

Umweltforschungsplan  
des Bundesministeriums für Umwelt,  
Naturschutz und Reaktorsicherheit

Förderkennzeichen 200 42 264

"Verbesserung der bundesweiten Immissionsdaten durch Entwicklung und Anwendung  
einer operationellen Methode zur Validierung und Korrektur kontinuierlich ermittelter  
Immissionswerte nach einheitlichen objektiven Kriterien"

Schlussbericht

von

Dr. Wolfgang Enke

I.D.E.A:  
Immissionsdaten Erfassung und Analyse  
Karin Seewöster  
Wandelsweg 89  
4594 Gelsenkirchen

IM AUFTRAG  
DES UMWELTBUNDESAMTES

November 2002

# Berichts-Kennblatt

1. Berichtsnummer UBA-FB	2.	3.
4. Titel des Berichts Verbesserung des bundesweiten Immissionsdaten durch Entwicklung und Anwendung einer operationellen Methode zur Validierung und Korrektur kontinuierlich ermittelter Immissionswerte nach einheitlichen objektiven Kriterien		
5. Autor(en), Name(n), Vorname(n) Dr. Enke, Wolfgang Waltraudstr. 20 D – 14532 Güterfelde		8. Abschlussdatum Oktober 2002
		9. Veröffentlichungsdatum November 2002
6. Durchführende Institutionen (Name, Anschrift) I.D.E.A., Karin Seewöster Wandelsweg 89 D – 45894 Gelsenkirchen-Buer		10. UFOPLAN-Nr. 200 42 264
		11. Seitenzahl 67
7. Fördernde Institution (Name, Anschrift) Umweltbundesamt Postfach 33 00 22 D – 14191 Berlin		12. Literaturangaben 10
		13. Tabellen und Diagramme 14
		14. Abbildungen 40
15. Zusätzliche Angaben		
16. Zusammenfassung Ziel des Vorhabens war die Entwicklung und Anwendung einer Methode zur Validierung von Ozondaten nach einheitlichen Prüfkriterien. Das vorliegende Validierungsmodul ist in der Lage, zufällige Fehler (Ausreißer) wie auch systematische Fehler (Abdriften von Messgeräten usw.) aufzuspüren, zu erkennen und je nach vorgegebener Zielrichtung diese als fehlerhaft erkannte Werte zu markieren oder zu ersetzen. Das Gesamtsystem ist in drei Prüfstufen untergliedert: - <b>Routineprüfung</b> der Ozondaten und anderer Luftschadstoffkomponenten. Die täglich einlaufenden Daten werden vor ihrer Weiterverarbeitung in operationellen Anwendungen von nicht plausiblen Werten bereinigt. Ferner fungiert das System als „Fehlerfrühwarnsystem“, indem vom System angezeigte Signale über mögliche systematische Fehler oder Ausreißer an die betroffenen Messnetzzentralen der Länder weiter gegeben werden. - <b>Die interaktive Arbeitsweise</b> erlaubt eine komplexe Beurteilung der Datenreihen, da im Gegensatz zur echtzeitnahen Validierung mögliche Fehler besser aus dem zeitlichen und räumlichen Kontinuum der Ozonreihen beurteilt werden können.. Dieses Modul stellt dem Anwender eine Vielzahl graphisch aufbereiteter Informationen zur Verfügung und erlaubt eine Reihe von Interaktionsmöglichkeiten zur besseren Bewertung angezeigter Fehler. Besonders bei zufälligen Fehlern bietet sich die Möglichkeit einer subjektiven Korrektur an. - <b>Die finale Datenvalidierung</b> hat die Aufgabe, Jahresdatensätze einer komplexen Wertung zu unterziehen. Es werden sowohl systematische Fehler als auch nicht plausible Werte untersucht. Das dazu eingesetzte Validierungsmodul beinhaltet eine vollautomatische Kontrolle und eine interaktive Variante.		
17. Schlagwörter Ozondaten, Echtzeit Datenvalidierung , interaktive Datenvalidierung		
18. Preis	19.	20.

## Report Cover Sheet

1. Report No. UBA-FB	2.	3.
4. Report Title A realtime and interactive validation and correction method of air quality data using objective criteria.		
5. Autor(s), Family Name(s), First Name(s) Dr. Enke, Wolfgang Waltraudstr. 20 D – 14532 Güterfelde	8. Report Date Oktober 2002	
	9. Publication Date November 2002	
6. Performing Organisation (Name, Address) I.D.E.A., Karin Seewöster Wandelsweg 89 D – 45894 Gelsenkirchen-Buer	10. UFOPLAN-Ref. No. 200 42 264	
	11. No. of Pages 67	
7. Funding Agency (Name, Address) Umweltbundesamt (Federal Environmental Agency) Postfach 33 00 22, 14191 Berlin	12. No. of Reference 10	
	13. No. of Tables, Diagrams 14	
	14. No. of Figures 40	
15. Supplementary Notes		
16. Abstract The aim of the research project was the development and application of a data validation method using unified test criteria. The validation module checks ozone time series for systematic errors and implausible values. Implausible data can be marked or replaced depending on the further use of the tested data set. The complete system consists of three parts: - <b>The real time validation</b> of ozone time series and other components of air quality. Implausible data are set to nil before using this data e.g. for presentation purpose. This module can be used as an early warning system for systematic errors and implausible values. In case of assigned errors information is sent to the Länder agency concerned. - <b>The interactive validation</b> allows a more complex procedure because complete time series are available. In addition the module produces a lot of complex information for a better evaluation of errors and allows interactions. For implausible values there is a graphical correction tool for the user. - <b>The purpose of the final data validation</b> is the complex validation of complete yearly data sets. The time series are checked for systematic errors as well as for implausible values. A fully automatically validation module as well as an interactive variant is available.		
17. Keywords ozone data, realtime data validation, interactive data validation		
18. Price	19.	20.

	<b>Inhaltsverzeichnis</b>	<b>Seite</b>
	<b>Allgemeines</b>	3
<b>1</b>	<b>Methodische Grundlagen</b>	5
1.1	Stationarität und Homogenität von Zeitreihen	5
1.2	Einige Methoden zur Stationaritätsprüfung von Zeitreihen	5
1.2.1	Abbe-Test	6
1.2.2	Buishand-Test	6
1.2.3	Craddock-Test	7
1.2.4	Doppelsummenanalyse	7
1.2.5	Alexandersson-Test	7
1.3	Methodische Grundlagen zur raum-zeitlichen Kontrolle bei der Prüfung von Einzelwerten	8
1.4	Methodische Grundlagen zur Erkennung systematischer Fehler	9
<b>2</b>	<b>Zeitnahe Erkennung nicht plausibler Werte</b>	11
2.1	Beschreibung der Vorgehensweise	11
2.1.1	Schwellenwerte	11
2.1.2	Die Prüfschranken	13
2.1.3	Der Programmablauf	14
2.1.3.1	Validierung - unsinnige Werte	14
2.1.3.2	Validierung – Ausreißerprüfung	14
2.1.3.3	Validierung - Prüfung des Anstiegs bzw. des Rückgangs zwischen aufeinander folgenden Halbstundenwerten	15
2.1.3.4	Validierung - Prüfung in der Fläche	15
2.1.3.5	Validierung - Prüfung mit einer approximierten Zeitreihe	15
2.1.3.6	Abwägung der bemängelten Messwerte	16
2.1.4	Das log-File	16
2.1.4.1	Rücklesen von Daten aus dem log-File	17
2.1.5	Anmerkungen zur Dateistruktur	17
2.2	Beschreibung der Eingriffsmöglichkeiten	18
<b>3</b>	<b>Zeitnahe Erkennung systematischer Fehler</b>	19
3.1	Grundlagen der Validierung zum Auffinden systematischer Fehler	19
3.2	Beschreibung des Bildschirminhaltes zur zeitnahen Erkennung systematischer Fehler	20
3.3	Bedeutung der Button	21
3.4	Demonstration der Wirkungsweise des Validierungsverfahrens durch Aufprägen systematischer Fehler	23
<b>4</b>	<b>Interaktive Datenvalidierung</b>	25
4.1	Interaktive Validierung systematischer Fehler	26
4.1.1	Beschreibung des Bildschirminhaltes	27
4.1.1.1	Beschreibung des Bildschirminhaltes oberer Teil.	27
4.1.1.2	Beschreibung des Bildschirminhaltes unterer Teil.	28
4.1.2	Bedeutung und Funktion der Button	30
4.1.2.1	Oberer Teil der interaktiven Grafik	30
4.1.2.2	Unterer Teil der interaktiven Grafik	39
4.1.3	Typisches Beispiel eines systematischen Fehlers	39
4.2	Erkennung und interaktive Validierung nicht plausibler Werte	40

4.2.1	Interaktive Erkennung und Korrektur nicht plausibler Messwerte	40
4.2.2	Beschreibung des Bildschirminhaltes	42
4.2.3	Möglichkeiten einfacher Interaktionen	43
4.2.3.1	Auswahl eines Zoom-Gebietes	43
4.2.3.2	Ablezen der Ozonwerte an den Kurven der Tagesgänge	44
4.2.4	Möglichkeit der Interaktion zur manuellen Korrektur von Stundenwerten	44
4.2.5	Funktionsweise der Button	45
<b>5</b>	<b>Finale Datenvalidierung</b>	<b>49</b>
5.1	Finale Validierung systematischer Fehler	49
5.2	Dokumentation der Validierungsstufe zur Erkennung systematischer Fehler	57
<b>6</b>	<b>Erste Untersuchungen zur Validierung systematischer Fehler anderer Komponenten</b>	<b>58</b>
6.1	Methodik	58
6.2	Ergebnisse	58
<b>7</b>	<b>Zusammenfassung</b>	<b>63</b>
<b>8</b>	<b>Ausblick</b>	<b>66</b>
<b>9</b>	<b>Literatur</b>	<b>67</b>
	<b>Anhang: Schnittstellen</b>	

## Allgemeines

Erste Erfahrungen zur rechnergestützten Validierung von Ozondatenreihen liegen aus dem F+E-Vorhaben 297 42 848 des UBA „Analyse historischer Immissionsmessreihen und Entwicklung einer Methode zur quasi-wetterbereinigten Trendanalyse von bodennahem Ozon“ vor.

Im laufenden Vorhaben zur „Verbesserung der bundesweiten Immissionsdaten durch Entwicklung und Anwendung einer Methode zur Validierung und Korrektur kontinuierlich ermittelter Immissionswerte nach einheitlichen Kriterien“ wurden wir beauftragt, in der ersten Arbeitsphase ein operationell einsetzbares System zu entwickeln, das die von den Messnetzen der Bundesländer und vom Immissionsmessnetz des UBA dem Fachgebiet II 6.2 im UBA übergebenen Ozondaten nach einheitlichen Kriterien überprüft. Das System soll in der Lage sein, zufällige Fehler (Ausreißer) wie auch systematische Fehler (Abdriften von Messgeräten usw.) aufzuspüren und zu erkennen und je nach vorgegebener Zielrichtung diese

- fehlerhaften Werte zu markieren,
- als fehlerhaft erkannte Werte durch Ausfallkennungen zu ersetzen,
- als fehlerhaft erkannte Werte und Ausfallwerte durch Schätzwerte zu ersetzen.

Als erstes zu realisierendes Ziel wurde uns vom UBA vorgegeben, das System so zu gestalten, dass realtime-Ozondaten, so wie sie täglich im UBA eingehen, relativ echtzeitnah geprüft werden können.

Diese Datenprüfungen sind wichtig, weil fehlerhafte Daten vor ihrer Weiterverarbeitung in operationellen Anwendungen (Ozonprognoserechnung, Erstellung von Karten zur aktuellen Ozonsituation) aus den Datensätzen ausgefiltert werden müssen.

Ferner fungiert das System auch als „Fehlerfrühwarnsystem“, indem vom System angezeigte Signale über mögliche Fehler an die betroffenen Messnetzzentralen der Länder weiter gegeben werden. Insbesondere dürfte dies zur Früherkennung und zu relativ frühzeitigem Gegensteuern bei systematischen Abweichungen von großer Bedeutung sein, da solche Fehler von diesem System objektiv und recht zuverlässig erkannt werden können.

Bei dieser echtzeitnahen Validierung erfolgen zunächst noch keinerlei Veränderungen der Originaldaten. Es wird lediglich für jede Station, für die mögliche Unplausibilitäten gefunden werden, eine Fehlerdatei erstellt. Im operationellen Validierungslauf werden zudem diese Auffälligkeiten in der grafischen Oberfläche des Systems visualisiert. Außerdem können mittels einer speziellen Präsentationssoftware einzelne Stationen mit möglichen fehlerhaften Messwerten, die in den genannten Fehlerdateien abgelegt sind, aufgeblendet und sichtbar gemacht werden.

Der eigentliche Validierungslauf kann in mehreren Modi erfolgen:

- vollautomatische Version zum Finden von zufälligen Fehlern
- vollautomatische Version zum Finden von systematischen Fehlern
- vollautomatische Version zum Finden von zufälligen und systematischen Fehlern
- interaktive Version zum Finden von zufälligen Fehlern
- interaktive Version zum Finden von systematischen Fehlern
- interaktive Version zum Finden von zufälligen und systematischen Fehlern

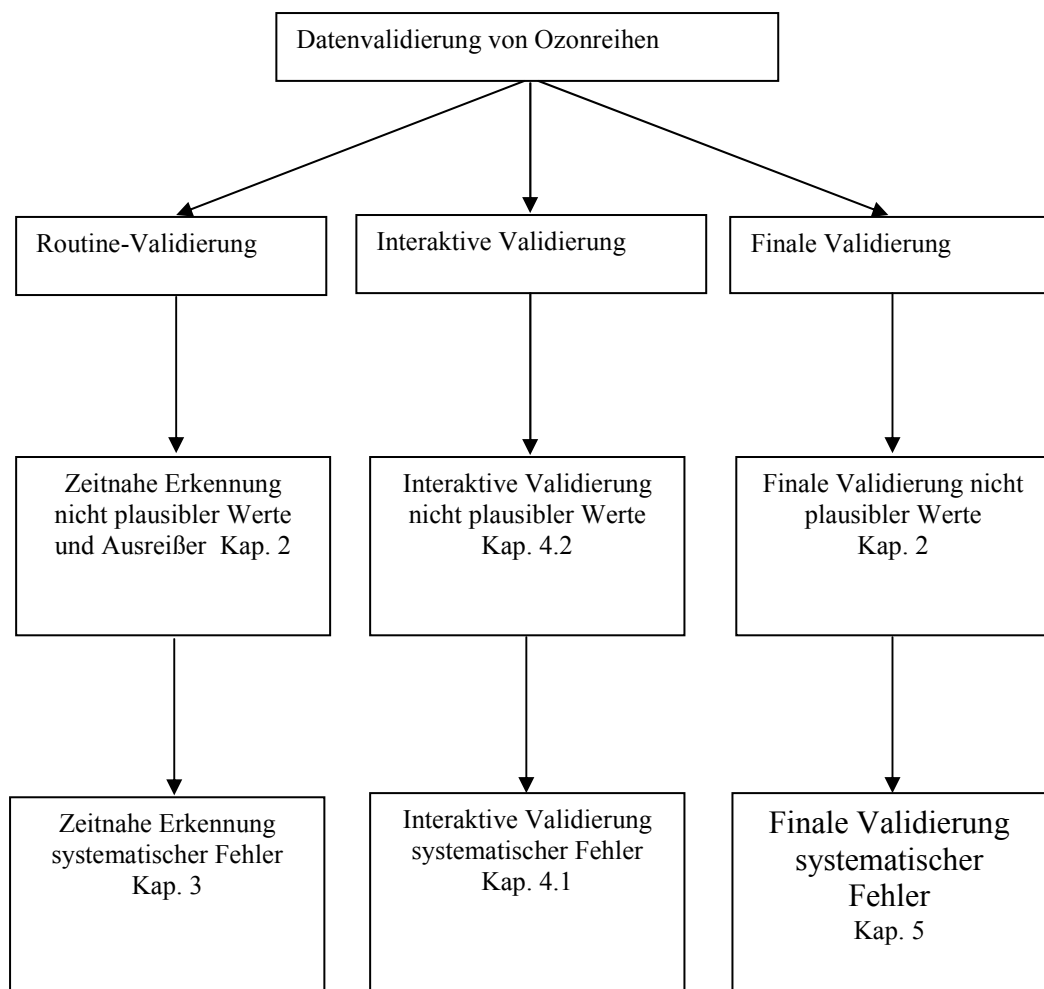
Die interaktive Arbeitsweise erlaubt u.U. besonders bei zufälligen Fehlern die Möglichkeit einer subjektiven Korrektur.

Über die echtzeitnahe Ozondaten Validierung hinaus gehend, ist es auch möglich, mit diesem System komplette Ozondatensätze eines ganzen Jahres zu prüfen (also jeden Einzelwert eines Jahres, für jede Stunde und jeden Tag). Im Gegensatz zur echtzeitnahen Datenvalidierung werden jedoch nicht nur Fehlermarkierungen durchgeführt. Vielmehr wird parallel zum Originaldatensatz ein von systematischen und zufälligen Fehlern bereinigter Datensatz erzeugt, in welchem alle Daten die verändert wurden, eine Markierung erhalten. Diese Jahresdatensatz- bzw. finale Datenvalidierung bedeutet entsprechend der wesentlich größeren zu validierenden Datenmenge einen zeitlichen Mehraufwand. Ebenfalls wie bei der echtzeitnahen Datenvalidierung ist hier eine automatische oder interaktive Arbeitsweise wählbar.

Auf die beschriebene Weise validierte Daten, sind eine wichtige Hilfestellung für die nach EU-Richtlinien dem Mitgliedstaat obliegende Qualitätssicherung der Daten.

Eine Übersicht der vorhandenen Validierungsmodule mit den dazugehörigen Verweisen auf die entsprechenden Kapitel ist in Abb. 1.1 zu finden.

**Abb. 1.1: Strukturbild der vorhandenen Validierungsmodule für Ozon**



# 1 Methodische Grundlagen

## 1.1 Stationarität und Homogenität von Zeitreihen

Ein grundlegendes Problem bei der Analyse klimatologischer Zeitreihen ist die Feststellung ihrer Homogenität bzw. Inhomogenität. Im Folgenden werden einige Verfahren zur Homogenitätsprüfung dahingehend bewertet, ob sie zur Auffindung von Messfehlern in Ozonreihen geeignet sind.

Stationarität bedeutet, dass die statistischen Charakteristika einer Messreihe zeitlich invariant sind, d.h. keine zeitlichen Abhängigkeiten aufweisen (Schönwiese, Malcher (1985). Bernhofer (2001) untersuchte die Stationarität und Inhomogenität von Niederschlagsreihen für Sachsen und konnte zeigen, dass Inhomogenitäten unterschiedlichste Ursachen haben können.) Die Instationarität meteorologischer Zeitreihen und hierzu können auch Luftschadstoffe, wie Ozon gerechnet werden, kann zwei prinzipielle Ursachen haben: meteorologische und nicht meteorologische. In der vorliegenden Arbeit geht es um die Erkennung nicht meteorologischer Inhomogenitäten. Diese können ihrerseits verschiedene Ursachen haben. Eine Ursache ist eine Verlegung der Messstelle, die sich durch einen Bruch (Intervention, einzelne Stufe) in der Zeitreihe bemerkbar macht. Das gleiche gilt für einen Wechsel des Messgeräts oder des Messverfahrens. Für Luftschadstoffe, die anthropogenen Einflüssen unterliegen, kommen noch Variationen der Quellen hinzu, d.h. Änderungen des Schadstoffausstoßes einer Quelle oder die Verlagerung einer Quelle. Die einfachste und sicherste Aufdeckung von Inhomogenitäten dieser Art beruht auf der verlässlichen Dokumentation(en) der Stationsgeschichte. Auch die Statistik (Zeitreihenanalyse) liefert Verfahren zur Erkennung solcher Interventionen, die dann durch ein gut geführtes „Stationslogbuch“ verifiziert werden können (und müssen).

Es kommt jedoch eine weitere, schwieriger zu erkennende Inhomogenität hinzu, die auf schleichender oder plötzlicher Änderung der Empfindlichkeit des Messgerätes beruht. Diese Art der Inhomogenität ist bei Ozonmessreihen dominierend und zudem zeitlich begrenzt. Deshalb bedarf es der Entwicklung neuer Methoden zur Erkennung dieser Fehlerart.

Grundsätzlich gibt es zwei Möglichkeiten der Homogenitätsprüfung:

- a) absolute Homogenitätstests, die sich auf nur eine Zeitreihe stützen und entsprechend unsicher sind und
- b) relative Homogenitätstests, die auf Vergleichen mit benachbarten Stationen bzw. Gebietsmittelwerten (als Referenz-Datenreihen) beruhen.

## 1.2 Einige Methoden zur Stationaritätsprüfung von Zeitreihen

Eine sehr einfache Prüfung auf Stationarität einer Zeitreihe ist mittels folgendem Schnelltest möglich.

$$st \approx \pm s \cdot \sqrt{\frac{n}{m \cdot (m-1)}}$$

wobei  $s$  die Standardabweichung,  $n$  der Stichprobenumfang und  $m$  die Anzahl der Filtergewichte der genutzten Tiefpassfilterung sind (mit  $m \ll n$ ). Tiefpassfilterung werden durchgeführt, wenn ein vorwiegend niederfrequentes Signal (langsame Schwingungen) in einer Zeitreihe von vorwiegend hochfrequenten Rauschen (störende rasche Fluktuationen) befreit werden sollen.



Als erster Schritt wird die Reihe mit  $m$  Filtergewichten Tiefpass-gefiltert (abhängig von der zu filternden Reihe). Diese gefilterte Reihe wird zusammen mit der gemessenen Reihe in einem Diagramm über der Zeit aufgetragen. In diesem Diagramm werden die vom Mittelwert abgezogene sowie hinzugezogene st-Linie dargestellt. Wenn sich die Tiefpass-gefilterte Reihe innerhalb der vorgegebenen Grenzen „st“ bewegt, kann die Aussage getroffen werden, dass die Reihe stationär ist. Dies beinhaltet keine Aussage über die Homogenität! Verlässt die Tiefpass-gefilterte Reihe den durch die „st“-Grenzen gegebenen Rahmen, so kann die Reihe als nicht-stationär bezeichnet werden.

### 1.2.1 Abbe-Test

Dieser Test berücksichtigt sowohl die Vorzeichen der Datenänderungen als auch die Abweichungen vom Mittelwert. Nach Abbe muss im Fall von Homogenität, die folgende Ungleichung erfüllt sein:

$$1 - \frac{1}{\sqrt{n-1}} \leq \frac{2 \cdot A}{B} \leq 1 + \frac{1}{\sqrt{n-1}}$$

mit

$$A = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 - \frac{1}{2}[(x_1 - \bar{x})^2 - (x_n - \bar{x})^2]$$

$$B = [(x_1 - \bar{x}) - (x_2 - \bar{x})]^2 + [(x_2 - \bar{x}) - (x_3 - \bar{x})]^2 + \dots + [(x_{n-1} - \bar{x}) - (x_n - \bar{x})]^2$$

In dieser Form diagnostiziert der Abbe-Test jedoch auch langfristig auftretende Fluktuationen, die klimatisch real sind, als Inhomogenitäten, ist also überkritisch. Außerdem kann mit diesem Verfahren der Betrag der Inhomogenitäten nicht bestimmt werden.

### 1.2.2 Buishand-Test

Der Buishand-Test (Buishand 1982), verwendet in seinem relativen Homogenitätstest ebenfalls kumulative Abweichungen. Er verwendet den Vergleich von zwei Testreihen, was für den Homogenitätstest von Ozonmessreihen schon eher geeignet ist. Der Test bestimmt drei Testgrößen:

$$Q = \max |S_{k*}| \text{ aus } 0 \leq k \leq n$$

$$R = \max(S_{k*}) - \min(S_{k*}) \text{ , jeweils aus } 0 \leq k \leq n$$

$$U = \frac{1}{1(n+1)} \cdot \sum_{k=1}^{n-1} (S_{k*})^2$$

dabei ist

$$S_{k*} = \frac{\sum_{i=1}^k (d_i - d)}{\sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - d)^2}}$$

mit  $k = 1, 2, \dots, n$ ;

$d_i$  sind die Differenzwerte zwischen Test- und Referenzreihe und  $d$  ist deren Mittelwert. Der Test setzt Normalverteilung und Datenunabhängigkeit (d.h. keine signifikante Autokorrelation) voraus. Die Testgrößen  $Q, \dots, U$  müssen bestimmte Tabellenwerte, siehe Tabelle 1.1, überschreiten, um auf dem betreffenden Niveau Inhomogenität vermuten zu lassen. Kurzzeitige Sprünge und Brüche in den Reihen lassen sich jedoch nicht ermitteln.

Tabelle 1.1 Signifikanzgrenzen zum Buishand-Test

	$\frac{Q}{\sqrt{n}}$			$\frac{R}{\sqrt{n}}$			U		
n	90%	95%	99%	90%	95%	99%	90%	95%	99%
20	1,10	1,22	1,42	1,34	1,43	1,60	0,343	0,447	0,662
50	1,14	1,27	1,52	1,44	1,55	1,78	0,342	0,452	0,718
100	1,17	1,29	1,55	1,50	1,62	1,86	0,341	0,457	0,712
$\infty$	1,22	1,36	1,63	1,62	1,75	2,00	0,347	0,461	0,743

### 1.2.3 Craddock-Test

Craddock (1979) verwendet für seinen grafischen Homogenitätstest ebenso den Vergleich zweier Reihen, wobei die Verteilung keinen wesentlichen Einfluss hat. Ausgeprägte Maxima und Minima weisen in den Summen  $S_i$ ,  $i=1,n$  auf Inhomogenitäten hin und geben den ungefähren Zeitpunkt der Strukturbrüche an. Der Test eignet sich für eine interaktive Datenkontrolle, ermöglicht aber keine definitiven Signifikanzaussagen.

$$S_1 = \frac{\bar{b}}{a} a_1 - b_1; S_2 = \frac{\bar{b}}{a} a_1 - b_1 + \frac{\bar{b}}{a} a_2 - b_2; \dots S_n..$$

$a_1$  :homogene Referenzreihe (ggf. auch Gebietsmittel-Reihe),  $b_i$  die zu testende Reihe.

### 1.2.4 Doppelsummenanalyse

Dyck (1980) verwendet Summenkurven der fortlaufend aufsummierten Werte der zu prüfenden Beobachtungsreihe gegen die im gleichen Zeitrhythmus aufsummierten Werte einer Bezugsreihe. Bei signifikanter Korrelation lässt sich die Doppelsummenkurve in ein Diagramm eintragen. Ist eine signifikante Änderung des Anstieges gegeben, deutet dieses auf eine Inhomogenität hin. Im üblichen Rahmen erfolgt nur eine optische Auswertung. Für die Untersuchung der Signifikanz der Änderung des Anstieges wird der F-Test angewendet.

### 1.2.5 Alexandersson - Test

Die bisher untersuchten Homogenitätstests sind nicht in der Lage, erkannte Inhomogenitäten zu bereinigen. Mit dem von Alexandersson (1986) entwickelten Test ist dies möglich.

Im ersten Schritt wird die Homogenität der vorliegenden Reihe geprüft. Hierzu sind zwei Methoden wählbar. Die erste nutzt den Quotienten zweier Reihen, die zweite nutzt deren Differenzen. Vor dem eigentlichen Test wird eine Normierung der Differenzen bzw. Quotientenreihe vorgenommen.

$$z_i = \frac{q_i - \bar{q}}{s_q}$$

Dabei ist  $q_i$  der Quotient bzw. die Differenz,  $\bar{q}$  der Mittelwert und  $s_q$  die Standardabweichung mit  $1 \leq i \leq n$ . Mit Hilfe der normierten Zeitreihe werden nun rechts und links in der Zeitreihe beginnende Summenkurven  $z_1$  und  $z_2$  berechnet.

$$z_1 = \frac{1}{i_0} \cdot \sum_{i=0}^{i_0} z_i$$

$$\bar{z}_2 = \frac{1}{n-i_0} \cdot \sum_{i=i_0+1}^n z_i$$

Die Testgröße wird daraus wie folgt ermittelt:

$$T_0 = \max[i_0 \cdot \bar{z}_1^2 + (n-i_0) \cdot \bar{z}_2^2] \text{ mit } 1 \leq i_0 \leq n$$

Überschreitet der Wert  $T_0$  die entsprechende Signifikanzgrenze, so ist auf dem entsprechenden Niveau Inhomogenität zu erwarten.

Tabelle 1.2: Signifikanzgrenzen zum Alexandersson-Test;  $T_{90}$  stellt die Signifikanzgrenze von 90 % dar,  $T_{95}$  entsprechend 95 %

n	$T_{90}$	$T_{95}$
25	6,55	7,75
50	7,25	8,55
75	7,65	8,95
100	7,80	9,15
150	8,05	9,35
200	8,20	9,55

Der Zeitpunkt der Inhomogenität ist  $n_0$  [=n für  $T_n = \max$ ]. Das relative Ausmaß des Sprunges beträgt:

$$\frac{\frac{1}{n-n_0} \cdot \sum_{i=n_0+1}^n q_i}{\frac{1}{n} \cdot \sum_{i=1}^{n_0} q_i}$$

### 1.3 Methodische Grundlagen zur raum-zeitlichen Kontrolle bei der Prüfung von Einzelwerten

Die Prüfung von Einzelwerten wird sowohl durch plausible Schranken als auch durch statistische Kenngrößen aus der Zeitreihe selbst und durch die Prüfung der räumlichen Konsistenz vorgenommen.

- Die Sinnfälligkeit eines Messwertes wird durch die Vorgabe plausibler Grenzen geprüft.
- Auf der Basis stationsspezifischer statistischer Kenngrößen (Mittelwert und Streuung) werden die Überschreitung von Schwellenwerten (n-Faches der stationsspezifischen Streuung um den Mittelwert) getestet. Im allgemeinen Fall wird hier Normalverteilung der Zeitreihe vorausgesetzt. Ist dies nicht der Fall, sollte eine Transformation der Zeitreihe in eine angenäherte Normalverteilung vorgenommen werden, um das entsprechende „n“ zu bestimmen.
- Weiterhin wird die Kontinuität des Tagesganges der Messgröße untersucht. Für jede Prüfkomponeute werden stationsspezifische Schwellen berechnet.
- Die räumliche Konsistenz der Messwerte wird über die Berechnung der simulierten Stationszeitreihe unter Verwendung benachbarter Stationen geprüft.
- Als nicht plausibel erkannte Werte werden, je nach Programmmodus, automatisch ersetzt, mit Ausfallkennung versehen oder können interaktiv am Bildschirm verändert werden.

## 1.4 Methodische Grundlagen zur Erkennung systematischer Fehler

Für die Erkennung systematischer Fehler, d.h. für die Erkennung von zeitlich begrenzten Inhomogenitäten wird ein abgewandeltes Verfahren nach Alexandersson (1986) verwendet.

Der erste Unterschied besteht darin, dass die Differenzreihe (Zeitreihe der Residuen) verwendet wird, wie sie sich als Differenz zwischen den beobachteten Werten und einer aus den umliegenden Stationen approximierten Zeitreihe ergibt. Dieses Verfahren wurde im Rahmen der Fehlwertersetzung von Ozonreihen entwickelt (F+E-Vorhaben 297 42 848 des Umweltbundesamtes Berlin: „Analyse historischer Immissionsmessreihen und Entwicklung einer Methode zur quasi-wetterbereinigten Trendanalyse von bodennahem Ozon“ – 1. Etappe: „Trendanalyse der täglichen Ozonmaxima der Jahre 1990 bis 1997“).

Ein weiterer Unterschied zu dem Andersson-Verfahren resultiert aus der Besonderheit der Inhomogenitäten von Ozonreihen. Diese sind gegenüber langen Klimareihen der Temperatur oder des Niederschlages häufig mit kurzzeitigen Sprüngen und Brüchen (Interventionen) behaftet, die ihre Ursache im Betrieb der Messtechnik haben.

Das gleitende Mittel dieser Residuenreihen liefert einen guten Anhaltspunkt über mögliche systematische Fehler in diesen Zeitreihen, da man mit Recht davon ausgehen kann, dass über eine längere Zeitspanne gemittelt die statistischen Charakteristika der Abhängigkeiten zwischen den Stationen nur geringfügig schwanken. Treten Sprünge, langsame Verschiebungen oder Wellen in diesen gleitenden Mitteln auf, so deutet dies auf mögliche systematische Fehler in den Messreihen hin. Dies kann die unterschiedlichsten Ursachen haben und sollte einer subjektiven Bewertung unterzogen werden.

Zur Erkennung systematischer Fehler wird in folgenden Schritten vorgegangen:

- Der Gesamtdatensatz (mindestens 3 Jahre) wird jeweils in Sommer- und Winterhalbjahr (April bis September bzw. Oktober bis März) unterteilt. Dies trägt dem Umstand Rechnung, dass die meteorologischen Bedingungen zur Ozonbildung bzw. zum Ozonabbau zwischen Sommer- und Winterhalbjahr sehr verschieden sind und somit auch die statistischen Relationen zwischen den Ozonmessreihen deutlichen jahreszeitlichen Abhängigkeiten unterliegen.
- Zur Stabilisierung der Aussagen über mögliche systematische Fehler in den Zeitreihen, werden um die zu untersuchende Tageszeit fünfstündige Mittelwerte gebildet. Dadurch werden zufällige Schwankungen in den Messreihen stärker geglättet, so dass systematische Fehler deutlicher hervortreten.
- Alle Datensätze werden vor der Berechnung der räumlichen Abhängigkeiten zwischen den Messstationen vom Jahresgang und (linearen) Trend bereinigt. Aus der klassischen Statistik ist bekannt (Taubenheim 1969), dass systematische Schwankungen (zyklische Schwankungen oder Trends stochastischer Zeitreihen Scheinkorrelationen hervorrufen können, die sich störend auf die wirklichen Abhängigkeiten zwischen den Messgrößen auswirken.
- Berechnung der Residuenreihe als Differenz zwischen der beobachteten Ozonreihe (5-stündiges Mittel) und der mittels multipler Regression simulierten Reihe.
- Erneute Trendbereinigung der Residuenreihe.
- Für die realtime-Datenvalidierung (Kap. 3) wird innerhalb des Zeitintervalls der zurückliegenden 50 Tage die maximale Änderung der Residuenreihe mittels linearer Regression gesucht. Überschreitet dieser berechnete Wert eine stationsspezifische Schwelle, so liegt möglicherweise ein systematischer Fehler vor. Die stationsspezifischen Schwellenwerte werden über die entsprechende Streuung der

Residuenreihe berechnet. D.h. je ungenauer die Simulation einer Ozonzeitreihe ist, um so höher bestimmen sich die Schwellenwerte, die zum Auslösen eines Warnsignals führen.

- Zur finalen und interaktiven Datenvalidierung dienen die von rechts und von links kommenden gleitenden Mittel ( $\pm 15$  Tage) der Residuenreihe als Validierungsgröße (Alexandersson verwendet im Gegensatz dazu Summen). Überschreiten diese gleitenden Mittel einen stationsspezifischen Schwellenwert, der gleichfalls aus der Streuung der Residuenreihe berechnet wird, so stellt der Schnittpunkt zwischen den gleitenden Mitteln den Schwerpunkt des Bereiches eines möglichen systematischen Fehlers dar. Die maximale Differenz der gleitenden Mittel rechts und links vom Schnittpunkt markiert den Beginn und das Ende des so beanstandeten Intervalls (nähere Ausführung dazu in Kap. 4 und 5).

## 2 Zeitnahe Erkennung nicht plausibler Werte

Im Folgenden wird ein Modul beschrieben, das automatisch stündliche Daten prüft, die dem UBA von den Messnetzen der Bundesländer und des UBA geliefert werden. Es wird auf negative Werte, Ausreißer sowie die zeitliche und räumliche Konsistenz hin geprüft. Diese Prüfung erfolgt auf der Basis von historischen Datenreihen, wie sie in den Datenarchiven des UBA vorhanden sind. Mit Hilfe dieser Datenreihen werden stationsspezifische Schwellenwerte ermittelt. Die Änderungen werden in einer log-Datei protokolliert.

### 2.1 Beschreibung der Vorgehensweise

Für die Erstellung des Validierungsmoduls wurden folgende Schritte unternommen.

Als erstes wurden Schwellenwerte für jeden Stofftyp ( $O_3$ ,  $SO_2$ ,  $NO_2$ ,  $PM_{10}$ ) ermittelt. Diese werden während der Validierung als Prüfgrößen herangezogen. Abgesehen von Schwefeldioxid erfolgt bei allen anderen Größen zusätzlich eine Einteilung in Stationskategorien (z.B. für Ozon: Stadt, Land, Verkehr und Berg).

Im Prozess der Validierung werden unter Verwendung dieser Schwellenwerte folgende fünf Prüfschritte durchgeführt.

- Prüfung auf Werte kleiner Null
- Ausreißerprüfung
- Prüfung des Anstiegs bzw. des Rückgangs zwischen aufeinander folgenden Halbstundenwerten
- Prüfung in der Fläche
- Prüfung mit einer approximierten Zeitreihe

Wird ein Wert als falsch befunden, so wird er auf die Nil-Kennung (-999.00) gesetzt. Für jeden dieser Fälle wird in einem log-File ein Eintrag mit der Begründung für dessen Ablehnung abgelegt.

#### 2.1.1 Schwellenwerte

Mit den vom UBA für jeden Stofftyp bereitgestellten Datenarchiven wurden für jeden Stofftyp Schwellenwerttabellen berechnet. Der Umfang der Datenarchive, die für die Ableitung der Schwellenwerte zur Verfügung standen, war bei den einzelnen Stofftypen sehr unterschiedlich.

- **Ozon** 10 Jahre mit insgesamt 480 Stationen
- **$SO_2$**  4 Jahre mit insgesamt 375 Stationen
- **$NO_2$**  10 Jahre mit insgesamt 615 Stationen
- **$PM_{10}$**  3 Jahre mit insgesamt 192 Stationen

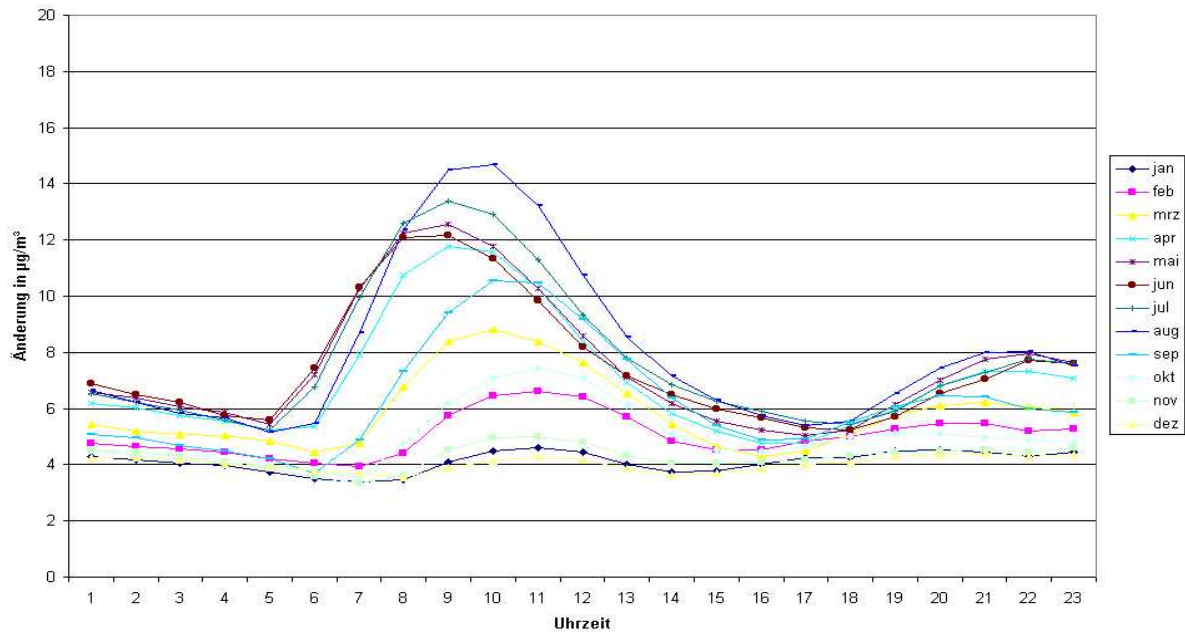
Insbesondere bei  $PM_{10}$  ist die Datenlage sehr problematisch. So haben im ersten Jahr (1998) nur drei Länder  $PM_{10}$  gemessen. Erst ab 2000 ist die Datenlage mit mehr als 13 meldenden Ländern (inkl. UBA) besser. Die Tage zum Jahreswechsel (31.12. und 1.1.) wurden generell nicht genutzt. Hier werden auf Grund pyrotechnischer Aktivitäten extrem hohe Konzentrationen gemessen.

Ermittelt wurden jeweils der Mittelwert und die Standardabweichung getrennt für:

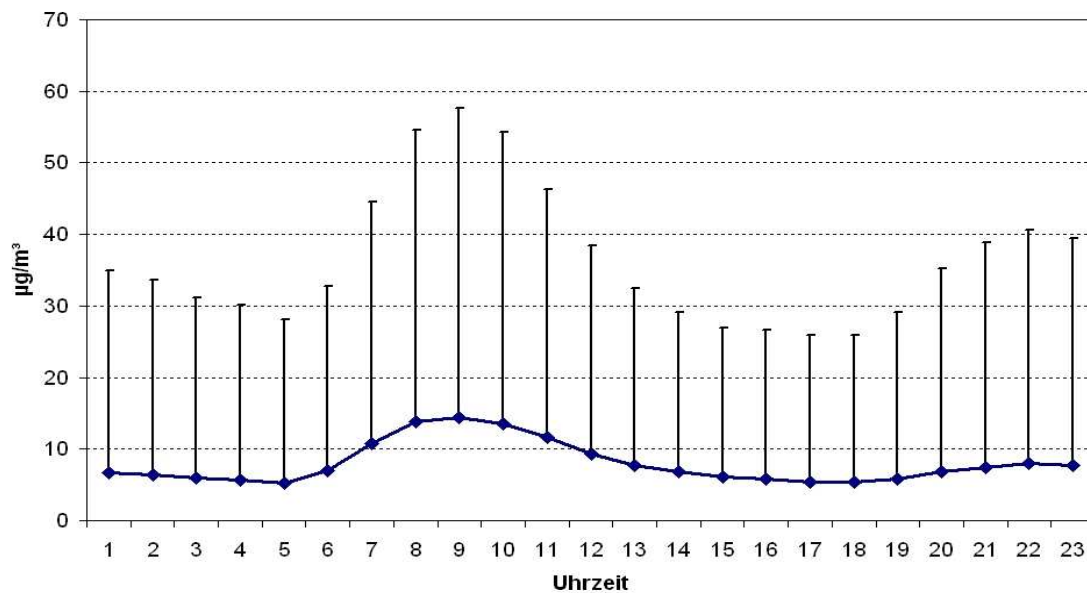
- den Stofftyp
- den Anstieg bzw. die Reduktion innerhalb einer Stunde

- die Differenz zwischen Teststation und den Nachbarstationen gleichen Typs.

Es erfolgt eine getrennte Mittelung und Schwellenwertberechnung über alle Stationskategorien, Monate und Stunden.



**Abb. 2.1:** Beispielfhafte Darstellung der mittleren stündlichen Veränderungen der Ozonwerte, gemittelt über alle Stationen. Jede Kurve stellt einen Monat dar.



**Abb. 2.2:** Beispielfhafte Darstellung von Prüfschwellen für Ozon. Dargestellt wurden die Monatsmittelwerte für Juli mit dem Mittelwert plus 4-mal die Standardabweichung für den Anstieg von Stunde zu Stunde (Stationstyp: Stadt).

### 2.1.2 Die Prüfschranken

Im Rahmen der Entwicklung dieses Prüfmoduls im Herbst 2001 wurden feste Prüfschranken für jede Prüfung bei allen Stoffen festgelegt. Das heißt zum Beispiel, dass bei der Ausreißerprüfung jeder Messwert mit einer Summe aus Mittelwert und der F-fachen Standardabweichung verglichen wird.

Gleichung 2.1: Prüfwert = Mittelwert +  $F$  \* Standardabweichung

Der Mittelwert und die Standardabweichung entsprechen den oben berechneten Schwellenwerten für jeden Stofftyp, Monat, Uhrzeit und Stationskategorie. Für die zu prüfenden Stoffe wurden aus den Datenreihen folgende in der Tabelle 2.1 in den ersten fünf Zeilen aufgelisteten Faktoren ( $F$ ) für die Standardabweichung ermittelt.

Tabelle 2.1 Der Inhalt der Datei *Testwert.dat* im Verzeichnis *\konstdat*. (Die Prüfung von NO ist in der Routine implementiert, aber abgeschaltet.)

Zeile		Faktor F in Gl. 2.1
1	Stdabw. Ozon:	4
2	Stdabw. SO <sub>2</sub> :	5
3	Stdabw. NO <sub>2</sub> :	5
4	Stdabw. NO:	5
5	Stdabw. PM <sub>10</sub> :	10
6	MaxWert Ozon:	360
7	MaxWert SO <sub>2</sub> :	500
8	MaxWert NO <sub>2</sub> :	300
9	MaxWert NO:	400
10	MaxWert PM <sub>10</sub> :	500
11	approx GW1:	30
12	approx GW2:	50

Die Zeilen 6 bis 10 der Tabelle 2.1 beinhalten die Werte der als maximal möglich angesehenen Messwerte in  $\mu\text{g}/\text{m}^3$  für jeden Stoff. In den letzten zwei Zeilen sind die Prüfschranken für die Prüfung gegenüber der approximierten Reihe eingetragen.

Im Zeitraum 17. bis 20. Juni 2002 trat jedoch eine Ozonepisode auf, die Änderungen an der bisher vorliegenden Prüfroutine nötig machte. Während dieser Episode, bei der das Hereinströmen von tropischer und subtropischer Festlandsluft in Verbindung mit hoher Einstrahlung eine extreme Hitzeperiode verursachte, traten Ozonwerte auf, die seit einigen Jahren nicht mehr gemessen wurden. Hinzu kam, dass es nur sehr schwacher regionaler Luftmassenaustausch stattfand. Die Ozonchemie erfolgte daher sehr lokal. Somit war die üblicherweise gute Korrelation zwischen den Stationen nicht mehr gegeben. Als Folge dessen wurde eine Reihe von Messwerten fälschlicherweise verworfen. Aus dieser Erfahrung heraus waren Änderungen am Validierungsmodul nötig.

Als Ansatz wurde die Idee verfolgt, dass ab einer Temperatur von  $\geq 25^\circ\text{C}$  die festen Prüfschranken aufgegeben werden. Für Temperaturen über  $25^\circ\text{C}$  werden die Prüfgrenzen nun an die für den aktuellen Tag prognostizierte Maximum-Temperatur angepasst. Mittels einer Polynomfunktion 2. Grades erhöhen sich die Grenzen entsprechend Gleichung 2.1.



Die Prognose der Maximum-Temperatur wird vom Deutschen Wetterdienst (DWD) für die Ozonprognose bereitgestellt. Jede zu prüfende Station wird zu diesem Zweck einem der Prognosegitterpunkte zugeordnet. Überschreitet die Maximum-Temperatur den oben genannten Wert so wird der vorgeschriebene Faktor unter Nutzung der Gleichung 2.2 erhöht.

Gleichung 2.2: 
$$F_{Stdabw,var} = F_{Stdabw,fix} + 2 * (0,004 * T^2 - 0,124 * T + 0,58)$$

Hierbei ist T die prognostizierte Maximum-Temperatur für einen in der Nähe liegenden Gitterpunkt der GME-Prognose des Deutschen Wetterdienstes.

Auch die beiden Grenzwerte, die man bei der Prüfung mit der approximierten Reihe nutzt, werden mit einer analogen Funktion (Gleichung 2.3) verändert. Deren Nutzung wird im Kapitel 2.1.3.5 erläutert.

Gleichung 2.3: 
$$GW_{var} = GW_{fix} + GW_{fix} * (0,004 * T^2 - 0,124 * T + 0,58)$$

Hierbei ist T wieder die prognostizierte Maximum Temperatur für einen in der Nähe liegenden Gitterpunkt der GME-Prognose des Deutschen Wetterdienstes.

### 2.1.3 Der Programmablauf

Nach dem Start des Validierungsprogramms werden alle Schwellenwerte in den Arbeitsspeicher geladen.

Danach wird die zu bearbeitende Datei geöffnet. In der ersten Zeile dieser Datei ist das Datum abgelegt. Entsprechend dieses Datums erfolgt die Auswahl der entsprechenden Schwellenwerte. Beim Einlesen der oben genannten Datei werden die Daten in einem Feld mit folgender Struktur abgelegt.

- getrennt nach Tag und Vortag
- getrennt nach Stofftyp

Gleichzeitig werden auf der Grundlage der Stationsdatenbank des UBA für jeden Stofftyp Stationslisten für aktuell vorhandene Stationen erstellt. Anhand dieser Liste können die jeweils acht am nächsten gelegenen (je zwei in jedem Quadranten) Stationen gesucht werden. Die maximale Entfernung zur jeweiligen Station ist mit 150 km vorgegeben. Auch hier erfolgt eine Unterteilung nach Stationskategorien.

Tritt beim Auslesen der csv-Datei (Datei mit den aktuellen Immissionswerten aus den Ländermessnetzen) das Problem auf, dass die entsprechende Station nicht im Stationsverzeichnis ist, so wird die Station ungeprüft in die neue Werte-Datei geschrieben. In der log-Datei wird dieses aufgelistet.

#### 2.1.3.1 Validierung - unsinnige Werte

Als erster Schritt erfolgt die Ersetzung aller negativen Werte durch –999.00 Nil-Kennung. Weiterhin wird geprüft, ob der Wert einen stoffspezifischen Maximalwert überschreitet. Wenn ja, so erfolgt eine Ersetzung durch die Nil-Kennung.

#### 2.1.3.2 Validierung - Ausreißerprüfung

Monats- und Uhrzeit-spezifisch erfolgt eine Prüfung aller Einzelwerte. Als nächstes wird auf Ausreißer bzw. Extremwerte geprüft. Die Prüfung erfolgt nur für den positiven Bereich, da alle negativen Werte schon auf Ausfallkennung gesetzt wurden. Übersteigt ein Wert die Schranke 'Mittelwert plus F-mal Standardabweichung', so wird dieser speziell gekennzeichnet.

Zusätzlich gibt es für jede Stoffklasse eine Datei, in der Stationen aufgelistet werden können, bei denen ein markierter Prüfalgorithmus nicht durchgeführt werden soll. Diese Dateien sind im Verzeichnis */konstdat* als z.B. *pm10\_sonder.dat* zu finden. Hiermit können Stationen, deren Messwerte durch Besonderheiten der lokalen Emissionssituation stark geprägt sind, von einer Prüfung ausgeschlossen werden. Vom normalen Änderungsverhalten der Ozonkonzentration abweichend treten von einer zur anderen Stunde z.B. an der Station Hürth hervorgerufen durch lokale Emittenten von Ozonvorläuferstoffen – überproportional hohe Anstiege auf.

#### 2.1.3.3 Validierung - Prüfung des Anstiegs bzw. des Rückgangs zwischen aufeinander folgenden Halbstundenwerten

Jeder Wert wird mit seinem zeitlich vorhergehenden Wert verglichen. Entsprechend der Änderungsrichtung wird der Betrag der Änderung geprüft. Übersteigt die Änderung die Schranke 'Mittelwert plus F-mal Standardabweichung' (siehe Abbildung 2.2), erfolgt eine Markierung des Wertes. Fehlt der vorhergehende Wert, so werden die Werte bis zu zwei Stunden zurück genutzt. Bei der Verwendung von Werten mit einem zeitlichen Abstand von mehr als einer Stunde wird die Schranke entsprechend modifiziert (es erfolgt eine Interpolation über den entsprechenden Zeitraum).

#### 2.1.3.4 Validierung - Prüfung in der Fläche

Mit den aus der Stationsliste extrahierten nächstliegenden Stationen kann eine Prüfung in der Fläche erfolgen. Hierbei wird der zu validierende Wert mit den aktuellen Werten der Nachbarstationen verglichen und die minimale Differenz gesucht. Diese wird dann gegen eine für den Schadstofftyp und die Stationskategorie spezifische Schranke geprüft. Wird die Schranke überschritten, so wird der Wert durch die Nil-Kennung ersetzt. Da nicht immer alle acht Stationen zu allen Zeitpunkten mit Werten besetzt sind, wurde die Bedingung eingebaut, dass bei mindestens drei Nachbarstationen Werte vorhanden sein müssen.

#### 2.1.3.5 Validierung - Prüfung mit einer approximierten Zeitreihe

Diese Prüfung erfolgt z.Z. nur bei Ozon. Dabei wird der zu prüfende Wert mit einem approximierten Wert verglichen, wie er durch multiple Regression mittels ausgewählter Nachbarstationen erhalten wird. Übersteigt die absolute Differenz zwischen beiden den Wert  $30\mu\text{g}/\text{m}^3$ , so wird der Wert gekennzeichnet. Ist die Differenz größer als  $50\mu\text{g}/\text{m}^3$ , so ist dieser Wert nicht mehr als sinnvoll anzusehen und er wird durch die Nil-Kennung ersetzt. Diese Grenzwerte gelten für Temperaturen unter  $25^\circ\text{C}$ . Darüber werden diese Grenzwerte entsprechend der Gleichung 2.3 erhöht.

Zusätzlich zu der Veränderung der Grenzwerte wurde noch eine weitere Bedingung eingefügt. Ab Temperaturen von  $\geq 25^\circ\text{C}$  erfolgt eine Markierung und Nil-Setzung nur, wenn auch die Prüfung des Anstieges bzw. der Reduktion einen möglichen Fehler erkannt hat.

Basis für die Approximierung einer Zeitreihe ist eine Screening-Regressions-Analyse. Hierbei wird für jede Station, auf Basis des Datenarchivs, die beste Kombination an Stationen gesucht, mit denen der Wert an der Station approximiert werden kann. Vor der Screening-Prozedur werden alle Messwerte vom Jahres- und Tagesgang bereinigt. Somit erfolgt eine Berechnung der Abweichung vom Jahres- und Tagesgang.

Fehlt ein Messwert bei einer der für die Regression genutzten Stationen, so wird deren Regressionsfunktion genutzt um wiederum deren Wert zu approximieren. Fehlt auch dort ein Messwert, so wird das Mittel der Abweichungen über insgesamt bis zu 8 Nachbarstationen angesetzt. Mit diesen approximierten Vergleichswerten, wird der Messwert geprüft.

### 2.1.3.6 Abwägung der bemängelten Messwerte

Bei drei der fünf Prüfungen werden bemängelte Werte nur markiert (Ausreißerprüfung, Prüfung des Anstiegs bzw. des Rückgangs zwischen aufeinander folgenden Halbstundenwerten und Prüfung mit einer approximierten Zeitreihe). Als letzter Schritt erfolgt nun eine Abwägung der markierten Fehlertypen. Tritt bei zwei Tests die Kennzeichnung *fragwürdig* auf, so wird dieser Wert durch die Nil-Kennung ersetzt.

Zum Abschluss wird das geprüfte Werte-Feld in eine dem csv-Dateityp entsprechende Datei geschrieben.

### 2.1.4 Das log-File

Im log-File werden alle Ergebnisse der Validierung aufgelistet. Hier wird neben Stationsnummer, Schadstofftyp, Datum, ursprünglicher Wert und Uhrzeit auch die Fehlerart aufgeschrieben (siehe Tabelle 2.2).

**Tabelle 2.2: log-File (Beispiel)**

Intern Code *	Stationscode	Stofftyp	Datum	Intern Code	Messwert	Uhrzeit **	Prüfwert	Fehlerart
0	NI052	PM10	20011005	F2	-33,00	18		negativer Wert
0	HE018	PM10	20011005	F6	205,00	30	151,93	Ausreißer nahe Stationen 8
0	SN006	PM10	20011005	F3	528,00	21	500,00	Unsinniger Wert

\* siehe Bemerkungen zum Rücklesen von Daten aus dem log-File

\*\* Es wird mit Halbstundenwerten gerechnet. Durch eine Division durch 2 kann die Uhrzeit errechnet werden.

Wird ein Wert durch die Nil-Kennung ersetzt, so wird dieses in das log-File geschrieben. Im o.g. Beispiel bedeutet das für die erste Zeile: An der Station NI052 trat bei der PM<sub>10</sub>-Messung am 05.10.2001 um 09:00 Uhr ein negativer Wert auf. In der zweiten Zeile trat folgender Fall ein: Am selbigen Tag war für PM<sub>10</sub> die kleinste Differenz des Messwertes der Station HE018 zum Messwert der Nachbarstationen größer als 151,93 µg/m<sup>3</sup> (die für diese Uhrzeit und den Monat zum Vergleich aus „Mittelwert und F-mal Standardabweichung“ berechneten Schranke; es wurden im aktuellen Fall 8 Stationen zu Vergleich genutzt). Bei dem Messwert in der dritten Zeile überstieg der aktuell gemessene Wert den Grenzwert (hier mit 500 µg/m<sup>3</sup> angenommen).

Neben dem großen log-File werden für jedes Bundesland komprimierte log-Files geschrieben. Hier werden nur die Station, der Stofftyp, die Uhrzeit und die Fehlerart aufgelistet. Die Auflistung eines Tages für den jeweiligen Stofftyp und die Station erfolgt in einer Zeile. Somit sind diese log-Files deutlich kleiner als das detaillierte log-File. Für den Fall, dass in einem Bundesland bei keiner Station ein Wert auf die Nil-Kennung gesetzt wurde, wird am Ende des Programms diese Datei gelöscht. Für die Verteilung der log-Files via Mail stehen damit nur Dateien zur Verfügung bei denen wirklich Informationen enthalten sind.

### 2.1.4.1 Rücklesen von Daten aus dem log-File

Nachdem das Modul automatisch die csv-Datei geprüft hat, kann eine subjektive Prüfung der bemängelten Werte angeschlossen werden. Hierbei hilft das log-File. Am Anfang einer jeden Zeile des log-Files wird während des Schreibens eine '0' gesetzt. Wird diese bei einer Handkontrolle durch eine '1' ersetzt, so kann bei Nutzung des Programms *neu-einlesen* jeder einzelne von der UMEG ursprünglich gemeldete Wert in die csv-Datei rückübertragen werden. Somit ist der bemängelte Wert wieder für die grafische Darstellung nutzbar. Die Prüfung mit den Daten des Vortages wird zweimal durchgeführt. Am Folgetag morgens um 7 Uhr und abends um 19 Uhr. Damit wird den Ländern die Möglichkeit eingeräumt, bemängelte Werte zu prüfen und gegebenenfalls zu korrigieren oder zu sagen, der Wert ist richtig. Wird von den Ländern gemeldet, dass einer der bemängelten Werte als richtig angesehen wird, und mittels der *neu-einlesen* Routine zurück gelesen, so wird der Wert bei der Prüfung am Abend nicht noch mal geprüft. Es wird also die Prüfung bei diesem Wert deaktiviert. Dazu wird im Verzeichnis *\log* eine Datei mit dem Datum angelegt und darin die Stationsnummer, der Stofftyp, das Datum und Uhrzeit eingetragen. Das Validierungsmodul sucht bei jedem Durchlauf ob für den Tag oder den Vortag eine solche Datei vorhanden ist, und wenn ja, liest es die entsprechenden Zeilen aus und überspringt dann die eingetragenen Werte.

### 2.1.5 Anmerkungen zur Dateistruktur

Das Modul umfasst drei Programm-Dateien sowie ein untergeordnetes Verzeichnis *\daten* mit drei Unterverzeichnissen (siehe Tab.: 2.3). **Anmerkung:** das Modul befindet sich auf den UBA-Rechnern im Verzeichnis *\Validierung*.

**Tabelle 2.3: Dateistruktur**

---

validierung.exe			
neu-einlesen.exe			
SortierenLogFile.exe			
	/daten	/konstdat	/ Dateiname.txt
	/daten	/konstdat	/ TestWert.dat
	/daten	/konstdat	/ Steuer.dat
	/daten	/konstdat	/ alle Sonderdateien
	/daten	/log	/ alle ReVal-Dateien
	/daten	/aktdat	/ die DWD Prognose
	/daten	/statlist	/ alle Stationslisten

---

In die Datei *Dateiname.txt* wird der Dateiname der zu bearbeitenden csv-Datei eingetragen (z.B.: uba.csv). Diese muss sich im Verzeichnis *daten/konstdat* befinden. Während der Bearbeitung werden im Verzeichnis */daten* zwei neue Dateien erstellt. Zum einen die geprüfte Datei (hier wird an die Endung des Dateinamens eine '1' angefügt (z.B.: uba.csv1)). Weiterhin wird hier das log-File abgelegt (Namensbeispiel: uba.log).

Wird das Programm *neu-einlesen* verwendet, so sind drei zusätzliche Dateien zu finden. Zum einen eine neue csv-Datei (Endung: \*.csv1), ein neues log-File (Endung: \*.log1) und im Verzeichnis */daten/log* eine Datei mit dem aktuellen Datum und der Endung \*.ReVal.

Die Datei *TestWert.dat* beinhaltet für jeden Stofftyp den Faktor F (Gleichung 2.1) für die Standardabweichung bei der Schrankenberechnung, den stoffspezifischen Maximalwert und die beiden Grenzwerte für die Prüfung mit den approximierten Zeitreihen. Der Inhalt und die Bedeutung der Datei *Steuer.dat* werden am Ende des nachfolgenden Kapitels beschrieben.

## 2.2 Beschreibung der Eingriffsmöglichkeiten

Eingriffe ins Programm sind an mehreren Stellen möglich. Zum ersten kann der Faktor für die Standardabweichungen und der stoffspezifische Maximalwert für den jeweiligen Stofftyp geändert werden. Auch die Grenzwerte für die approximierten Zeitreihen können hier geändert werden. Diese sind in der Datei *TestWert.dat* im Verzeichnis */daten/konstdat* aufgelistet. Hiermit können Anpassungen bei der Prüfschärfe des Moduls vorgenommen werden.

Als zweite Eingriffsmöglichkeit sind die Sonderdateien zu werten. Diese sind im selbigen Verzeichnis. Für jeden Stofftyp ist eine Datei angelegt worden. Bis auf Ozon sind in ihnen nur Platzhalter-Kennungen aufgelistet. In der Form UB001 können dort zeilenweise Stationen aufgelistet werden. Eintragungen von Stationen in diese Dateien sollten nur mit Bedacht durchgeführt werden! Grund dafür ist das Überspringen wichtiger Prüfabschnitte (Prüfung des Anstiegs bzw. des Rückgangs zwischen aufeinander folgenden Halbstundenwerten, Prüfung in der Fläche, Prüfung mit der approximierten Reihe).

Eine weitere Möglichkeit des Eingriffes ist die Datei *Sonder-Datum.dat* im Verzeichnis */daten/konstdat*. Dort können die Einstellungen der Datei *TestWert.dat* für einzutragende Tage geändert werden (zum Beispiel für den 31. Dezember). Beispielhaft ist in der Datei eine Zeile eingefügt worden, um die Formatkonventionen aufzuzeigen. Entspricht der aktuell zu bearbeitende Tag bzw. der Vortag einem Datum in der Datei, so werden die dort festgesetzten Prüfwerte genutzt.

Der wirkungsvollste Eingriff ist durch die Datei *Steuer.dat* möglich. Hier kann für jeden Stofftyp separat eingestellt werden, mit welchem Testmodul geprüft werden soll. Auch eine generelle Abschaltung der Prüfung eines Stoffes ist möglich. Am Ende der Datei steht noch die *Notfall*-Temperatur. Fällt die DWD-Prognose aus, kann dort eine mittlere Maximum-Temperatur für das gesamte Bearbeitungsgebiet eingetragen werden. Dieses ist nur bei Temperaturen deutlich über 25 °C nötig.

Der Aufbau der Datei ist Tabelle 2.4 zu entnehmen.

In der ersten Spalte (Typ) steht das Stoffkürzel bzw. das Temperaturkürzel. In der zweiten Spalte (Gesamt) steht, ob der jeweilige Stoff geprüft werden soll oder nicht. In der untersten Zeile steht die Ersatztemperatur in °C für den Fall das keine Temperaturprognose verfügbar ist. Die nachfolgenden Spalten (1-6) zeigen an, welche der Prüfungen beim jeweiligen Stoff angewendet werden soll. Die Spalte zwei (Gesamt) hat dabei die höhere Priorität. Die Nummern der Prüfkürzel bedeuten 1: Werte kleiner Null, 2: Prüfung auf konstante Werte (noch nicht implementiert), 3: Ausreißer- und Extremwertprüfung, 4: Prüfung des Anstieges bzw. der Reduktion, 5: Nahe Stationen und 6: Prüfung mit der approximierten. Reihe.

**Tabelle 2.4 Der Inhalt der Datei *Steuer.dat* im Verzeichnis */konsdat*. Die Prüfung von NO ist in der Routine implementiert, aber abgeschaltet.**

Typ	Gesamt	1	2	3	4	5	6
O3	J	J	N	J	J	J	J
SO2	J	J	N	J	J	J	J
NO2	J	J	N	J	J	J	J
NO	N	N	N	N	N	N	N
PM10	J	J	N	J	J	J	J
Temp	00						

### 3 Zeitnahe Erkennung systematischer Fehler

Die frühzeitige Erkennung möglicher systematischer Fehler in den Ozonmessreihen dient der Qualitätssteigerung schon vor Ort, d.h. durch eine zeitige Information der Messnetzzentralen über mögliche systematische Fehler sollen die Ursachen für diese systematischen Verschiebungen bei den Ozonmessgeräten selbst aufgespürt und beseitigt werden. Dies bedarf natürlich einer subjektiven Bewertung der angebotenen grafischen Ergebnisse, da hier die langjährige Erfahrung der Messtechniker an den Stationen ein unverzichtbarer Faktor ist.

#### 3.1 Grundlagen der Validierung zum Auffinden systematischer Fehler

Auch wenn ein Datenkollektiv der Ozonkonzentration von nicht plausiblen Werten und Ausreißern nach der im Kap. 2 dargestellten Validierungsstufe bereinigt ist, können im zeitlichen Verlauf einer Reihe trotzdem Trends und Strukturbrüche auftreten, die sowohl reale Ursachen (meteorologische Bedingungen, Änderungen der Emission der Vorläuferstoffe u.a.m.) haben als auch artifiziell (z.B. Änderungen der Probenahme- und Messtechnik) bedingt sein können. Deshalb ist es notwendig, die meteorologisch oder emissionsseitig bedingten Einflüsse jahreszeitspezifisch und stündlich aus der Messreihe zu eliminieren. Eine Möglichkeit hierzu bietet die Ersetzung der zu untersuchenden Zeitreihe mit Erwartungswerten, die aus anderen Ozonmessreihen berechnet werden (simulierte Zeitreihe). Für die operationelle Datenprüfung setzt man dabei voraus, dass die meteorologischen Bedingungen sich in den Ozonkonzentrationen umliegender Stationen genügend abbilden. Um nicht nur die Situation eines konkreten Tages für die Interpolation zu verwenden, wie dies bei rein räumlichen Interpolationsverfahren (Shepard-Verfahren, Thin Plate Spline, Triangulation oder IDW-Verfahren) der Fall ist, sondern auch die statistischen Beziehungen zwischen verschiedenen Beobachtungsreihen und der zu ersetzenden Reihe, wird ein stufenweises Screening-Regressionsverfahren eingesetzt. In Anlehnung an den Alexandersson-Test (Alexandersson 1986), wird die Differenz zwischen der insgesamt simulierten Messreihe (Erwartungswerte) und den Originalwerten einer Station verwendet (in den späteren Abbildungen die schwarze Kurve). Die Zeitreihe der Erwartungswerte selbst, ist dagegen nicht dargestellt. In dieser Differenzenreihe treten durch die weitgehende Eliminierung meteorologischer und emissionsseitig bedingter Einflüsse mögliche Strukturbrüche, Trends oder Messfehler deutlicher hervor, als dies bei den Originalreihen der Fall ist.

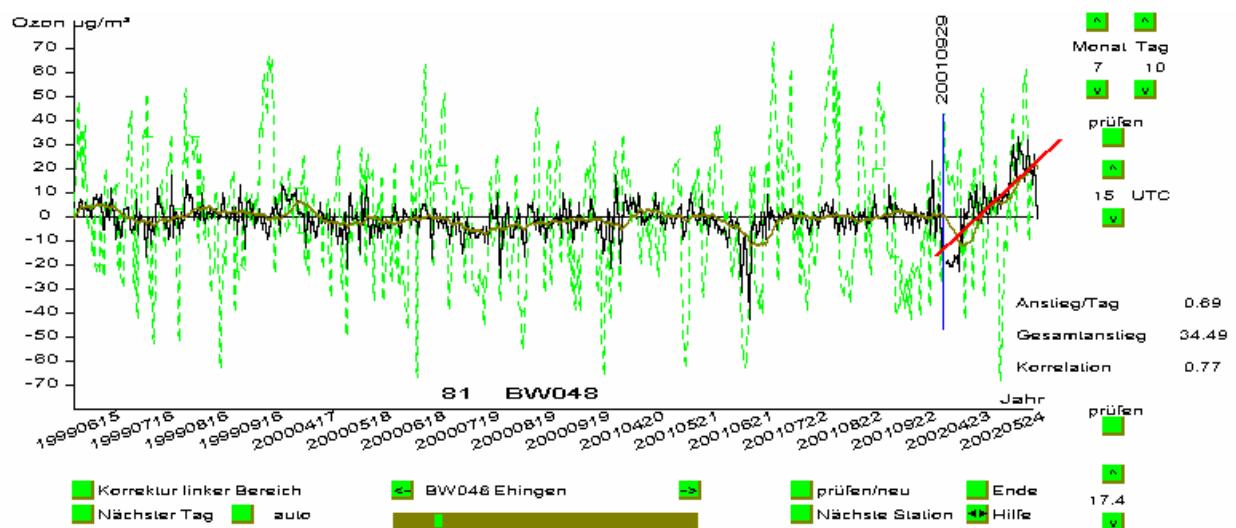
Eine korrekte Eliminierung meteorologischer Einflüsse, wie sie im Rahmen des F+E Vorhabens:297 42 848 „Analyse historischer Datenreihen und Entwicklung einer Methode zur quasi-wetterbereinigten Trendanalyse von bodennahem Ozon“ Teilbericht 3 „Wetterbereinigung der Jahre 1980 bis 1997“ unter Einbeziehung meteorologischer

Informationen durchgeführt wurde, ist hier aufgrund der Komplexität und des großen Aufwandes eines solchen Vorgehens nicht möglich. Man kann jedoch davon ausgehen, dass sich in den täglichen Schwankungen der Ozonmessreihen ein Großteil der Variabilität des Wettergeschehens widerspiegelt. Die Synchronität der Tagesgänge zwischen nahe beieinander liegenden Stationen drückt also die Ähnlichkeit des Wettergeschehens auf regionaler Ebene aus. Andersherum betrachtet, ist in der Differenzenreihe zwischen zwei oder mehreren nahe beieinander liegenden Ozonreihen der Wettereinfluss größtenteils eliminiert. Zur Erledigung dieser Aufgaben wurde ein Programmpaket entwickelt, dessen Funktionsweise in den nachfolgenden Punkten beschrieben wird.

Dieses Programm läuft einmal täglich vollautomatisch im UBA und überprüft die von den Messnetzen der Bundesländer und des UBA eingegangenen Ozondaten. Die vom Programm als fehlerhaft erkannten und signalisierten Bereiche werden in speziellen Files dokumentiert und abgelegt. Diese Dateien wiederum werden in einem mit den Bundesländern abgestimmten Modus via e-mail den betroffenen Bundesländern übermittelt und können in den Messnetzzentralen mittels eines von uns zur Verfügung gestellten Präsentationsmoduls visualisiert werden.

### 3.2 Beschreibung des Bildschirminhaltes zur zeitnahen Erkennung systematischer Fehler

Vor der näheren Erläuterung der Funktionsweise der zeitnahen Validierung systematischer Fehler sollen an Hand der Messstation BW048 – Ehingen (Abb.3.1) der Bildschirminhalt und die Funktionsweise der einzelnen Button erläutert werden.



**Abb. 3.1: Grundbild zur Erläuterung der Funktionalität der zeitnahen Erkennung systematischer Fehler an Hand der Station BW048 – Ehingen 15 UTC. Mit dem Wechsel der Ozonsaison 2001/2002 wird ein deutlicher Sprung angezeigt.**

Die **grün gestrichelte Kurve** zeigt die fünfständigen Mittel der Ozonkonzentration als Abweichung vom stationsspezifischen mittleren Jahresgang.

Die **schwarze Kurve** zeigt die Differenz zwischen beobachteten Ozonwerten (fünfständiges Mittel) und der unter Verwendung der umliegenden Stationen simulierten Ozonreihe.

Die **ocker-farbige Kurve** stellt das von links kommende gleitende Mittel der Differenz zwischen beobachteter und simulierter Zeitreihe (schwarze Kurve) dar. Diese Kurve stellt die Basis für die Beurteilung der systematischen Fehler dar.

Die **blaue senkrechte Linie** markiert den Beginn eines möglichen systematischen Fehlers. Das entsprechende Datum ist oberhalb der blauen Linie zu finden. Im vorliegenden Fall (Abb.3.1) wird der Beginn einer möglichen Verwerfung für den 29.09.2001 signalisiert.

Die **rote Gerade ist die lineare Regression**, berechnet über die Differenzkurve (schwarze Kurve). Zur Berechnung der linearen Regressionsgeraden werden **maximal 50** Tage rückwärts berücksichtigt, bezogen auf den aktuellen Tag. Mit Hilfe linearer Regression wird die **maximale Differenz** zwischen Beginn und Ende des beanstandeten Intervalls bestimmt. Diese als Gesamtanstieg bezeichnete maximale Differenz, ist auf der rechten Seite der Abbildung zu finden (in Abb. 3.1  $-34,49\mu\text{g}/\text{m}^3$ ).

Gleichfalls auf der rechten Seite der Abbildungen ist der **Anstieg der Regressionsgeraden pro Tag** ( $0,69\mu\text{g}/\text{Tag}$  in Abb.3.1) und der lineare **Korrelationskoeffizient** ( $0,77$  in Abb.3.1) zu finden.

Im interaktiven Modus kann durch Drücken der rechten Maustaste neben der blauen Linie der **beanstandete Bereich manuell geändert werden**. Sowohl die Regressionsgerade als auch Anstieg und Korrelationskoeffizient werden entsprechend angepasst.

Der Name der aktuell **bearbeiteten Station ist unten in der Mitte** der Grafik zu finden.

Auf der **Zeitachse** (X-Achse) sind **Jahr, Monat, und Tag** aufgetragen. Der letzte rechte Wert markiert hier den letzten Termin, der zur Validierung verwendet wurde. Das aktuelle Datum ist rechts oben unterhalb von „Monat“ und „Tag“ zu sehen. Es hat für die automatische, zeitnahe Erkennung systematischer Fehler keine Bedeutung.

### 3.3 Bedeutung der Button

#### Hilfe

Es wird eine kurze online-Hilfe gestartet, die die wesentlichen Funktionen der interaktiven Grafik erläutert. Durch Betätigen der Pfeiltasten im Button der Hilfe Funktion kann im Hilfetext geblättert werden.

#### Ende

Das Programm wird verlassen und alle geänderten Reihen werden gespeichert.

Die beanstandeten Stationen werden im Verzeichnis `c:\routwarn\temp\warn\syssterg\prl Stationskürzel.dat` abgelegt. Diese Datei enthält die Zeitreihe der Residuen sowie die von rechts und von links kommenden gleitenden Mittel der Residuenreihe und die jahresgangbereinigte Originalreihe. Sie können mit dem Programm- `praeswarn.exe`- erneut betrachtet werden. Mit jedem Aufruf des Programms `routwarnsteuer.exe` werden die alten Warndateien gelöscht.

#### Nächste Station

Es wird entsprechend der Stationsliste die nachfolgende Station bearbeitet.

Die beanstandeten Stationen werden im Verzeichnis `c:\routwarn\temp\warn\syssterg\prl Stationskürzel.dat` abgelegt. Sie können mit den Programm `praeswarn.exe` erneut betrachtet werden.

#### Ockerfarbiger Balken unter der Stationsbezeichnung

Durch Drücken der linken Maustaste auf den ockerfarbigen Balken erfolgt eine Grobeinstellung der Stationsauswahl. Eine **Feineinstellung** der gewünschten Station kann mit den darüber liegenden rechten und linken **Pfeiltasten** erfolgen.

Die Prüfung der Station wird mit dem Button „nächste Station“ oder „prüfen/neu“ aktiviert. Der Button „nächste Station“ bewirkt eine schnellere Reaktion als der Button „prüfen/neu“.



Letzterer setzt jedoch die dahinter liegende Datenbasis wieder auf den Ausgangszustand. (Der Button „nächste Station“ vererbt Eigenschaften vorangegangener Aktionen, z.B. geändertes Datum/Uhrzeit.)

### **Nächste Station**

Es wird die im Speicher folgende Station bearbeitet.

Wurde vorher auf dem ockerfarbigen Balken eine Grobeinstellung der Stationsauswahl vorgenommen oder mittels der darüber liegenden Pfeiltasten eine Feineinstellung vorgenommen, so aktiviert der Button „nächste Station“ die über dem ocker farbigen Balken angezeigte Station.

### **Datum einstellen**

In der Grafik rechts oben lassen sich mit Hilfe der Pfeiltasten **Monat und Tag einstellen**.

Danach ist der darunter liegende Button "prüfen" zu drücken, um die Änderung wirksam werden zu lassen. (Diese Funktion ist nur im interaktiven Modus aktiv.)

### **Uhrzeit einstellen**

In der Grafik rechts oben lässt sich mit Hilfe der Pfeiltasten die Uhrzeit einstellen.

Danach ist der darüber liegende Button "prüfen" zu drücken, um die Änderung wirksam werden zu lassen.

### **Prüfen (ganz rechts oben)**

Dieser Button führt die eingestellten Änderungen des Datums und der Uhrzeit aus.

### **Prüfen (rechts unten)**

Ganz unten rechts im Bild ist die **aktuelle Schwelle des Gesamtanstieges** angegeben, bei deren **Über- oder Unterschreitung** der angezeigte Bereich als möglicherweise fehlerhaft eingestuft wird.

Diese **Schwelle kann manuell** durch Betätigung der darüber bzw. darunter liegenden Pfeiltasten **modifiziert werden**.

Danach ist der darüber liegende Button "prüfen" zu drücken, um die Änderung wirksam werden zu lassen.

### **Nächster Tag**

Durch Drücken dieses Buttons wird der Folgetag geprüft. Dies ist nur möglich, wenn das Validierungsprogramm nicht im Routinebetrieb angewandt wird, d.h. die zu prüfenden Daten aus den Zyklusarchiven geholt werden.

### **auto**

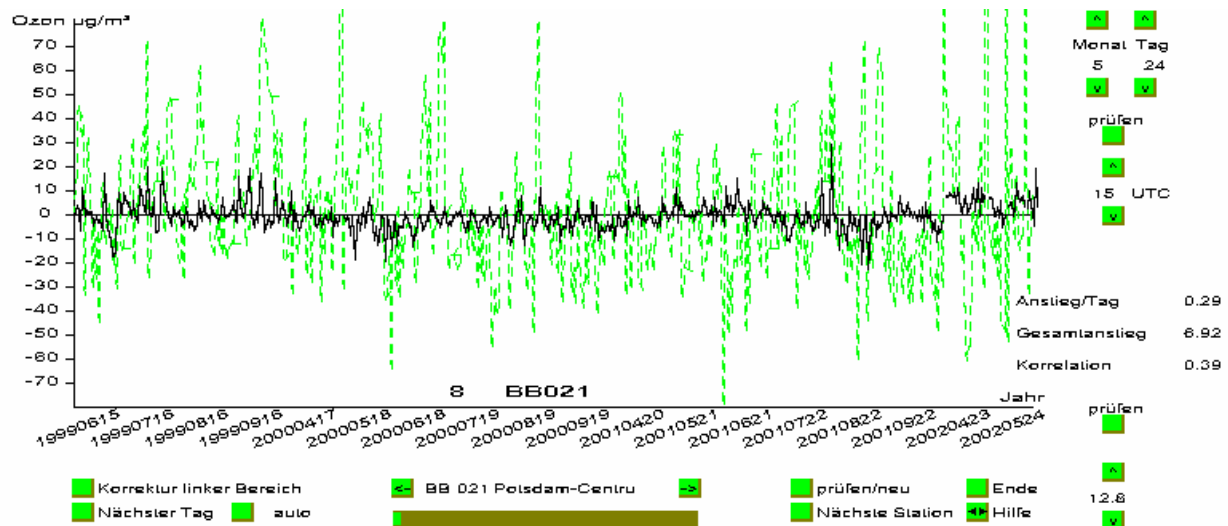
Durch Drücken dieses Buttons werden die 10 Folgetage in die Prüfung einbezogen. Dies ist nur möglich, wenn das Validierungsprogramm nicht im Routinebetrieb angewandt wird, d.h. die zu prüfenden Daten aus den Zyklusarchiven geholt werden.

### **Korrektur linker Bereich**

Es wird der systematische Fehler links von der blauen senkrechten Linie (beanstandeter Bereich) aus der Reihe eliminiert. (diese Interaktion hat nur kosmetische Bedeutung und dient dazu, den beanstandeten Bereich besser hervortreten zu lassen).

### 3.4 Demonstration der Wirkungsweise des Validierungsverfahrens durch Aufprägen systematischer Fehler

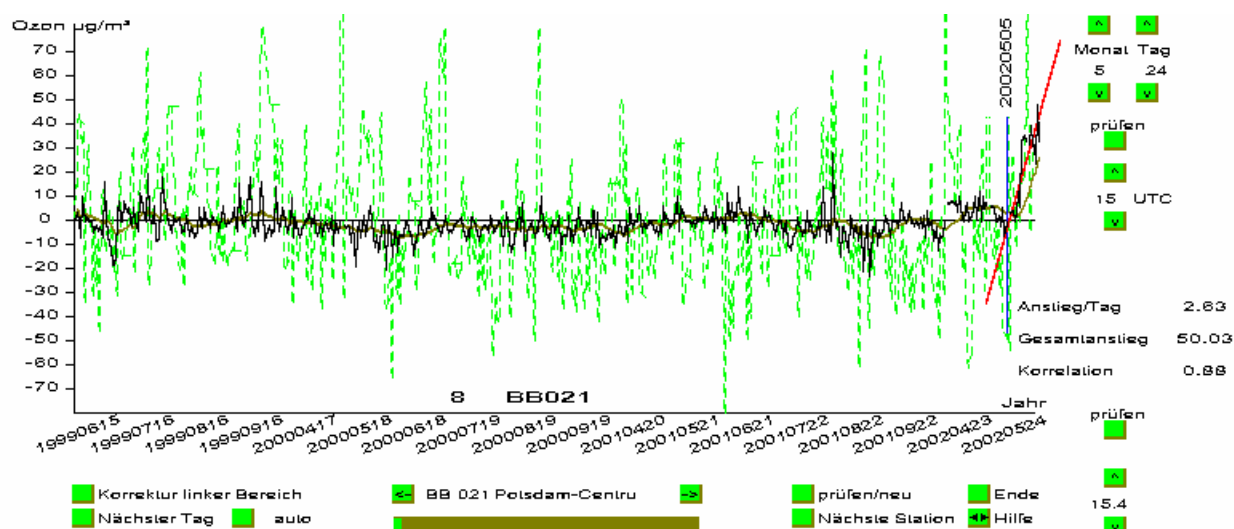
Um die Wirksamkeit und das Verhalten des entwickelten Prüfverfahrens zu demonstrieren, wird auf die Originalreihe Reihe der Station BB021 – Potsdam Zentrum (Abb. 3.2), ein systematischer Fehler von  $30 \mu\text{g}/\text{m}^3$  ab dem 14.05.2002 aufgeprägt (Abb.3.3). Im ungestörten Verlauf zeigt die Reihe BB021 keine Auffälligkeiten und eignet sich deshalb besonders für die Demonstration einer künstlich aufgeprägten Störung.



**Abb. 3.2: Grundbild zur Erläuterung der Funktionalität der zeitnahen Erkennung systematischer Fehler an Hand der Station BB021 - Potsdam-Zentrum 15 UTC. Die Messreihe zeigt keine Auffälligkeiten.**

Trotz größerer Änderungen des fünfständigen Mittels der Ozonkonzentration von Tag zu Tag des 15-Uhr-Termins (grüne Kurve Abb. 3.2) zeigt sich über den gesamten Zeitraum ein glatter Verlauf der Differenz zwischen simulierter und beobachteter Zeitreihe (schwarze Kurve Abb. 3.2).

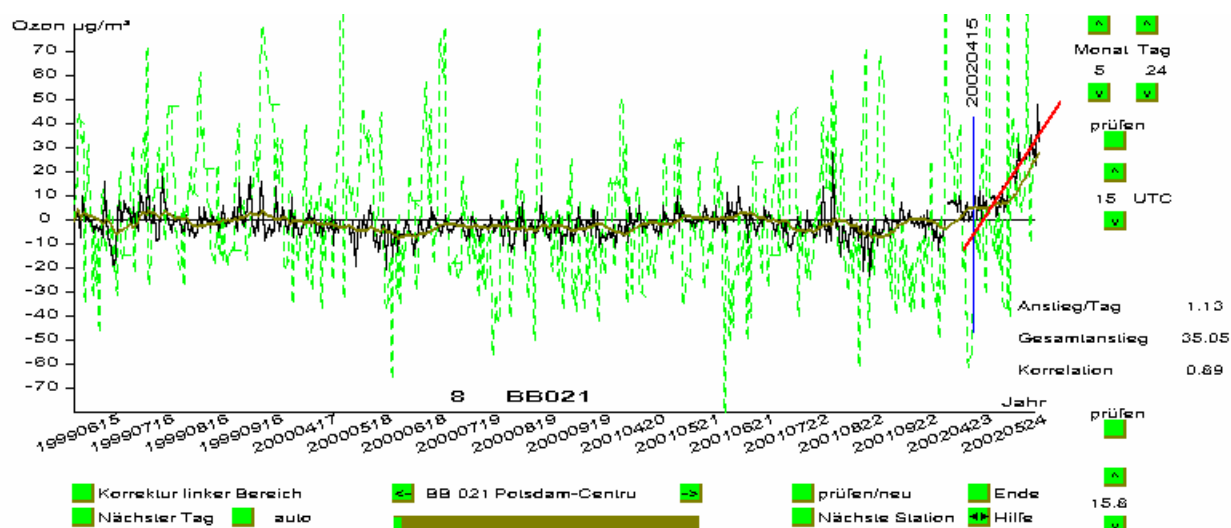
Wird zu der Messreihe der Station BB021 - Potsdam-Zentrum ab dem 14. Mai 2002 ein konstanter Wert von  $30 \mu\text{g}/\text{m}^3$  addiert (Abb.3.3), so signalisiert das Programm diese Verwerfung durch eine rot gezeichnete Regressionsgerade über die schwarze Kurve (Differenz zwischen simulierter und beobachteter Zeitreihe). Die Höhe der Verwerfung im angezeigten Intervall (rechts von der blauen senkrechten Kurve) wird als Anstieg pro Tag und als Gesamtanstieg abgeschätzt. Der für die untersuchte Station maximal zulässige Gesamtanstieg ist unter diesem Textfeld mit  $12.6 \mu\text{g}/\text{m}^3$  (Abb. 3.2) angegeben. Liegt der Gesamtanstieg unter dieser Schranke, wie in Abb. 3.2 mit  $6.92 \mu\text{g}/\text{m}^3$ , so wird diese Reihe als unauffällig eingestuft. Sowohl die Schwelle als auch das beanstandete Zeitintervall kann subjektiv verändert werden. Siehe dazu Kapitel.3.2 und 3.3. Das Validierungsmodul selbst sucht innerhalb des zu testenden Zeitintervalls (aktuelles Datum minus 50 Tage) den maximalen Gesamtanstieg. Im Falle des Aufprägens einer systematischen Abweichung von  $30 \mu\text{g}/\text{m}^3$ , ab dem 14.05.2002 (Abb. 3.3) wird der Gesamtanstieg ( $50.03 \mu\text{g}/\text{m}^3$ ) überschätzt. Dies resultiert aus leicht negativen Werten (schwarze Kurve) zu Beginn des beanstandeten Intervalls und der Abschätzung durch eine lineare Regression. Der Beginn der Verwerfung wird ab dem 05.05.2002 angezeigt. Bei einem täglichen Einsatz des Validierungsmoduls würde eine solch kräftige Verwerfung schon 3 Tage nach Beginn des Bruches erkannt.



**Abb. 3.3:** Grundbild zur Erläuterung der Funktionalität der zeitnahen Erkennung systematischer Fehler an Hand der Station BB021 - Potsdam-Zentrum 15 UTC. Ab dem 14.05.2002 wird eine künstliche Erhöhung um  $30 \mu\text{g}/\text{m}^3$  vorgenommen.

Prägt man auf die Zeitreihe der Station BB021 (Abb. 3.2) ab dem 24.04.2002 einen von  $1 \mu\text{g}/\text{m}^3$  auf  $30 \mu\text{g}/\text{m}^3$  stetig ansteigenden systematischen Fehler auf (Abb. 3.4), so wird dieser schon ab dem 02. Mai erkannt, obwohl die Störung zu diesem Zeitpunkt erst ca.  $10 \mu\text{g}/\text{m}^3$  betrug. Der Gesamtanstieg wird mit  $35.05 \mu\text{g}/\text{m}^3$  gut abgeschätzt. Das in Abb. 3.4 an der blauen Linie aufgetragene Datum stellt nur eine Abschätzung des ungefähren Beginns der Störung dar. Bei schleichenden systematischen Fehlern ist dieser geschätzte Beginn entsprechend Unsicher und weicht im konkreten Fall um 9 Tage vom realen Beginn ab.

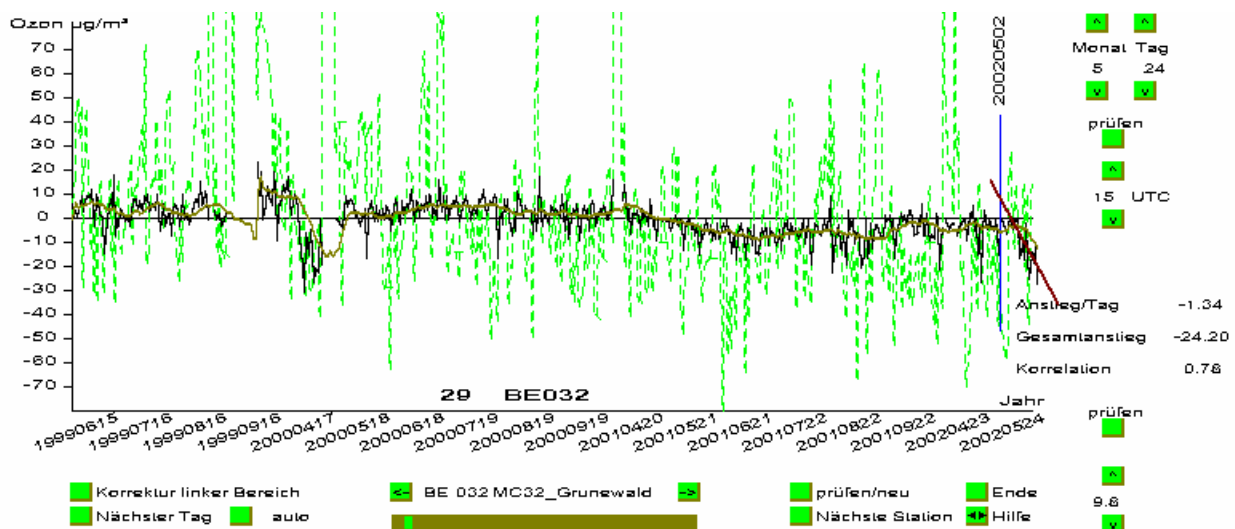
In den real untersuchten Zeitreihen kommen beide Arten systematischer Fehler (sprunghafte Veränderung und schleichender Anstieg bzw. Abfall) als typische systematische Abweichungen vor.



**Abb. 3.4:** Grundbild zur Erläuterung der Funktionalität der zeitnahen Erkennung systematischer Fehler an Hand der Station BB021 - Potsdam-Zentrum 15 UTC. Ab dem 24.04.2002 wird eine künstliche Erhöhung von  $1 \mu\text{g}/\text{m}^3$  kontinuierlich zunehmend auf  $30 \mu\text{g}/\text{m}^3$  vorgenommen.

Die Frage bleibt noch offen, ob der beanstandete systematische Fehler in der untersuchten Station selbst oder durch eine benachbarte Station verursacht wird, die zur Simulation der untersuchten Zeitreihe herangezogen wurde. Um diese Frage zu beantworten, bleibt die im konkreten Fall mit der Station BB021 am stärksten korrelierte Ozonmessreihe BE034 (Berlin-Neukölln) in einem zweiten Schritt unberücksichtigt. D.h. die Simulation der Ozonreihe BB021 erfolgt im zweiten Schritt ohne Verwendung der Station BE024. Tritt der systematische Fehler auch dann zu Tage, so wird angenommen, dass der systematische Fehler der zu untersuchenden Station zuzuordnen ist.

Zur Demonstration wird auf die Ozonmessreihe der Station BB021 ein systematischer Fehler ab dem 20.04.2001 von  $1 \mu\text{g}/\text{m}^3$  auf  $30 \mu\text{g}/\text{m}^3$  stetig ansteigend aufgeprägt. Die Station BB021 liefert wieder die beste Korrelation zur Simulation der Zeitreihe der Station BE032 – Berlin-Grünwald. Der systematische Fehler der Station BB021 wird in der Zeitreihe der Station BE032–Berlin-Grünwald durch eine stetige Abnahme der Differenz zwischen simulierter und beobachteter Ozonmessreihe sichtbar (Abb. 3.5). Die Regressionsgerade in Abb. 3.5 ist allerdings dunkel-rot eingefärbt. Dies signalisiert, dass der beanstandete Bereich durch eine benachbarte Station, in diesem Fall durch die Station BB021 verursacht wurde.



**Abb. 3.5: Grundbild zur Erläuterung der Funktionalität der zeitnahen Erkennung systematischer Fehler an Hand der Station BE032 – Berlin-Grünwald 15 UTC. Der Einfluss des künstlich aufgeprägten Fehlers auf die Zeitreihe der Station BB021 - Potsdam-Zentrum an der Station BE032 als scheinbarer Fehler sichtbar (vgl. Abb. 3.4).**

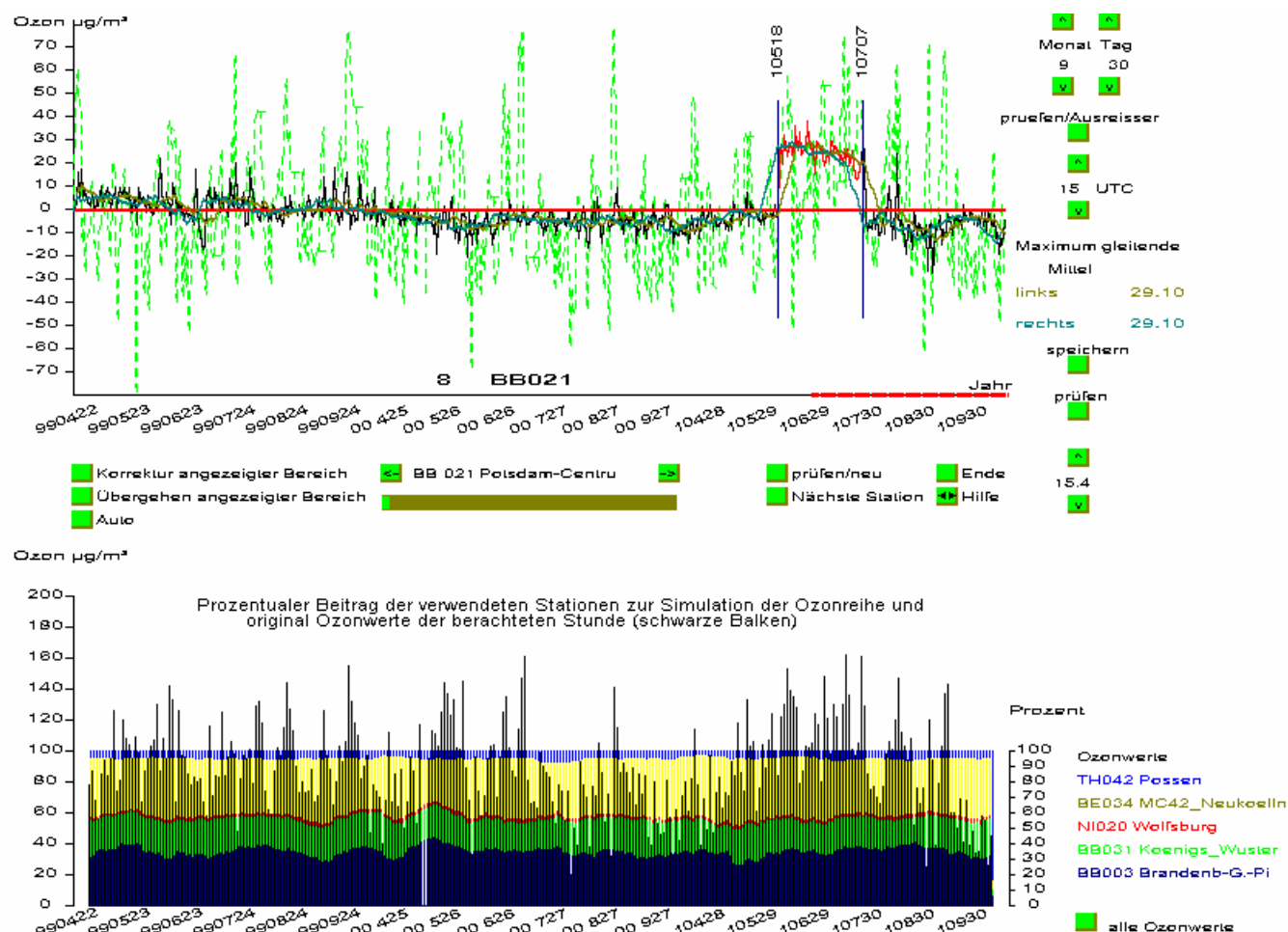
## 4 Interaktive Datenvalidierung

Zusätzlich zur Version für den vollautomatischen Lauf der Validierung wurde ein Modul entwickelt, dass eine interaktive Arbeitsweise erlaubt. Dabei ist vom UBA vorgesehen, dieses interaktive Datenvalidierungsmodul den Messnetzzentralen der Bundesländer zu übergeben, um bereits dort Monats- und Jahresdatensätze einer komplexen Wertung unterziehen zu können. Im UBA selbst kann dieses Modul zusätzlich auch zur Prüfung der täglich einlaufenden Daten genutzt werden, da in den, dem Programm hinterlegten Zykluspeicher zunächst die täglich eingehenden Datensätze eingelesen werden, die dann zu einem späteren Zeitpunkt mit den von den Ländern bereitgestellten Monatsdatensätzen und einmal jährlich mit den Jahresdatensätzen überschrieben werden.

Diese Datenvalidierung stützt sich auf länderübergreifende statistische Relationen zwischen den Ozonmessreihen. Ausreißer und mögliche systematische Fehler werden dokumentiert und können je nach Aufgabenstellung bei der Weiterverarbeitung (Präsentation oder Berechnung aggregierter Größen) eingearbeitet werden. Dabei bleiben die Originalreihen erhalten. Die interaktive Datenvalidierung besteht zum einen aus dem Modul zur Identifikation möglicher systematischer Fehler (Kap. 4.1) und zum anderen aus einem Modul zur Auffindung nicht plausibler Werte (Kap. 2).

#### 4.1 Interaktive Validierung systematischer Fehler

Im Gegensatz zur realtime-Erkennung systematischer Fehler (Kap. 3), deren Ziel die möglichst frühzeitige Warnung der Messnetze ist, kommt es bei der interaktiven Validierung zu einer Beurteilung der gesamten Messreihe. Schwerpunktmäßig, d.h. mit allen Interaktionsmöglichkeiten, wird jedoch nur das letzte Jahr geprüft. Das dazu eingesetzte Validierungsmodul beinhaltet sowohl eine vollautomatische Kontrolle und Korrektur der Datenreihen (Kap. 5) als auch die Möglichkeit einer interaktiven Kontrolle der Einzelreihen



**Abb. 4.1:** Grundbild zur Erläuterung der Funktionalität der interaktiven Kontrolle systematischer Fehler an Hand der Station Potsdam Zentrum 15 UTC. Zwischen dem 18.05.2001 und dem 07.07.2001 wurde zu der Ozonmessreihe eine systematische Abweichung von  $30 \mu\text{g}/\text{m}^3$  aufgeprägt.

Zur Erläuterung des Bildschirminhaltes im Grundzustand und zur Erläuterung der Interaktionsmöglichkeiten am Bildschirm, wurde der ursprünglich fehlerlosen Reihe Potsdam-Zentrum zwischen dem 18.05.2001 und dem 07.07.2001 eine systematische Abweichung von  $30 \mu\text{g}/\text{m}^3$  aufgeprägt. (Abb. 4.1)

### 4.1.1 Beschreibung des Bildschirminhaltes

Der gesamte Bildschirminhalt besteht aus zwei Teilbereichen: Der obere Teil dient zur Präsentation und zur Erläuterung der detektierten systematischen Fehler (Kap. 4.1.1.1). Der untere Teil der Abbildung dient zur näheren Erläuterung der Hintergründe, die zu einer systematischen Abweichung dieser Zeitreihe geführt haben könnten, und ermöglicht es, weitgehend auszuschließen, dass der dargestellte systematische Fehler von einer benachbarten Station verursacht wurde (Kap. 4.1.1.2)

#### 4.1.1.1 Beschreibung des Bildschirminhaltes oberer Teil.

Die Y-Achse zeigt die um den stationsspezifischen mittleren Jahresgang variierenden Ozonwerte zwischen  $-70 \mu\text{g}/\text{m}^3$  und  $+70 \mu\text{g}/\text{m}^3$ .

Auf der Zeitachse (X-Achse) sind Jahr, Monat und Tag aufgetragen. Der letzte rechte Wert markiert den letzten Termin, der in die Validierung einbezogen wurde. In Abb. 4.1 handelt es sich um den 30.09.2001. Das Datum des zuletzt geprüften Tages ist ebenso rechts oben unterhalb der Ausschrift „Monat Tag“ zu finden.

Die grün gestrichelte Kurve zeigt die fünfständigen Mittel der Ozonkonzentration als Abweichung vom stationsspezifischen mittleren Jahresgang. Die Kennzeichnung der untersuchten Stunde (fünfständiges Mittel) ist auf der rechten Seite der Abbildung zu finden (in diesem Fall 15 UTC).

Die schwarze Kurve zeigt die Differenz zwischen den beobachteten Ozonwerten (fünfständiges Mittel) und der unter Verwendung umliegender Stationen simulierten Ozonreihe.

Die zur multiplen Regression verwendeten Regressionsbeziehungen wurden anhand des Datensatzes der Jahre 1999, 2000 und 2001 mittels eines erwartungstreuen Screening-Regressionsansatzes abgeleitet (Enke 1988). Damit wird deutlich, dass nur solche Stationen in die Validierung einbezogen werden können, die einen Messzeitraum von mindestens 3 Jahren abdecken.

Die ocker-farbige Kurve stellt das von links kommende gleitende Mittel (t-30Tage) der Differenzen zwischen beobachteter und simulierter Zeitreihe (schwarze Kurve) dar.

Die türkis-farbige Kurve stellt das von rechts kommende gleitende Mittel (t+30Tage) der Differenzen zwischen beobachteter und simulierter Zeitreihe (schwarze Kurve) dar.

Überschreitet das von rechts kommende gleitende Mittel (türkis-farbige Kurve) bzw. das von links kommende gleitende Mittel (ocker-farbige Kurve) einen stationsspezifischen Schwellenwert, so wird ein systematischer Fehler detektiert. Die Maxima der gleitenden Mittelkurven sind rechts neben der Grafik in entsprechender Farbgebung zu finden. Die stationsspezifische Prüfschwelle ist unmittelbar darunter zwischen den vertikal angeordneten Pfeiltasten angegeben. Dieser kann subjektiv verändert werden, indem mit der rechten Maustaste ein Klick auf die darüber bzw. darunter liegenden Pfeiltasten vorgenommen wird.

Zur Abgrenzung des Bereiches eines möglichen systematischen Fehlers wird der Schnittpunkt der gleitenden Mittel zwischen den Maxima beider Kurven bestimmt. Rechts bzw. links von diesem Schnittpunkt markiert der maximale Betrag der Differenz zwischen rechts- und links gleitendem Mittel den Beginn bzw. das Ende des beanstandeten Intervalls. Die blauen senkrechten Linien markieren den Beginn und das Ende eines möglichen systematischen Fehlers. Das entsprechende Datum ist oberhalb der blauen Linien zu finden. Am konkreten Beispiel in Abb. 4.1 wird zwischen dem 18.05.2001 und dem 07.07.2001 ein systematischer Fehler angezeigt (in diesem Zeitintervall wurden künstlich  $30 \mu\text{g}/\text{m}^3$  addiert.). Die im

Normalfall schwarze Kurve der Differenzen zwischen beobachteter und simulierter Zeitreihe wird in diesem Intervall hellrot eingefärbt.

Ist ein Abschnitt auf der X-Achse rot markiert, so bedeutet dies, dass für eine Station, die zur Simulation verwendet wurde, keine Messwerte vorlagen und somit eine Ersatzstation verwendet werden musste. In den meisten Fällen ist dies ohne Bedeutung. Wird aber gerade für diesen Zeitraum ein systematischer Fehler angezeigt, sollte man sich diese Station näher ansehen.

Durch einen Klick mit der linken Maustaste neben eine der blauen Linien kann der beanstandete Bereich manuell geändert werden. Dies ist von Bedeutung, wenn eine Korrektur der Reihe vorgenommen wird.

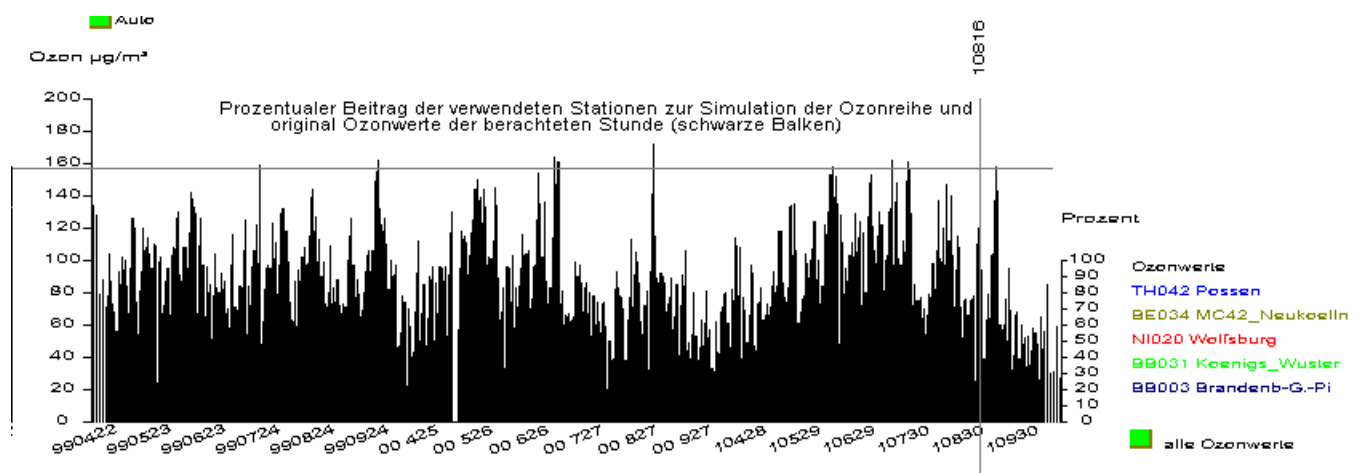
Die Kurzbezeichnung der bearbeiteten Station ist unten in der Mitte der Grafik zu finden, der Name der Station befindet sich darunter, oberhalb des ockerfarbigen Balkens.

#### 4.1.1.2 Beschreibung des Bildschirminhaltes unterer Teil

Die Achsenbeschriftung des unteren Teiles der Bildschirmgrafik zur Validierung systematischer Fehler ist vergleichbar mit der des oberen Bildausschnittes. Zur Darstellung der täglichen Ozonwerte (schwarze Balken) der oben angezeigten Stunde ist auch die Y-Achse entsprechend skaliert.

Durch einen Klick mit der linken Maustaste auf den Button „alle Ozonwerte“ rechts unten, werden alle Werte der untersuchten 15 UTC-Reihe dargestellt (Abb. 4.2). Durch Betätigen der Button „prüfen“ oder „prüfen/neu“ wird der Ausgangszustand wieder hergestellt.

Durch einen Klick auf den unteren Teil der Abbildung erscheint ein Fadenkreuz, mit dessen Hilfe sowohl die Höhe des Ozonwertes an der Y-Achse abgelesen werden kann als auch das entsprechende Datum, das an der senkrechten Linie eingeblendet wird. Am Beispiel in Abb. 4.2 wird für den 16.08.2001 ein Ozonwert von ca. 158  $\mu\text{g}/\text{m}^3$  abgelesen.



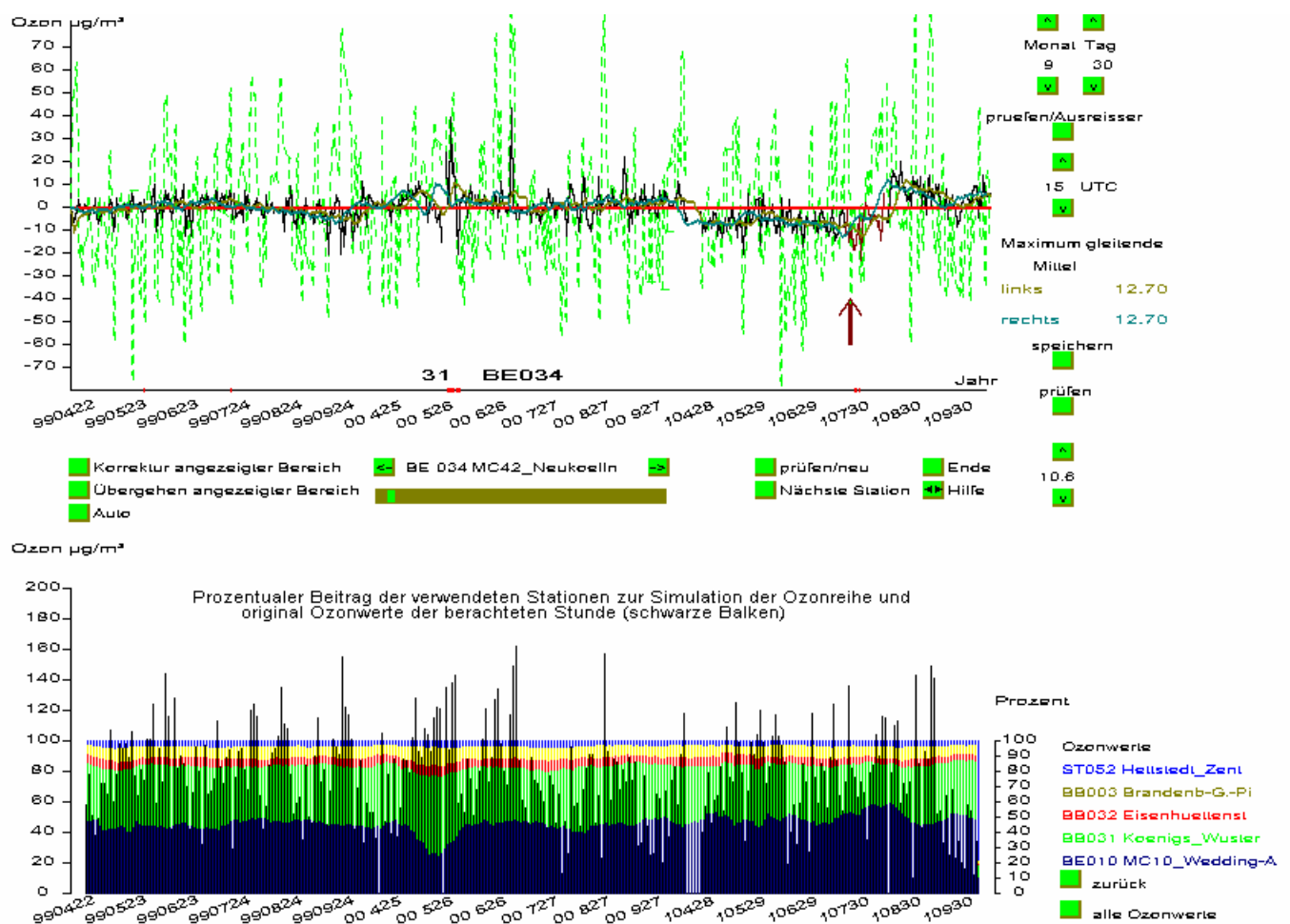
**Abb. 4.2: Grundbild zur Erläuterung der Funktionalität der interaktiven Kontrolle systematischer Fehler an Hand der Station Potsdam Zentrum 15 UTC. (unterer Teil der Bildschirmgrafik, Darstellung aller täglichen Ozonwerte 15 UTC )**

Die farbig hinterlegten Flächen im unteren Teil der Abb. 4.1, geben den prozentualen Beitrag der zur Simulation verwendeten Ozonmessreihen an. Die Differenz zwischen simulierter und gemessener Reihe ist als schwarze Linie in Abb. 4.1 oben dargestellt. Wie schon weiter oben ausgeführt, bildet diese Kurve die Basis zur Erkennung der systematischen Fehler der untersuchten Zeitreihe. Systematische Fehler in den zur Simulation verwendeten Stationen



können auch Auswirkungen auf die zu untersuchende Station haben, wie später noch gezeigt wird. Die rechts neben der Grafik aufgeführten Ozonmessreihen entsprechen in ihrer Farbgebung dem prozentualen Beitrag, den diese Station zum Gesamtergebnis leistet. Die prozentualen Beiträge unterliegen leichten Schwankungen von Tag zu Tag. Eine größere Änderung der Abhängigkeiten von einzelnen Stationen deutet auf Unstimmigkeiten in den verwendeten Zeitreihen hin.

Es besteht die Möglichkeit, durch einen Klick mit der linken Maustaste auf eine der aufgeführten Ozonmessstationen diese Reihe auf den Bildschirm zu präsentieren. Im vorliegenden Beispiel liefern die Stationen BE034 - MC42\_Neukölln und BB003 - Brandenburg-G.Pi mit ca. je 30 % die größten Anteile zur simulierten Reihe. Um sicherzustellen, dass der angezeigte systematische Fehler in Abb. 4.1 nicht durch eine andere Station verursacht wurde, kann hier beispielsweise die Station BE034 - MC42\_Neukölln aufgeblendet werden. (Abb. 4.3)



**Abb. 4.3: Grundbild zur Erläuterung der Funktionalität der interaktiven Kontrolle systematischer Fehler an Hand der Station Potsdam Zentrum 15 UTC ( zur Simulation der Station Potsdam-Zentrum verwendete Ozonmessreihe Berlin-Neukölln).**

Für den beanstandeten Zeitraum (18.05.2001 bis 07.07.2001) ist in der Ozonreihe Berlin-Neukölln kein Anzeichen für einen systematischen Fehler zu finden, der einen nennenswerten Einfluss auf die untersuchte Station Potsdam-Zentrum haben könnte. Auch bei den anderen zur Simulation verwendeten Stationen findet sich für diesen Zeitraum kein systematischer



Fehler. Damit erhärtet sich der "Verdacht", dass der in Abb. 4.1 angezeigte systematische Fehler in der Station selbst zu finden ist. (Der Station Potsdam-Zentrum wurde eine systematische Abweichung künstlich aufgeprägt.)

Mit dem Button „zurück“ springt das Programm zurück zur untersuchten Ausgangsstation.

## **4.1.2 Bedeutung und Funktionsweise der Button**

### **4.1.2.1 Oberer Teil der interaktiven Grafik**

#### **Hilfe**

Es wird eine kurze Beschreibung der Funktion und Interaktionsfähigkeit der Grafikoberfläche erläutert. Durch Drücken der linken Maustaste auf den rechten bzw. linken Pfeil im Hilfe-Button ist ein Blättern im Text möglich.

#### **Ende**

Das Programm wird verlassen.

#### **Ockerfarbiger Balken unter der Stationsbezeichnung**

Durch Drücken der linken Maustaste auf den ockerfarbigen Balken erfolgt eine Grobeinstellung der Stationsauswahl. Eine Feineinstellung der gewünschten Station kann mit den darüber liegenden rechten und linken Pfeiltasten erfolgen.

Die Prüfung der Station wird mit dem Button „nächste Station“ oder „prüfen/neu“ aktiviert. Der Button „nächste Station“ bewirkt eine schnellere Reaktion als der Button „prüfen/neu“. Letzterer setzt jedoch die dahinter liegende Datenbasis wieder auf den Ausgangszustand. (Der Button „nächste Station“ vererbt Eigenschaften vorangegangener Aktionen, z.B. geändertes Datum/Uhrzeit).

#### **Nächste Station**

Es wird die im Speicher folgende Station bearbeitet.

Wurde vorher auf dem ockerfarbigen Balken eine Grobeinstellung der Stationsauswahl vorgenommen oder mittels der darüber liegenden Pfeiltasten eine Feineinstellung vorgenommen, so aktiviert der Button „nächste Station“ die über dem ockerfarbigen Balken angezeigte Station.

#### **Prüfen/neu**

Durch Drücken dieses Buttons mit der linken Maustaste wird die Prüfung der in der Mitte zwischen den Pfeiltasten angezeigten Station vorgenommen. Es erfolgt ein Rücksetzen der Datenbasis, so dass alle vorhergehenden Änderungen wieder rückgängig gemacht werden.

#### **Datum einstellen**

In der Grafik rechts oben lassen sich mit Hilfe der Pfeiltasten Monat und Tag einstellen, bis zu welchem Datum die Prüfung erfolgen soll. Danach ist der Button "prüfen/Ausreisser" unterhalb der Datumseinstellung zu drücken, um die Änderung wirksam werden zu lassen. Das Datum wird auch am rechten Ende der X-Achse angezeigt.

Das Ausgangsdatum lässt sich mit dem Button „prüfen/neu“ wieder herstellen.

#### **Uhrzeit einstellen**

In der Grafik rechts oben lässt sich mit Hilfe der Pfeiltasten die Uhrzeit einstellen, für die die Validierung durchgeführt werden soll. Danach ist der Button "prüfen/Ausreisser" oberhalb der Uhrzeit Einstellung zu drücken, um die Änderung wirksam werden zu lassen.

Im Falle einer automatischen Abarbeitung werden alle Stunden durchgängig getestet und selbstständig korrigiert (siehe Button „auto“).

### Subjektive Wahl der Schwelle für die zu prüfende Station

Im oberen Teil des Bildes und dort unten rechts ist die aktuelle stationsspezifische Prüfschwelle für das Überschreiten der Maxima der links bzw. rechts gleitenden Mittel angegeben. Bei Über- oder Unterschreiten der gleitenden Mittel (ocker- bzw. türkis-farbige Kurve) wird der untersuchte Zeitraum als möglicherweise fehlerhaft eingestuft.

Diese Schwelle kann manuell durch Betätigung der Pfeiltasten modifiziert werden.

Danach ist der darüber liegende Button "prüfen" zu drücken, um die Änderung wirksam werden zu lassen.

Das aktuell berechnete Maximum des rechts bzw. links gleitenden Mittels ist unterhalb der Textausgabe „Maximum gleitendes Mittel“ angezeigt.

### Korrektur angezeigter Bereich

Es wird der systematische Fehler aus der Reihe im angezeigten Intervall eliminiert.

Abb. 4.4 zeigt die korrigierte Zeitreihe der Station Potsdam-Zentrum.

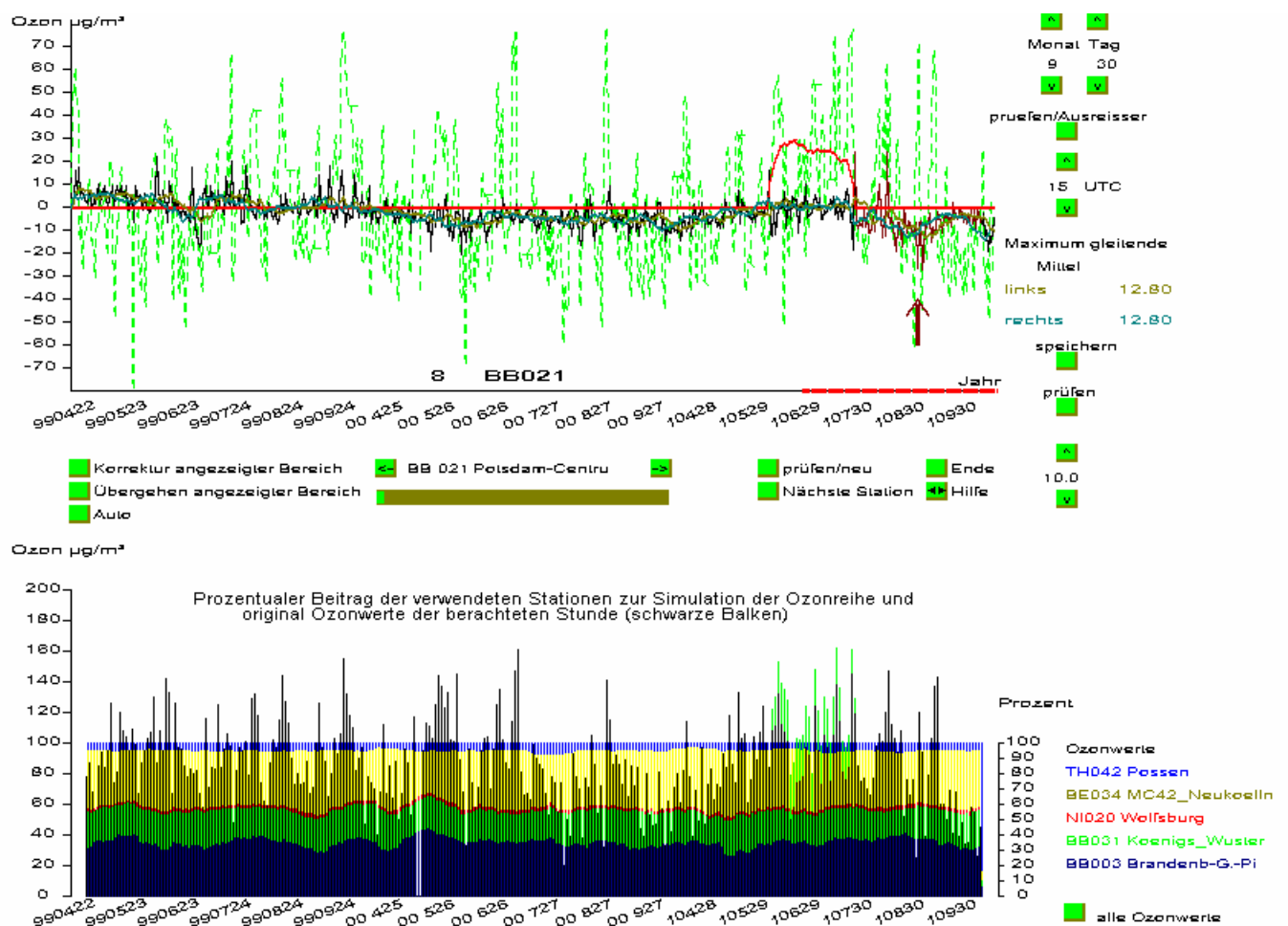
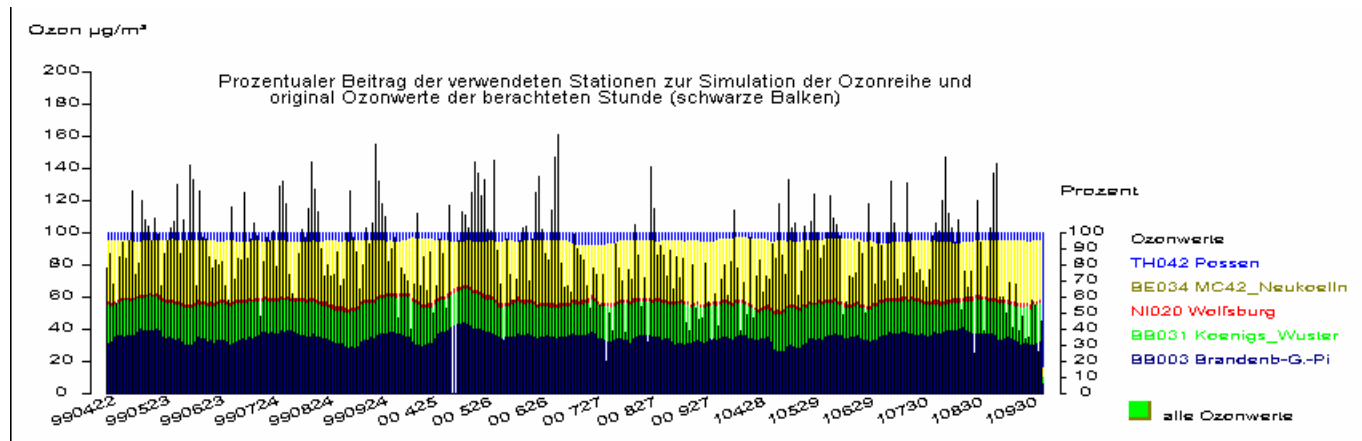


Abb. 4.4: Grundbild zur Erläuterung der Funktionalität der interaktiven Kontrolle systematischer Fehler an Hand der Station Potsdam Zentrum 15 UTC. (Korrektur des systematischen Fehlers)

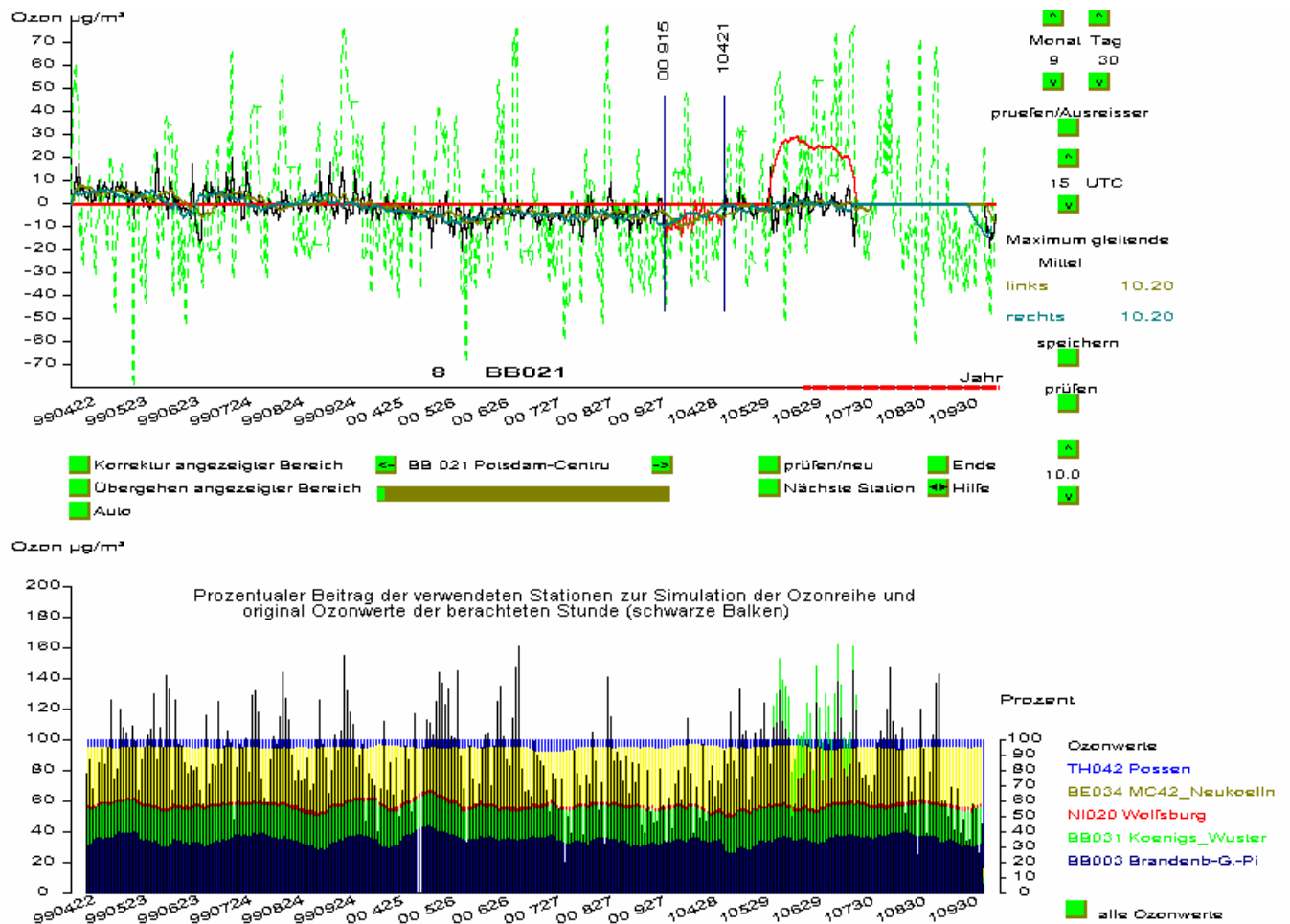


**Abb. 4.5: Grundbild zur Erläuterung der Funktionalität der interaktiven Kontrolle systematischer Fehler an Hand der Station Potsdam Zentrum 15 UTC. (Originalreihe)**

Die an der Zeitreihe durchgeführte Änderung ist im oberen Teil der Abb. 4.4 durch eine rote Kurve dokumentiert. Im unteren Teil der Abbildung werden die Änderungen an den realen Ozonwerten grün markiert, wenn es sich um eine Verminderung der Ozonwerte handelt und rot, wenn eine Erhöhung der Ozonwerte vorgenommen wurde. Vergleicht man die Originalreihe der Station Potsdam-Zentrum, d.h. ohne künstlich aufgeprägten systematischen Fehler (Abb. 4.5) mit der durch das Validierungsmodul wieder korrigierten Ozonmessreihe (Abb. 4.3), so stellt man fest, dass durch die Korrektur die Ausgangsreihe nahezu identisch wieder hergestellt wurde.

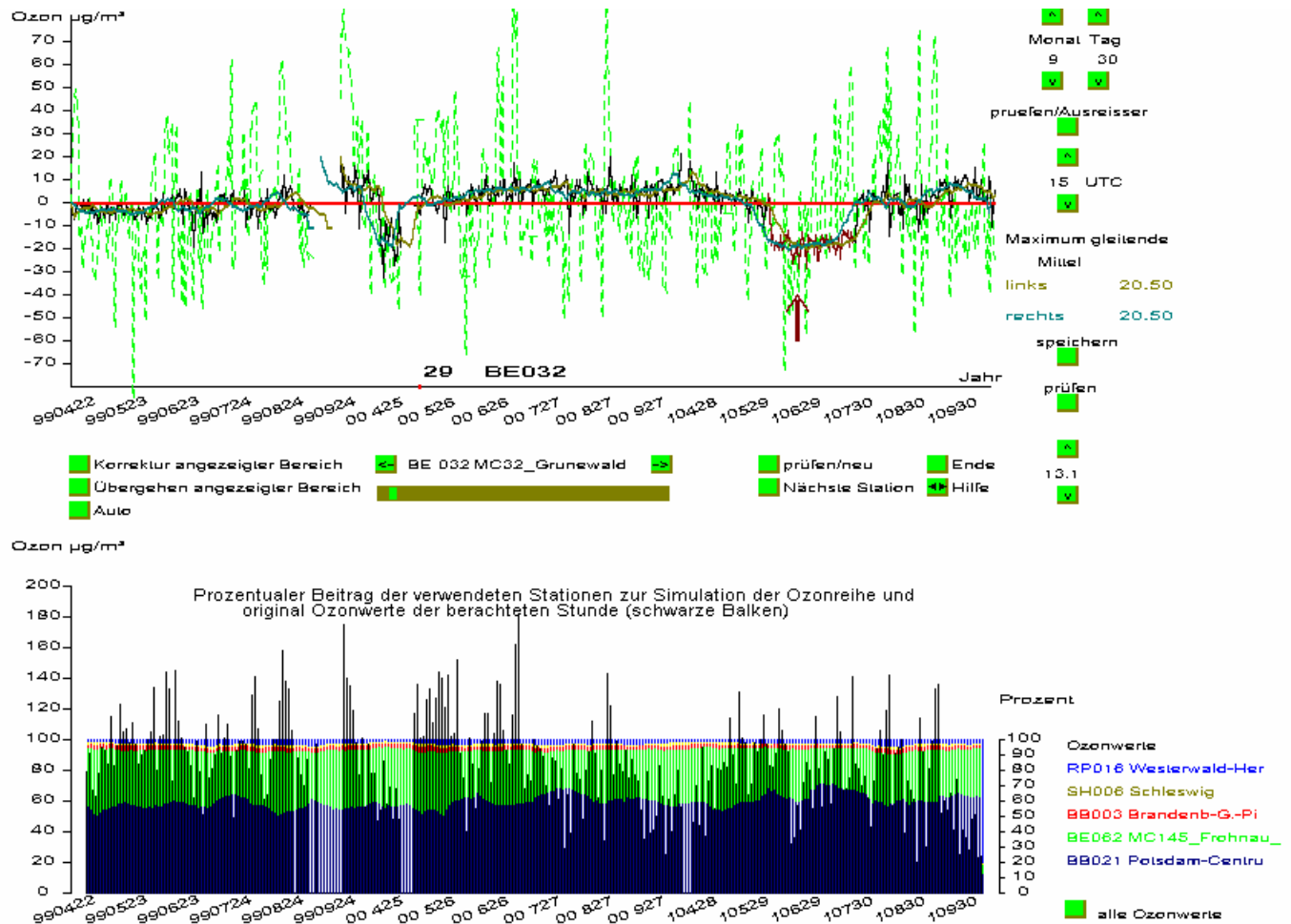
Treten weitere auffällige Episoden auf, so wird durch eine erneute Betätigung des Button „Korrektur angezeigter Bereich“ auch dieser Fehler eliminiert.

Wird ein möglicher systematischer Fehler durch eine Station verursacht, die zur Simulation verwendet wurde, so wird der beanstandete Bereich braun eingefärbt und mit einem braunen Pfeil markiert (Abb. 4.4). Durch Aktivieren der Button „Korrektur angezeigter Bereich“ oder „Übergehen angezeigter Bereich“ wird dieser Bereich von der weiteren Prüfung ausgeschlossen. Das nächste beanstandete Intervall wird angezeigt (Abb. 4.6)



**Abb. 4.6: Grundbild zur Erläuterung der Funktionalität der interaktiven Kontrolle systematischer Fehler an Hand der Station Potsdam Zentrum 15 UTC. (Ausschließen des braun markierten Bereiches von der weiteren Prüfung)**

Ein markantes Beispiel, wie sich der an der Station Potsdam-Zentrum künstlich aufgeprägte Fehler auf die Station BE032 - Berlin-Grunewald auswirkt, ist in Abb. 4.7 zu finden.



**Abb. 4.7: Grundbild zur Erläuterung der Funktionalität der interaktiven Kontrolle systematischer Fehler an Hand der Station Potsdam Zentrum 15 UTC. (Auswirkung eines künstlich aufgeprägten Fehlers der Station Potsdam-Zentrum auf die Station BE032 - Berlin-Grünwald)**

Zur Simulation der Zeitreihe Berlin-Grünwald wurde die Messreihe Potsdam-Zentrum verwendet (Abb. 4.7 unteres Bild) und diese liefert mit ca. 60 % den größten Anteil im Vergleich zu den anderen zur Simulation herangezogenen Stationen. Da aufgrund des systematischen Fehlers von  $30 \mu\text{g}/\text{m}^3$  bei der Messreihe Potsdam-Zentrum die simulierte Reihe für Berlin-Grünwald für dieses Zeitintervall um ca.  $20 \mu\text{g}/\text{m}^3$  (60 % von  $30 \mu\text{g}/\text{m}^3$ ) zu hoch ausfällt, erscheint der systematische Fehler als Abweichung der simulierten Reihe zur Messreihe Berlin-Grünwald als negative Abweichung (Abb. 4.6).

Vom Validierungsmodul wird jedoch erkannt, dass es sich hier um den Einfluss einer benachbarten Station handeln muss, so dass das Zeitintervall entsprechend braun mit einem Pfeil markiert wird. Zu diesem Zweck wird die Simulation der Zeitreihe einmal mit Verwendung der am stärksten korrelieren Reihe durchgeführt und ein weiteres mal unter Ausschluss dieser Reihe. Wird der systematische Fehler bei beiden Simulationen beobachtet, so wird geschlussfolgert, dass der systematische Fehler der untersuchten Zeitreihe zuzuordnen ist. Im vorliegenden Fall zeigte sich bei Ausblendung der Station Potsdam-Zentrum zur Simulation bei der untersuchten Station Berlin-Grünwald kein systematischer Fehler mehr. Der Bereich wurde braun markiert.

## Übergehen angezeigter Bereich

Das nächste, möglicherweise mit einem systematischen Fehler behaftete Intervall in Abb. 4.6 (15.09.2000 bis 21.04.2001) liegt nur knapp über der Nachweisgrenze. Das Maximum der gleitenden Mittel rechts und links (vgl. Kap. 4.1.1.1) beträgt  $10.2 \mu\text{g}/\text{m}^3$  und der Schwellenwert wurde mit  $10.0 \mu\text{g}/\text{m}^3$  angegeben (Abb. 4.6 oben rechte Seite). Bei subjektiver Prüfung würde man dieses Intervall übergehen. Diese Möglichkeit ist durch die Aktivierung des Buttons „Übergehen angezeigter Bereich“ möglich. Es wird der systematische Fehler im angezeigten Intervall unverändert gelassen und das nächste Intervall auf systematische Fehler hin untersucht (vergleiche Abb. 4.6 und 4.8).

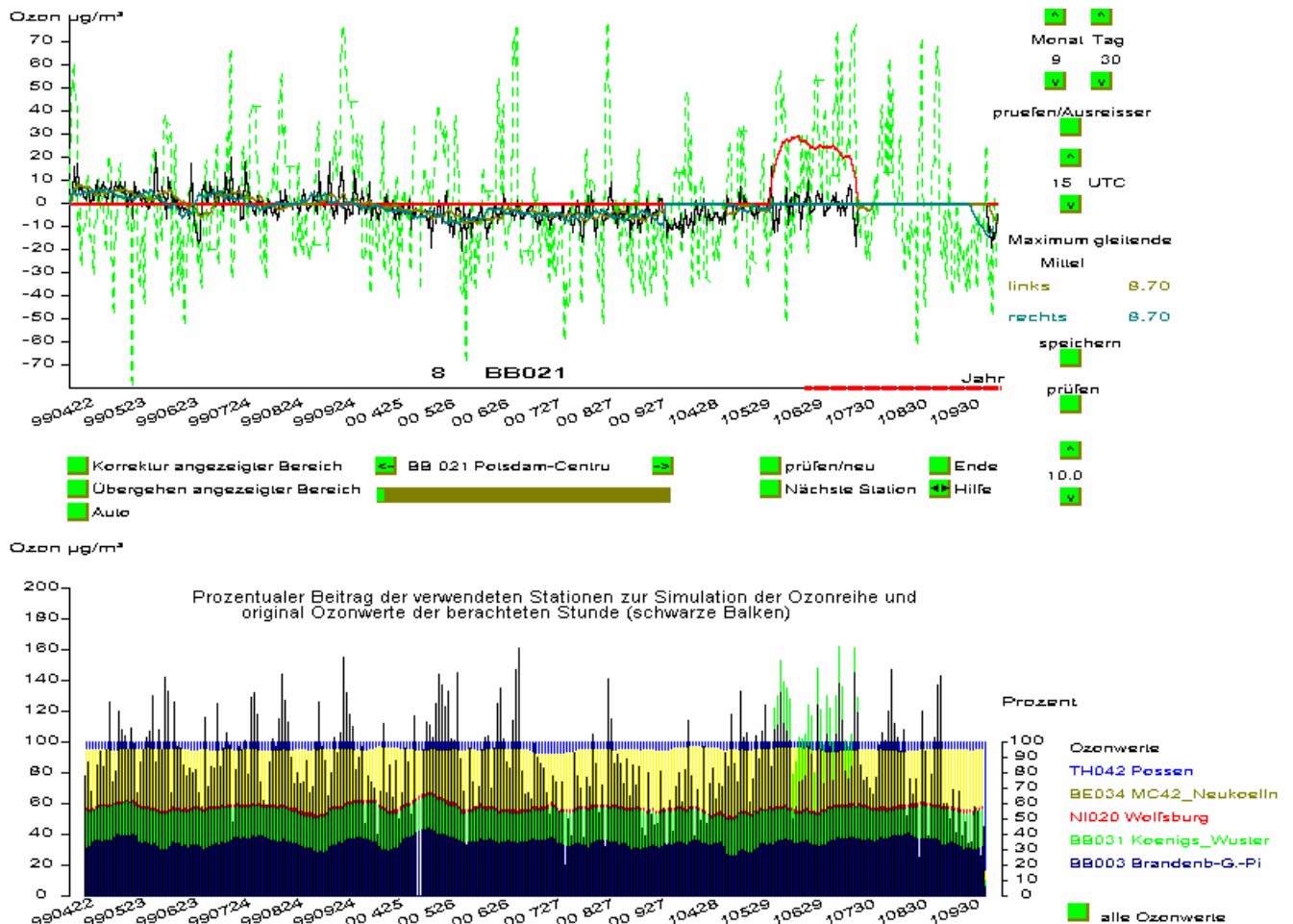
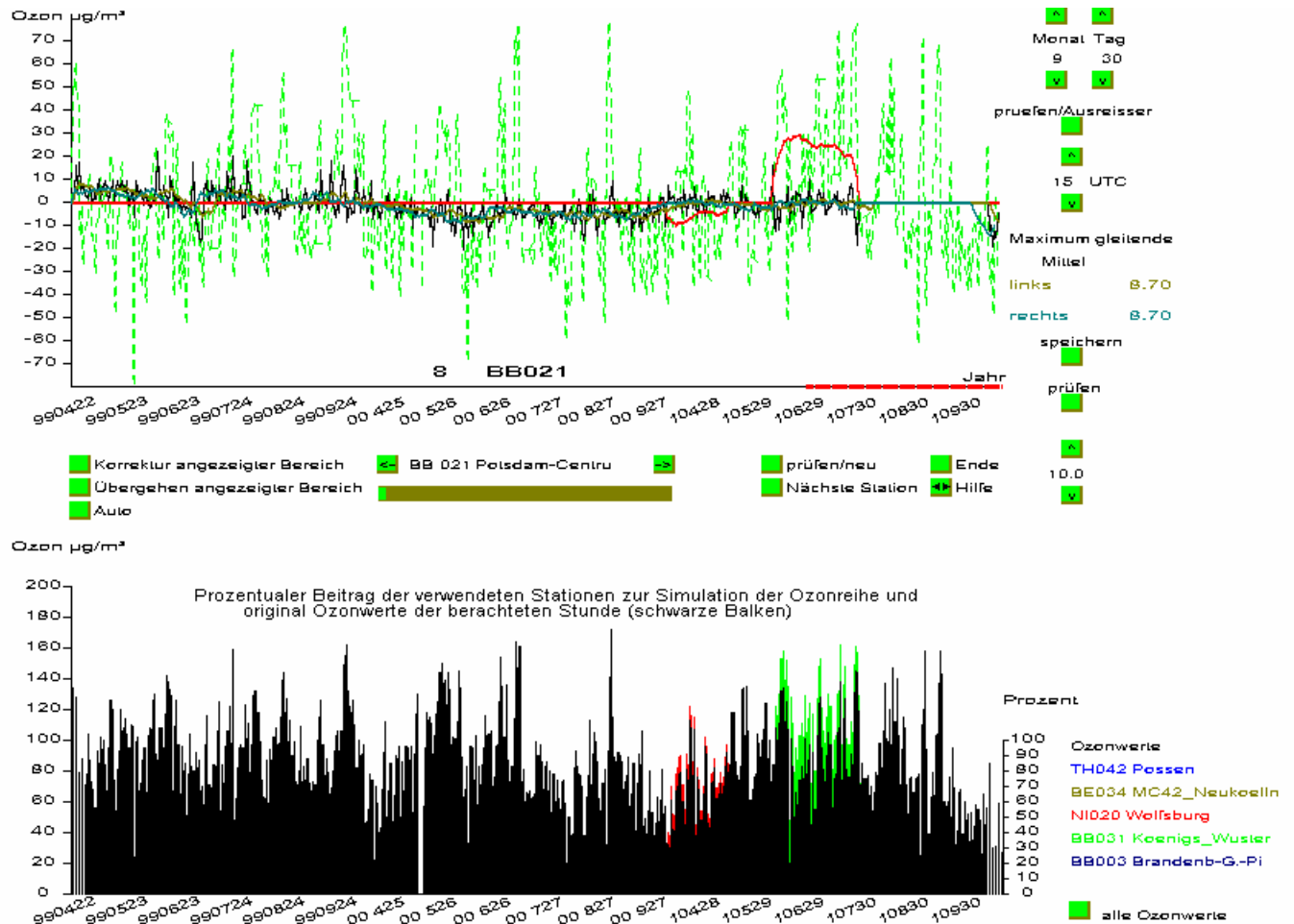


Abb. 4.8: Grundbild zur Erläuterung der Funktionalität der interaktiven Kontrolle systematischer Fehler an Hand der Station Potsdam-Zentrum 15 UTC. (Übergehen des in Abb. 4.6 rot markierten Zeitintervalls)

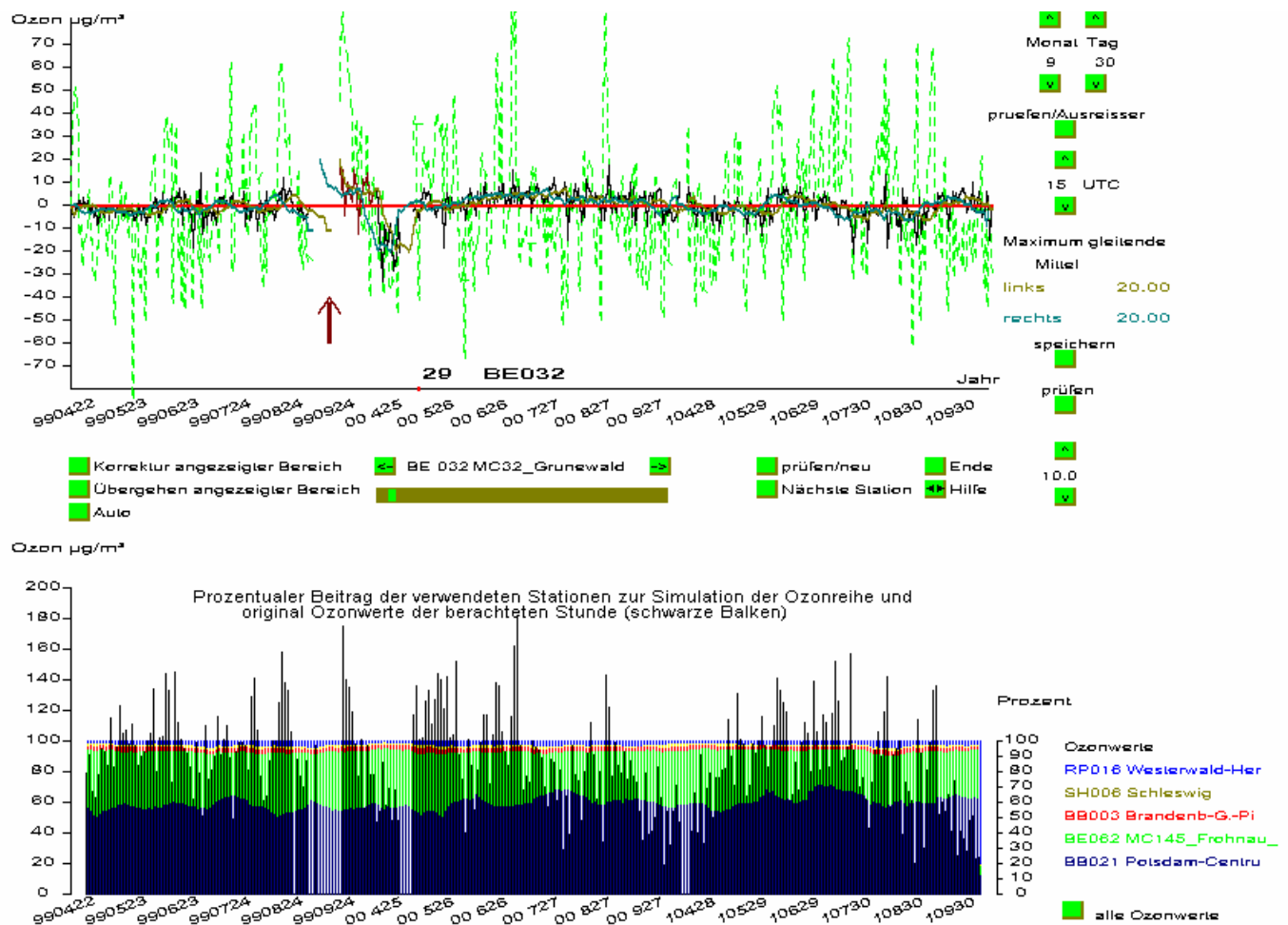
## Auto

Durch Betätigen dieses Buttons werden alle systematischen Fehler der untersuchten Reihe für die angezeigte Uhrzeit nacheinander selbstständig bereinigt. Die durchgeführten Änderungen werden in Abb. 4.9 im unteren Teil als rote bzw. grüne Markierungen sichtbar.



**Abb. 4.9: Grundbild zur Erläuterung der Funktionalität der interaktiven Kontrolle systematischer Fehler an Hand der Station Potsdam-Zentrum 15 UTC. (Automatische Korrektur der systematischen Fehler)**

Prüft man die Station BE032 - Berlin-Grunewald nach durchgeführter Korrektur erneut, so zeigt sich eindeutig, dass kein systematischer Fehler mehr im vorher beanstandeten Intervall zu beobachten ist. (Abb. 4.10 im Vergleich mit Abb. 4.7). Dadurch wird deutlich, dass der in Abb. 4.7 braun markierte Abschnitt auf den systematischen Fehler in der Station Potsdam-Zentrum zurückzuführen ist.



**Abb. 4.10: Grundbild zur Erläuterung der Funktionalität der interaktiven Kontrolle systematischer Fehler an Hand der Station Potsdam Zentrum 15 UTC. (Prüfung der Station Berlin-Grünwald nach Beseitigung des systematischen Fehlers der Station Potsdam-Zentrum)**

### Speichern

Die vorgenommenen Änderungen werden gespeichert, sobald danach der Button "Nächste Station" oder "Ende" aktiviert wird. Hierzu validiert das Programm automatisch alle Stunden und speichert die durchgeführten Änderungen unter

c:\routwarn\dat\systfehlval1.som\sF-Stationsnummer ab. Dabei müssen folgende zusätzliche Bedingungen erfüllt sein, damit ein systematischer Fehler der untersuchten Station gespeichert wird:

1. Der systematische Fehler muss an mindestens 5 aufeinander folgenden Stunden aufgetreten sein
2. Der Betrag der Änderung soll mindestens für eine Stunde  $5 \mu\text{g}/\text{m}^3$  überschritten haben
3. Es wird eine 3-stündige gleitende Glättung der systematischen Fehler vorgenommen

Tabelle 1 zeigt einen Ausschnitt aus der Tabelle der gespeicherten systematischen Fehler für die Station BB021 – Potsdam-Zentrum. Der aufgeprägte Fehler zwischen dem 18.05.2001 und dem 07.07.2001 wurde gut erkannt. Kleinere systematische Abweichungen wurden noch im April 2001 gefunden.



Datum	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Uhr
930	-3	-4	-5	-5	-5	-5	-6	-7	-7	-7	-7	-7	-7	-7	-7	-5	-6	-7	-7	-6	-7	-7	-6	-7	-6
10401	-3	-4	-5	-5	-5	-5	-6	-7	-6	-7	-7	-7	-7	-7	-7	-5	-5	-5	-4	-4	-5	-6	-5	-6	-5
10402	-2	-3	-4	-4	-4	-4	-5	-6	-6	-6	-6	-6	-6	-6	-6	-4	-4	-5	-4	-3	-4	-5	-4	-5	-4
10403	-4	-4	-4	-5	-5	-5	-6	-6	-6	-6	-6	-6	-6	-6	-5	-6	-4	-4	-5	-4	-3	-4	-4	-4	-4
10404	-4	-4	-4	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-4	-3	-4	-4	-3	-4	-4	-4	-4	-4
10405	-4	-4	-4	-4	-4	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-4	-3	-3	-3	-4	-4	-4	-4	-4	-4
10406	-5	-5	-4	-5	-5	-4	-5	-5	-5	-5	-5	-5	-5	-5	-4	-4	-2	-2	-3	-5	-5	-5	-4	-5	-5
10407	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-6	-5	-5	-4	-3	-2	-1	-1	-3	-4	-5	-4	-4	-4
10408	-5	-5	-5	-5	-5	-4	-4	-4	-4	-5	-5	-6	-5	-4	-4	-2	0	0	-1	-1	-2	-2	-3	-2	-3
10409	-5	-4	-4	-4	-3	-3	-4	-4	-4	-5	-6	-6	-6	-5	-4	-3	-1	1	1	0	-1	-2	-2	-3	-3
10410	-4	-4	-4	-3	-3	-3	-3	-4	-4	-5	-6	-6	-6	-6	-4	-4	-2	0	0	-1	-2	-2	-3	-3	-3
10411	-5	-5	-4	-4	-4	-4	-4	-4	-4	-5	-6	-7	-7	-6	-4	-4	-2	-1	-1	-1	-2	-2	-3	-3	-3
10412	-6	-5	-5	-4	-4	-3	-3	-3	-3	-4	-5	-6	-5	-4	-4	-3	-1	-1	-1	-2	-2	-3	-3	-4	-4
10413	-6	-6	-5	-5	-4	-3	-3	-4	-4	-5	-6	-7	-7	-6	-5	-4	-4	-2	-1	-2	-2	-3	-3	-3	-3
10414	-5	-4	-4	-3	-3	-2	-2	-3	-3	-5	-6	-7	-7	-6	-5	-4	-4	-2	-2	-2	-2	-3	-3	-4	-4
10415	-4	-4	-4	-3	-3	-3	-3	-3	-4	-5	-6	-7	-7	-7	-6	-5	-6	-4	-3	-2	-2	-3	-3	-3	-3
10416	-4	-4	-4	-3	-2	-2	-3	-4	-4	-6	-7	-8	-8	-7	-6	-6	-6	-4	-3	-2	-2	-3	-2	-2	-2
10417	-2	-2	-2	-1	-1	-1	-2	-3	-4	-6	-7	-8	-8	-7	-6	-6	-7	-5	-3	-2	-2	-3	-2	-2	-2
10418	-2	-2	-2	-2	-1	-1	-2	-3	-4	-5	-6	-7	-7	-6	-5	-5	-7	-5	-3	-1	0	-1	-1	-1	-1
10419	-3	-3	-2	-1	-1	-1	-1	-2	-3	-5	-6	-7	-7	-6	-5	-5	-6	-4	-3	-2	-2	-2	-2	-3	-3
10420	-3	-3	-2	-1	-2	-2	-3	-5	-6	-7	-7	-6	-4	-4	-5	-4	-2	-2	-1	-1	-1	-2	-2	-2	-2
10421	-3	-3	-3	-2	-2	-1	-2	-3	-4	-5	-6	-7	-7	-6	-4	-4	-5	-4	-2	-1	-1	-1	-1	-1	-1
10422	-1	-1	-1	-1	0	-1	-1	-3	-4	-6	-6	-7	-6	-6	-4	-4	-5	-4	-2	-1	0	-1	-1	-1	-1
10423	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.....																									
10518	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10519	11	9	6	7	8	10	11	13	13	14	13	12	12	11	10	10	11	11	10	11	11	11	13	12	12
10520	8	9	10	10	11	12	14	16	16	16	15	15	14	14	13	13	13	13	13	14	15	15	16	15	15
10521	12	13	13	13	13	14	16	17	18	18	17	17	17	17	16	15	15	14	15	15	15	15	16	18	18
10522	17	17	17	16	16	17	18	19	20	20	19	19	19	19	18	18	17	17	17	17	17	18	20	20	20
10523	19	19	19	18	18	19	20	21	22	22	21	21	21	21	20	20	20	20	20	20	20	20	22	22	22
10524	22	22	22	21	21	21	22	24	25	26	25	25	25	24	24	23	23	23	23	23	23	25	25	25	25
10525	24	24	24	23	24	24	26	27	29	29	28	28	28	28	27	26	26	26	26	26	26	26	28	27	27
10526	22	23	24	23	23	24	25	26	28	28	27	27	27	27	27	26	27	27	27	27	27	27	29	27	27
10527	23	24	25	24	24	24	25	26	28	28	27	28	28	28	27	27	27	28	28	27	27	27	28	27	27
10528	24	25	26	25	25	25	26	27	29	29	28	28	28	28	27	27	27	28	28	28	29	29	30	30	30
10529	27	28	28	28	28	28	28	28	28	28	28	28	28	28	27	26	27	27	27	26	28	28	30	30	30
10530	31	31	30	30	30	30	30	30	30	30	29	28	28	28	28	27	27	27	27	26	27	28	30	30	30
10531	33	32	31	31	31	30	30	30	30	30	29	29	29	29	29	28	28	28	28	27	25	26	27	29	29
10601	31	31	30	30	30	29	29	29	30	30	29	29	29	29	29	28	28	28	29	28	26	27	28	29	29
10602	29	29	29	29	29	29	29	29	30	30	29	29	29	29	29	28	28	29	30	29	27	28	28	29	29
10603	29	29	29	29	30	30	30	30	30	30	31	30	30	30	30	29	29	29	29	29	27	28	28	30	30
10604	30	30	30	30	31	30	30	31	31	29	28	28	29	28	28	28	29	30	28	27	28	28	29	29	29
10605	30	29	29	30	30	30	29	28	28	26	25	26	26	26	27	28	29	30	29	28	28	28	29	29	29
10606	30	30	30	31	31	31	30	30	29	27	27	27	27	27	27	28	28	29	29	28	29	29	30	30	30
10607	31	31	30	31	32	31	30	29	28	26	26	26	26	26	27	27	28	29	29	29	29	29	30	31	31
10608	30	30	30	31	31	31	30	29	29	26	26	26	26	26	27	28	28	30	30	29	30	30	31	30	30
10609	26	27	28	29	30	29	28	28	29	27	27	27	26	26	27	28	28	31	31	30	31	31	32	30	30
10610	24	26	27	28	28	28	28	28	28	29	26	26	26	26	26	27	28	30	31	31	32	32	32	30	30
10611	24	26	27	28	28	28	28	27	27	28	26	26	26	25	25	25	25	26	29	30	31	33	33	33	32
10612	25	27	28	29	29	29	29	29	29	26	25	24	24	24	24	25	26	28	29	30	32	33	34	32	32
10613	24	26	28	28	28	28	28	27	28	25	24	23	23	23	24	25	25	27	28	30	32	32	34	32	32
10614	24	26	28	28	28	27	27	27	27	24	24	23	23	23	24	25	25	27	29	30	32	32	33	31	31
10615	23	25	27	27	27	27	27	27	27	25	24	24	23	23	24	25	25	27	29	30	32	33	34	32	32
10616	24	26	28	27	27	27	27	28	29	27	27	26	26	25	25	25	25	27	28	30	32	33	34	32	32
10617	24	26	28	27	27	27	27	28	28	26	26	25	25	25	25	25	25	27	28	29	31	32	33	31	31
10618	23	25	27	26	26	26	26	27	28	26	25	25	25	25	24	25	24	24	26	27	28	30	31	32	31
10619	23	25	27	26	26	26	26	27	27	25	25	25	24	24	24	24	24	25	26	27	30	32	33	31	31
10620	24	26	28	27	27	27	27	28	28	25	25	24	24	24	24	24	23	24	24	25	27	29	30	32	31
10621	26	27	28	28	27	28	28	28	28	26	25	24	24	24	24	23	23	24	24	25	27	29	30	29	29
10622	23	25	26	25	25	25	26	26	26	24	24	24	24	24	25	24	23	24	24	25	27	28	30	29	29
10623	23	25	26	25	25	25	25	25	25	25	25	24	24	24	25	25	24	23	22	22	23	25	26	28	28
10624	28	28	27	27	27	27	26	25	24	25	25	25	25	25	25	23	23	23	23	23	24	25	27	27	27
10625	29	28	28	27	26	26	26	25	25	24	24	24	24	24	25	24	24	23	22	23	23	24	25	26	27
10626	28	27	27	26	25	25	25	24	24	23	24	24	24	24	24	23	22	21	22	23	23	24	25	27	27
10627	28	27	27	26	26	25	24	24	23	23	24	24	24	24	23	22	21	22	22	22	23	23	25	25	25
10628	26	25	25	24	23	23	22	23	22	22	23	23	23	22	21	20	21	21	21	21	21	21	22	23	23
10629	24	23	22	22	21	21	21	22	22	22	22	22	22	21	21	20	1								

### **Prüfen/Ausreißer (ganz rechts oben)**

Dieser Button hat zwei Funktionen. Zum einen aktiviert er die eingestellten Änderungen des Datums und der Uhrzeit der darüber bzw. darunter liegenden Pfeiltasten, zum anderen wird er zur Aktivierung des Moduls zur interaktiven Prüfung von nicht plausiblen Einzelwerten oder Tagesgängen verwendet. Dazu wird durch einen Klick auf den unteren Teil der Abbildung das schon erwähnte Fadenkreuz zum Ablesen eines Stundenwertes (Abb. 4.2) aktiviert. Ist das Fadenkreuz aktiviert, so wird durch Drücken des Buttons „prüfen/Ausreißer“ die Ozonmessreihe für das am Fadenkreuz angezeigte Datum zur interaktiven Bearbeitung aufgeblendet.

### **4.1.2.2 Unterer Teil der interaktiven Grafik**

#### **alle Ozonwerte**

Im Grundzustand der interaktiven Grafik werden nur für jeden zweiten Tag die Ozonwerte der angezeigten Stunde dargestellt. Durch Betätigen des Buttons „alle Ozonwerte“, werden die Ozonwerte für jeden Tag dargestellt (Abb. 4.9 bzw. 4.12).

#### **Stationsbezeichnung rechts neben der Abbildung unterer Teil**

Durch einen Klick mit der rechten Maustaste auf eine der farbigen Stationsbezeichnungen wird das Grundbild der entsprechenden Station aufgeblendet. Einige Interaktionsmöglichkeiten sind für diese Präsentation gesperrt. Es erscheint unterhalb der farbigen Stationsbezeichnungen der Button „zurück“. Durch Betätigen dieses Buttons gelangt man zu der ursprünglich bearbeiteten Station zurück.

### **4.1.3 Typisches Beispiel eines systematischen Fehlers**

Abb. 4.11 zeigt die Zeitreihe der Ozonstation BE032 – Berlin-Grunewald 16 Uhr, bei der der bemängelte systematische Fehler in der Reihe selbst zu finden ist.

Die Station BE032 wurde aus einer Vielzahl von Beispielen systematischer Fehler eher zufällig ausgewählt. Sie demonstriert aber bei sonst ruhigem Verlauf der Kurve ein von Ausfällen begrenztes Intervall mit sprunghaften Veränderungen in der Größenordnung von +10 und -20  $\mu\text{g}/\text{m}^3$ . Weder bei den zur Simulation verwendeten Stationen waren systematische Fehler in diesem Zeitintervall erkennbar noch wurden Ersatzstationen (rote Markierung auf der x-Achse) zur Simulation herangezogen.

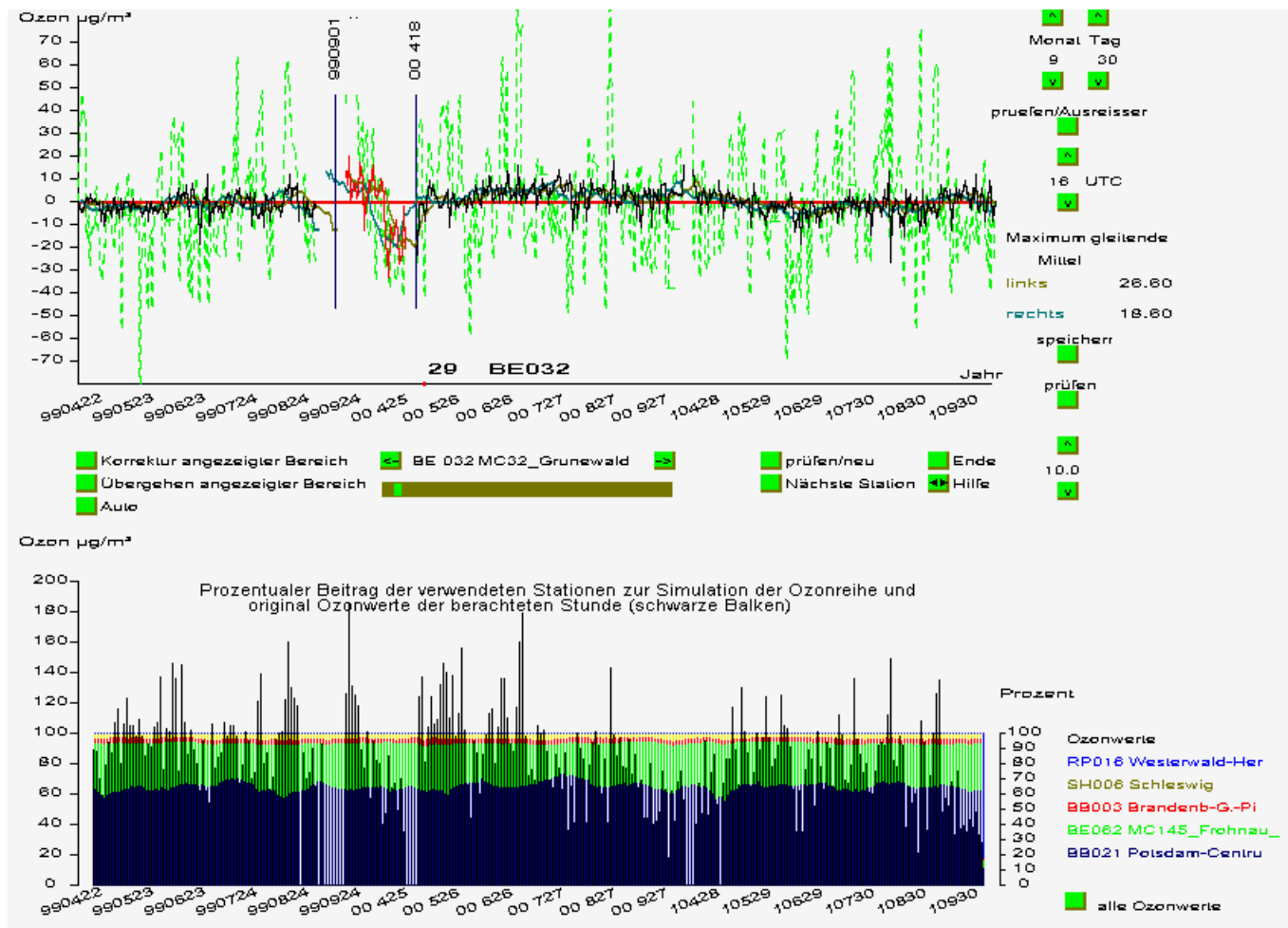


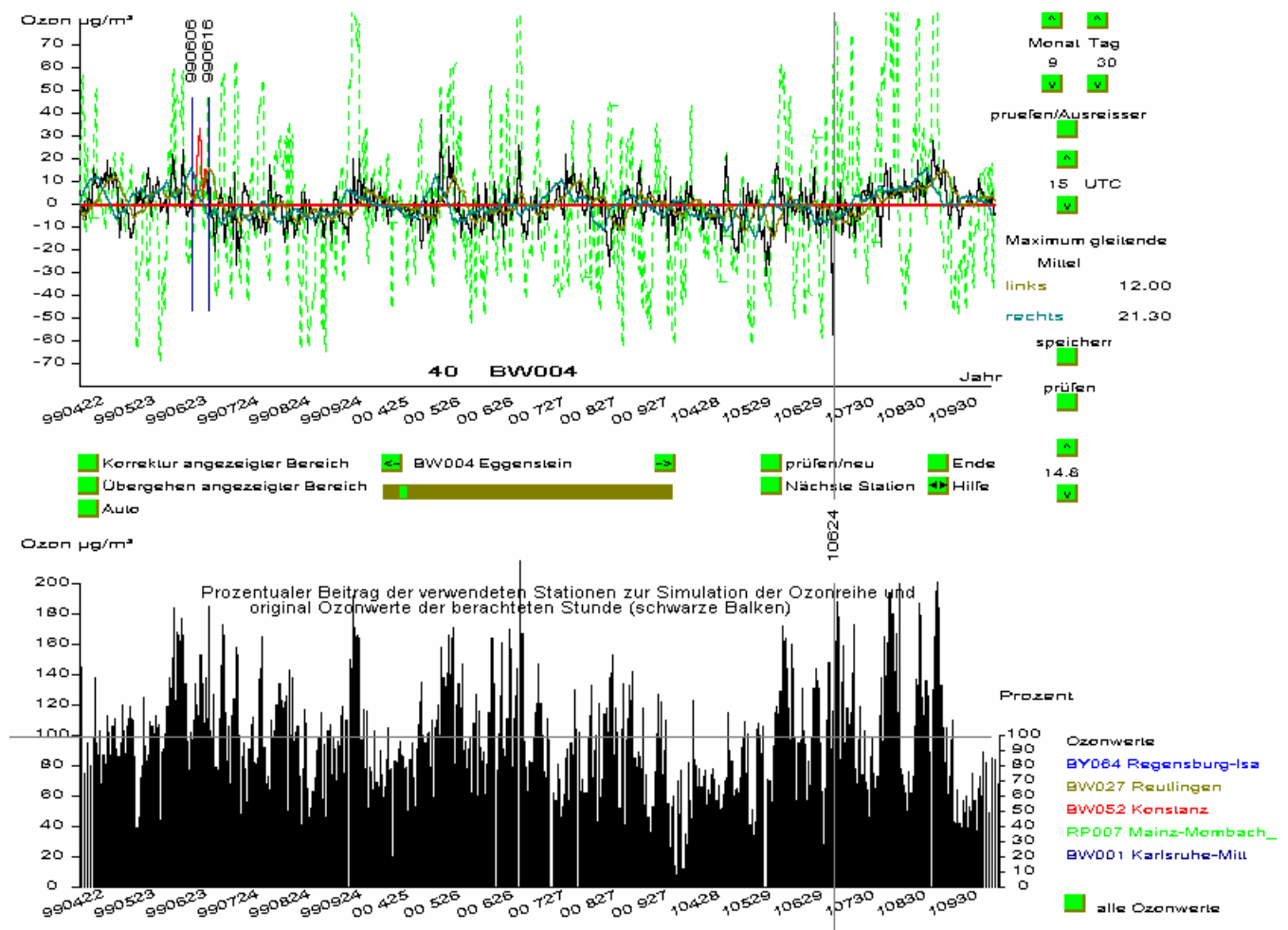
Abb. 4.11 Beispiel eines systematischen Fehlers anhand der Ozonmessstelle Berlin-Grunewald.

## 4.2 Erkennung und interaktive Validierung nicht plausibler Werte

Neben der im Kap. 3 beschriebenen realtime-Validierung zur Erkennung von Ausreißern und nicht plausiblen Werten steht ein interaktives Grafikmodul zur Verfügung, das eine Bewertung der am Bildschirm ausgewählten Tagesgänge im Kontext mit umliegenden Stationen, der simulierten Zeitreihe und dem Tagesgang der Flächendarstellung der Ozonverteilung erlaubt.

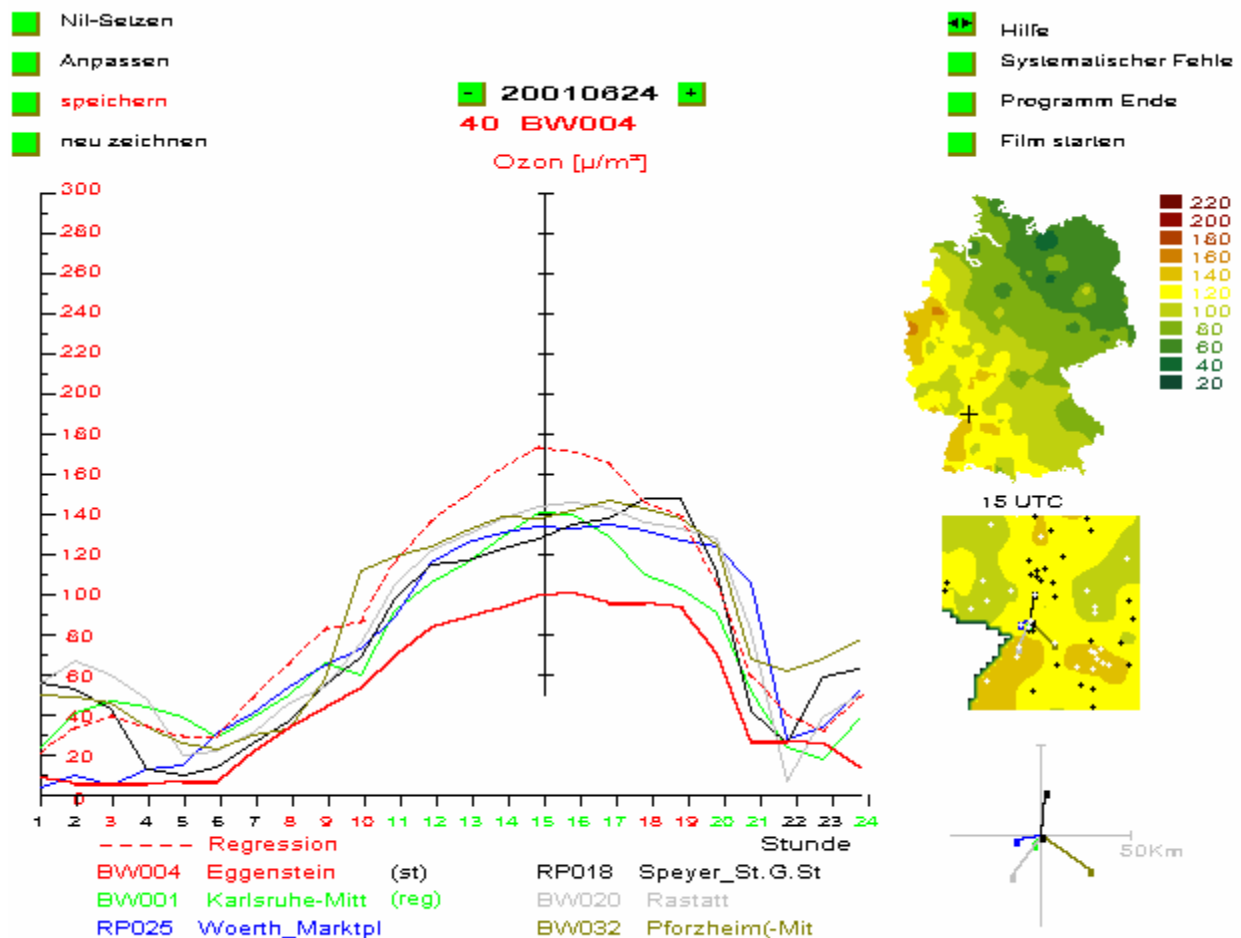
### 4.2.1 Interaktive Erkennung und Korrektur nicht plausibler Messwerte

Die Auswahl der Station und des zu bewertenden Tages erfolgt vom Grundzustand des Moduls zur Validierung systematischer Fehler aus durch Aktivieren des Fadenkreuzes (Klick mit der linken Maustaste auf den unteren Teil der Abbildung, siehe Abb. 4.12)



**Abb. 4.12 Beispiel zur Auswahl nicht plausibler Messwerte anhand der Station BW004 - Eggenstein, 24.06.2001.**

Der Messwert am 24.06.2001 an der Station BW004 - Eggenstein ist auffällig durch eine starke Abweichung von mehr als  $50 \mu\text{g}/\text{m}^3$  zwischen simulierter und beobachteter Reihe. Dies ist aus dem Grundbild zur Validierung systematischer Fehler deutlich durch einen Peak in der schwarzen Kurve erkennbar. Aktiviert man nun den Button „prüfen/Ausreisser“, so wird das interaktive Grafikmodul zur Kontrolle nicht plausibler Werte (Abb. 4.13) aktiviert.



**Abb. 4.13: Grundbild der interaktiven Datenvalidierung zur Erkennung und Korrektur nicht plausibler Ozonmesswerte anhand des Tagesganges vom 24.06.2001 an der Station - BW004 Eggenstein.**

#### 4.2.2 Beschreibung des Bildschirminhaltes

Allgemein ist in Abb. 4.12 der zeitliche Verlauf ausgewählter Ozonstationen eines Tages und die dazugehörige geografische Verteilung der Ozonkonzentration dargestellt.

Die Speichernummer und die Stationskennung der untersuchten Station ist in der Bildmitte oben, rot markiert, ausgegeben. Das Datum (Jahr, Monat, Tag) des untersuchten Termins ist unmittelbar darüber zu finden.

Die farbigen Kurven stellen den Tagesgang der Ozonkonzentration der untersuchten Station (rote Kurve) und den Tagesgang der 5 nächstliegenden Stationen gleichen Typs (verkehrsnahe: vk, städtisch: st, ländlich: la oder Bergstation: bg) dar. Diese 5 nächstliegenden Stationen sind in der Regel nicht mit den zur Simulation der gesamten Zeitreihe genutzten Stationen identisch. Lediglich die grüne Kurve ist die am stärksten mit der untersuchten Reihe korrelierte Ozonmessreihe (mit '(reg)' gekennzeichnet).

Der Stationstyp ist rechts neben der Stationsbezeichnung der untersuchten Messreihe (erste Zeile rote Schrift) zu finden. Im vorliegenden Beispiel handelt es sich um die Station BW004 - Eggenstein, eine städtische Station.

Die Bezeichnung der Vergleichsstationen steht in farblicher Übereinstimmung mit dem dargestellten Tagesgang der entsprechenden Station. Die Entfernung der Vergleichsstationen, relativ zur untersuchten Station, ist rechts unten in analoger Farbgebung zu den dargestellten Tagesgängen der Ozonkonzentration zu finden.

Die rot gestrichelte Kurve ist der mittels multipler Regression simulierte Tagesgang der Ozonreihe. Alle zur Simulation verwendeten Stationen sind im Modul zur Erkennung systematischer Fehler (Abb. 4.12 rechts unten) zu finden.

Die geografische Verteilung der Ozonkonzentration ist in Abb. 4.12 rechts oben für die unter der Abbildung angezeigte Uhrzeit dargestellt. Die farbliche Zuordnung der Ozonwerte lässt sich anhand der Legende zu dieser Abbildung erkennen. Die Konzentrationen sind in  $\mu\text{g}/\text{m}^3$  dargestellt. Die geografische Lage der aktuell untersuchten Station ist auf dieser Teilabbildung durch ein schwarzes oder weißes Kreuz, je nach Hintergrundfarbe zu finden.

Die Farbgebung der Stundenwerte auf der X-Achse charakterisiert die Fehlerart des beanstandeten Ozonwertes (Tab. 4.2).

Farbe	Art des Fehlers
schwarz	Keine Beanstandung
rot	Der betragsmäßige Abstand des Ozonwertes zu allen anderen aufgeblendeten Messwerten der entsprechenden Stunde überschreitet einen vorgegebenen Schwellenwert.
grün	Der betragsmäßige Abstand des Ozonwertes der entsprechenden Stunde überschreitet eine vorgegebene Schwelle zu dem mittels multipler Regression simulierten Stundenwert..
ocker	Es liegen keine Messwerte vor (Nil-Kennung).

**Tabelle 4.2: Charakterisierung der Fehlerart eines beanstandeten Ozonwertes**

Bei der ausgewählten Station Eggenstein (Abb. 4.13) werden eine Reihe von Stunden durch rote bzw. grüne Markierungen auf der X-Achse beanstandet. Da die Schwellenwert-Überschreitung des gemessenen Ozonwertes zu benachbarten Stationen gleichen Typs höhere Priorität hat als die Schwellenwert-Überschreitung zum simulierten Ozonwert, sind die Stunden 7 bis 17 Uhr grün markiert. Beide Überschreitungskriterien waren für dieses Zeitintervall zutreffend. Inwieweit der Tagesgang an der Station Eggenstein durch besondere meteorologische Bedingungen geprägt war oder ein Fehler am Messgerät vorlag, bedarf der subjektiven Bewertung. Wie aus Abb. 4.12 ersichtlich, ist die betragsmäßige Abweichung zwischen simulierter Reihe und Messreihe (schwarze Kurve) in den allermeisten Tagen kleiner als  $20 \mu\text{g}/\text{m}^3$ .

## 4.2.3 Möglichkeiten einfacher Interaktionen

### 4.2.3.1 Auswahl eines Zoom-Gebietes

Durch einen Klick mit der linken Maustaste auf die Karte der geografischen Verteilung der Ozonkonzentration erscheint unterhalb dieser Karte ein Zoom des ausgewählten Gebietes (Abb. 4.13). Durch einen erneuten Klick auf diese Abbildung kann ein anderer Gebietsausschnitt gewählt werden.

Auf der Zoomkarte sind die aktuell verwendeten Ozonmessstellen als schwarze oder weiße Kreuze, je nach Hintergrundfarbe, eingetragen. Die geografische Zuordnung der

Vergleichsstationen ist mit analoger Farbgebung zu den dargestellten Tagesgängen der Ozonkonzentration gleichfalls in der Zoomkarte zu finden .

#### 4.2.3.2 Ablesen der Ozonwerte an den Kurven der Tagesgänge

Durch einen Klick mit der linken Maustaste auf den weißen Hintergrund innerhalb der Grafik der Tagesgänge erscheint eine graue Linie, an der der Ozonwert auf der linken Skala abgelesen werden kann (Abb. 4.14).

Durch einen Klick mit der linken Maustaste auf eine der farbigen Kurven werden der an dieser Stelle eingetragene Ozonwert und die dazu gehörige Uhrzeit unter der zweiten Y-Achse (Bildmitte) in der dazu gehörigen Farbe eingeblendet (In Abb. 4.14 ist der 16-Uhr-Wert der Station Karlsruhe–Mitte, grün mit  $140 \mu\text{g}/\text{m}^3$  angezeigt).

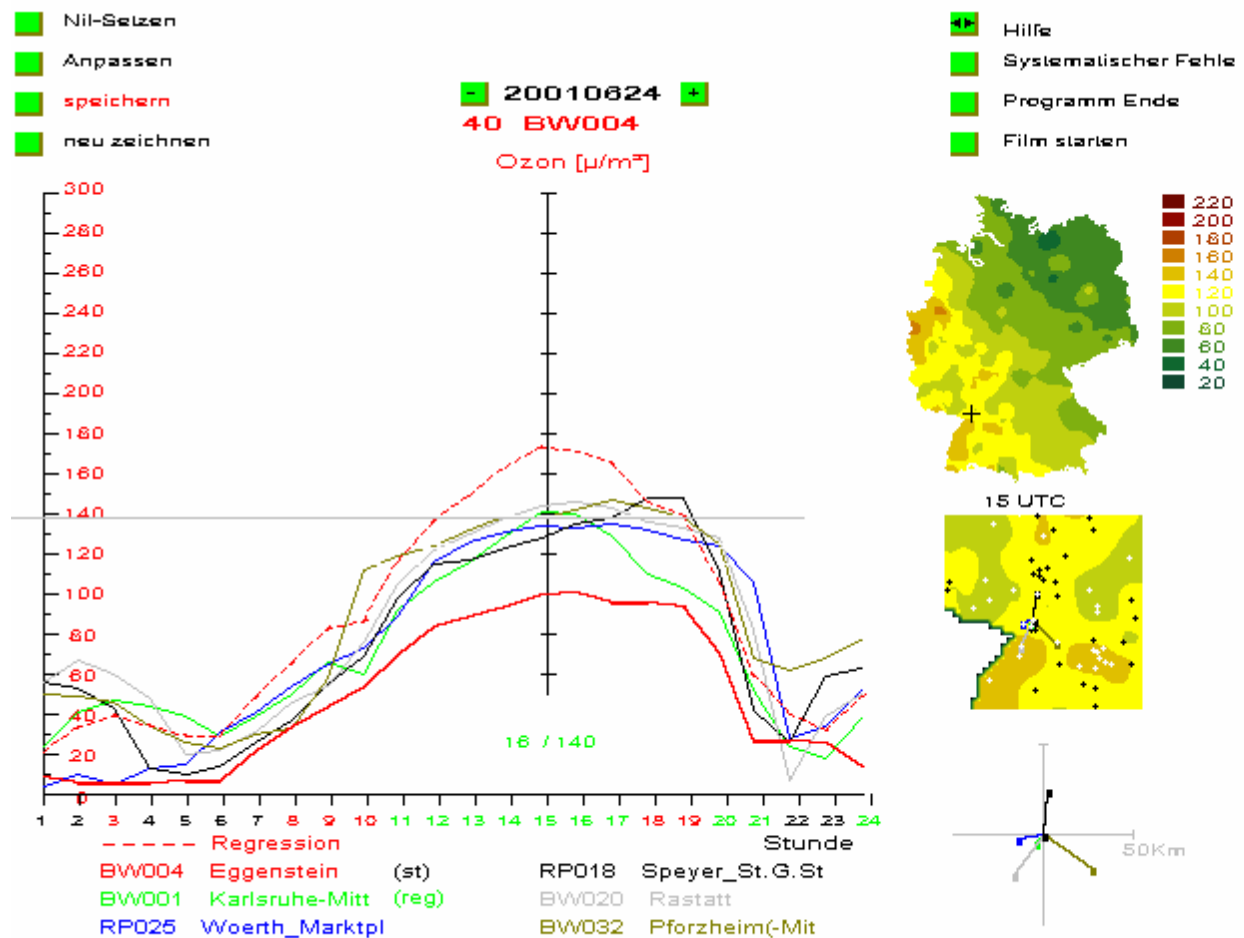


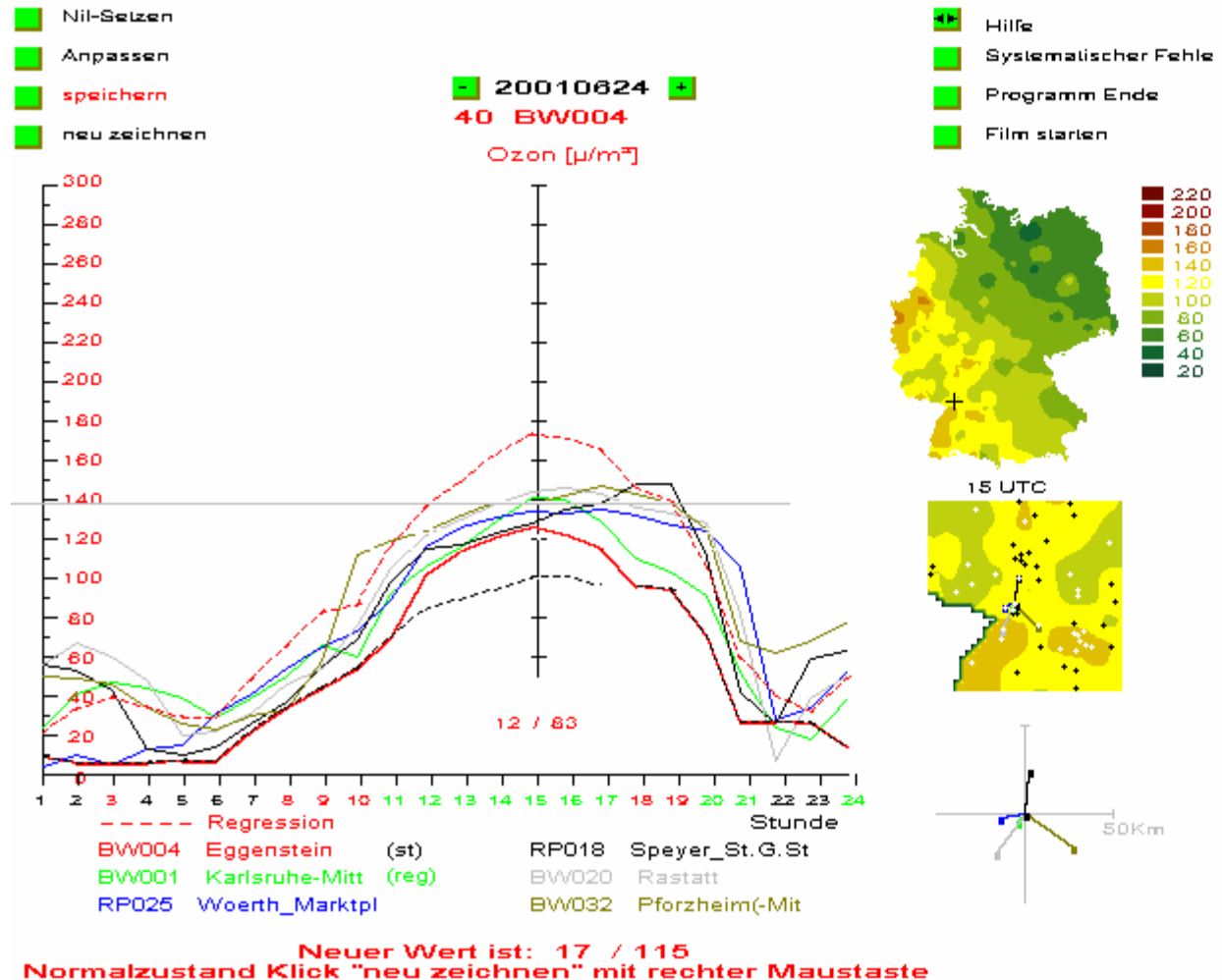
Abb. 4.14: Darstellung einfacher Interaktionsmöglichkeiten anhand des Tagesganges vom 24.06.2001 an der Station BW004 - Eggenstein.

#### 4.2.4 Möglichkeit der Interaktion zur manuellen Korrektur von Stundenwerten

Durch einen Klick auf die zu untersuchende Station (rote Kurve in Abb. 4.15) kann anschließend durch einen Klick mit der rechten Maustaste ein konkreter Stundenwert manuell verändert werden. Ein weiterer Klick mit der rechten Maustaste verändert diesen Wert erneut, oder je nach Cursorposition einen anderen Ozonwert. Der genaue Wert mit der dazugehörigen

Uhrzeit wird in der Abbildung unten Mitte rot eingeblendet. Die ursprüngliche Kurve bleibt als gestrichelte schwarze Kurve bestehen.

Als Beispiel wurden in Abb. 4.15 die Stundenwerte 11 bis 17 Uhr der Station BW004 manuell verändert (vgl. mit Abb. 4.13). Der zuletzt geänderte Wert von 17 Uhr wird in der Abbildung mit  $115 \mu\text{g}/\text{m}^3$  rot eingeblendet.



**Abb. 4.15: Möglichkeit der Interaktion zur manuellen Korrektur von Stundenwerten**

Um wieder in den normalen Arbeitsmodus zu gelangen, muss zur Beendigung der manuellen Manipulation ein Klick mit der rechten Maustaste auf den Button 'neu zeichnen' erfolgen. Danach sind alle anderen Button mit der linken Maustaste wieder aktiv. Durchgeführte Änderungen können durch einen Klick auf 'neu zeichnen' wieder verworfen werden.

## 4.2.5 Funktionsweise der Button

### Hilfe

Es wird eine kurze online - Hilfe gestartet, die wesentliche Funktionen der interaktiven Grafik erläutert. Durch Betätigen der Pfeiltasten im Button der Hilfe-Funktion kann im Hilfetext geblättert werden.



### Film starten

Durch Drücken dieses Buttons kann die stündliche Entwicklung der geografischen Verteilung der Ozonkonzentration zwischen 6 Uhr und 22 Uhr betrachtet werden. Bei eingeschaltetem Zoom läuft die Vergrößerung automatisch mit.

### Neu zeichnen

Mit diesem Button wird die gesamte Abbildung neu aufgebaut. Alle bisher durchgeführten Änderungen werden rückgängig gemacht.

### Speichern

Sollen die geänderten Werte übernommen werden, so ist dieser Button zu drücken. Die durchgeführten Änderungen werden in der Datei c:\routwarn\temp\warn\ausrerg\*(Station, Datum)*.dat gespeichert

### Programm Ende

Soll die Arbeit an dieser Stelle beendet werden, so ist dieser Button zu drücken.

### Systematische Fehler

Dieser Button bewirkt die Rückkehr zur Validierung der systematischen Fehler.

### Wechsel des zu untersuchenden Tages

Durch eine Aktivierung der Pfeiltasten unmittelbar neben dem Datum kann jeweils um einen Tag vor oder zurück geblättert werden.

### Anpassen

Durch Betätigen dieses Buttons werden folgende Aktivitäten ausgelöst:

- Fehlwerte (Der Termin ist auf der X-Achse ocker eingefärbt) werden durch die simulierte Reihe (rotgestrichelte Kurve) ersetzt.
- Vorhandene, jedoch als evtl. fehlerhaft erkannte Werte werden entweder durch den nächstliegenden Messwert der Vergleichsstationen oder durch den Wert der simulierten Reihe ersetzt, wenn die entsprechenden Kriterien erfüllt sind.

Die Art der Korrektur eines bestimmten Termins ist aus der Färbung der Uhrzeit auf der X-Achse ersichtlich:

Farbe	Art der Korrektur
schwarz	Es wurde keine Korrektur vorgenommen.
ocker	Der Wert wurde auf Ausfallkennung gesetzt.
blau	Es erfolgte eine zeitliche Interpolation von Ausfallwerten.
rot	Es erfolgte eine Anpassung an den Regressionswert (rot gestrichelte Kurve).
grün	Es wurde der nächstliegende Stationswert der aufgeblendeten Kurven verwendet.
hellgrau	Es wurden alle Stundenwerte der aufgeblendeten Stationen des entsprechenden Termins gemittelt.

### Tab. 4.3: Kennzeichnung der Art der Fehlerkorrektur

In Abb. 4.16 wurde eine automatische Anpassung aller beanstandeten Werte an den nächstliegenden Messwert einer der aufgeblendeten Vergleichsstationen angepasst. Die Werte der ursprünglichen Kurve des Tagesganges sind als rot gestrichelte Linie weiterhin erkennbar. Die Änderungen können mit dem Button „neu zeichnen“ wieder rückgängig gemacht werden.

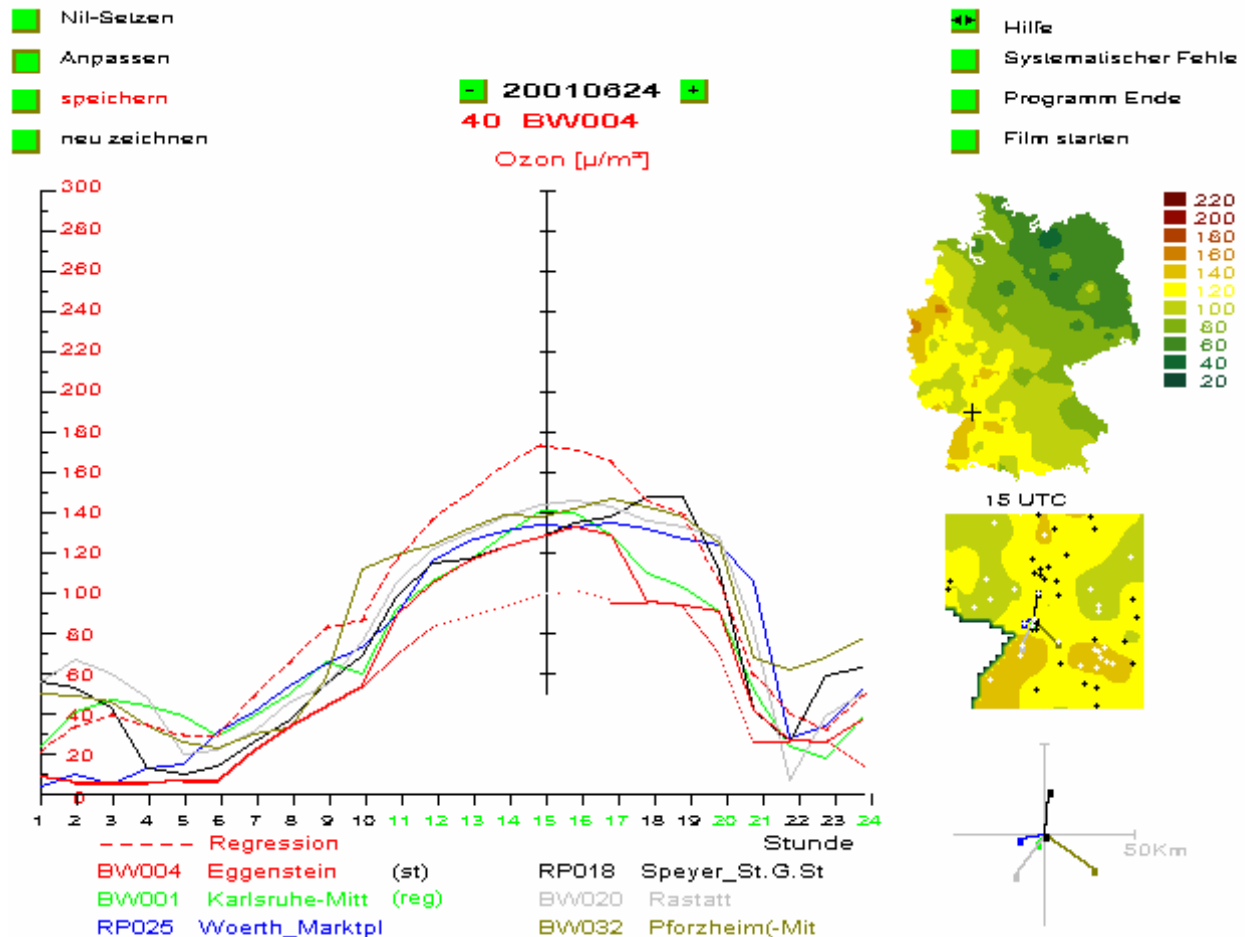


Abb. 4.16: Darstellung der automatischen Korrektur anhand des Tagesganges vom 24.06.2001 an der Station BW004 - Eggenstein.

### NIL-Setzen

Durch diesen Button werden die als fehlerhaft erkannten Werte auf Ausfallkennung gesetzt. Wurden Werte manuell markiert, so werden diese gleichfalls auf Ausfallkennung gesetzt.

### Speichern

In dieser Programmversion, die im Rahmen einer realtime-Datenvalidierung Verwendung findet, wird neben dem systematischen Fehler, der hier einen Zeitraum von 50 Tagen betrachtet (vgl. Kap. 3.2 und Kap. 4), nur der aktuelle Tag auf nicht plausible Daten getestet. Die Untersuchungsergebnisse werden in zweierlei Weise dokumentiert:

- Stationsspezifische Ablage der Ausgangsdaten, wodurch eine spätere Betrachtung und Bearbeitung der bemängelten Datenreihe analog Abb. 3.1 und Kap. 3.2 bzw. Abb. 4.15 in Kap. 4.2.4 möglich ist. Die dazugehörigen Dateien werden im Verzeichnis `c:\routwarn\temp\warn\ausrerg\` abgelegt und mit jedem Lauf aktualisiert. Dies bedeutet, dass nur die Ergebnisse des letzten Validierungslaufes zur erneuten Bearbeitung und Betrachtung archiviert werden.

- Zyklusspeicherorientierte Archivierung der Validierungsergebnisse:

Zum Zwecke der dauerhaften Archivierung der durchgeführten Änderungen werden die korrigierten Daten in einem direct-access-File „zykoz01k.dat“ datumsorientiert gespeichert. Die Daten können mit dem Programm „qualaus.exe“ gelesen und im Temp-Verzeichnis unter den Namen „zykkorrlaus.dat“ im ASCII-Format abgelegt werden. Die Zahl „01“ im Dateinamen `zykoz01k.dat` kennzeichnen das Jahr 2001 des Zyklusspeichers.

Im direct-access-File `zykoz01q.dat` werden die an den Daten vorgenommenen Veränderungen dokumentiert. Die Ergebnisse werden tagesspezifisch für jede Station und Stunde mit einer zweistelligen Zahl dokumentiert. Die erste Zahl bedeutet die Art des erkannten Fehlers, beschrieben in Tabelle 1 und die zweite Zahl die Art der vorgenommenen Korrektur, dokumentiert in Tabelle 2. Der Inhalt des Zyklusspeichers kann durch Abarbeitung des Programms „qualaus.exe“ in einer Tabelle im Tempverzeichnis unter dem Namen `qualaus.dat` im ASCII Format bereit gestellt werden.

## 5 Finale Datenvalidierung

Die finale Datenvalidierung hat die Aufgabe, die von den Ländern und dem Messnetz des UBA zur Verfügung gestellten Jahresdatensätze einer komplexen Wertung zu unterziehen. Diese Datenvalidierung stützt sich auf länderübergreifende statistische Relationen zwischen den Ozonmessreihen. Ausreißer und mögliche systematische Fehler werden dokumentiert und können je nach Aufgabenstellung bei der Weiterverarbeitung (Präsentation oder abgeleitete Größen) eingearbeitet werden. Dabei bleiben die Originalreihen erhalten. Die finale Datenvalidierung besteht zum einen aus dem Modul zur Identifikation möglicher systematischer Fehler (Kap.3) und zum anderen aus einem Modul zur Auffindung nicht plausibler Werte (Kap.2).

### 5.1 Finale Validierung systematischer Fehler

Im Gegensatz zur realtime Erkennung systematischer Fehler (Kap. 3), deren Ziel die möglichst frühzeitige Warnung der Messnetze ist, wobei ein Fehlersignal am Ende bzw. am rechten Rand der Messreihe ausgegeben wird, kommt es bei der finalen und interaktiven Validierung zu einer Beurteilung der gesamten Messreihe, wobei schwerpunktmäßig, d.h. mit allen Interaktionsmöglichkeiten nur das letzte Jahr geprüft wird. Das dazu eingesetzte Validierungsmodul beinhaltet sowohl eine vollautomatische Kontrolle und Korrektur der Datenreihen als auch die Möglichkeit einer interaktiven Kontrolle der Einzelreihen.

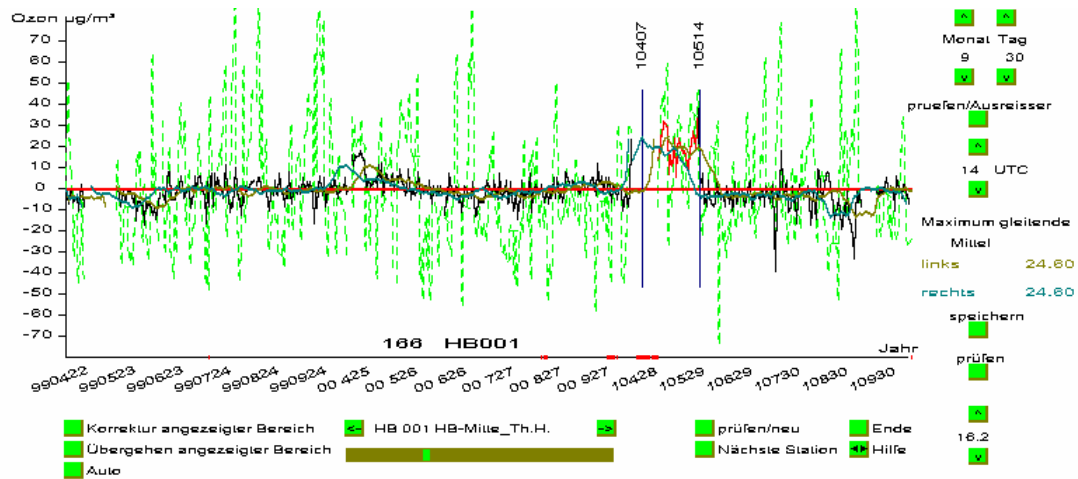
Während bei der interaktiven Validierung systematischer Fehler einzelne Reihen subjektiv beurteilt werden können (vgl. Kap. 4), ist bei der automatischen finalen Validierung systematischer Fehler kein manueller Eingriff während der Erkennung und Korrektur systematischer Abweichungen in den Datenreihen möglich. Diese Validierung erfolgt iterativ in 3 Stufen. Dadurch wird gewährleistet, dass systematische Fehler, die fälschlicherweise von einer anderen Messreihe verursacht wurden, und als Fehler der zu prüfenden Messreihe interpretiert wurden, in der nächsten Iteration wieder korrigiert werden können. Weiterhin ist zu beobachten, dass das Validierungsmodul Abschnitte in den Datenreihen, die mit systematischen Fehlern behaftet sind, erst mit der zweiten oder dritten Iteration vollständig bereinigt oder dass Überreaktionen durch eine nachfolgende Validierungsstufe wieder korrigiert werden.

An Hand von zwei Beispielen sollen die Wirkungsweise und die Besonderheiten der automatischen Validierung systematischer Fehler demonstriert werden.

In Abb. 5.1a ist ein deutlicher systematischer Fehler bei der Station HB001 (Bremen-Mitte) mit dem Wechsel der Ozonsaison 2000/2001 zu erkennen, der bis zum 14.5.2001 reicht und mit einem Datenausfall Anfang April 2001 einhergeht. In der ersten Validierungsstufe wird zwischen dem 14. April und dem 2. Mai eine systematische Abweichung gefunden, die in den Nachmittagsstunden ca.  $20 \mu\text{g}/\text{m}^3$  beträgt und einen deutlichen Tagesgang aufweist (Tab. 5.1a). Die durchgeführte Korrektur erweist sich in einer zweiten Iterationsstufe (Abb. 5.1b) als zu gering. Deshalb wird für diesen Zeitraum (1.April bis 13.Mai) ein weiteres Korrekturfeld bestimmt (Tab. 5.1b). Jetzt liegt der Schwerpunkt der Korrekturen um den 19. April und dem 9. Mai. In einem weiteren Iterationsschritt wird festgestellt, dass die am 19. April durchgeführte Korrektur zu hoch war. (Abb. 5.1c). Deshalb wird ein entsprechendes Korrekturfeld berechnet, das letztendlich eine Zeitreihe erzeugt, die weitgehend frei von systematischen Fehlern ist.

Abb. 5.2a zeigt an der Station HE030 - Marburg einen deutlichen systematischen Fehler zwischen dem 21. Juli 2000 und dem 29. September 2000, der auch durch das Validierungsmodul angezeigt wird. Dieser ist aber nicht während des gesamten Zeitraumes hinweg über den ganzen Tag als systematischer Fehler nachweisbar, sondern nur im Zeitraum vom 7. bis 21. August 2000 (Tab. 5.2a). Mindestens bei 6 der überprüften Stunden je Tag

muss ein systematischer Fehler gefunden werden, bevor er vom Modul korrigiert wird. Erst in einer zweiten Iterationsstufe (Abb. 5.2b) wird auch der Zeitraum vom 20. Juli bis 8. August korrigiert (Tab.5.2b). Eine dritte Validierungsstufe (Abb. 5.2c) zeigt keine systematischen Fehler in der Zeitreihe mehr an, obwohl der optische Eindruck auch nach dem 20. August einen systematischen Fehler vermuten lässt. Setzt man die Prüfschwelle von  $17.3 \mu\text{g}/\text{m}^3$  auf  $13.3 \mu\text{g}/\text{m}^3$  herab, so wird auch dieses Zeitintervall erkannt und korrigiert (Abb. 5.2d und 5.2e).



**Abb. 5.1a: Finale Datenvalidierung systematischer Fehler 1. Iteration an Hand der Station HB001 - Bremen-Mitte 15 Uhr**

**Tab. 5.1a: Systematischer Fehler, 1. Iteration an Hand der Station HB001 - Bremen-Mitte geglätteter Tagesverlauf vom 01.04.2001 bis 17.05.2001**

Uhrzeit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
Datum	1 4 1	3.	2.	0.	0.	0.	0.	1.	-1.	-3.	-3.	-3.	-5.	-5.	0.	2.	1.	3.	4.	3.	1.	-1.	-2.	-3.	-1.
1 4 2	5.	4.	2.	2.	0.	1.	1.	1.	0.	-3.	-4.	-4.	-5.	-6.	2.	4.	2.	4.	5.	5.	3.	1.	-1.	-1.	1.
1 4 3	8.	7.	5.	4.	3.	3.	3.	3.	0.	-3.	-5.	-5.	-6.	-7.	2.	4.	1.	4.	7.	7.	6.	4.	2.	1.	4.
1 4 4	13.	11.	8.	7.	6.	6.	5.	2.	-2.	-5.	-6.	-8.	-8.	5.	7.	4.	5.	6.	7.	7.	6.	4.	3.	7.	
1 4 5	20.	17.	13.	12.	12.	11.	9.	3.	-1.	-2.	-4.	-6.	-8.	12.	11.	6.	8.	8.	8.	8.	8.	2.	2.	8.	
1 4 6	28.	23.	18.	15.	0.	0.	0.	2.	-2.	-3.	-5.	-8.	-10.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 4 7	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 4 8	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 4 9	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 410	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 411	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 412	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 413	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	14.	12.	
1 414	8.	9.	10.	9.	6.	4.	1.	0.	1.	2.	4.	10.	12.	15.	16.	15.	12.	14.	15.	15.	14.	16.	13.	12.	
1 415	7.	9.	10.	8.	5.	3.	1.	0.	1.	3.	7.	13.	15.	18.	19.	17.	16.	18.	19.	20.	19.	18.	13.	12.	
1 416	7.	8.	9.	7.	6.	4.	3.	3.	4.	6.	10.	16.	18.	20.	21.	20.	18.	20.	21.	21.	21.	20.	15.	13.	
1 417	6.	8.	10.	9.	9.	8.	7.	6.	7.	9.	12.	17.	20.	22.	23.	22.	20.	21.	21.	20.	20.	19.	15.	14.	
1 418	8.	9.	11.	9.	8.	7.	6.	6.	8.	11.	14.	19.	22.	24.	24.	23.	21.	21.	19.	18.	18.	15.	14.		
1 419	11.	12.	13.	10.	9.	7.	5.	5.	7.	11.	14.	20.	23.	25.	24.	24.	22.	22.	22.	20.	18.	17.	14.	14.	
1 420	13.	13.	13.	10.	9.	7.	5.	5.	7.	10.	13.	19.	21.	22.	22.	21.	20.	20.	20.	19.	17.	17.	15.	14.	
1 421	13.	13.	14.	10.	9.	7.	5.	5.	7.	10.	13.	18.	21.	22.	20.	20.	19.	20.	19.	18.	16.	16.	14.	13.	
1 422	13.	13.	13.	10.	8.	6.	4.	4.	7.	9.	12.	17.	19.	21.	20.	19.	18.	20.	20.	19.	17.	17.	14.	13.	
1 423	9.	10.	11.	8.	7.	5.	3.	3.	6.	9.	12.	16.	19.	20.	19.	19.	19.	20.	20.	19.	17.	16.	14.	13.	
1 424	9.	10.	11.	9.	8.	7.	4.	5.	7.	10.	13.	17.	20.	21.	20.	20.	20.	21.	20.	19.	18.	16.	14.	12.	
1 425	8.	9.	10.	9.	7.	6.	4.	4.	7.	9.	12.	16.	19.	20.	20.	20.	20.	21.	21.	20.	19.	17.	15.	14.	
1 426	8.	9.	11.	10.	8.	7.	5.	5.	8.	10.	12.	16.	18.	19.	19.	19.	20.	20.	19.	19.	18.	16.	15.	13.	
1 427	9.	10.	11.	10.	8.	7.	5.	5.	8.	10.	12.	15.	18.	19.	19.	19.	20.	20.	19.	19.	18.	16.	15.	14.	
1 428	8.	9.	11.	9.	6.	5.	3.	4.	7.	9.	11.	14.	17.	18.	18.	19.	19.	19.	18.	18.	17.	16.	14.	13.	
1 429	8.	9.	10.	9.	7.	5.	4.	4.	7.	8.	10.	12.	15.	17.	17.	18.	18.	18.	18.	17.	15.	14.	13.		
1 430	7.	8.	10.	8.	7.	6.	5.	6.	8.	9.	10.	12.	15.	16.	17.	17.	17.	17.	16.	16.	15.	14.	12.		
1 5 1	5.	7.	9.	8.	6.	6.	5.	7.	9.	9.	11.	12.	15.	17.	18.	18.	18.	18.	17.	17.	16.	14.	13.	11.	
1 5 2	3.	5.	7.	6.	6.	5.	5.	6.	8.	9.	11.	12.	15.	16.	18.	19.	18.	18.	17.	17.	16.	14.	13.	9.	
1 5 3	-9.	-4.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 5 4	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 5 5	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 5 6	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 5 7	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 5 8	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 5 9	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 510	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 511	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 512	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 513	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 514	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
1 515	5.	4.	3.	4.	4.	4.	4.	4.	6.	8.	9.	11.	11.	11.	11.	10.	10.	10.	11.	10.	8.	7.	6.	5.	
1 516	-2.	0.	1.	2.	3.	3.	3.	3.	5.	6.	7.	9.	9.	10.	9.	8.	7.	7.	8.	8.	6.	5.	5.	3.	
1 517	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	

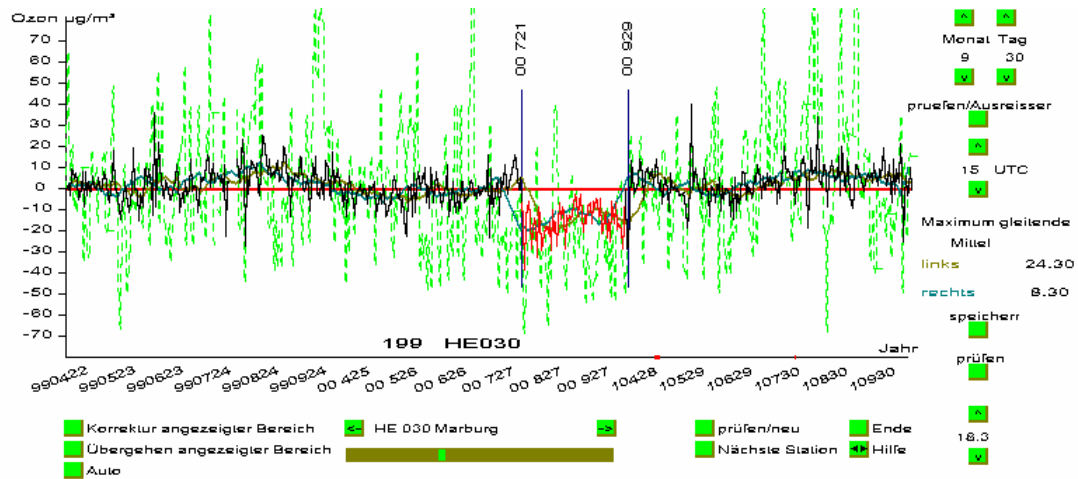
**Tab. 5.1b: Systematischer Fehler, 2. Iteration an Hand der Station HB001 - Bremen-Mitte geglätteter Tagesverlauf vom 01.04.2001 bis 17.05.2001**

Uhrzeit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24			
Datum	1	4	4	17.	14.	12.	10.	9.	8.	7.	5.	1.	-1.	-2.	-4.	-5.	7.	8.	3.	3.	4.	5.	5.	9.	6.	7.	10.
	1	4	5	25.	21.	18.	16.	15.	14.	12.	7.	2.	3.	1.	-3.	-4.	13.	12.	5.	6.	6.	6.	7.	10.	5.	7.	12.
	1	4	6	33.	28.	22.	14.	0.	0.	0.	8.	2.	2.	0.	-4.	-6.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
	1	4	7	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
	1	4	8	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
	1	4	9	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
	1	4	10	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
	1	4	11	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
	1	4	12	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
	1	4	13	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
	1	4	14	7.	8.	8.	8.	6.	4.	0.	0.	1.	1.	0.	2.	2.	2.	4.	4.	4.	6.	8.	11.	12.	14.	11.	10.
	1	4	15	4.	6.	7.	6.	5.	3.	1.	1.	1.	2.	2.	5.	7.	6.	7.	6.	8.	11.	12.	16.	17.	17.	12.	10.
	1	4	16	3.	5.	6.	6.	6.	5.	3.	3.	4.	4.	4.	7.	8.	7.	9.	8.	9.	12.	14.	17.	18.	18.	13.	11.
	1	4	17	2.	4.	6.	7.	8.	9.	7.	7.	7.	7.	7.	9.	11.	9.	11.	10.	11.	13.	14.	15.	16.	18.	14.	12.
	1	4	18	4.	6.	8.	8.	8.	8.	6.	7.	8.	8.	9.	11.	12.	11.	12.	12.	12.	13.	13.	14.	14.	16.	13.	12.
	1	4	19	8.	9.	10.	9.	9.	8.	5.	6.	7.	8.	10.	11.	13.	12.	13.	12.	13.	13.	14.	14.	14.	15.	12.	12.
	1	4	20	10.	10.	10.	9.	9.	8.	6.	6.	8.	8.	9.	10.	12.	10.	10.	10.	10.	11.	12.	13.	13.	14.	12.	12.
	1	4	21	9.	10.	11.	9.	8.	7.	5.	6.	8.	8.	9.	10.	11.	9.	9.	8.	9.	10.	11.	12.	12.	13.	11.	11.
	1	4	22	9.	10.	10.	8.	8.	7.	4.	5.	6.	7.	8.	9.	10.	8.	8.	7.	8.	10.	11.	13.	12.	13.	11.	10.
	1	4	23	6.	7.	8.	7.	6.	6.	3.	4.	6.	6.	7.	8.	9.	7.	8.	7.	8.	10.	10.	12.	12.	12.	10.	9.
	1	4	24	5.	6.	7.	6.	7.	7.	4.	5.	7.	7.	8.	9.	9.	8.	8.	7.	8.	10.	10.	11.	11.	11.	9.	8.
	1	4	25	4.	5.	6.	5.	6.	6.	4.	4.	6.	6.	7.	7.	8.	6.	7.	7.	8.	9.	10.	11.	11.	10.	9.	9.
	1	4	26	5.	6.	7.	6.	7.	7.	4.	4.	6.	6.	6.	7.	7.	5.	6.	6.	6.	7.	8.	8.	8.	8.	8.	8.
	1	4	27	6.	6.	7.	6.	6.	6.	4.	4.	5.	5.	6.	6.	6.	4.	4.	4.	5.	6.	7.	7.	7.	7.	8.	8.
	1	4	28	7.	7.	7.	5.	4.	4.	2.	2.	4.	4.	5.	5.	5.	5.	5.	5.	5.	6.	6.	7.	6.	6.	7.	7.
	1	4	29	7.	7.	7.	5.	4.	4.	2.	2.	4.	4.	4.	5.	4.	5.	4.	5.	5.	6.	6.	7.	7.	6.	7.	7.
	1	4	30	6.	7.	7.	5.	5.	5.	4.	4.	5.	5.	5.	5.	5.	4.	5.	5.	5.	6.	6.	6.	7.	6.	7.	7.
	1	5	1	5.	5.	6.	4.	4.	5.	4.	4.	5.	5.	6.	6.	6.	6.	7.	7.	7.	8.	8.	7.	8.	6.	7.	6.
	1	5	2	4.	4.	5.	4.	4.	5.	4.	4.	5.	5.	5.	6.	7.	8.	9.	9.	9.	8.	8.	8.	7.	8.	7.	
	1	5	3	4.	5.	5.	4.	4.	5.	5.	4.	6.	6.	7.	7.	8.	8.	10.	9.	9.	9.	7.	7.	7.	6.	8.	7.
	1	5	4	6.	6.	7.	5.	4.	5.	5.	5.	7.	7.	8.	8.	9.	10.	11.	12.	11.	11.	9.	8.	8.	7.	9.	8.
	1	5	5	7.	7.	8.	5.	4.	5.	5.	5.	7.	7.	9.	10.	11.	11.	13.	13.	13.	13.	11.	10.	9.	9.	10.	9.
	1	5	6	8.	8.	9.	7.	6.	6.	6.	6.	8.	9.	11.	12.	14.	15.	16.	16.	15.	15.	12.	11.	9.	9.	10.	9.
	1	5	7	8.	8.	9.	7.	6.	6.	7.	6.	9.	10.	12.	14.	16.	17.	18.	18.	18.	17.	15.	13.	12.	11.	12.	11.
	1	5	8	7.	8.	9.	7.	6.	7.	8.	8.	11.	12.	14.	16.	17.	18.	19.	20.	19.	19.	16.	14.	12.	11.	11.	10.
	1	5	9	5.	6.	7.	6.	5.	6.	8.	8.	10.	12.	15.	17.	19.	20.	21.	21.	21.	21.	19.	16.	14.	12.	10.	9.
	1	5	10	4.	5.	6.	6.	5.	6.	6.	7.	9.	11.	14.	16.	16.	17.	18.	19.	19.	19.	17.	15.	13.	11.	10.	8.
	1	5	11	2.	4.	5.	5.	4.	4.	5.	5.	7.	9.	11.	12.	13.	15.	15.	16.	16.	17.	16.	15.	12.	11.	10.	8.
	1	5	12	2.	3.	5.	4.	3.	4.	4.	4.	7.	8.	10.	11.	13.	14.	14.	14.	14.	15.	14.	13.	11.	11.	9.	8.
	1	5	13	2.	4.	5.	4.	3.	4.	4.	4.	6.	7.	9.	10.	11.	12.	12.	12.	12.	12.	11.	9.	9.	8.	7.	7.
	1	5	14	1.	2.	4.	4.	3.	3.	3.	3.	5.	6.	8.	9.	9.	10.	11.	10.	10.	11.	11.	11.	9.	9.	7.	5.
	1	5	15	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
	1	5	16	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
	1	5	17	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.

**Tab. 5.1c: Systematischer Fehler, 3. Iteration an Hand der Station HB001 - Bremen-Mitte geglätteter Tagesverlauf vom 01.04.2001 bis 17.05.2001**

[illegible]





**Abb. 5.2a: Finale Datenvalidierung systematischer Fehler 1. Iteration an Hand der Station HE030 - Marburg 15 Uhr**

**Tab. 5.2a: Systematischer Fehler, 1. Iteration an Hand der Station HE030 - Marburg geglätteter Tagesverlauf vom 20.07.2000 bis 30.08.2000**

Uhrzeit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
Datum																									
720	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
721	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
722	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
723	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
724	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
725	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
726	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
727	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
728	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
729	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
730	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
731	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
801	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
802	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
803	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
804	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
805	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
806	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
807	-11.	-9.	-7.	-6.	-5.	-4.	-5.	-5.	-7.	-10.	-14.	-16.	-19.	-20.	-19.	-19.	-19.	-19.	-22.	-20.	-18.	-15.	-13.	-11.	
808	-4.	-6.	-7.	-7.	-5.	-4.	-5.	-6.	-7.	-10.	-14.	-16.	-19.	-19.	-19.	-20.	-21.	-21.	-22.	-21.	-18.	-14.	-12.	-11.	
809	-6.	-7.	-8.	-8.	-6.	-5.	-6.	-7.	-9.	-11.	-15.	-18.	-20.	-20.	-21.	-21.	-21.	-21.	-19.	-16.	-12.	-11.	-11.	-11.	
810	-8.	-9.	-9.	-9.	-7.	-6.	-6.	-7.	-9.	-11.	-14.	-15.	-17.	-18.	-19.	-19.	-20.	-20.	-19.	-18.	-15.	-11.	-10.	-10.	
811	-8.	-8.	-9.	-9.	-7.	-5.	-6.	-6.	-8.	-10.	-13.	-14.	-17.	-17.	-18.	-19.	-19.	-19.	-19.	-18.	-14.	-9.	-8.	-8.	
812	-6.	-6.	-7.	-7.	-5.	-4.	-4.	-5.	-7.	-9.	-12.	-14.	-16.	-17.	-17.	-18.	-18.	-18.	-18.	-17.	-13.	-8.	-7.	-6.	
813	-5.	-5.	-5.	-6.	-4.	-3.	-3.	-4.	-7.	-8.	-11.	-13.	-15.	-15.	-17.	-18.	-19.	-19.	-20.	-19.	-15.	-10.	-8.	-7.	
814	-5.	-5.	-6.	-6.	-4.	-4.	-4.	-5.	-8.	-10.	-13.	-14.	-16.	-16.	-17.	-18.	-18.	-18.	-19.	-18.	-14.	-10.	-8.	-7.	
815	-5.	-6.	-6.	-7.	-6.	-5.	-6.	-6.	-9.	-11.	-13.	-14.	-16.	-16.	-17.	-18.	-18.	-18.	-19.	-18.	-14.	-10.	-8.	-8.	
816	-6.	-6.	-7.	-8.	-6.	-6.	-6.	-7.	-9.	-11.	-13.	-13.	-14.	-14.	-15.	-16.	-17.	-17.	-18.	-18.	-15.	-10.	-9.	-8.	
817	-7.	-7.	-7.	-8.	-6.	-6.	-6.	-7.	-10.	-12.	-13.	-14.	-15.	-15.	-15.	-16.	-16.	-17.	-18.	-17.	-14.	-10.	-8.	-8.	
818	-7.	-8.	-8.	-9.	-6.	-6.	-6.	-7.	-9.	-11.	-13.	-13.	-14.	-14.	-15.	-15.	-16.	-16.	-17.	-17.	-14.	-10.	-8.	-8.	
819	-6.	-7.	-7.	-8.	-6.	-5.	-6.	-6.	-9.	-11.	-13.	-13.	-14.	-14.	-14.	-14.	-14.	-14.	-16.	-16.	-13.	-10.	-9.	-8.	
820	-5.	-6.	-7.	-7.	-6.	-5.	-5.	-6.	-8.	-10.	-11.	-12.	-12.	-12.	-12.	-12.	-12.	-12.	-12.	-14.	-14.	-12.	-9.	-8.	-8.
821	-6.	-6.	-7.	-7.	-6.	-5.	-6.	-7.	-8.	-10.	-11.	-12.	-12.	-12.	-12.	-12.	-13.	-13.	-14.	-15.	-13.	-11.	-10.	-7.	
822	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
823	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
824	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
825	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
826	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
827	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
828	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
829	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	
830	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	

**Tab. 5.2b: Systematischer Fehler, 2. Iteration an Hand der Station HE030 - Marburg geglätteter Tagesverlauf vom 20.07.2000 bis 30.08.2000**

[illegible]

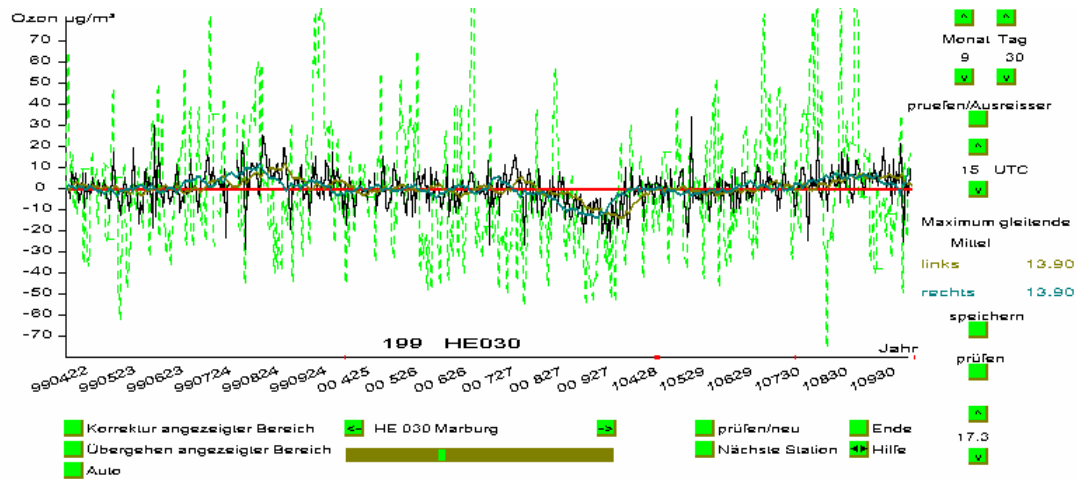


Abb. 5.2c: Finale Datenvalidierung systematischer Fehler, 3. Iteration an Hand der Station HE030 - Marburg 15 Uhr

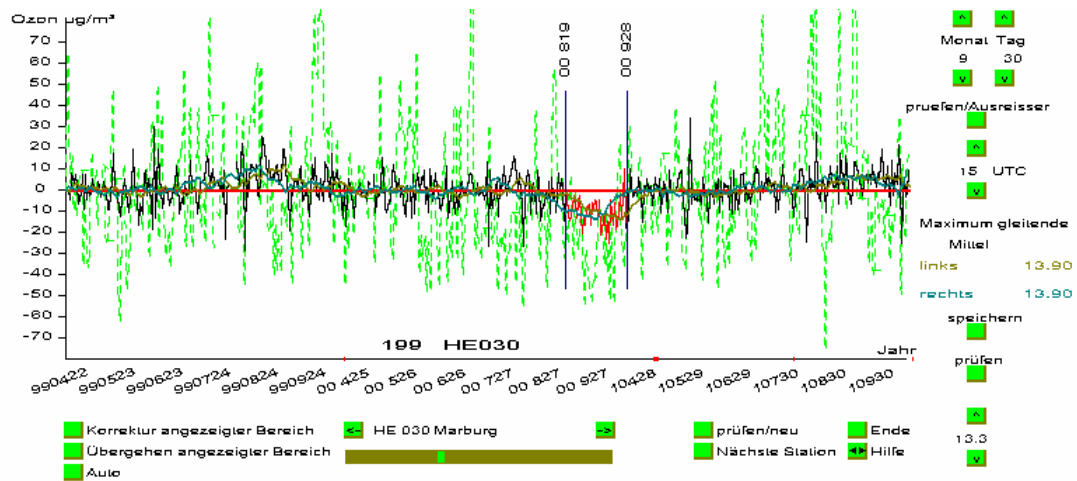


Abb. 5.2d: Finale Datenvalidierung systematischer Fehler, 3. Iteration an Hand der Station HE030 - Marburg 15 Uhr, Prüfschwelle auf  $13.3 \mu\text{g}/\text{m}^3$  herabgesetzt

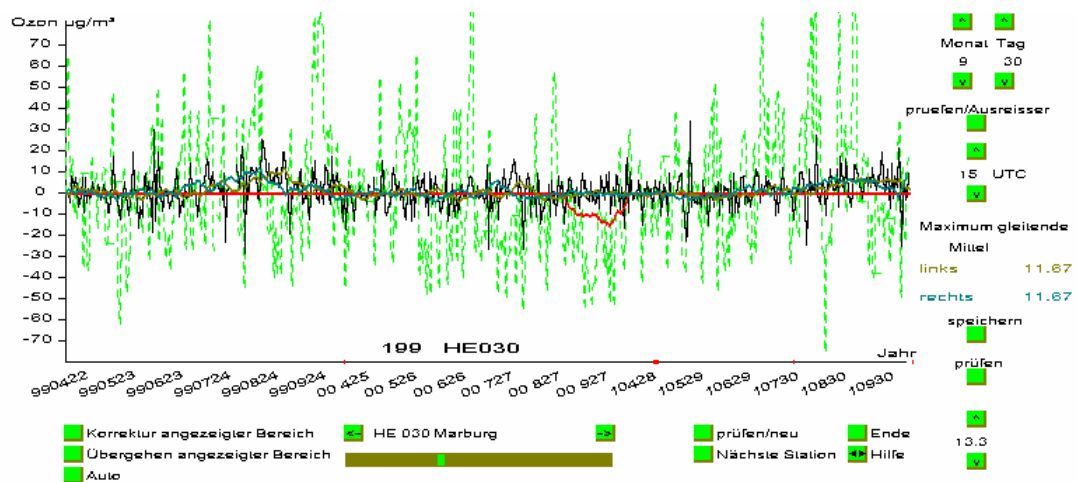


Abb. 5.2e: Finale Datenvalidierung systematischer Fehler, 3. Iteration an Hand der Station HE030 - Marburg 15 Uhr, Prüfschwelle auf  $13.3 \mu\text{g}/\text{m}^3$  herabgesetzt und im interaktiven Modus korrigiert

## **5.2 Dokumentation der Validierungsstufe zur Erkennung systematischer Fehler**

In dieser Programmversion, die im Rahmen einer finalen Datenvalidierung Verwendung findet, werden die Reihen auf nicht plausible Messwerte hin überprüft und der dargestellte Untersuchungszeitraum auf systematische Fehler untersucht. Die Untersuchungsergebnisse werden stationsspezifisch dokumentiert. Die systematischen Fehler werden im Verzeichnis:

c:\routwarn\dat\systfehl\sf.-*Stationskennung*-dat abgelegt und werden zur Korrektur der Datenreihen verwendet.

## 6 Erste Untersuchungen zur Validierung systematischer Fehler anderer Komponenten (neben Ozon)

Die realtime-Validierung von Ausreißern und nicht plausiblen Messwerten der Stoffe Schwefeldioxid ( $\text{SO}_2$ ), Stickstoffmonoxid ( $\text{NO}$ ), Stickstoffdioxid ( $\text{NO}_2$ ) und Feinstaub ( $\text{PM}_{10}$ ) wurde im Kapitel 2 näher ausgeführt. Ungeachtet dieser operationellen Datenkontrolle hat sich aus den Erfahrungen mit der finalen und interaktiven Validierung der Ozonmessreihen gezeigt, dass auch in den von den Ländern als Monats- und Jahresdatensätzen gelieferten Reihen sowohl unplausible Werte als auch systematische Fehler enthalten sind. Deshalb wurden für die Stofftypen  $\text{SO}_2$ ,  $\text{NO}$ ,  $\text{NO}_2$  und  $\text{PM}_{10}$  Voruntersuchungen durchgeführt, die darauf zielten zu prüfen, inwieweit das für Ozon entwickelte und operationell eingesetzte Verfahren auch auf andere Stofftypen übertragen werden kann.

### 6.1 Methodik

Die entscheidende Frage, ob die für Ozon entwickelten Verfahren ebenso für andere Stofftypen anwendbar sind, ist die Frage nach den räumlichen Abhängigkeiten zwischen den Messreihen. Mit anderen Worten: Lässt sich eine Messreihe genügend genau durch umliegende Messreihen simulieren? Genügend genau heißt in diesem Zusammenhang, dass die Streuung der Residuenreihe als Differenz zwischen simulierter und beobachteter Messreihe klein genug ist, um systematische Fehler in den Beobachtungsreihen signifikant herausfiltern zu können. Die Streuung der Originalreihe ist in der Regel so groß, dass systematische Fehler einer Messreihe überdeckt werden. Die Größe der Streuung in der Originalreihe ist aber für das Problem der Suche nach systematischen Fehlern mittels des beschriebenen Verfahrens ohne Bedeutung. Aus den Erfahrungen mit Ozonmessreihen sollte eine multiple Korrelation von 0.7 bzw. ein RV (prozentuale Reduktion der Varianz einer Reihe gegenüber ihrer Ausgangsvarianz) von  $\geq 50\%$  zwischen simulierter und beobachteter Reihe bestehen.

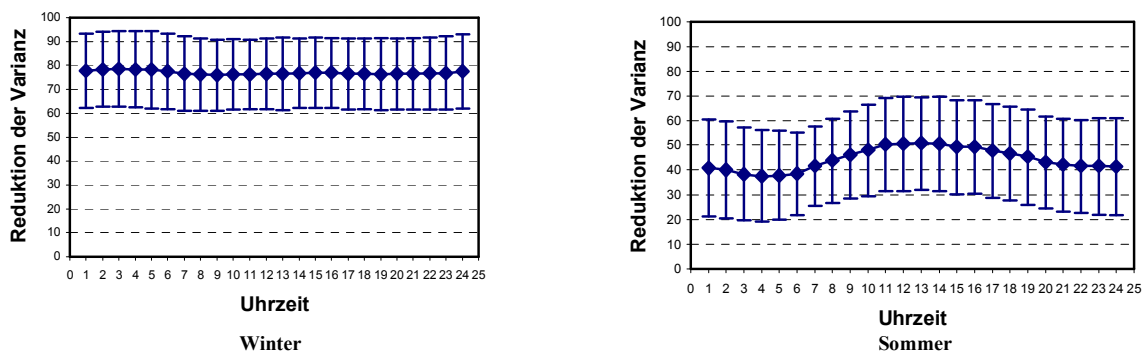
Zur Berechnung der multiplen Interkorrelation wurde die schon mehrfach beschriebene Screening-Regressionsanalyse (Enke 1988) mit allen zur Verfügung stehenden Stationen des gleichen Stofftyps durchgeführt. Dazu wurden für jede Stunde die stationsspezifischen Jahresgänge eliminiert. Nach der Selektion der multiplen Abhängigkeiten zwischen den Messreihen des entsprechenden Stofftyps wurde in einem zweiten Schritt eine Vereinheitlichung der zur multiplen Regression verwendeten Stationsmessreihen vorgenommen. Im Regelfall führt eine Screening-Regressionsanalyse stundenspezifisch zu Regressionsgleichungen, in denen unterschiedliche Stationskombinationen einen maximalen multiplen Korrelationskoeffizienten hervorbringen. Für eine ganztägige Prüfung der einzelnen Stofftypen ist es jedoch wichtig, dass zwischen den für die Einzelstunde simulierten Zeitreihen keine Sprünge auftreten. Dies wird erreicht, indem man für alle Stunden des Tages dieselbe Stationskombination festlegt, auch wenn damit im Einzelfall, d.h. auf eine einzige Stunde bezogen, ein geringer Teil an erklärter Varianz „verschenkt“ wird.

Die Untersuchungen wurden getrennt für Sommer- und Winterhalbjahr durchgeführt, da sowohl die Belastungssituation als auch die räumlichen Abhängigkeiten aufgrund der unterschiedlichen meteorologischen Gegebenheiten sehr unterschiedlich sein können.

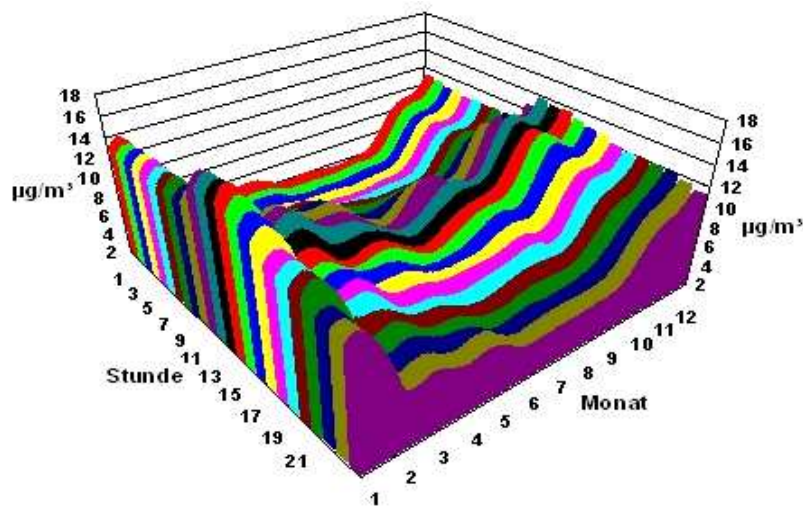
### 6.2 Ergebnisse

Betrachtet man Abb. 6.1, so fällt der deutliche Unterschied zwischen Sommer- und Winterhalbjahr für den Stofftyp  $\text{SO}_2$  auf. Während im Winterhalbjahr die räumlichen Interkorrelationen sehr gut sind und hier auch die höchsten  $\text{SO}_2$  Werte auftreten (Abb. 6.2), liegen im Sommer die RV-Werte, die in diesem Fall auch mit dem Bestimmtheitsmaß ( $R^2$ ) gleichgesetzt werden können, als Maßzahl für die räumlichen Abhängigkeiten der Messreihen

untereinander meist unter 50 %. Nur in den Mittags- und Nachmittagsstunden erreicht man die Marke  $RV \geq 50 \%$ , für das eine Kontrolle der Messreihen auf systematische Fehler noch als sinnvoll erscheint. Die hohen räumlichen Interkorrelationen im Winter, verbunden mit einer geringen Schwankung im Tagesverlauf spiegeln die im Winterhalbjahr überwiegenden großräumigen Advektionsverhältnisse wieder. Eine Prüfung auf systematische Fehler ist hier gut durchführbar. Dagegen ist die beschriebene Methode zur Prüfung von systematischen Fehlern bei sommerlichen  $SO_2$ -Reihen nicht ohne weiteres möglich. Hierzu bedarf es noch weiterführender Untersuchungen, die aber nicht mehr Bestandteil dieser Untersuchungen sind.

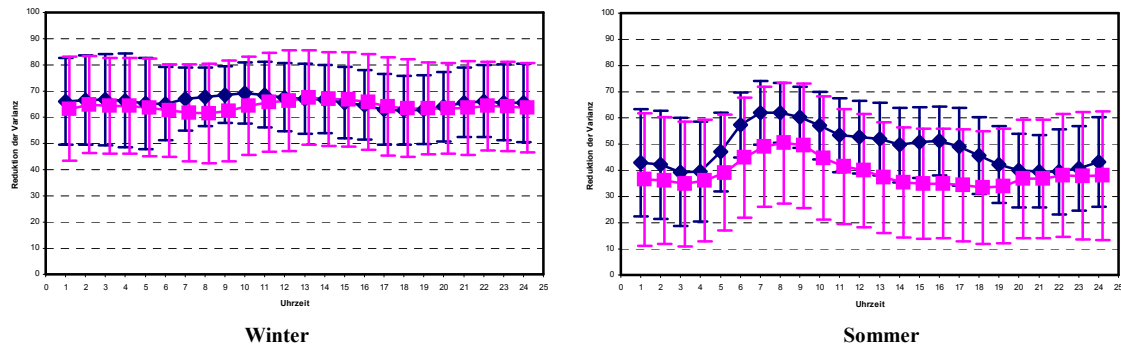


**Abb. 6.1:** Erklärte Varianz der mittels umliegender Stationen simulierten stündlichen  $SO_2$  - Zeitreihen (Mittelwerte und Streuung der Erklärten Varianz (RV) als Mittel über ca. 370 Stationen)

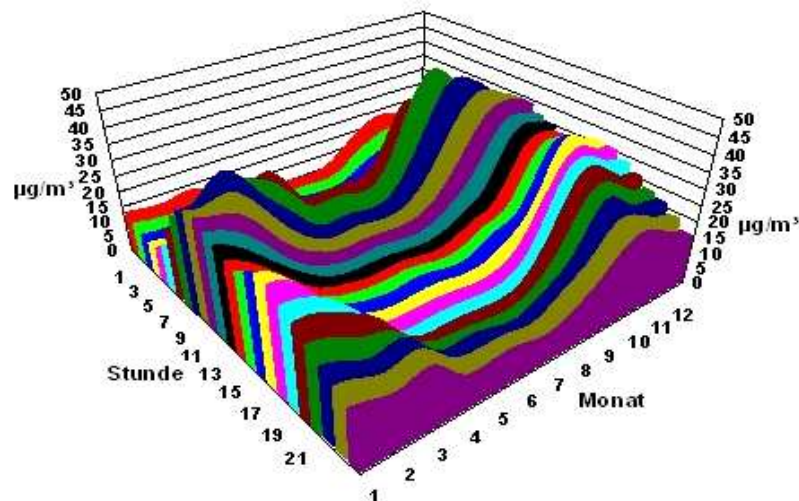


**Abb. 6.2:** Jahresgang der  $SO_2$ -Konzentration am Beispiel der Station BE034 – Berlin-Neukölln für jede Stunde, Mittel 1997 bis 2000

Bemerkenswert bei Stickstoffmonoxid ist auch hier der Unterschied zwischen Sommer- und Winterhalbjahr. Während im Winterhalbjahr, wo generell die höhere Belastungssituation zu finden ist (Abb. 6.3), keine wesentlichen Unterschiede in den räumlichen Abhängigkeiten zwischen stark belasteten und schwach belasteten Stationen auftreten, wird der Unterschied zwischen beiden Stationstypen im Sommerhalbjahr deutlich. Hier sind auch durchweg die höchsten Interkorrelationen in den Morgenstunden zu finden, wo im Mittel ein rascher Anstieg von den Tiefstständen der Nacht zu den Höchstständen zwischen 7 und 9 Uhr erfolgt. Mit RV- Werten von ca. 60 % bietet sich für das Sommerhalbjahr und für diese Stunden eine Kontrolle auf systematische Fehler an.

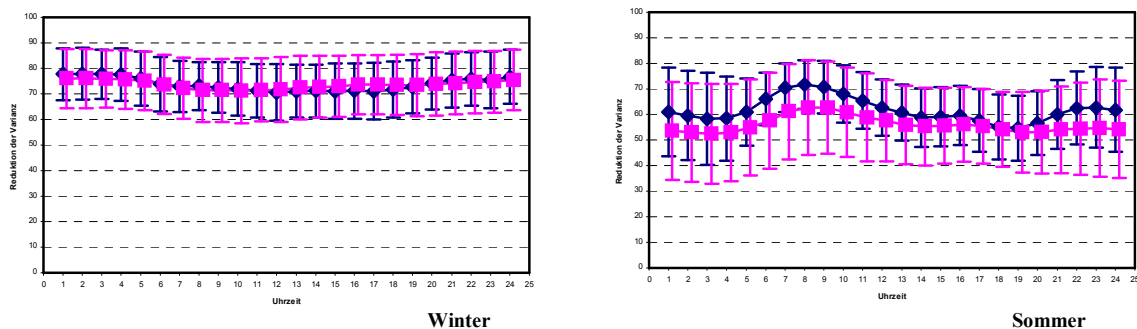


**Abb. 6.3: Erklärte Varianz der mittels umliegender Stationen simulierten stündlichen NO-Zeitreihen** (Mittelwerte und Streuung der Erklärten Varianz (RV) als Mittel über ca. 600 Stationen) Typ1 (schwarz): stark belastete Stationen; Typ2 (rot): schwach belastete Stationen

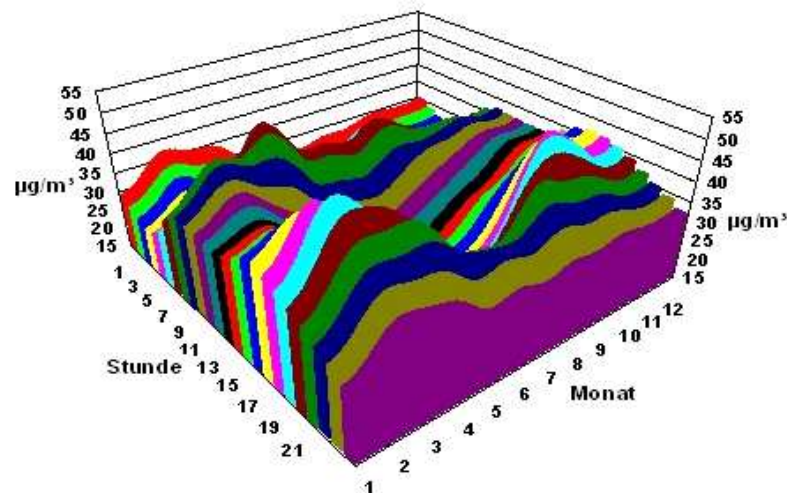


**Abb. 6.4: Jahresgang der NO Konzentration am Beispiel der Station BE034 – Berlin-Neukölln für jede Stunde, Mittel 1990 bis 1999**

Für Stickstoffdioxid ( $\text{NO}_2$ ) sind die räumlichen Interkorrelationen sowohl im Sommer- als auch im Winterhalbjahr deutlich höher als bei Stickstoffmonoxid. Man beobachtet auch hier, dass im Winter praktisch keine tageszeitlichen Schwankungen der räumlichen Abhängigkeiten zwischen den Messreihen zu beobachten sind. Die Höhe des RV- Maßes hängt deutlich vom Betrag der mittleren  $\text{NO}_2$ -Belastung selbst ab (Abb. 6.5). Im Sommer ist ein vergleichbarer Tagesgang der räumlichen Abhängigkeiten wie bei NO vorhanden, wenn auch die Welle in ihrer Amplitude geringer ausfällt. Auch der Unterschied zwischen stark- und schwach belasteten Stationen fällt im Sommer deutlich geringer aus als bei NO. Im Winter existiert kein Unterschied zwischen Stationstyp1 und Stationstyp2.



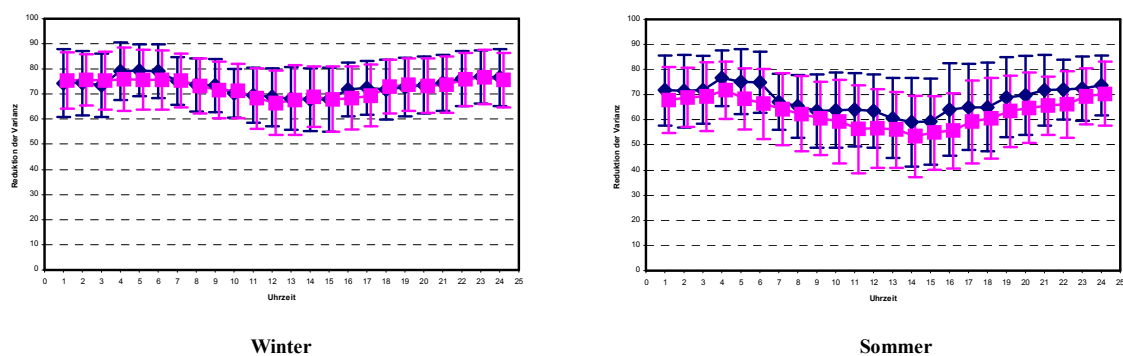
**Abb. 6.5: Erklärte Varianz der mittels umliegender Stationen simulierten stündlichen  $\text{NO}_2$ -Zeitreihen** (Mittelwerte und Streuung der erklärten Varianz (RV) als Mittel über ca. 190 Stationen )



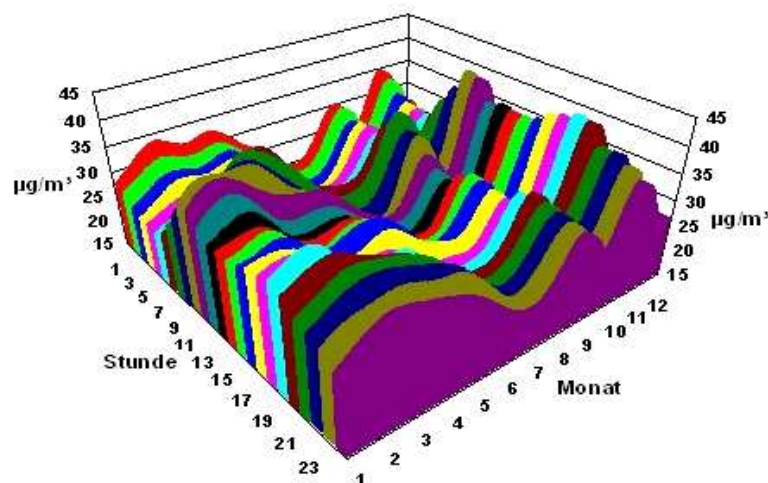
**Abb. 6.6: Jahresgang der  $\text{NO}_2$ -Konzentration am Beispiel der Station BE034 – Berlin-Neukölln für jede Stunde, Mittel 1990 bis 1999**



Für Feinstäube ( $\text{PM}_{10}$ ) ist die derzeitige Datenlage noch unbefriedigend. Es standen nur drei Jahre zur Auswertung zur Verfügung (1998 – 2000). Bei aller Vorsicht kann man jedoch aus Abb. 6.7 und Abb. 6.8 ableiten, dass eine recht gute räumliche Interkorrelation zwischen den Messreihen sowohl im Sommerhalbjahr als auch im Winter vorhanden ist. Zwischen den Stationstypen besteht kein nennenswerter Unterschied. Die Tageswelle mit ihrem Minimum der räumlichen Abhängigkeit zwischen den Messreihen in den Mittagsstunden ist im Sommer stärker ausgeprägt als im Winter. Interessant ist ebenso die ausgeprägte Doppelwelle im jahreszeitlichen Verlauf (Abb. 6.8). Dies ist ein Hinweis für weiterführende Untersuchungen, die Berechnung der simulierten Zeitreihen in Hinblick auf diese Doppelwelle vorzunehmen. Trotz der Einschränkungen hinsichtlich der zur Auswertung zur Verfügung stehenden  $\text{PM}_{10}$ -Datensätze kann mit einiger Sicherheit geschlussfolgert werden, dass sich die beschriebene Methode auch für  $\text{PM}_{10}$  eignet.



**Abb. 6.7: Erklärte Varianz der mittels umliegender Stationen simulierten stündlichen  $\text{PM}_{10}$  - Zeitreihen** (Mittelwerte und Streuung der Erklärten Varianz (RV) als Mittel über ca. 600 Stationen)



**Abb. 6.8: Jahresgang der  $\text{PM}_{10}$  Konzentration am Beispiel der Station BE034 Berlin Neukölln für jede Stunde, Mittel 1998 bis 2000**

Zusammenfassend kann gesagt werden, dass alle hier über Ozon hinausgehenden untersuchten Luftschadstoffkomponenten für eine Validierung auf systematische Fehler mittels der dargestellten Methode geeignet sind. Gewisse Einschränkungen sind gegenwärtig noch für sommerliche  $\text{SO}_2$ - und  $\text{NO}$ -Werte vorhanden. Weiterführende Untersuchungen sollten aber auch dafür Problemlösungen liefern.

## 7 Zusammenfassung

Im abgeschlossenen Vorhaben zur „Verbesserung der bundesweiten Immissionsdaten durch Entwicklung und Anwendung einer Methode zur Validierung und Korrektur kontinuierlich ermittelter Immissionswerte nach einheitlichen Kriterien“ wurde in der ersten Arbeitsphase ein operationell einsetzbares System entwickelt, das die von den Messnetzen der Bundesländer und vom Immissionsmessnetz des UBA dem Fachgebiet II 6.2 im UBA übergebenen Ozondaten nach einheitlichen Kriterien überprüft. Das System ist in der Lage, zufällige Fehler (Ausreißer) wie auch systematische Fehler (Abdriften von Messgeräten usw.) aufzuspüren und zu erkennen und je nach vorgegebener Zielrichtung diese

- fehlerhaften Werte zu markieren,
- als fehlerhaft erkannte Werte durch Ausfallkennungen zu ersetzen,
- als fehlerhaft erkannte Werte und Ausfallwerte durch Schätzwerte zu ersetzen.

Aus dem Strukturbild der vorhandenen Validierungsmodule (Abb.1.1) ist der generelle Aufbau des Validierungskomplexes ersichtlich. Er gliedert sich einerseits in die Validierung nicht plausibler Werte und Ausreißer und andererseits in die Validierung systematischer Fehler (vertikale Gliederung). Die horizontale Gliederung verdeutlicht die Nähe der Validierung zur Echtzeit. Während die Routine Validierung sehr echtzeitnah abläuft, befasst sich die finale Datenvalidierung mit Monats- und Jahresdatensätzen, wie sie von den Messnetzzentralen der Länder und des Umweltbundesamtes geliefert werden.

**Die Routine Prüfung** der Ozondaten und anderer Luftschadstoffkomponenten, so wie sie täglich im UBA eingehen sind wichtig, da fehlerhafte Daten vor ihrer Weiterverarbeitung in operationellen Anwendungen (Ozonprognoserechnung, Erstellung von Karten zur aktuellen Ozonsituation) aus den Datensätzen ausgefiltert werden müssen.

Ferner fungiert das System auch als „Fehlerfrühwarnsystem“, indem vom System angezeigte Signale über mögliche Fehler an die betroffenen Messnetzzentralen der Länder weiter gegeben werden. Insbesondere dürfte dies zur Früherkennung und zu relativ frühzeitigem Gegensteuern bei systematischen Abweichungen von großer Bedeutung sein, da solche Fehler von diesem System objektiv und recht zuverlässig erkannt werden können.

Bei dieser echtzeitnahen Validierung erfolgen zunächst noch keinerlei Veränderungen der Originaldaten. Es wird lediglich für jede Station, für die mögliche Unplausibilitäten gefunden werden, eine Fehlerdatei erstellt. Im operationellen Validierungslauf werden zudem diese Auffälligkeiten in der grafischen Oberfläche des Systems visualisiert. Außerdem können mittels einer speziellen Präsentationssoftware einzelne Stationen mit möglichen fehlerhaften Messwerten, die in Fehlerdateien abgelegt sind, aufgeblendet und sichtbar gemacht werden.

Der eigentliche Validierungslauf kann in mehreren Modi erfolgen:

- vollautomatische Version zum Finden von zufälligen Fehlern
- vollautomatische Version zum Finden von systematischen Fehlern
- vollautomatische Version zum Finden von zufälligen und systematischen Fehlern

**Die interaktive Arbeitsweise** erlaubt eine komplexere Beurteilung der Datenreihen, da im Gegensatz zur echtzeitnahen Validierung mögliche Fehler besser aus dem zeitlichen und räumlichen Kontinuum der Ozonreihen beurteilt werden können. Besonders bei zufälligen Fehlern bietet sich die Möglichkeit einer subjektiven Korrektur an.

Der Validierungslauf kann in mehreren Modi erfolgen:

- interaktive Version zum Finden von zufälligen Fehlern

- interaktive Version zum Finden von systematischen Fehlern
- interaktive Version zum Finden von zufälligen und systematischen Fehlern

Über die echtzeitnahe Ozondaten Validierung hinaus gehend, ist es auch möglich, mit diesem System im Rahmen der „**Finalen Datenvalidierung**“ komplette Ozondatensätze eines ganzen Jahres zu prüfen (also jeden Einzelwert eines Jahres, für jede Stunde und jeden Tag). Im Gegensatz zur echtzeitnahen Datenvalidierung werden jedoch nicht nur Fehlermarkierungen durchgeführt. Vielmehr wird parallel zum Originaldatensatz ein von systematischen und zufälligen Fehlern bereinigter Datensatz erzeugt, in welchem alle Daten die verändert wurden, eine Markierung erhalten. Ebenfalls, wie bei der echtzeitnahen Datenvalidierung, ist hier eine automatische oder interaktive Arbeitsweise wählbar.

### **Methodik zur raum-zeitlichen Kontrolle bei der Prüfung von Einzelwerten**

Die Prüfung von Einzelwerten wird sowohl durch plausible Schranken als auch durch statistische Maßzahlen aus der Zeitreihe selbst und durch die Prüfung der räumlichen Konsistenz vorgenommen.

- Die Sinnfälligkeit eines Messwertes wird durch die Vorgabe plausibler Grenzen geprüft.
- Auf der Basis stationsspezifischer statistischer Kenngrößen (Mittelwert und Streuung) werden die Überschreitung von Schwellenwerten (n-Faches der stationsspezifischen Streuung um den Mittelwert) getestet.
- Weiterhin wird die Kontinuität des Tagesganges der Messgröße untersucht. Für jede Prüfkomponeente werden auch hier stationsspezifische Schwellen berechnet.
- Die räumliche Konsistenz der Messwerte wird über die Berechnung der simulierten Stationszeitreihe unter Verwendung benachbarter Stationen geprüft.
- Als nicht plausibel erkannte Werte werden, je nach Programmmodus, automatisch ersetzt, mit Ausfallkennung versehen oder können interaktiv am Bildschirm verändert werden.

### **Methodik zur Erkennung systematischer Fehler**

Für die Erkennung systematischer Fehler, d.h. für die Erkennung von zeitlich begrenzten Inhomogenitäten wird ein abgewandeltes Verfahren nach Alexandersson (1986) verwendet.

Zu diesem Zweck wird die Differenzreihe (Zeitreihe der Residuen) verwendet, wie sie sich als Differenz zwischen den beobachteten Werten und einer aus den umliegenden Stationen approximierten Zeitreihe ergibt. Das gleitende Mittel dieser Residuenreihen liefert einen guten Anhaltspunkt über mögliche systematische Fehler in diesen Zeitreihen, da man mit Recht davon ausgehen kann, dass über eine längere Zeitspanne gemittelt die statistischen Charakteristika der Abhängigkeiten zwischen den Stationen nur geringfügig schwanken. Treten Sprünge, langsame Verschiebungen oder Wellen in diesen gleitenden Mitteln auf, so deutet dies auf mögliche systematische Fehler in den Messreihen hin.

Zur Erkennung systematischer Fehler wird in folgenden Schritten vorgegangen:

- Der Gesamtdatensatz (mindestens 3 Jahre) wird jeweils in Sommer- und Winterhalbjahr (April bis September bzw. Oktober bis März) unterteilt. Dies trägt dem Umstand Rechnung, dass die meteorologischen Bedingungen zur Ozonbildung bzw. zum Ozonabbau zwischen Sommer- und Winterhalbjahr sehr verschieden sind und somit auch die statistischen Relationen zwischen den Ozonmessreihen deutlichen jahreszeitlichen Abhängigkeiten unterliegen.
- Zur Stabilisierung der Aussagen über mögliche systematische Fehler in den Zeitreihen,

werden um die zu untersuchende Tageszeit fünfstündige Mittelwerte gebildet. Dadurch werden zufällige Schwankungen in den Messreihen stärker geglättet, so dass systematische Fehler deutlicher hervortreten.

- Alle Datensätze werden vor der Berechnung der räumlichen Abhängigkeiten zwischen den Messstationen vom Jahresgang und linearen Trend bereinigt.
- Berechnung der Residuenreihe als Differenz zwischen der beobachteten Ozonreihe (5-stündiges Mittel) und der mittels multipler Regression simulierten Reihe.
- Erneute Trendbereinigung der Residuenreihe.

Für die **realtime- Datenvalidierung** wird innerhalb des Zeitintervalls der zurückliegenden 50 Tage die maximale Änderung der Residuenreihe mittels linearer Regression gesucht. Überschreitet dieser berechnete Wert eine stationsspezifische Schwelle, so liegt möglicherweise ein systematischer Fehler vor. Die stationsspezifischen Schwellenwerte werden über die entsprechende Streuung der Residuenreihe berechnet. D.h. je ungenauer die Simulation einer Ozonzeitreihe ist um so höher bestimmen sich die Schwellenwerte, die zum Auslösen eines Warnsignals führen.

Zur **finalen und interaktiven Datenvalidierung** dienen die von rechts und von links kommenden gleitenden Mittel ( $\pm 15$  Tage) der Residuenreihe als Validierungsgröße. Überschreiten diese gleitenden Mittel einen stationsspezifischen Schwellenwert, der gleichfalls aus der Streuung der Residuenreihe berechnet wird, so stellt der Schnittpunkt zwischen den gleitenden Mitteln den Schwerpunkt des Bereiches eines möglichen systematischen Fehlers dar. Die maximale Differenz der gleitenden Mittel rechts und links vom Schnittpunkt markieren den Beginn und das Ende des so beanstandeten Intervalls

### **Interaktive Datenvalidierung**

Zusätzlich zur Version für den vollautomatischen Lauf der Validierung wurde ein Modul entwickelt, dass eine interaktive Arbeitsweise erlaubt. Dabei können sowohl Monats- als auch Jahresdatensätze einer komplexen Wertung unterzogen werden. Diese Datenvalidierung stützt sich auf länderübergreifende statistische Relationen zwischen den Ozonmessreihen.

Ausreißer und mögliche systematische Fehler werden bei Anwendung dieses Moduls dokumentiert und können je nach Aufgabenstellung bei der Weiterverarbeitung (Präsentation oder Berechnung aggregierter Größen) eingearbeitet werden. Dabei bleiben die Originalreihen erhalten. Die interaktive Datenvalidierung besteht zum einen aus dem Modul zur Identifikation möglicher systematischer Fehler und zum anderen aus einem Modul zur Auffindung nicht plausibler Werte

### **Finale Datenvalidierung**

Die finale Datenvalidierung hat die Aufgabe, die von den Ländern und dem Messnetz des UBA zur Verfügung gestellten Jahresdatensätze einer komplexen Wertung zu unterziehen. Sie besteht zum einen aus dem Modul zur Identifikation möglicher systematischer Fehler und zum anderen aus einem Modul zur Auffindung nicht plausibler Werte

Das dazu eingesetzte Validierungsmodul beinhaltet sowohl eine vollautomatische Kontrolle und Korrektur der Datenreihen als auch die Möglichkeit einer interaktiven Kontrolle der Einzelreihen. Während bei der interaktiven Validierung systematischer Fehler einzelner Reihen subjektiv beurteilt werden können, führt die automatische finale Validierung systematischer Fehler eine automatische Erkennung und Korrektur systematischer Abweichung in den Datenreihen durch. Diese Validierung erfolgt iterativ in 3 Stufen. Dadurch wird gewährleistet, dass systematische Fehler, die fälschlicherweise von einer anderen Messreihe verursacht wurden, und als Fehler der zu prüfenden Messreihe interpretiert

wurden, in der nächsten Iteration wieder korrigiert werden können. Weiterhin ist zu beobachten, dass das Validierungsmodul Abschnitte in den Datenreihen, die mit systematischen Fehlern behaftet sind, erst mit der zweiten oder dritten Iteration vollständig bereinigt oder dass Überreaktionen durch eine nachfolgende Validierungsstufe wieder korrigiert werden.

## **8 Ausblick**

Das Gesamtmodul gibt dem Anwender ein Validierungswerkzeug in die Hand, das von vollautomatischer Datenkontrolle bis hin zu hoch komplexer interaktiver Datenkontrolle eine breite Palette von Möglichkeiten bietet. Dieses Validierungswerkzeug steht derzeit jedoch nur für die Luftschadstoffkomponente Ozon zur Verfügung. Voruntersuchungen zeigen, dass auch für Stickoxide, Feinstaub, Schwefeldioxid und VOCs das Verfahren geeignet ist.

Neben der Optimierung des interaktiven und automatischen Validierungsmoduls bietet sich eine Verbesserung der zugrundeliegenden statistischen Methodik an.

Bei Luftschadstoffen liefern univariate Modelle häufig nur einen relativ geringen Erklärungsanteil. Zusätzliche Erklärung sind durch multivariate Modelle, wie z.B. multiple Regression, zu erzielen. Die erklärenden Variablen können andere Messgrößen der gleichen Station sein (endogene Variable) oder auch die gleichen oder auch andere Messgrößen anderer Stationen (exogene Variable) sein. Im vorliegenden Modell wurden als exogene Variablen, nur die gleichen Messgrößen anderer Stationen verwendet

Eine weitere Verbesserung des Verfahrens lassen Transfer-Funktions-Modelle oder ARMAX-Modelle erwarten, die sowohl endogene als auch exogene Informationen verwenden.

## 9 Literatur

**Alexandersson, H. (1986):** A homogeneity test applied to precipitation data. J. Climatol. 6 S. 661 - 675.

**Bernhofer, Ch. (2001):** CLISAX – Statistische Untersuchungen regionaler Klimatrends in Sachsen, Abschlussbericht AZ: 3-8802.3521/48 Technische Universität Dresden.

**Buishand, J. A. (1982):** Some Methods for Testing the Homogeneity of Rainfall Records. J. Hydrol. 58, S. 11-27 1982.

**Craddock, J. M. (1979):** Methods for Comparing Annual Rainfall Records for Climatic Purposes. Weather 34, S. 322-346 1979.

**Dyck, S. u. a. (1980):** Angewandte Hydrologie. Teil 1: Berechnung und Regelung des Durchflusses der Flüsse. 2. Auflage. Berlin: Verlag für Bauwesen 1980.

**Enke, W., 1988** Ein erwartungstreues lineares Selbstorganisationsmodell Z. f. Meteorol., 35, 5, S. 308-314

**F+E-Vorhaben 297 42 848** des Umweltbundesamtes Berlin: „Analyse historischer Immissionsmessreihen und Entwicklung einer Methode zur quasi -wetterbereinigten Trendanalyse von bodennahem Ozon“ – 1. Etappe: Trendanalyse der täglichen Ozonmaxima der Jahre 1990 bis 1997)

**F+E Vorhaben 297 42 848** „Analyse historischer Datenreihen und Entwicklung einer Methode zur quasi–wetterbereinigten Trendanalyse von bodennahem Ozon“ Teilbericht 3 „Wetterbereinigung der Jahre 1980 bis 1997

**Schönwiese, C.-D.; Malcher, J. (1985):** Nicht-Stationarität oder Inhomogenität? Ein Beitrag zur statistischen Analyse klimatologischer Zeitreihen. Wetter und Leben 37, S. 181 - 193 1985.

**Taubenheim, J. (1969):** Statistische Auswertung geophysikalischer und meteorologischer Daten, Akademische Verlagsgesellschaft Geest&Portig K.G. Leipzig

Für die Betreuung des Vorhabens und viele hilfreiche Diskussionen möchten wir den Herren Berndt Kahra, Dr. Wolf Dieter Garber, Rainer Lacombe und Frau Susanne Grittner herzlichen Dank sagen.