zeitaufwandes bei der Beseitigung von Fehlcodierungen konnte dieser und alle nachfolgenden Schritte nur für ein Jahr durchgeführt werden. Es wurde 1997 ausgewählt, weil es sich um ein relativ aktuelles Jahr handelt und alle Stoffe (CO, NO, NO₂, O₃, PM10, SO₂) vorliegen.

4.2 NO-Daten

24 sächsische NO-Messreihen von 1997 sind vermutlich fehlerhaft kodiert.

In der ersten Stunde eines Tages war an verschiedenen Stellen "999" in einer Reihe von "-999" oder "5??" in einer Reihe von einstelligen Konzentrationen eingetragen. Anstelle dieser vermutlichen Fehleinträge werden Fehlwerte eingetragen. Tabelle 2 gibt den Dateinamen und die Anzahl der Fehleinträge wieder.

Tabelle 2 Auffällige Codierungen in 24 sächsischen Stationen, näheres siehe Text

Lfd.Nr.	Dateiname	"999"	"5??"
1	SN001s97.NO	6	9
2	SN002s97.NO	4	8
3	SN004s97.NO	3	6
4	SN005s97.NO	14	7
5	SN006s97.NO	12	7
6	SN011s97.NO	27	5
7	SN012s97.NO	18	10
8	SN014s97.NO	23	9
9	SN017s97.NO	9	7
10	SN019s97.NO	13	6
11	SN020s97.NO	55	3
12	SN024s97.NO	17	9
13	SN025s97.NO	31	4
14	SN028s97.NO	25	6
15	SN034s97.NO	3	9
16	SN038s97.NO	12	8
17	SN045s97.NO	14	7
18	SN047s97.NO	10	4
19	SN048s97.NO	11	8
20	SN050s97.NO	20	7
21	SN051s97.NO	38	5
22	SN057s97.NO	25	6
23	SN059s97.NO	5	8
24	SN061s97.NO	22	5

Zusammen handelt es sich um 580 Fehleinträge. Sie wurden korrigiert, weil sonst die sächsischen Stationen in dem Jahr 1997 massiv unterrepräsentiert wären.

4.3 NO₂-Daten

Beim der visuellen Inspektion der NO₂-Daten gibt es Ausreißer in der Station Wittenberg-Zimmermann (ST066S97.NO2). Diese Station weist offensichtliche Editierfehler (farbige Einträge) auf, die durch Fehlwerte ersetzt werden.

:															
267	11	4	11	7	18	41	38	38	27	30	16	4	6	2	...
268	-999	-999	1	3	24	30	30	17	-999	-999	-999	-999	-999	-999	...
269	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	1	1	2	...
270	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
271	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
272	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
273	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
274	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
275	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
276	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
277	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
278	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
279	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
280	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	23	24	24	...
281	32	24	26	36	40	52	58	52	42	31	48	60	38	36	...
:															
333	30	28	28	30	34	34	34	36	38	30	32	34	38	37	...
334	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
335	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
336	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	16	28	...
337	28	28	30	32	33	38	33	30	32	31	46	31	26	28	...

BY092S97.NO2 ist in der 310. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

309	14	10	9	11	8	15	10	10	18	19	36	60	50	16	...
310	0	0	0	0	6980	0	0	0	0	0	0	0	0	0	...
311	59	62	62	62	60	60	56	56	50	42	34	28	32	34	...

ST002S97.NO2 ist in der 153. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

151	1	1	1	2	3	7	8	6	3	2	1	1	1	1	...
152	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
153	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
154	-999	-999	-999	-999	-999	-999	-999	-999	2	2	2	1	1	1	...
155	4	5	7	8	24	27	12	10	8	6	3	2	2	2	...

ST025S97.NO2 ist in der 363. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

362	7	6	6	6	8	12	12	12	11	16	16	-999	-999	-999	...
363	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
364	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...

ST028S97.NO2 ist in der 62. und 63. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

061	12	10	9	10	9	10	-999	-999	-999	-999	-999	-999	-999	-999	...
062	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
063	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
064	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	19	14	18	...

ST030S97.NO2 ist in der 289. Zeile offensichtlich fehlerhaft kodiert, durch Fehlwert ersetzt:

288	2	3	4	4	1	1	5	10	7	12	10	6	6	5	...
289	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
290	-999	-999	-999	-999	-999	-999	-999	-999	31	18	9	8	12	8	...
291	22	22	20	22	24	24	24	30	31	32	36	37	39	37	...

ST033S97.NO2 ist in der 354. und 355. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

352	16	16	16	20	24	24	26	28	30	28	30	30	28	28	...
353	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
354	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
355	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
356	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
357	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
358	16	16	16	17	18	18	18	20	24	29	28	28	28	30	...

ST078S97.NO2 ist in der 166. und 167. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

163	24	30	22	32	26	34	38	37	44	37	26	17	-999	-999	...
164	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
165	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	18	6	-999	-999	...
166	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
167	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
168	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
169	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
170	-999	-999	-999	-999	-999	-999	-999	-999	-999	22	6	8	8	12	...
171	15	16	14	17	24	35	34	22	18	17	21	20	22	21	...

ST081S97.NO2 ist in der 27., 96, 97., 361., 362. und 365. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

026	30	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
027	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
028	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	87	94	77	82	...
029	41	32	36	35	44	74	82	76	90	82	88	71	77	62	...
:															
095	16	12	28	20	66	72	84	94	66	56	48	50	48	-999	...
096	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
097	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
098	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	69	53	78	...
099	34	40	38	36	60	90	86	74	70	75	74	56	61	64	...
:															
360	18	10	12	10	8	12	11	10	11	15	20	26	23	20	...
361	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
362	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
363	999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
364	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	52	70	...
365	20	19	16	22	18	30	31	34	34	46	43	36	33	44	...

Diese Analyse der NO₂-Zeitreihen von 1997 (das ist circa eine 1%-Stichprobe des gesamten Datenbestandes) lässt vermuten, dass ähnliche Fehler auch in den sonstigen Reihen enthalten sein werden.

4.4 CO-Daten

TH031S97.CO ist in der.24., 25. und 26. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

4	8	7	7	8	-999	-999	-999	-999	-999	-999	-999	-999	...
990	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
990	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
990	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
1	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
3	6	6	5	5	6	12	29	28	12	15	11	10	...

4.5 SO₂-Daten

ST066S97.SO2 ist in der 269. bis 278. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

10	12	12	13	14	15	16	15	-999	-999	-999	-999	-999	-999	...		
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	2	2	2	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	4	3	2	2	...
1	1	1	1	1	1	2	1	1	-999	4	5	2	2	...		

ST030S97.SO2 ist in der 288. Zeile offensichtlich fehlerhaft kodiertv durch Fehlwert ersetzt:

1	1	1	1	1	1	1	1	1	4	4	4	4	29	4	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	3	2	1	1	1	1	...
2	2	2	2	2	5	8	8	10	19	23	12	12	11	12	...

ST028S97.SO2 ist in der 61. und 62. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

5	6	6	6	6	6	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	9	8	8	9	...
9	8	8	8	8	9	10	14	12	11	10	13	-999	14	13	...	

ST039S97.SO2 ist in der 138. und 143. und 145. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

40	23	15	14	19	20	18	26	24	24	24	24	24	24	24	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	7	6	7	8	8	8	8	10	8	...
6	6	7	7	8	8	10	10	11	11	11	10	10	9	8	...	
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	2	3	-999	-999	...
2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	...

ST039S97.SO2 ist in der 159. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

17	20	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
1	2	5	4	5	4	4	4	2	2	3	5	8	15	3	...	

ST066S97.SO2 ist in der 233. und 234. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

58	50	50	48	49	44	39	33	34	32	34	39	44	36	30	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	10	10	10	...	
6	6	6	4	4	2	2	4	12	12	25	12	12	14	10	...	

ST078S97.SO2 ist in der 165. und 166. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

10	2	1	-999	-999	-999	-999	-999	19	73	26	2	-999	-999	-999	...	
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	8	5	6	4	4	1	...
6	6	6	6	5	8	9	10	12	11	8	4	28	65	38	...	

ST081S97.SO2 ist in der 26. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

20	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	69	70	41	75	54	...

ST081S97.SO2 ist in der 95. und 96. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

1	2	2	2	8	7	8	12	9	7	6	8	7	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	1	1	1	...
2	2	1	2	14	39	33	25	20	21	76	75	34	24	39	...

ST081S97.SO2 ist in der 360. und 362. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

1	-999	-999	-999	-999	-999	1	1	4	24	7	7	19	6	8	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
5	4	3	2	2	4	6	8	8	25	22	20	28	24	36	...

4.6 TSP¹/PM₁₀²-Daten

ST002S97.TSP ist in der 152. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999

ST066S97.TSP ist in der 269. bis 278. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

7	14	21	22	24	28	35	39	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	68	56	34	34
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	41	44	58	58	53	54
64	61	57	52	51	54	58	69	86	82	66	62	66	67	75	...

ST066S97.TSP ist in der 333. und 334. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

40	42	42	44	44	46	50	56	63	69	70	67	66	66	68	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	6	6	8	...

¹ total suspended particles
² particulate matter < 10µm

ST078S97.TSP ist in der 165. und 166. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...

ST081S97.TSP ist in der 26. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

40	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	54	62	82	66	...

ST081S97.TSP ist in der 95. und 96. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

2	2	10	4	12	24	13	19	24	50	64	72	58	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	14	15	12	...

ST081S97.TSP ist in der 360. bis 362. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

8	10	6	4	2	3	3	4	5	16	34	19	3	6	18	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...	
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	6	16	29	...
12	8	4	6	10	18	17	21	24	34	38	38	40	44	58	...	

ST028S97.TSP ist in der 61. und 62. Zeile offensichtlich fehlerhaft kodiert (farbige Einträge) und wurde durch Fehlwert ersetzt:

8	10	14	20	20	20	18	14	13	19	26	27	26	30	30	...	
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	15	17	17	17	...
10	9	10	14	14	12	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	...
-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	18	21	20	22	...
29	27	30	32	30	31	32	30	28	26	24	24	-999	30	31	...	

5 Erzeugung von Zeitreihen

5.1 NO_x-Zeitreihen

Es gibt in dem LIMBA-Dateisystem 4'479 NO-Dateien und 5'235 NO₂-Dateien. Dateien von dem selben Ort und dem selben Jahr wurden zu NO_x, ausgewiesen als NO₂, mittels folgender Vorschrift zusammengefasst: $[\text{NO}_x] = [\text{NO}_2] + \frac{46}{30} \cdot [\text{NO}]$.

Für sieben der 4'479 NO-Zeitreihen gibt es keine entsprechende NO₂-Zeitreihe, das heißt es wurden 4'472 NO_x-Zeitreihen erzeugt.

5.2 Erzeugung von PM10-Zeitreihen

PM10-Zeitreihen gab es nur in den Jahren 1997 und 1998 in Baden-Württemberg. Alle anderen Bundesländer haben, wenigstens bis 1998, nur Gesamtstaub (TSP) gemeldet. Um zu einer vergleichbaren Datenbasis zu kommen, wurden die TSP-Zeitreihen in die entsprechenden PM10-Zeitreihen umgerechnet. Der Umrechnungsfaktor von 0.83 wurde einer Veröffentlichung von Bruckmann (Bruckmann, P., 1998) entnommen.

6 Fehlwertersetzung

6.1 Allgemeines

Ziel der Untersuchung ist, aus den Besonderheiten der zeitlichen Verläufe Gemeinsamkeiten der verschiedenen Zeitreihen herauszufinden. Fehlwerte stören den zeitlichen Verlauf, daher ist es wünschenswert lückenlose Zeitreihen vorliegen zu haben.

Die Notwendigkeit eine Fehlwertersetzung vorzunehmen, ergibt auch aus einer praktischen Erwägung. In Jahren mit einer großen Anzahl von Zeitreihen, sinkt die Anzahl der vollständigen Beobachtungen, d.h. Stunden, in denen jede der Zeitreihen besetzt ist. Besonders in den aktuellen Jahren wie 1997 sinkt die Anzahl der vollständigen Beobachtungen auf Null. Dieser Zusammenhang soll am Beispiel von NO_2 in Abbildung 3 verdeutlicht werden.

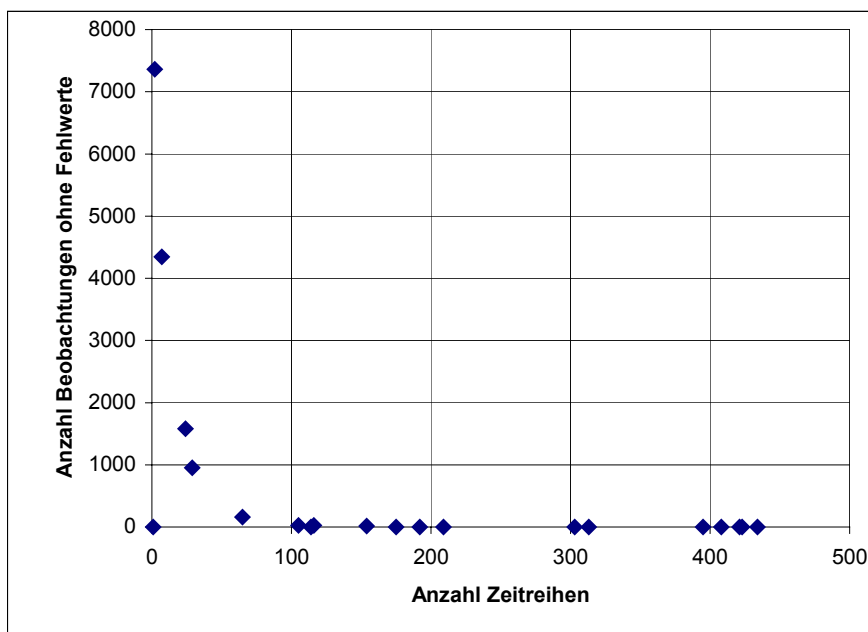


Abbildung 3 Anzahl der vollständigen Beobachtungen gegen Anzahl der Zeitreihen pro Jahr am Beispiel der NO_2 -Zeitreihen.

6.2 Fehlwertersetzungsverfahren

6.2.1 Allgemeines

Fehl- oder Ausfallwerte treten in den Original-Meßreihen von Immissionen unvermeidlich auf. Die Gründe dafür sind:

- × Messgeräte müssen regelmäßig kalibriert werden,
- × Messgeräte fallen aus wegen Wartung und
- × Messgeräte fallen aus wegen Störung der Funktionsweise.

Ausfälle, die auf das Kalibrieren der Messgeräte zurückzuführen sind, ereignen sich regelmäßig. Der Ausfall ist in der Regel kurz und zudem gleichmäßig über den Tag verteilt. Eine Auswirkung auf die statistischen Kennwerte der Reihe ist unwahrscheinlich.

Ereignet sich der Ausfall aufgrund von Wartungsarbeiten, so ist damit zu rechnen, dass diese Arbeiten nicht in der Periode geschehen, in denen in der Regel hohe Werte gemessen werden. Der Ausfall von niedrigen Werten kann zur Überschätzung der Lageparameter führen, da die Immissionen in der Regel einen Tagesgang aufweisen.

Fällt ein Messgerät wegen einer Störung seiner Funktionsweise aus, so kann zwar der Zeitpunkt des Ausfalls gleichverteilt sein, aber es ist wahrscheinlich, dass der Ausfall während der normalen Arbeitszeit, das heißt tagsüber, beseitigt wird. Hieraus kann sich eine Auswirkung auf die statistischen Kennwerte der Reihe ergeben, da die Immissionen in der Regel einen Tagesgang aufweisen.

Fehlwerte in den Zeitreihen machen den Einsatz von bestimmten statistischen Verfahren unmöglich. Die Theorie der Zeitreihenanalyse fordert z. B. lückenlos aufeinanderfolgende äquidistante Messwerte.

Es ist daher wünschenswert, ein statistisch in bestimmtem Sinne optimales und automatisches Ersetzungsverfahren für die Ausfallwerte zu entwickeln.

6.2.2 Fehlwerte als Elemente des raum-/zeitlichen Prozesses

6.2.2.1 Einführung

Die Hessische Landesanstalt für Umwelt (HLfU) begann im Jahre 1988, ein solches Fehlwert-Ersetzungsverfahren entwickeln zu lassen. Im Auftrag der HLfU wurde in den darauffolgenden Jahren von der Firma IVU GmbH ein Verfahren entwickelt, das die räumliche und zeitliche Inter-Korrelation im Messnetz ausnutzt, um die Fehlwerte zu ersetzen.

Die Messwerte eines Immissionsmessnetzes sind als Produkt eines raum-zeitlichen stochastischen Prozesses zu interpretieren. Gelingt es diesen Prozess befriedigend zu modellieren, hat man ein Instrument zur statistisch optimalen Fehlwertersetzung.

Die räumliche und die zeitliche Abhängigkeit eines Immissionswertes sind die beiden prinzipiellen Arten der Abhängigkeit.

Das bedeutet, dass z. B. die SO_2 -Immission von Grebenau in Nordhessen am 20.1.1985 um 8:30 Uhr nicht unabhängig ist von der SO_2 -Immission von Grebenau am gleichen Tag um 9:00 Uhr. Genauso wenig ist die SO_2 -Immission in Grebenau unabhängig von der gleichzeitig gemessenen SO_2 -Immission in z. B. Kassel-Mitte. Die Immissionssituation am 20. Januar 1985 war geprägt durch Schadstoff-Transporte aus dem Raum Halle/Leipzig. Aber auch beim überwiegenden Einfluss von lokalen Emittenten kann man die räumliche und zeitliche Persistenz von Immissionssituationen - wenn auch nicht so ausgeprägt - beobachten.

Die Modellierung dieses raum/zeitlichen Prozesses ist theoretisch in einem geschlossenen Modell möglich. Doch würde das den Einsatz von multivariater Zeitreihenanalyse erfordern. Das Verfahren der multivariaten Zeitreihenanalyse ist zu weiten Teilen noch in der Entwicklung, besonders die Software existiert dafür noch nicht in dem Umfang wie es erforderlich wäre.

Darum wurde der Weg über die Modellierung zweier getrennter Prozesse gewählt. Die Ergebnisse dieser zwei Prozesse werden über eine Gewichtungsfunktion zusammgeführt, die die Präzision der beiden Resultate berücksichtigt.

6.2.2.2 Zeitliches Aggregationsniveau

Fehlwerte sollten auf dem kleinsten zeitlichen Aggregationsniveau, das zur Verfügung steht, ersetzt werden. In diesem Anwendungsfall ist das der Stundenmittelwert (dazu auch Abbildung 4). Ausfallwerte auch in den zeitlichen Aggregaten wie Tages- oder Monatsmittelwerten zu ersetzen, führt unvermeidlich zu Inkonsistenzen, denn eine Aggregation aus den fehlwert-bereinigten Stundenwerten führt mit großer Wahrscheinlichkeit zu anderen Ergebnissen für die in den Tages- oder Monatsmittelwert-Messreihen ersetzten Fehlwerte.

6.2.2.3 Räumliche Korrelation

Die wichtigste Information, die die räumliche Ersetzung von Fehlwerten betrifft, sind die aktuellen Korrelationsmatrizen von Luftschadstoffmessungen der Stationen, die in die Fehlwertersetzung miteinbezogen werden.

Welche Messreihen mit der zu ersetzenden Luftschadstoff-Messreihe korreliert werden, ist sekundär, Es ist nur entscheidend, dass sie mit der in Frage stehenden Reihe korreliert sind. Die aktuelle Korrelationsmatrix reicht circa 500 Beobachtungen in die Vergangenheit.

Zunächst wird eine einfache lineare Regression von der zu ersetzenden Reihe auf die Reihe mit dem höchsten Eintrag in den Korrelationsmatrizen berechnet. Wenn der Messwert der Station mit dem höchsten Korrelationskoeffizienten vorhanden ist, wird der Schätzwert der Regression als räumliche Ersetzung genommen. Sollte die Reihe der Station mit dem höchsten Eintrag in den Korrelationsmatrizen auch ausgefallen sein, so wird mit dem zweithöchsten Eintrag in demselben Sinne verfahren, usw. Damit liegt ein räumlicher Schätzwert y_R vor.

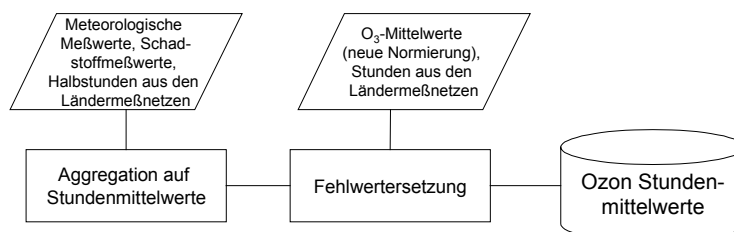


Abbildung 4 Fehlwertersetzung mit den Stundenmitteln Ozon und den auf Stundenmittelwerte aggregierten meteorologischen Daten aus den Ländermessnetzen

6.2.2.4 Zeitliche Korrelation

Einen Datenbestand, dessen Elemente in gleichen zeitlichen Abständen vorliegen, bezeichnet man als "Zeitreihe". So ist z. B. eine Reihe von stündlichen Immissionswerten eine solche Zeitreihe.

Diese Zeitreihen können bestimmte Strukturen haben. So ist es bei Stundenwerten sehr unwahrscheinlich, daß auf einen extrem hohen Wert in der nächsten Stunde ein extrem niedriger Wert folgt, dies ist durch die Trägheit des atmosphärische Systems bedingt. Diesen Zusammenhang zwischen aufeinanderfolgenden Werten kann man parametrisieren und quantifizieren. Das dafür geeignete Instrument der Statistik ist das sogenannte "univariate Box-Jenkins Modell". Diese Modellklasse heißt "univariat", weil zur Erklärung eines Wertes zu einem bestimmten Zeitpunkt nur vergangene Werte aus derselben Zeitreihe eingehen.

Die Box-Jenkins-Methode wurde in ihrer heute gebräuchlichen Form in dem Buch "Time Series Analysis, Forecast and Control" von Box und Jenkins (Box, G.E.P., Jenkins, G.M., 1970) dargestellt und erhielt daher ihren Namen.

Die Immissionszeitreihen haben eine ähnliche Struktur. Sie sind kurzfristig hoch korreliert und weisen eine hohe Korrelation mit den Werten des vergangenen Tages auf. Es ist daher möglich ein Standardmodell zu schätzen, das in nahezu allen Fällen eine befriedigende Erklärung der Reihe bietet. Das Standardmodell für Halbstundenwerte hat folgende Struktur:

$$(1 + a \cdot B + b \cdot B^2 + c \cdot B^3) \cdot (1 + d \cdot B^{\text{OPD}}) \cdot Y_t = \varepsilon_t$$

Mit	Y_t	Beobachtung zum Zeitpunkt t
	ε_t	Störprozess zum Zeitpunkt t
	a, b, c, d	Parameter des Modells
	B^n	backshift Operator (Box, G.E.P., Jenkins, G.M., 1970)
	OPD	Anzahl der Beobachtungen am Tag

Das zeitreihenanalytische Modell wird jedes mal, wenn ein Fehlwert auftritt, neu geschätzt. Hierbei wird auf Daten zurückgegriffen, die ungefähr 5 Tage in die Vergangenheit reichen. Bei der Schätzung werden sukzessive diejenigen Parameter entfernt, die nicht signifikant sind.

Mit dem geschätzten Modell wird dann der Fehlwert y_z prognostiziert. Die Beschreibung der Präzision der Prognose ist über die Varianz der Prognose möglich.

6.2.2.5 Synthese

Es liegen nun zwei Ersetzungswerte (y_R und y_Z) für einen Fehlwert vor, die in der Regel voneinander abweichen. Es liegen ebenfalls zwei Präzisionsmaße vor (r^2_R und r^2_Z). Das erlaubt, den Ersetzungswert über ein gewichtetes Mittel mit den jeweiligen Präzisionen zu ermitteln.

$$y = \frac{y_R \cdot r^2_R + y_Z \cdot r^2_Z}{r^2_R + r^2_Z}$$

Dieses gewichtete Mittel hat die wünschenswerten Eigenschaften, die man von ihm erwartet:

- ✗ Das räumliche Modell reagiert sehr schnell auf plötzliche Veränderungen in der Immissionssituation. Dafür wird es bei einer kurzen Lücke weniger gute Ersetzungsergebnisse liefern als das zeitliche Modell. Bei längeren Lücken ist es dagegen dem zeitlichen Modell überlegen.
- ✗ Das zeitliche Modell liefert gerade bei kurzen Lücken in den meisten Fällen bessere Fehlwertersetzungen als das räumliche Modell. Dagegen reagiert es auf plötzlichen Veränderungen in der Immissionssituation träge.

Das zeitliche Modell ist wesentlich rechenaufwendiger als das räumliche Modell. Bei größeren Lücken kommt das zeitliche Modell nicht zum Einsatz, da die maximale Prognosereichweite bei wenigen Tagen liegt.

6.2.2.6 Anwendbarkeit

Das Verfahren ist für alle Zeitreihen anwendbar, die mit anderen Zeitreihen korreliert sind und für die der Korrelationskoeffizient definiert ist.

Das gilt z. B. nicht für die Messung der Windrichtung, für die als zyklische Variable der Pearson'sche Korrelationskoeffizient nicht definiert ist. Weiterhin gilt es nicht für die Variable "Niederschlag", da das Auftreten von Niederschlag mit anderen Variablen kaum korreliert ist.

6.3 Anwendung des Fehlwertersetzungsverfahrens

Das oben beschriebene Fehlwertersetzungsverfahren wurde auf die

- 299 CO-Zeitreihen,
- 396 NO-Zeitreihen,
- 423 NO₂-Zeitreihen,
- 390 NO_x-Zeitreihen,
- 362 O₃-Zeitreihen,
- 408 PM10-Zeitreihen und
- 432 SO₂-Zeitreihen

angewandt.

Mit den wie beschrieben aufbereiteten und fehlerwertersetzten Zeitreihen wird eine multivariate Analyse zur Klassifizierung der Stationen für das Jahr 1997 durchgeführt.

Diese multivariate Analyse ist im 2. Teilbericht beschrieben (IVU Umwelt 2002.2).



7 Literatur

Box, G.E.P., Jenkins, G.M., 1970

Time Series Analysis, Forecasting and Control, ISBN 0-8162-1104-3

Bruckmann, P., 1998

Die neuen Richtlinie der Europäischen Union zur Luftqualität, Immissionsschutz 3, 1998, S.107

IVU Umwelt 2002.2

Automatische Klassifizierung der Luftschadstoff-Immissionsmessungen aus dem LIMBA-Meßnetz, Förderkennzeichen (UFOPLAN) 200 42 265, Klassifizierung · 2. 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)