



Technische Universität Graz

Bakkalaureatsprojekt TM (Statistik)

Leiter: Universitätsprofessor DI Dr. Ernst Stadlober

# Zeitreihen NO, SO<sub>2</sub> und CO für Graz-Don Bosco 2005/06

Doris Schadler (0630515)

Technische Mathematik (F033201)

23. April 2012

## **Vorwort**

Ich erkläre, dass ich diese Arbeit selbst verfasst, alle verwendeten Quellen zitiert und mich keiner unerlaubten Hilfsmittel bedient habe.

Graz, 23. April 2012



## Einleitung

Auf Grund der Schadstoffdaten, gemessen in Halbstundenwerten vom 1.1.2005 bis zum 31.12.2006, sollen mittels dieser Arbeit die Strukturen der Zeitreihen untersucht werden. Insbesondere sollen hiermit die Daten zu den Schadstoffen Stickstoffmonoxid (NO), Schwefeldioxid (SO<sub>2</sub>) und Kohlenmonoxid (CO) analysiert werden. Gemessen wurden diese Daten an der Messstelle Graz-Don Bosco. Mittels explorativer Analyse und Zeitreihenanalyse werden die Halbstundenwerte und auch die berechneten 4-Stunden-Mittelwerte und Tagesmittelwerte untersucht.

Bei der explorativen Analyse werden die Daten mittels statistischer Kenngrößen betrachtet. Um den Sachverhalt allerdings um einiges klarer und anschaulicher darzustellen, werden auch Boxplots und Boxplotserien analysiert. Bei diesen hat man einen Überblick über die interessanten Kenngrößen wie Median oder Schiefekoeffizient, kann sich aber allerdings noch ein Bild über die Streuung und sogenannte „Ausreißer“ machen. Weiters werden die Daten auf mögliche Zusammenhänge untersucht. Dies geschieht mittels Scatterplotmatrizen. Numerisch misst man die Stärke des linearen Zusammenhangs mittels des Korrelationskoeffizienten von Pearson.

Die Zeitreihenanalyse behandelt neben der Betrachtungen der Zeitreihen an sich auch Zerlegungen der Zeitreihen in verschiedene Zeitskalen, um mögliche saisonale Trends aufzuzeigen. Abschließend betrachtet man noch die Autokorrelation- und Kreuzkorrelationfunktionen.

## Abstract

This paper critically analyses structures in time series, based on the pollutant data, which was measured every half hour, from January 1st, 2005 to December 31th, 2006. Especially the measurements of the air pollutants NO, SO<sub>2</sub> and CO are evaluated. Additionally, this data has been generated at the control point Graz-Don Bosco. This half hour data, the calculated four-hour mean value as well as the daily mean value were investigated by using explorative analysis and time series analysis.

In the explorative analysis, the data is closely examined based on statistical parameters. To illustrate these parameters in an efficient way, this paper focuses on graphics called boxplot und boxplot series. They show the important measurements like median and skewness, but also data points, which have an extreme high or low value. These points are called „outliers“. Furthermore, possible coherence is analysed using scatterplot matrices and the strength of the linear coherence is computed by the coefficient of correlation from Pearson.

The time series analysis investigates the time series itself and also the decomposition of time series in different scales, to find out the possibility of a seasonal trend. Finally, this document deals with the autocorrelation- und cross-correlation-functions.

# Inhaltsverzeichnis

<b>1</b>	<b>Allgemeines</b>	<b>4</b>
1.1	Beschreibung der Messwerte . . . . .	4
1.1.1	Stickstoffmonoxid (NO) . . . . .	4
1.1.2	Schwefeldioxid (SO <sub>2</sub> ) . . . . .	4
1.1.3	Kohlenmonoxid (CO) . . . . .	4
1.2	Beschreibung des Messortes . . . . .	5
1.3	Angewandte Formeln . . . . .	6
<b>2</b>	<b>Explorative Analyse der einzelnen Schadstoffe</b>	<b>8</b>
2.1	Schadstoff: Stickstoffmonoxid - NO . . . . .	8
2.1.1	Vollständige Betrachtung . . . . .	8
2.1.2	Unterscheidung nach dem Jahr . . . . .	9
2.1.3	Unterscheidung nach der Saison . . . . .	11
2.1.4	Unterscheidung nach dem Monat . . . . .	13
2.1.5	Unterscheidung nach dem Tagestyp . . . . .	16
2.1.6	Unterscheidung nach dem Wochentag . . . . .	18
2.1.7	Unterscheidung nach der Uhrzeit . . . . .	20
2.2	Schadstoff: Schwefeldioxid - SO <sub>2</sub> . . . . .	23
2.2.1	Vollständige Betrachtung . . . . .	23
2.2.2	Unterscheidung nach dem Jahr . . . . .	24
2.2.3	Unterscheidung nach der Saison . . . . .	26
2.2.4	Unterscheidung nach dem Monat . . . . .	28
2.2.5	Unterscheidung nach dem Tagestyp . . . . .	31
2.2.6	Unterscheidung nach dem Wochentag . . . . .	33
2.2.7	Unterscheidung nach der Uhrzeit . . . . .	35
2.3	Schadstoff: Kohlenmonoxid - CO . . . . .	38
2.3.1	Vollständige Betrachtung . . . . .	38
2.3.2	Unterscheidung nach dem Jahr . . . . .	39
2.3.3	Unterscheidung nach der Saison . . . . .	41
2.3.4	Unterscheidung nach dem Monat . . . . .	43
2.3.5	Unterscheidung nach dem Tagestyp . . . . .	46
2.3.6	Unterscheidung nach dem Wochentag . . . . .	48
2.3.7	Unterscheidung nach der Uhrzeit . . . . .	50
<b>3</b>	<b>Zusammenhangsanalyse</b>	<b>54</b>
3.1	Zusammenhang der drei Schadstoffe . . . . .	54
3.1.1	Scatterplotmatrizen . . . . .	54
3.1.2	Scatterplotmatrizen bzgl. Tagestyp . . . . .	57
3.1.3	Scatterplotmatrizen bzgl. Saison . . . . .	60
3.2	Spezialfall: Zusammenhang von NO und CO . . . . .	63
<b>4</b>	<b>Zeitreihenanalyse</b>	<b>64</b>
4.1	Schadstoff: Stickstoffmonoxid - NO . . . . .	64
4.2	Schadstoff: Schwefeldioxid - SO <sub>2</sub> . . . . .	68
4.3	Schadstoff: Kohlenmonoxid - CO . . . . .	73
4.4	Kreuzkorrelation . . . . .	77
<b>5</b>	<b>Conclusio</b>	<b>79</b>

# 1 Allgemeines

## 1.1 Beschreibung der Messwerte <sup>1</sup>

### 1.1.1 Stickstoffmonoxid (NO)

Stickstoffmonoxid ist ein farbloses, erst in höheren Konzentrationen stechend riechendes Gas. Es gehört zur Gruppe der Stickoxide und bildet sich bei der Verbrennung von Erdölprodukten und Kohle. Hauptemittent ist der Verkehr. Stickstoffmonoxid ist nie in so hohen Konzentrationen vorhanden, dass Menschen, Tiere oder Pflanzen gefährdet werden, kann aber schleimhautreizend wirken bzw. durch die Bildung von Methämoglobin toxisch wirken. Da es nicht zu den wichtigsten „klassischen“ Luftschadstoffen europaweit gehört, wurde für die Stickstoffmonoxid-Belastung in Österreich kein Grenzwert festgelegt. In dem untersuchten Datensatz sind 34140 Messungen aufgezeichnet.

### 1.1.2 Schwefeldioxid (SO<sub>2</sub>)

Schwefeldioxid ist ein farbloses, stechend riechendes Reizgas. Es entsteht zum größten Teil bei der Verbrennung von Erdölprodukten und Kohle. Hauptemissionsquellen sind industrielle und gewerbliche Betriebe sowie Feuerungsanlagen des Kleinverbrauchs. In hohen Konzentrationen schädigt Schwefeldioxid Menschen durch Reizung der Atemwege bzw. der Bindehäute der Augen und Pflanzen, da es hauptverantwortlich für den Sauren Regen ist. Die Schwefeldioxid-Belastung zeigt in den letzten 15 Jahren einen deutlich rückläufigen Trend. Der Immissionsgrenzwert zum Schutz der Gesundheit liegt bei einem Tagesmittelwert vom 120 µg/m<sup>3</sup>. Die Anzahl der tatsächlichen Messwerte im Datensatz ist bei diesem Schadstoff am geringsten und liegt bei 33756.

### 1.1.3 Kohlenmonoxid (CO)

Kohlenmonoxid ist ein farbloses, geruchloses, sehr giftiges und brennbares Gas, das sich hauptsächlich bei unvollständiger Verbrennung von organischen Stoffen, Brenn- und Treibstoffen und bei Fäulnis entwickelt. Hausbrand und Verkehrsabgase zählen zu den größten Quellen. In beiden Bereichen gab es in den letzten Jahren jedoch eine starke Emissionsreduktion. Im Verkehrssektor wurde dies u.a. durch die Einführung des Katalysators erreicht. Im Bereich der Haushalte waren veraltete Heizungsanlagen für den hohen CO-Ausstoß verantwortlich. Da Kohlenmonoxid zur weltweiten Bildung von Ozon beiträgt, sollte die Emission weiter verringert werden. Zur Zeit sind 10 µg/m<sup>3</sup> der Immissionsgrenzwert zum Schutz der Gesundheit für den Gleitenden Achtstundenmittelwert. Kohlenmonoxid ist vor allem wegen der humantoxischen Wirkung von Bedeutung. Es beeinträchtigt nämlich die Sauerstoffaufnahmekapazität des Hämoglobins. Der zugrundeliegende Datensatz enthält 34185 Messwerte.

---

<sup>1</sup><http://www.umweltbundesamt.at>, <http://www.umwelt.steiermark.at>

## 1.2 Beschreibung des Messortes <sup>2</sup>

Die Messdaten wurden bei der Luftgütemessstation Graz-Don Bosco aufgenommen. Sie liegt in der Harter Straße 1, 8053 Graz und auf einer Seehöhe von 358m. Die Station mit Stationsnummer 164 wird vom Amt der Steiermärkischen Landesregierung, Fachabteilung 17 C, Referat Luftgüteüberwachung, betrieben. Das Ziel der Messungen ist die Erfassung der Immissionsstruktur im verkehrsnahen Bereich. Die Station wurde im Jänner 2000 in Betrieb genommen. Zu den erfassten Schadstoffkomponenten gehören Schwefeldioxid, Schwebstaub, Feinstaub, Stickstoffmonoxid, Stickstoffdioxid und Kohlenmonoxid, zu den erfassten Klimaelementen Lufttemperatur und relative Feuchte. Die gesetzlichen Grundlagen bieten das Bundesluftreinhaltegesetz und das Immissionsgesetz-Luft.

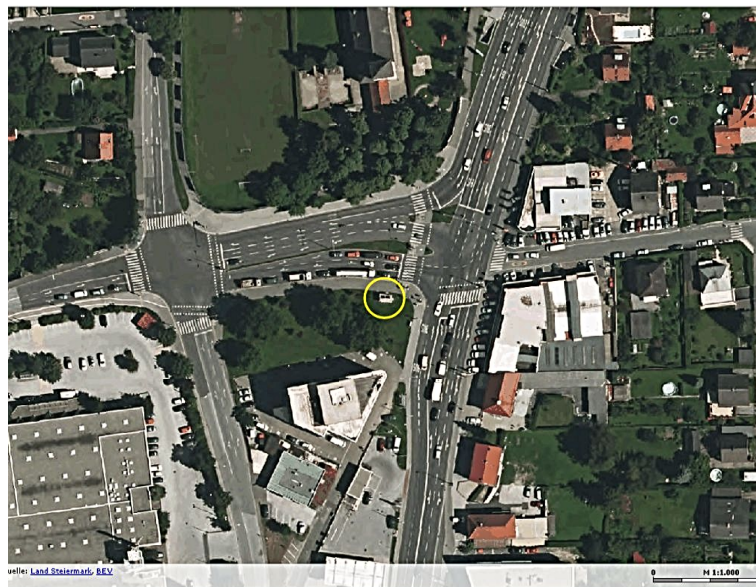


Abbildung 1.1: Lage der Luftgütemessstation Graz-Don Bosco



Abbildung 1.2: Lage der Luftgütemessstation Graz-Don Bosco

<sup>2</sup><http://www.umwelt.steiermark.at>, <http://www.gis.steiermark.at>

### 1.3 Angewandte Formeln

Für die Analyse der Messdaten mittels statistischer Kenngrößen werden folgende Maße verwendet:

- **Lagemaße:** Median(Zentralwert) und Arithmetisches Mittel(Mittelwert)

Sei  $(x_1, \dots, x_n)$  eine Stichprobe aus metrischen Daten, dann heißt

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

arithmetisches Mittel.

Sei  $(x_1, \dots, x_n)$  eine zumindest ordinal skalierte Stichprobe und  $(x_{(1)} \leq \dots \leq x_{(n)})$  die Transformation zu einer geordneten Stichprobe, dann unterscheidet man bei der Berechnung des Medians ( $x_{med}$ ) zwei Fälle:

$$n \text{ ungerade: } x_{med} = x_{(\frac{n+1}{2})}$$

$$n \text{ gerade: } x_{med} = \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)})$$

- **Minimum** ( $x_{min}$ ), **Maximum** ( $x_{max}$ ) und **Spannweite** ( $x_{max} - x_{min}$ )
- **Streuungsmaße:** Empirische Standardabweichung, 1. und 3. Quartil, und interquartiler Bereich

Sei  $(x_1, \dots, x_n)$  eine Stichprobe aus metrischen Daten, dann heißt

$$s_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

empirische Standardabweichung.

Der interquartile Bereich wird definiert als

$$iqr = q_{0.75} - q_{0.25},$$

wobei das  $q_{0.25}$ -Quartil auch als das 1. Quartil bezeichnet wird und das  $q_{0.75}$ -Quartil als 3. Quartil, die folgendermaßen berechnet werden:

$$q_p = (1-g)x_{(\lfloor (n-1)p \rfloor + 1)} + gx_{(\lfloor (n-1)p \rfloor + 2)},$$

wobei  $g = (n-1)p - \lfloor (n-1)p \rfloor$  und  $(x_{(1)} \leq \dots \leq x_{(n)})$  wiederum die Transformation der Stichprobe zu einer geordneten Stichprobe darstellen.

- **Schiefemaße:** Quantilkoeffizient der Schiefe

Der Quantilkoeffizient der Schiefe wird in dieser Weise berechnet:

$$g_1^q = \frac{q_{0.95} - q_{0.50}}{q_{0.50} - q_{0.05}} - 1,$$

wobei die einzelnen Quantile wie oben erwähnt berechnet werden.

Für die Analyse des linearen Zusammenhangs zwischen den einzelnen Schadstoffen wird der **Korrelationskoeffizienten von Pearson** verwendet, der wie folgt berechnet wird:

$$r = r_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Bei der Zeitreihenanalysen wird ein Blick auf die **Autokorrelations- und Kreuzkorrelationsfunktion** geworfen, deren Werte wie folgt berechnet werden:

Die Autokorrelationsfunktion ist definiert durch:

$$acf(k) = \frac{1}{n-k} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) / s_X^2, \quad 0 \leq k \leq n-1$$

und der Index  $k$  den lag der Variable indiziert.

Die Kreuzkorrelationsfunktion ist definiert durch:

$$ccf_{X,Y}(k) = \frac{cov_{X,Y}(k)}{s_X s_Y}, \quad 0 \leq k \leq n-1$$

wobei

$$cov_{X,Y}(k) = \frac{1}{n-k} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), \quad 0 \leq k \leq n-1$$

## 2 Explorative Analyse der einzelnen Schadstoffe

### 2.1 Stickstoffmonoxid - NO

#### 2.1.1 Vollständige Betrachtung

Um einen ersten Eindruck über die Verteilung des Schadstoffes Stickstoffmonoxid zu gewinnen, betrachtet man zunächst einige wichtige statistische Kenngrößen, die in der nachfolgenden Tabelle aufgezeichnet sind.

Tabelle 2.1: Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ )

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
NO	38.88	71.37	0.020	808.30	87.24	15.97	90.60	1.49

Man sieht, dass mit  $808.28 \mu\text{g}/\text{m}^3$  eine große Spannweite vorliegt, der interquartile Bereich aber nur ungefähr ein Zehntel dieser Differenz, nämlich  $74.63 \mu\text{g}/\text{m}^3$ , beträgt. Der maximale Wert ist mehr als zehnmal so groß wie der Mittelwert. Weiters ist eine Rechtsschiefe der Daten ablesbar.

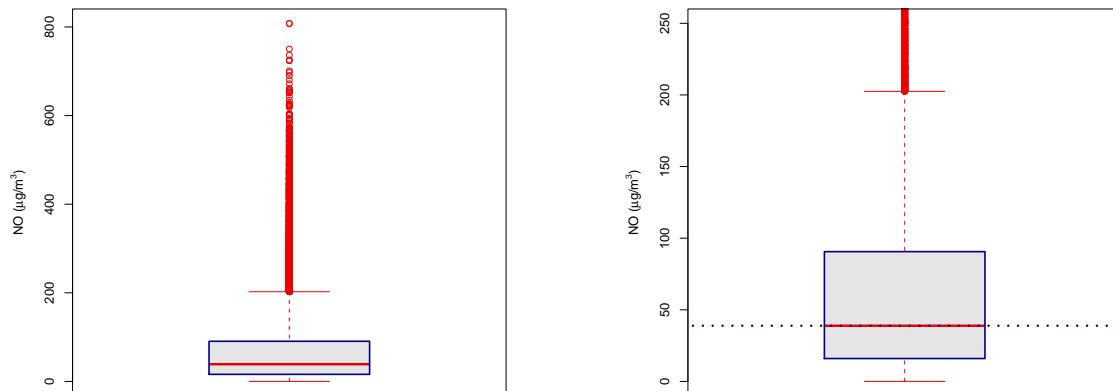


Abbildung 2.1: Boxplot der NO-Halbstundenwerte

Ganz klar zu erkennen ist, dass es keine Ausreißer nach unten gibt. Nach oben gibt es, auf Grund der auch zuvor erkannten großen Spannweite, eine große Anzahl von diesen. Nach oben nimmt die Dichte der Ausreißer jedoch ab. In den Abbildungen 2.1 bis 2.63 stellt die jeweils linke Grafik einen Gesamtüberblick der Daten dar, während in der rechten Grafik die Ausreißer abgeschnitten werden, um die zentralen Daten besser interpretieren zu können.

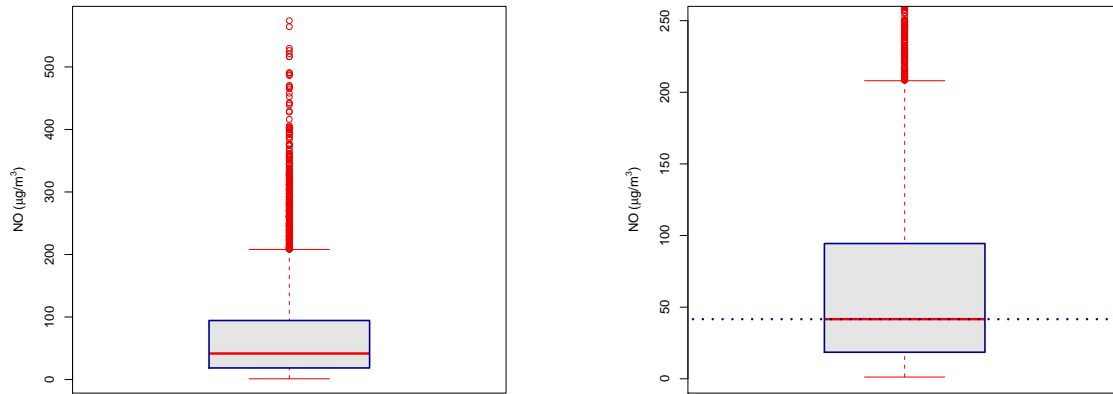


Abbildung 2.2: Boxplot der NO-4-Stunden-Mittelwerte

Da sich der Median innerhalb des interquartilen Bereichs leicht links befindet, kann auf eine Rechtsschiefe der Messdaten geschlossen werden. Der Median liegt bei  $38.88 \mu\text{g}/\text{m}^3$ . Die mittleren 50 % der Daten liegen zwischen  $15.97 \mu\text{g}/\text{m}^3$  und  $90.60 \mu\text{g}/\text{m}^3$ . Ein Viertel der Daten liegt also über einem Wert von  $90.60 \mu\text{g}/\text{m}^3$ .

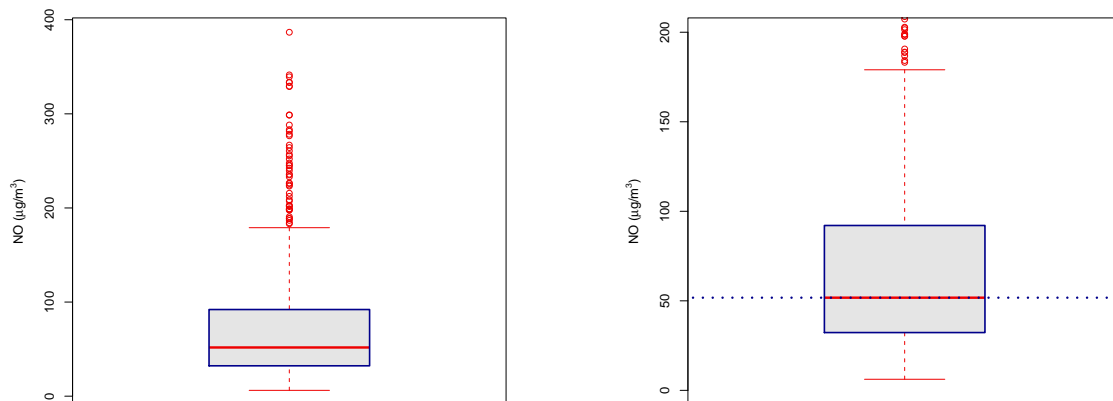


Abbildung 2.3: Boxplot der NO-Tagesmittelwerte

Auch die Analyse mittels Tagesmittelwerte spiegelt die oben erwähnten Erkenntnisse wider, wenn auch hier die Spannweite nur mehr die Hälfte der der Halbstundenwerte beträgt.

### 2.1.2 Unterscheidung nach dem Jahr

Nachdem man nun einen groben Überblick über die Verteilung des Schadstoffes erhalten hat, möchte man feststellen, ob es möglicherweise Auffälligkeiten bezüglich bestimmter Anschauungen gibt, um so ein detaillierteres Bild über die Schadstoffbelastung zu erhalten. Als erstes betrachtet man eine Aufteilung nach den Jahren 2005 und 2006.



Tabelle 2.2: Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Jahr

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
NO	38.88	71.37	0.020	808.30	87.24	15.97	90.60	1.49
- Jahr 2005	37.26	69.50	0.020	807.50	86.40	15.47	88.06	1.55
- Jahr 2006	40.73	73.10	0.180	808.30	88.04	16.43	93.85	1.43

Das Minimum der Gesamtdaten liegt im Jahr 2005, das Maximum im Jahr 2006. Allgemein sind die Größen im Jahr 2006 etwas höher, was darauf schließen lässt, dass die NO-Belastung im Jahr 2006 etwas stärker war. Jedoch scheint bei allen Kenngrößen der Unterschied zwischen den jährlichen Teildatensätzen nicht signifikant zu sein. Auch die nachfolgenden Boxplotserien zeigen dies.

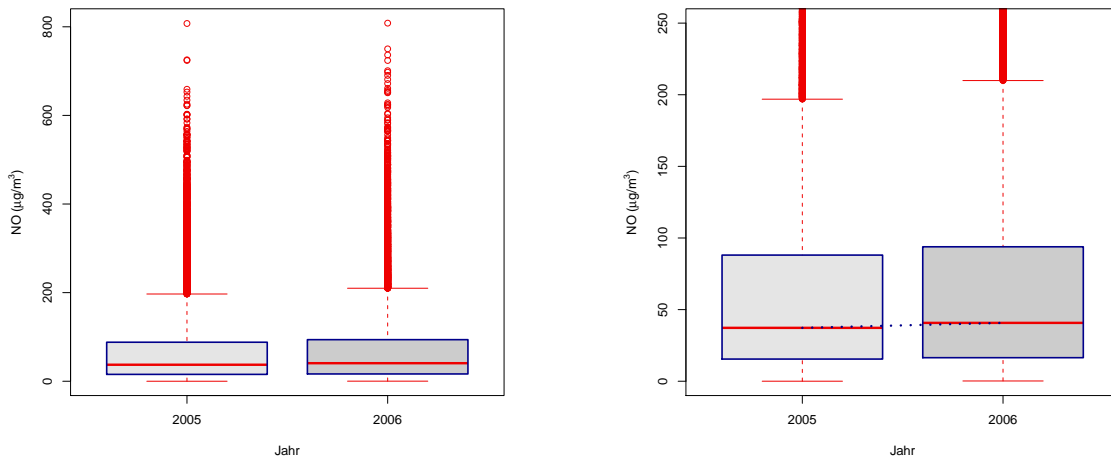


Abbildung 2.4: Boxplotserie der NO-Halbstundenwerte bzgl. Jahr

Die Boxplotserien bestätigen die Aussagen mittels Tabelle. Beide Boxplots sind sehr ähnlich. Der Blick auf die Streuung zeigt noch, dass die Dichte der höheren Ausreißer im Jahr 2006 etwas größer war als im Jahr 2005, was der Grund für den etwas höheren Mittelwert bzw. Median sein kann.

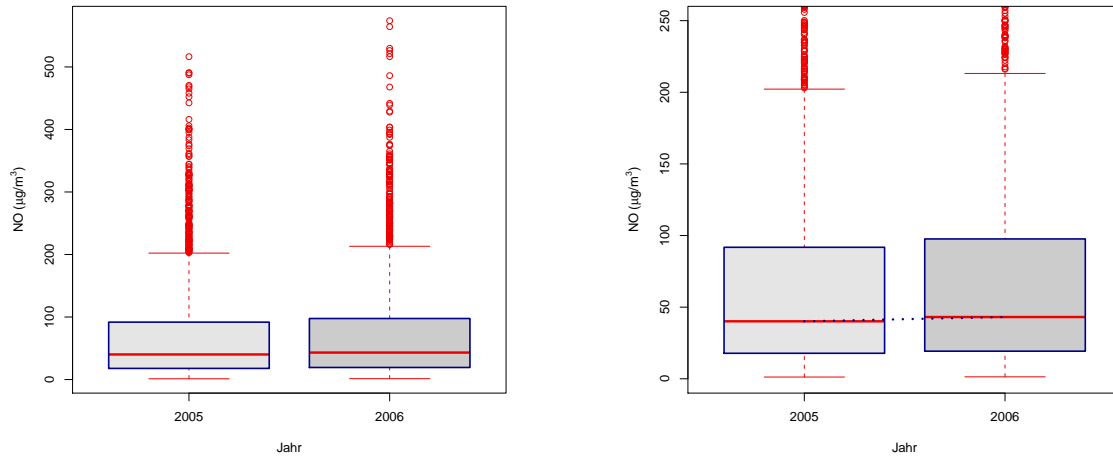


Abbildung 2.5: Boxplotserie der NO-4-Stunden-Mittelwerte bzgl. Jahr

Obige Abbildung zeigt wiederum, dass der Median der Daten im Jahr 2006 mit  $40.73 \mu\text{g}/\text{m}^3$  nur ein wenig höher ist als der im Jahr 2005 mit  $37.26 \mu\text{g}/\text{m}^3$ . Auch bezüglich der Schiefe sind keine großen Unterschiede erkennbar. Daraus folgt, dass auch das 1. und 3. Quartil ungefähr auf gleicher Höhe liegen.

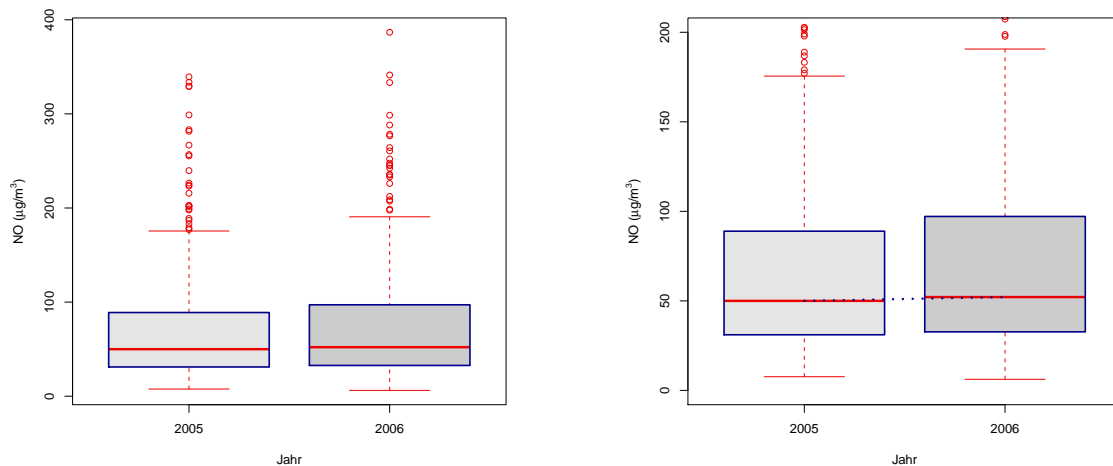


Abbildung 2.6: Boxplotserie der NO-Tagesmittelwerte bzgl. Jahr

Auch die Tagesmittelwerte zeigen dasselbe Bild und so kommt man mittels Kenngrößen und Boxplotserien einheitlich auf das Ergebnis, dass sich die Datensätze aufgeteilt nach den Jahren nicht auffallend unterscheiden.

### 2.1.3 Unterscheidung nach der Saison

Einen durchaus größeren Unterschied sieht man, wenn man sich die Daten geteilt bezüglich der beiden Saisonen, Sommer und Winter, ansieht. Zur Wintersaison gehören hierbei alle Werte die vom 1.1.2005 bis zum 31.3.2005, vom 1.10.2005 bis zum 31.3.2006 und vom 1.10.2006 bis zum 31.12.2006 aufgezeichnet wurden. Die übrigen Werte gehören zur Sommersaison. Hier erkennt man bereits anhand der Tabelle der Kenngrößen, dass es einen deutlichen Unterschied gibt.

Tabelle 2.3: Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Saison

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
NO	38.88	71.37	0.020	808.30	87.24	15.97	90.60	1.49
- Sommer	22.56	37.55	0.020	449.70	44.06	10.43	46.48	1.42
- Winter	71.14	105.40	0.058	808.30	104.94	31.45	142.30	0.90

Sowohl der Mittelwert als auch der Median sind im Winter dreimal so hoch wie im Sommer. Das Maximum wird eindeutig im Winter erreicht, das Minimum im Sommer. Auch der interquartile Bereich ist im Winter ( $110.85 \mu\text{g}/\text{m}^3$ ) um einiges größer als im Sommer ( $36.05 \mu\text{g}/\text{m}^3$ ). Die Daten sind im Sommer schiefer verteilt, streuen aber weniger als im Winter.

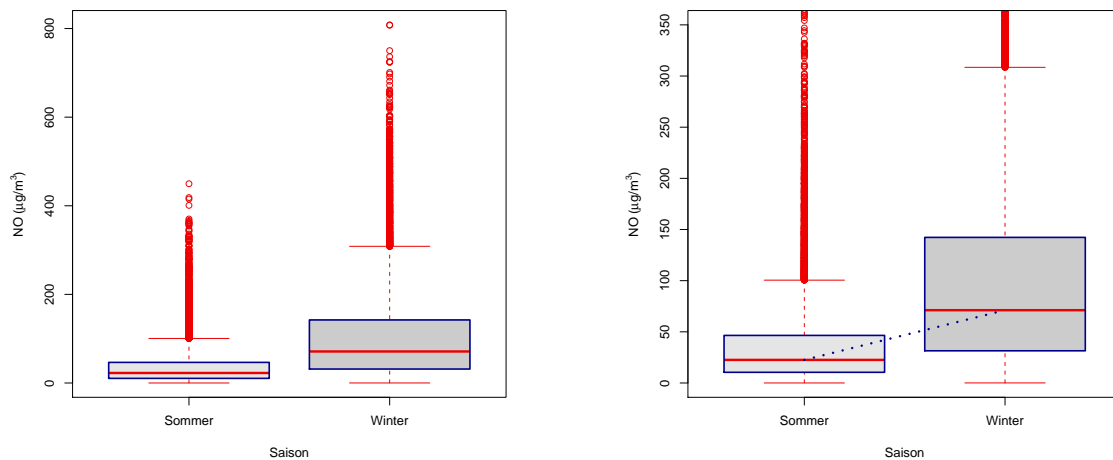


Abbildung 2.7: Boxplotserie der NO-Halbstundenwerte bzgl. Saison

Wirft man nun einen Blick auf die Verteilung der aufgeteilten Daten mittels Boxplot, so sieht man, dass es im Winter deutlich mehr Ausreißer gibt. Der extreme Unterschied wird sichtbar, wenn man das 3. Quartil der Sommer-Daten und den Median der Winter-Daten vergleicht. Im Sommer liegen 75 % unter der Grenze, unter der im Winter weniger als 50 % liegen.

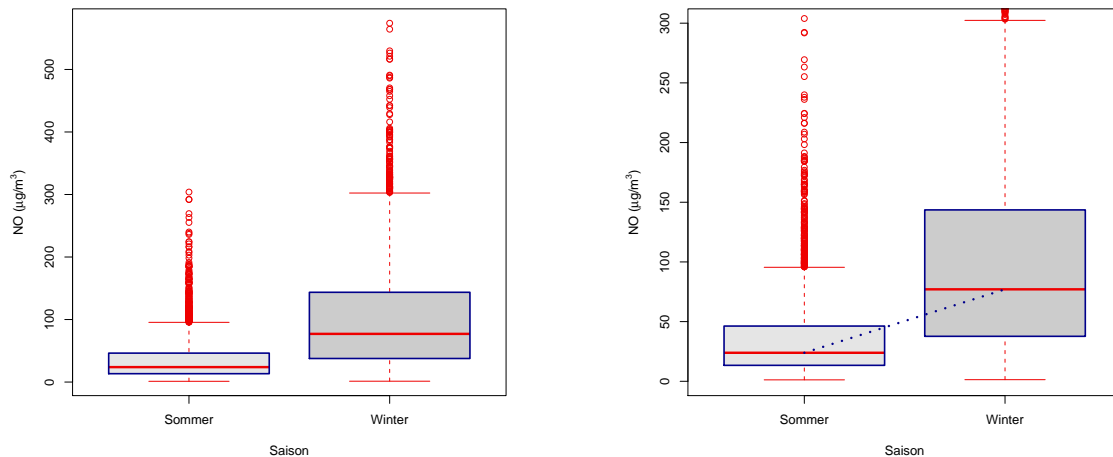


Abbildung 2.8: Boxplotserie der NO-4-Stunden-Mittelwerte bzgl. Saison

Anhand der rechten Grafik kann man erkennen, dass das 1. Quartil des Winter und das 3. Quartil des Sommers beinahe auf einer Höhe liegen. Durch die Mittelung zu 4-Stunden-Mittelwerten kann man also obige Aussage verstärken: Im Sommer liegen 75 % der Daten unter der Grenze, unter der im Winter ungefähr 25 % liegen. Der Unterschied ist also noch extremer.

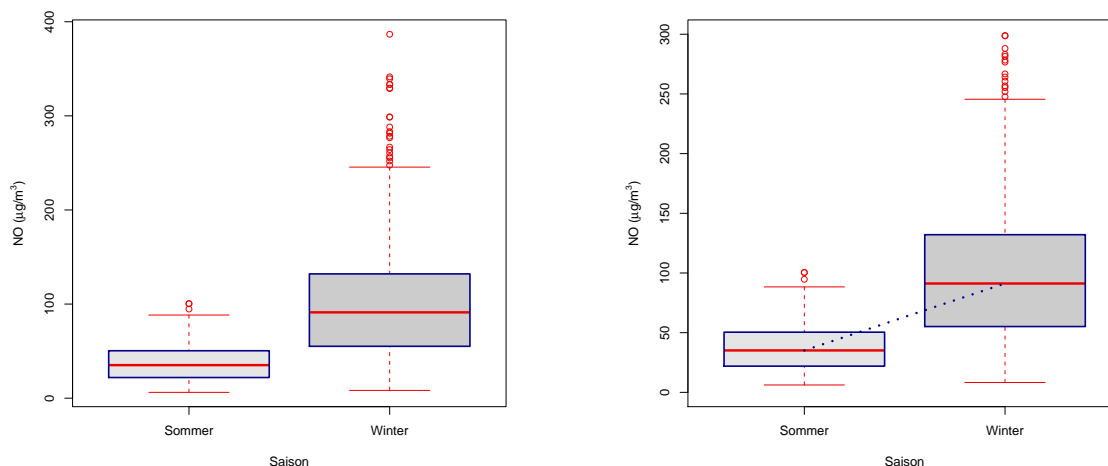


Abbildung 2.9: Boxplotserie der NO-Tagesmittelwerte bzgl. Saison

Auch die Analyse der Tagesmittelwerte zeigt einen deutlichen Unterschied der beiden Teildatensätze. Fazit ist, dass in beiden Wintersaisonen die NO-Belastung um einiges höher ist als in den Sommersaisonen.

### 2.1.4 Unterscheidung nach dem Monat

Da man nun gesehen hat, dass es einen deutlichen Unterschied bezüglich der Saison gibt, scheint es interessant, eine Aufteilung bezüglich der Monate zu wählen, um noch genauer zu sehen in welchen Wintermonaten bzw. Sommermonaten die Maxima bzw. Minima erreicht werden. Hier werden je Monat die aggregierten Werte von 2 Jahren betrachtet.

Tabelle 2.4: Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Monat

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
NO	38.88	71.37	0.020	808.30	87.24	15.97	90.60	1.49
- Jänner	93.98	135.20	0.404	807.50	125.24	45.16	185.40	0.76
- Februar	56.40	88.32	0.058	589.40	92.10	25.33	116.60	1.15
- März	37.09	61.02	0.180	505.90	68.16	17.19	76.64	1.39
- April	27.30	46.25	0.020	449.70	56.76	11.40	55.77	1.76
- Mai	20.76	33.65	0.244	330.50	37.94	9.20	41.85	1.47
- Juni	18.92	29.41	0.108	244.50	31.71	9.01	36.88	1.31
- Juli	17.13	28.83	0.125	245.00	31.48	8.24	36.10	1.57
- August	23.59	36.48	0.072	289.10	38.12	11.58	46.56	1.22
- September	31.32	51.25	0.789	415.10	56.51	15.75	63.62	1.55
- Oktober	57.45	84.16	0.343	583.20	82.47	30.27	108.30	0.97
- November	83.43	109.00	1.191	653.90	96.47	39.79	145.10	0.62
- Dezember	122.90	154.20	1.962	808.30	122.57	63.50	215.40	0.42

Das Maximum der Messwerte wird im Dezember erreicht. Im April gibt es die geringste Stickstoffmonoxidbelastung. Man sieht, dass die Belastung von Jänner bis Juli abfällt und ab August wieder zunimmt. Mediane, Mittelwerte und die Abweichungen fallen bzw. wachsen auf diese Weise. Die Schiefe unterscheidet sich leicht in den Monaten, es kann aber immer von einer Rechtsschiefe gesprochen werden.

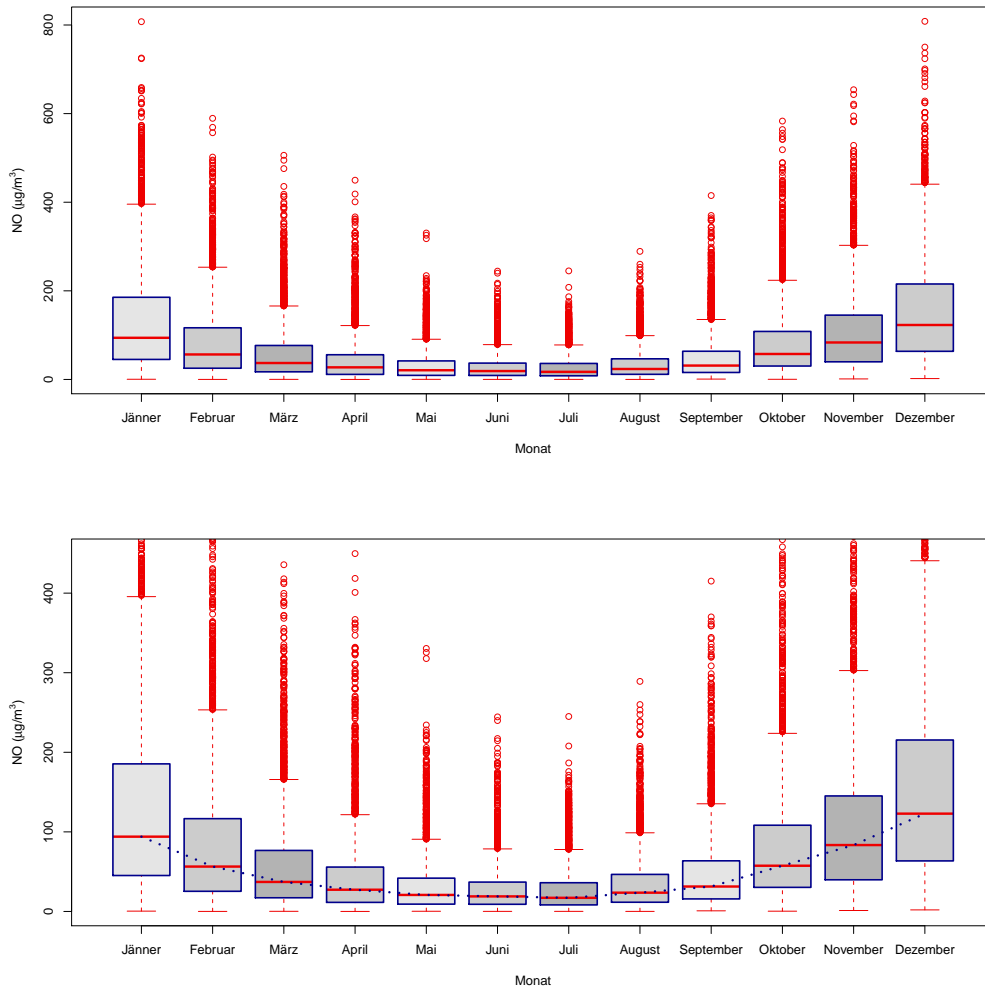


Abbildung 2.10: Boxplotserie der NO-Halbstundenwerte bzgl. Monat

Wie die Werte kontinuierlich von Jänner bis Juli fallen und dann wieder ansteigen, so nimmt auch die Anzahl der Ausreißer in Richtung Sommer ab und in Richtung Winter zu. In den Wintermonaten gibt es also eine größere Streubreite der Daten als in den Sommermonaten. Der Monat Mai zeigt ein paar sehr auffällige Ausreißer.

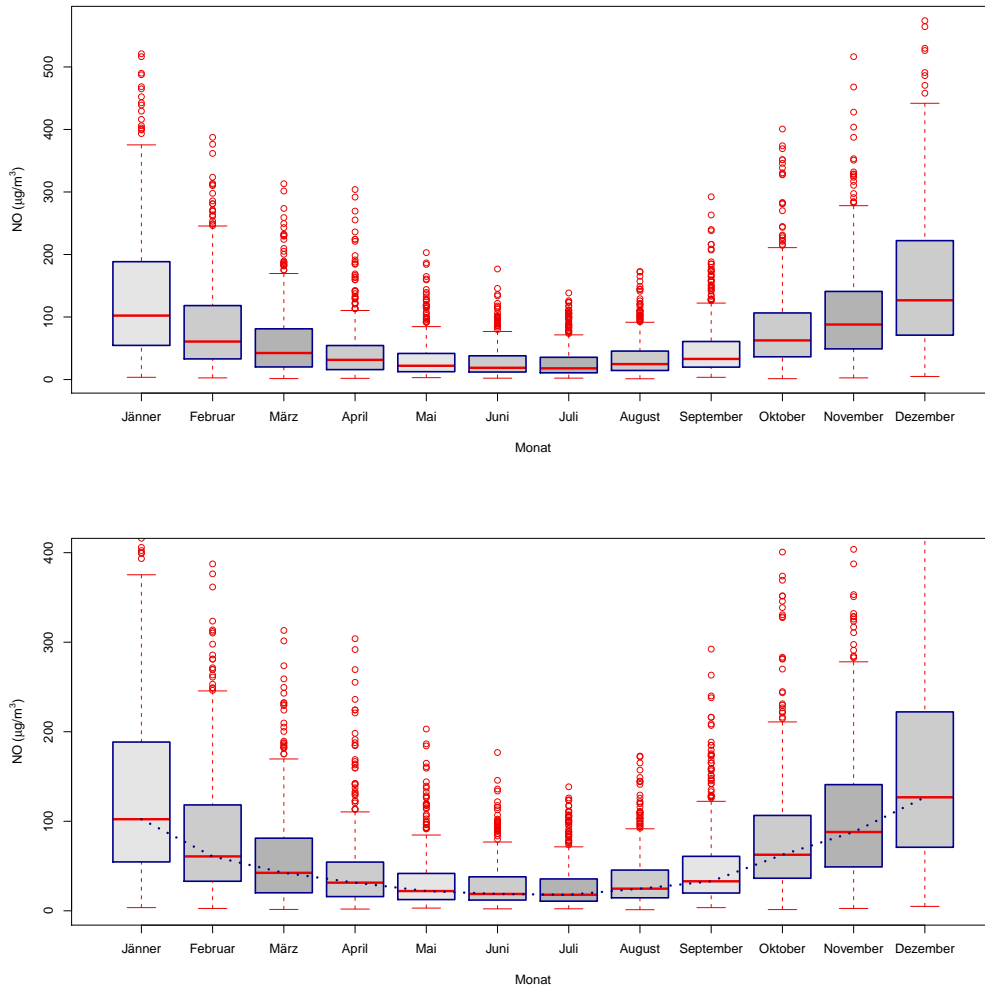


Abbildung 2.11: Boxplotserie der NO-4-Stunden-Mittelwerte bzgl. Monat

Die Mediane steigen bzw. sinken auf die schon zuvor erwähnte Art und Weise. Mit den Medianen ändert sich auch die Größe der interquartilen Bereiche. Je größer der Median, desto größer der Bereich. Auffällig ist weiters, dass die Minima sämtlicher Monate ungefähr auf der selben Höhe liegen, das heißt in jedem Monat gibt es einzelne 4-Stunden-Mittelwerte, die sehr niedrig sind.

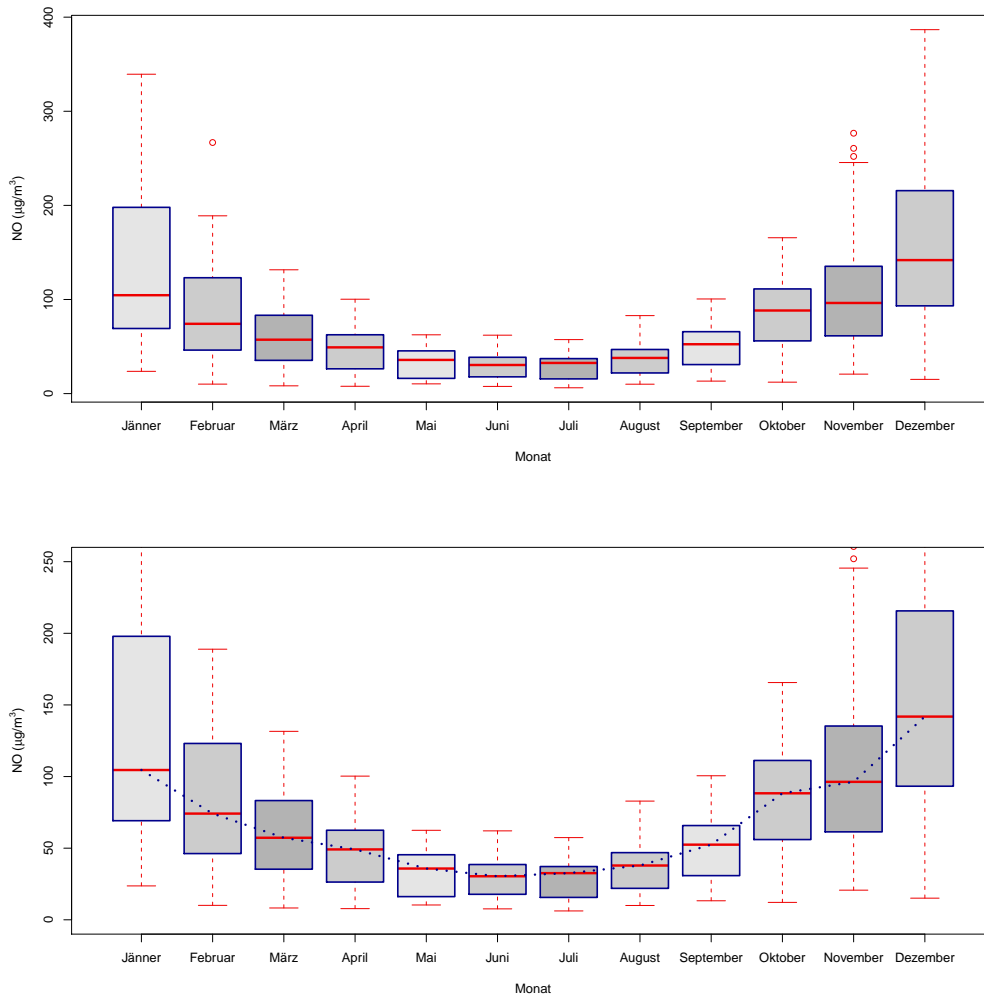


Abbildung 2.12: Boxplotserie der NO-Tagesmittelwerte bzgl. Monat

Auch wenn die Werte auf Tagesmittelwerte zusammengefasst sind, ändert sich das Bild nicht sonderlich. Deutlich zu sehen ist eben der Unterschied zwischen Sommer- und Wintermonaten und der monotone Anstieg bzw. Abfall. Dies ist nicht bei allen Analysen der Fall, wie noch zu sehen sein wird.

### 2.1.5 Unterscheidung nach dem Tagestyp

Als eine weitere wichtige Aufteilung erweist sich die mittels Tagestypen. Bei dieser Unterscheidung fasst man die Tage zu Werktagen, Samstagen und Sonn- bzw. Feiertagen zusammen. Zuerst beginnt man wiederum mit den Kenngrößen.

Tabelle 2.5: Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Tagestyp

Daten	Median	Mittelw.	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
NO	38.88	71.37	0.020	808.30	87.24	15.97	90.60	1.49
- Werktag	50.07	85.22	0.058	808.30	96.32	22.39	111.30	1.26
- Samstag	26.29	48.49	0.292	481.30	58.43	12.35	61.20	1.52
- Sonn-/Feiertag	17.74	36.44	0.020	455.30	46.87	8.86	42.72	1.91

Die Stickstoffmonoxidbelastung ist an den Werktagen am höchsten und an Sonn- bzw. Feiertagen am geringsten. So ist etwa der Median an den Werktagen mit  $50.07 \mu\text{g}/\text{m}^3$  beinahe dreimal so groß wie der an den Sonn- bzw. Feiertagen mit  $17.74 \mu\text{g}/\text{m}^3$ . Weiters wird diese Tatsache auch durch die „Platzierungen“ des Gesamtminimums bzw. Gesamtmaximums ersichtlich.

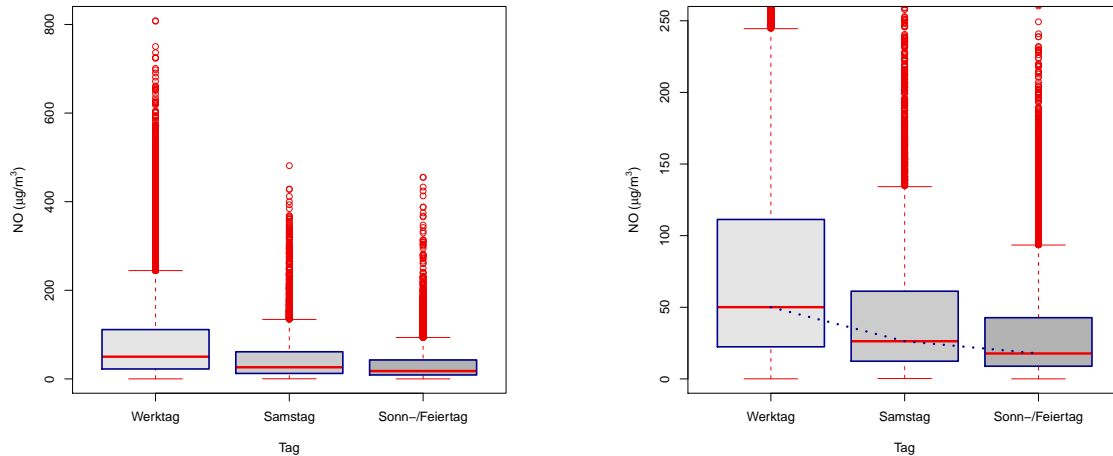


Abbildung 2.13: Boxplotserie der NO-Halbstundenwerte bzgl. Tagestyp

Alle drei Boxplots zeigen einen kleinen interquartilen Bereich und eine große Streubreite, wobei sowohl der interquartile Bereich als auch die Streubreite an den Werktagen am größten sind. Der maximale Wert an den Sonn- bzw. Feiertagen ist beinahe gleich hoch wie der an den Samstagen. Alle drei Boxplots der Serie zeigen eine rechtsschiefe Verteilung.

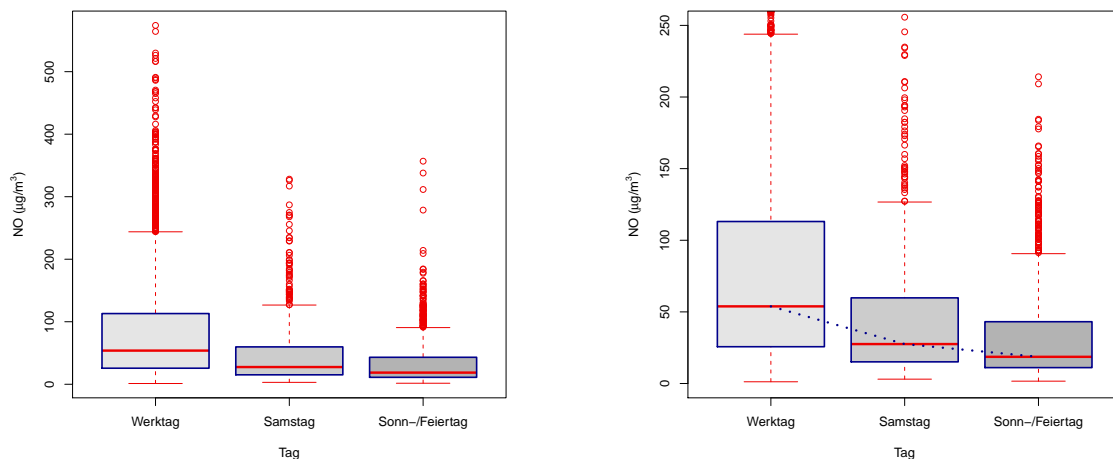


Abbildung 2.14: Boxplotserie der NO-4-Stunden-Mittelwerte bzgl. Tagestyp

Der Median der Werktagsdaten liegt etwas unter dem 3. Quartil der Samstagsdaten und über dem 3. Quartil der Sonn-/Feiertagsdaten. Insgesamt kann man also sagen, dass 50 % der Werte aufgezeichnet an Werktagen unter der Grenze liegen, unter der in etwa 75 % der Werte aufgezeichnet an den restlichen Tagen liegen. Es werden bei allen Teildatensätzen sehr kleine Minima erreicht. Das bedeutet, dass es auch an Werktagen sehr kleine 4-Stunden-Mittelwerte gibt.



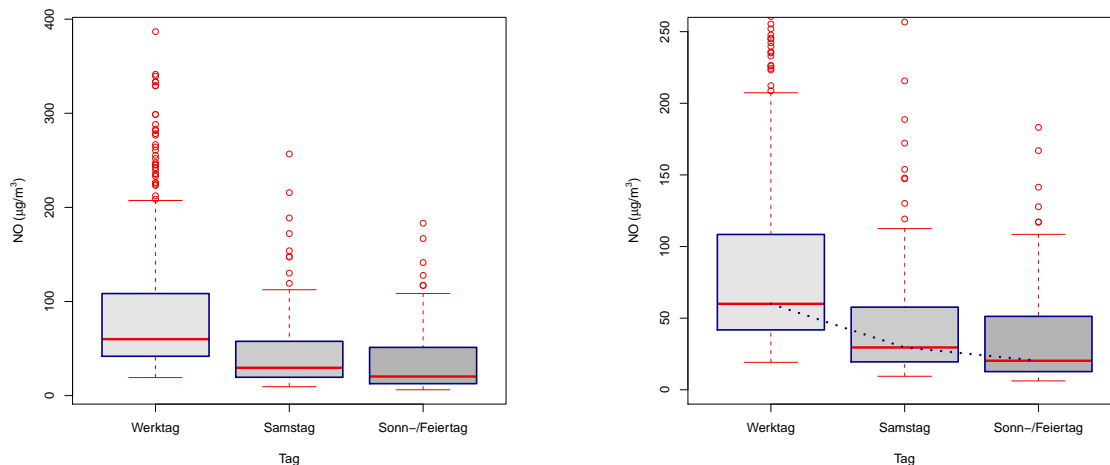


Abbildung 2.15: Boxplotserie der NO-Tagesmittelwerte bzgl. Tagestyp

Die Aufteilung bezüglich der Tage zeigt einen klaren Unterschied zwischen Werktagen und den restlichen Tagen. Bezüglich Samstag zu Sonn- bzw. Feiertagen ist der Unterschied nicht mehr ganz so groß. Dennoch ist die Belastung an den Sonntagen ein wenig kleiner und somit am geringsten.

### 2.1.6 Unterscheidung nach dem Wochentag

Da man nun zuvor gesehen hat, dass es gewisse Unterschiede bezüglich der Tagestypen gibt, möchte man nun wissen, ob es an bestimmten Werktagen besonders starke Belastungen gibt und wirft somit einen noch detaillierteren Blick auf die Sachlage. Die statistischen Kenngrößen der Daten nach dem Wochentag getrennt sehen wie folgt aus:

Tabelle 2.6: Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Wochentag

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
NO	38.88	71.37	0.020	808.30	87.24	15.97	90.60	1.49
- Montag	44.23	74.83	0.058	625.40	87.22	17.07	97.72	1.23
- Dienstag	50.03	85.52	0.125	807.50	94.92	21.94	114.80	1.10
- Mittwoch	52.03	87.15	0.218	653.90	97.17	23.87	112.50	1.34
- Donnerstag	47.15	84.87	0.072	736.40	100.39	21.00	109.00	1.35
- Freitag	50.91	85.09	0.139	808.30	95.25	23.38	110.40	1.30
- Samstag	26.57	49.24	0.292	481.30	58.98	12.44	61.86	1.56
- Sonntag	17.05	33.89	0.020	455.30	44.81	8.64	38.86	2.00

Die Kenngrößen von Montag bis Freitag liegen über den Kenngrößen des Gesamtdatensatzes, unterscheiden sich aber nicht allzu stark voneinander. Belastungsstärkster Tag ist der Mittwoch. Belastungsgeringster Tag bei den Werktagen ist der Montag, dessen Größen dennoch noch deutlich über denen der Sonntage liegen.

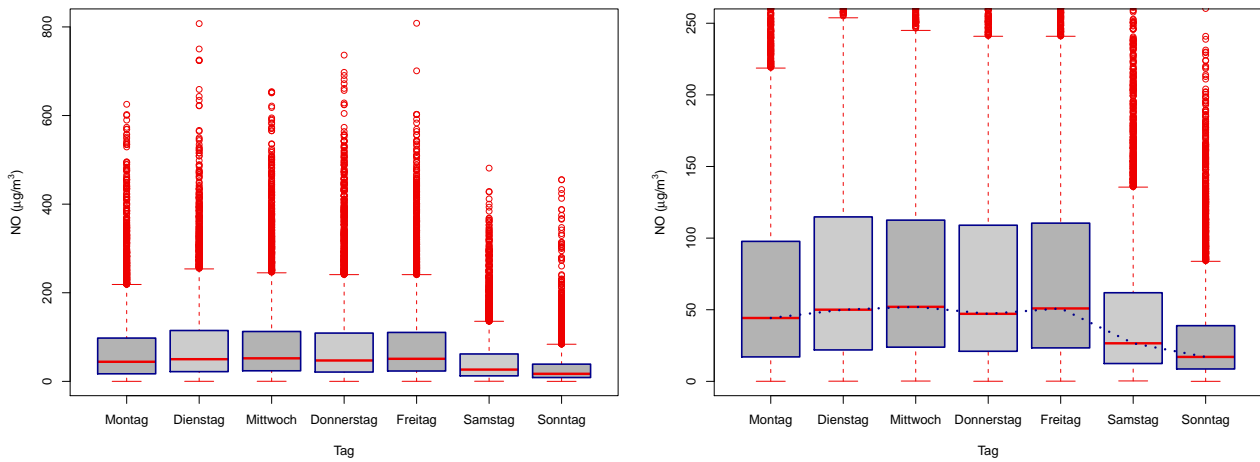


Abbildung 2.16: Boxplotserie der NO-Halbstundenwerte bzgl. Wochentag

Die Boxplotserie stellt oben erwähnte Ergebnisse grafisch dar. Die Boxplots der Werkzeuge unterscheiden sich nicht besonders. Die Dienstags- und Freitagsdaten zeigen ein paar auffallend hohe Ausreißer. Ansonsten ist das Bild an den Werktagen nahezu einheitlich.

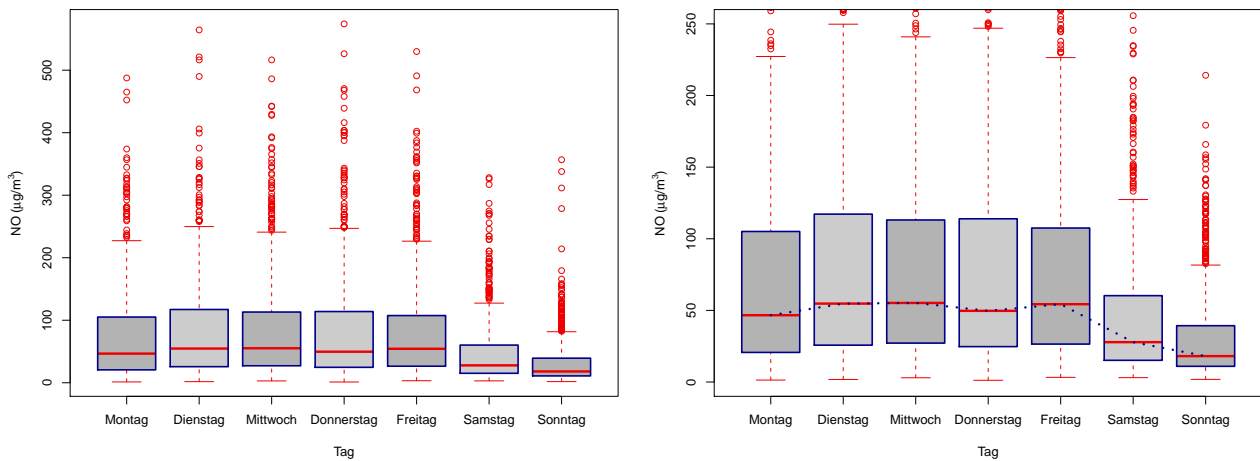


Abbildung 2.17: Boxplotserie der NO-4-Stunden-Mittelwerte bzgl. Wochentag

Die Mediane der Werkzeuge liegen fast auf gleicher Höhe wie die 75 %-Grenze des Samstag-Boxplots bzw. deutlich über der des Sonntag-Boxplots. Die Höhe der 25 %-Grenze ist nahezu überall gleich, von Dienstag bis Freitag etwas erhöht. Insgesamt bedeutet das, dass sich der Bereich in dem die mittleren Daten liegen, fast nur nach oben unterscheidet.

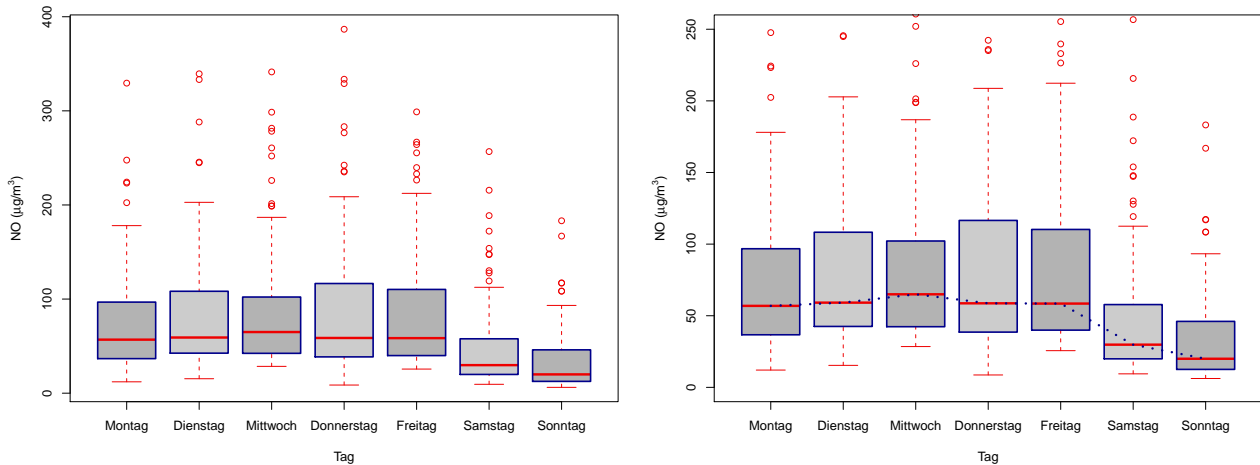


Abbildung 2.18: Boxplotserie der NO-Tagesmittelwerte bzgl. Wochentag

Die Tagesmittelwerte zeigen nach oben dasselbe Bild wie zuvor. Nach unten kann man ein wenig präzisere Aussagen machen. Tage wie Mittwoch und Freitag zeigen nämlich etwas höhere Tagesminima.

### 2.1.7 Unterscheidung nach der Uhrzeit

Untersucht wird zum Schluss noch ob es Unterscheidungen bezüglich der Uhrzeit gibt. Bei genauerer Betrachtung der Daten kommt man zur Aufteilung des Tages in 5 Zeitspannen, bei denen sich am deutlichsten ein Unterschied zeigt. Zu beachten ist, dass hier unterschiedliche Längen der Zeitspannen (4h, 7h, 6h, 4h, 3h) betrachten werden. Dies bedeutet, dass die Verteilungen der entsprechenden  $x$ -Stunden-Mittelwerte ( $x=4,7,6,4,3$ ) analysiert werden. Der Unterschied in den Zeitspannen ist bereits anhand der Kenngrößen erkennbar.

Tabelle 2.7: Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Uhrzeit

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
NO	38.88	71.37	0.020	808.30	87.24	15.97	90.60	1.49
- 0-4 Uhr	14.16	37.69	0.058	517.00	56.35	5.72	46.19	2.12
- 4-11 Uhr	81.39	113.20	0.150	808.30	104.45	37.86	152.90	0.74
- 11-17 Uhr	33.83	50.60	0.963	602.70	54.55	17.40	62.17	1.11
- 17-21 Uhr	38.02	74.63	0.390	750.20	94.53	19.09	87.36	1.96
- 21-0 Uhr	24.50	55.54	0.020	624.50	78.09	12.79	64.03	2.25

Die stärkste Belastung gibt es zwischen 4 und 11 Uhr. Am geringsten ist die Schadstoffkonzentration von 0 bis 4 Uhr. Zwischen den Zeitspannen 11-17 Uhr und 17-21 Uhr ist ein leichter Anstieg zu erkennen. Ab 21 Uhr nimmt die Belastung wieder ab.

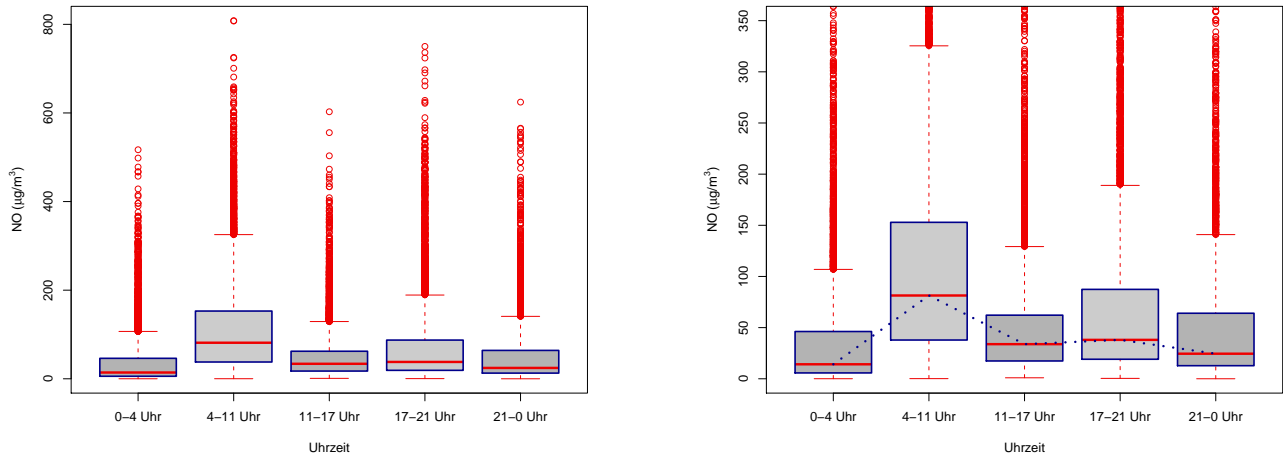


Abbildung 2.19: Boxplotserie der NO-Halbstundenwerte bzgl. Uhrzeit

Die höchsten Ausreißer gibt es in der Zeitspanne 4-11 Uhr. Der Boxplot der Daten gehörend zur Zeitspanne 17-21 Uhr hat zwar einen kleineren und niedriger liegenden interquartilen Bereich, die Ausreißer erreichen aber nahezu die gleiche Höhe. Die erste und die letzte Box der Serie zeigen eine starke Rechtsschiefe.

Um die Aufspaltung nach der Uhrzeit noch etwas genauer zu betrachten, sieht man sich Boxplotserien bezüglich zweier Variablen an, zuerst noch zusätzlich bezüglich nach der Saison, danach bezüglich der in 3 Kategorien zusammengefassten Tage.

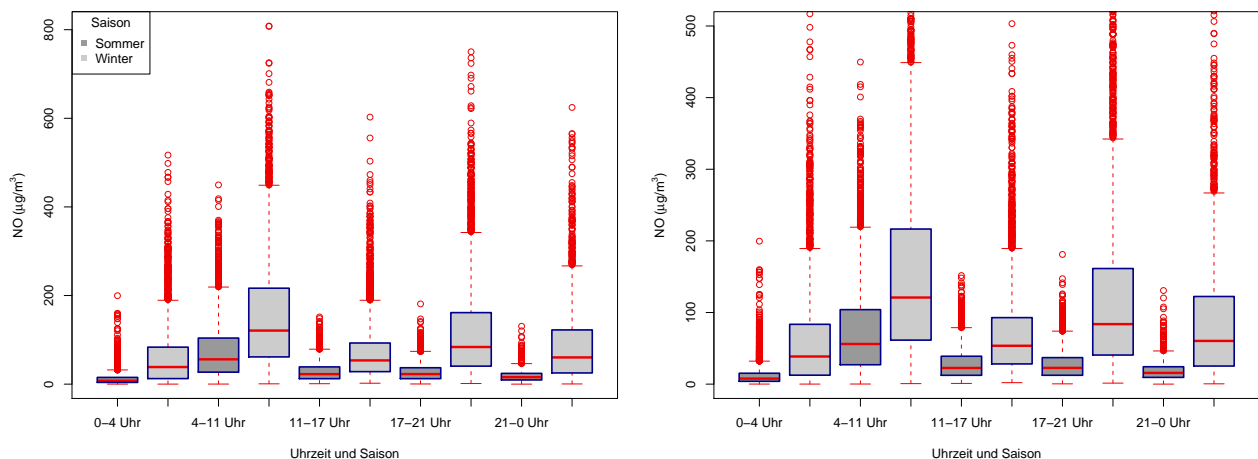


Abbildung 2.20: Boxplotserie der NO-Halbstundenwerte bzgl. Uhrzeit und Saison

Man sieht, dass in jeder Zeitspanne die Ergebnisse der Analyse bezüglich der Saison aufrecht erhalten bleiben. Je größer der Wertebereich der einzelnen Zeitspanne ist, desto größer auch der Unterschied bezüglich der Saison, erkennbar anhand der interquartilen Bereiche. Sämtliche Boxplots gehörend zur Wintersaison haben einen sehr großen Wertebereich. Am geringsten ist die Stickstoffmonoxidbelastung allerdings in der Sommersaison zwischen 0 und 4 Uhr. Am größten ist die Belastung in der Zeitspanne 4-11 Uhr an Wintertagen. Hier gibt es auch die höchsten Ausreißer.

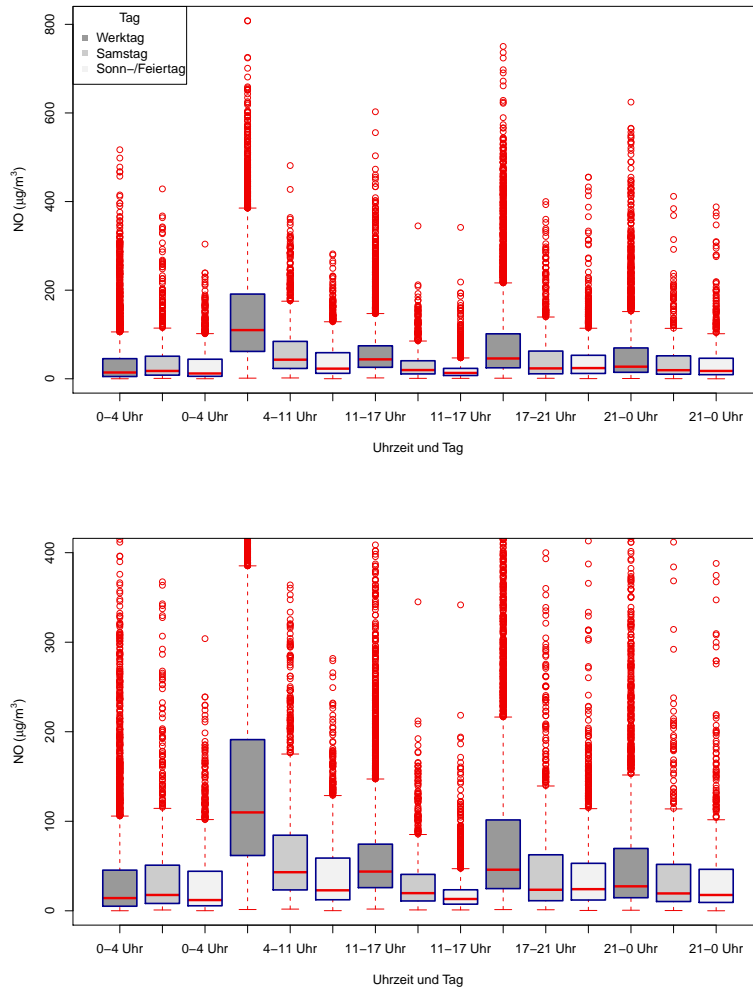


Abbildung 2.21: Boxplotserie der NO-Halbstundenwerte bzgl. Uhrzeit und Tagestyp

Der stärkste Unterschied bezüglich der Tagestypen zeigt sich bei der Zeitspanne 4-11 Uhr. Hier liegt auch das Belastungsmaximum dieser Boxplotserie, also an den Werktagen zwischen 4 und 11 Uhr ist die Belastung am größten. Die Stickstoffmonoxidbelastung ist an den Sonn- bzw. Feiertagen zwischen 11 und 17 Uhr am geringsten. Die Ausreißer sind umso größer, desto größer der Wertebereich ist. Außerdem zeigen so gut wie alle Boxplots dieser Serie eine rechtsschiefe Verteilung, wenn auch von unterschiedlicher Ausprägung.

## 2.2 Schadstoff: Schwefeldioxid - SO<sub>2</sub>

### 2.2.1 Vollständige Betrachtung

Zu allererst möchte man wiederum einen groben Überblick über die Verteilung der Daten bezüglich des Schadstoffes Schwefeldioxid (SO<sub>2</sub>) erhalten und betrachtet deswegen die wichtigsten Kenngrößen, die wie folgt aussehen:

Tabelle 2.8: Statistische Kenngrößen des Schadstoffes SO<sub>2</sub> (μg/m<sup>3</sup>)

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
SO <sub>2</sub>	6.95	8.31	0.003	58.72	6.14	3.93	11.31	0.29

Das Maximum ist mit 58.72 μg/m<sup>3</sup> mehr als achtmal so groß wie der Median mit 6.95 μg/m<sup>3</sup>. Die Spannweite ist mit 58.717 μg/m<sup>3</sup> beinahe das 8-fache des interquartilen Bereichs mit 7.38 μg/m<sup>3</sup>. Der Schiefekoeffizient ist nicht allzu signifikant, man kann aber trotzdem von einer Rechtsschiefe sprechen.

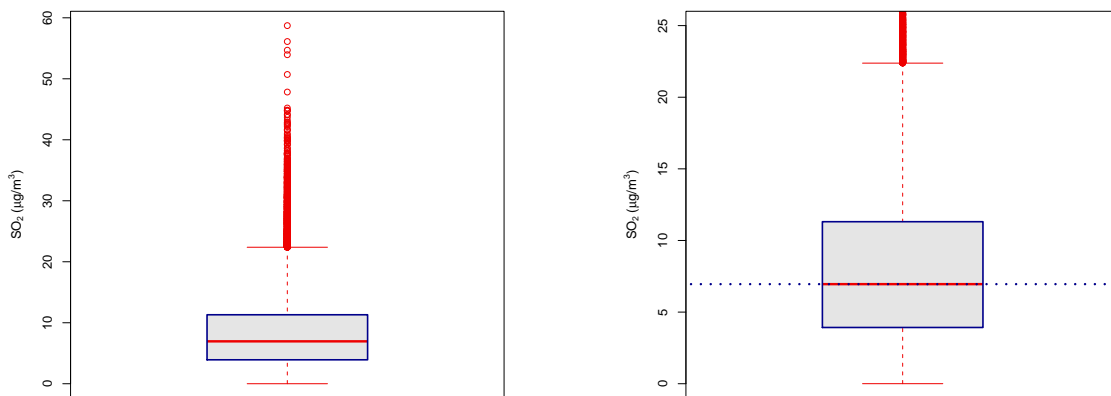


Abbildung 2.22: Boxplot der SO<sub>2</sub>-Halbstundenwerte

Wie auch bei den Stickstoffmonoxid-Daten sieht man hier, dass es nach unten keine Ausreißer gibt. Ab einem Wert von ungefähr 45 μg/m<sup>3</sup> nimmt die Anzahl der Ausreißer ab. Der Median liegt bei einem Wert von ungefähr 7 μg/m<sup>3</sup>.

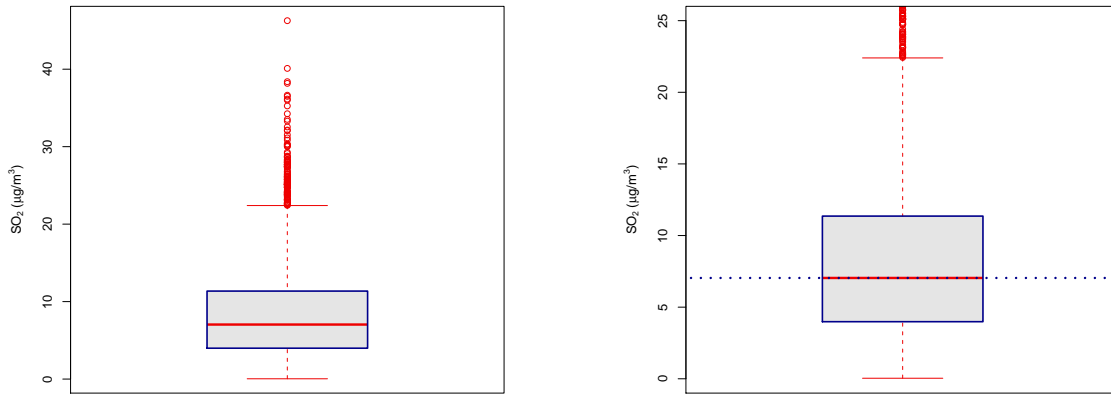


Abbildung 2.23: Boxplot der SO<sub>2</sub>-4-Stunden-Mittelwerte

Der Median liegt beinahe in der Mitte der Box. Dies bedeutet, dass die Daten annähernd symmetrisch verteilt sind, was sich anhand des Schiefekoeffizienten zeigte. Ein Blick auf das 1. und 3. Quartil zeigt: 25 % der Daten liegen unter einem Wert von  $3.93 \mu\text{g}/\text{m}^3$ , 75 % unter einem Wert von etwa  $11.31 \mu\text{g}/\text{m}^3$ .

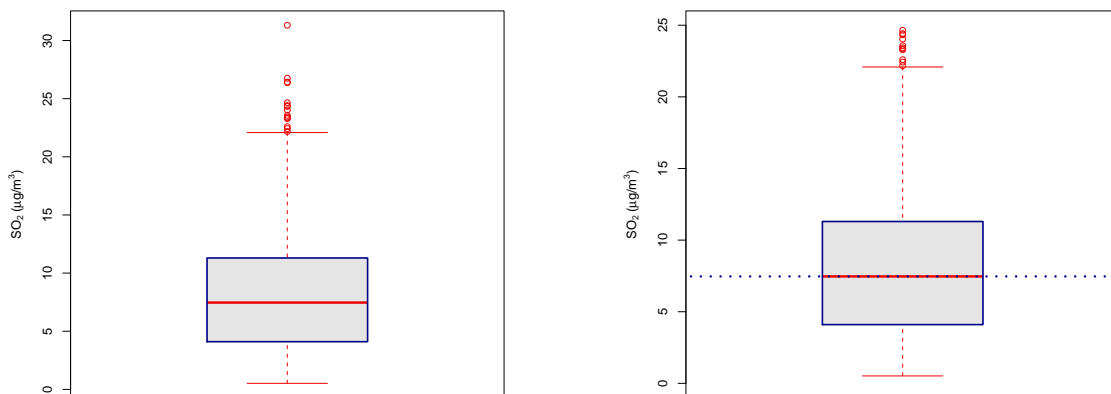


Abbildung 2.24: Boxplot der SO<sub>2</sub>-Tagesmittelwerte

Die Boxplots zu den Tagesmittelwerten des Schadstoffes Schwefeldioxid zeigen noch einmal die Erkenntnisse von zuvor. Auffällig ist jedoch die geringe Anzahl der Ausreißer bei der Tagesmitteilung.

## 2.2.2 Unterscheidung nach dem Jahr

Auch bei diesem Schadstoffdatensatz beginnt man mit der einfachsten Aufteilung, nämlich der bezüglich der Jahre. Die Berechnung der statistischen Kenngrößen zu den Messdaten gehörend zu den Jahren 2005 und 2006 liefert folgende Ergebnisse:

Tabelle 2.9: Statistische Kenngrößen des Schadstoffes SO<sub>2</sub> (µg/m<sup>3</sup>) bzgl. Jahr

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
SO <sub>2</sub>	6.95	8.31	0.003	58.72	6.14	3.93	11.31	0.29
- Jahr 2005	7.04	8.67	0.090	56.11	5.91	4.93	10.97	0.51
- Jahr 2006	6.82	7.95	0.003	58.72	6.35	2.74	11.64	0.19

Wie man sieht sind auch hier sämtliche Größen ungefähr gleich groß. Im Unterschied zum Stickstoffmonoxid sind sie aber im Jahr 2006 etwas kleiner als im Jahr 2005. So ist etwa der Mittelwert im Jahr 2005 um 0.72 µg/m<sup>3</sup> größer als im Jahr 2006.

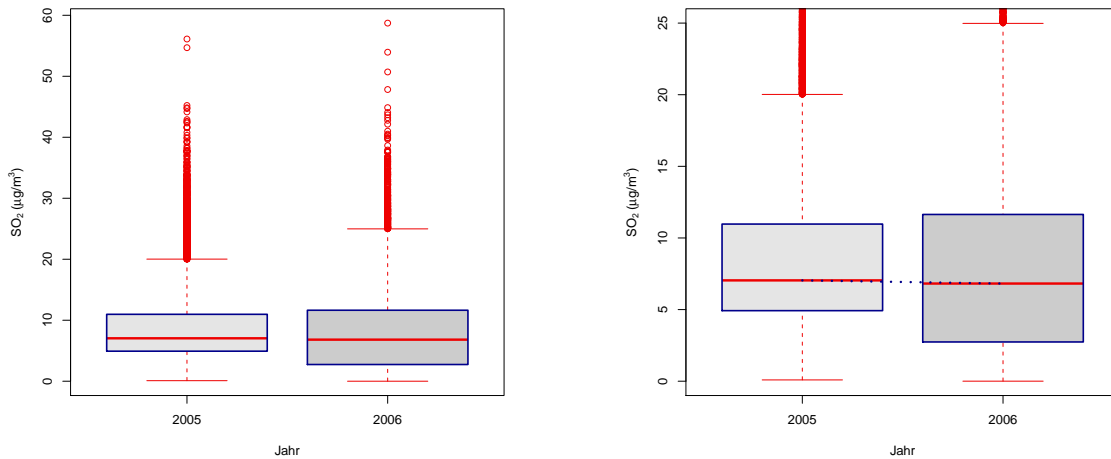


Abbildung 2.25: Boxplotserie der SO<sub>2</sub>-Halbstundenwerte bzgl. Jahr

Die maximale Höhe der Ausreißer ist ungefähr gleich groß. Ab einem Wert von ungefähr 45 µg/m<sup>3</sup> gibt es bei beiden Boxplots nur noch wenige aufgezeichnete Werte. Die Messdaten im Jahr 2006 scheinen annähernd symmetrisch verteilt zu sein.

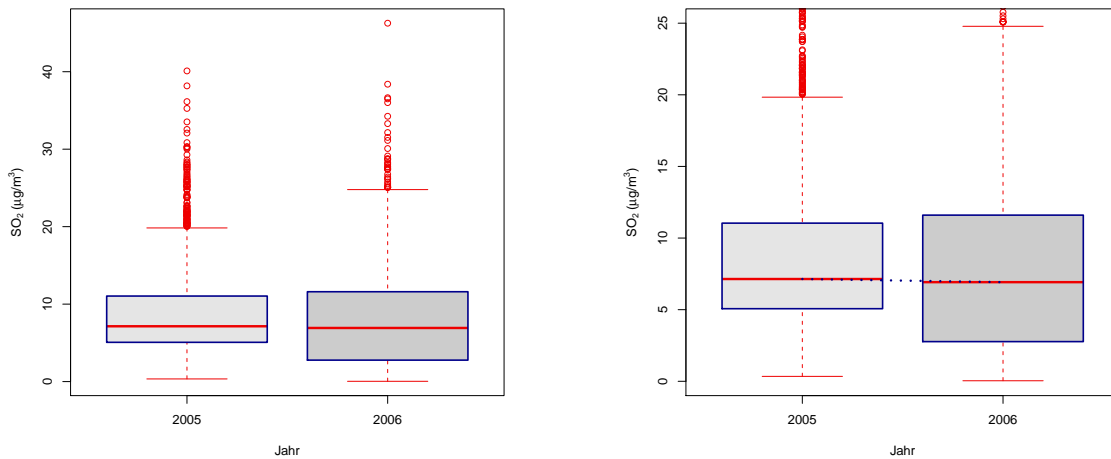


Abbildung 2.26: Boxplotserie der SO<sub>2</sub>-4-Stunden-Mittelwerte bzgl. Jahr



Obwohl die Mediane der Daten zu den beiden Jahren beinahe auf derselben Höhe liegen, gibt es einen signifikanten Unterschied bezüglich des interquartilen Bereichs. Dieser wird vor allem deutlich, wenn man das 1. Quartil betrachtet: Während im Jahr 2005 25 % der Daten unter einem Wert von  $4.93 \mu\text{g}/\text{m}^3$  liegen, liegt dieser Grenzwert im Jahr 2006 bei  $2.74 \mu\text{g}/\text{m}^3$ .

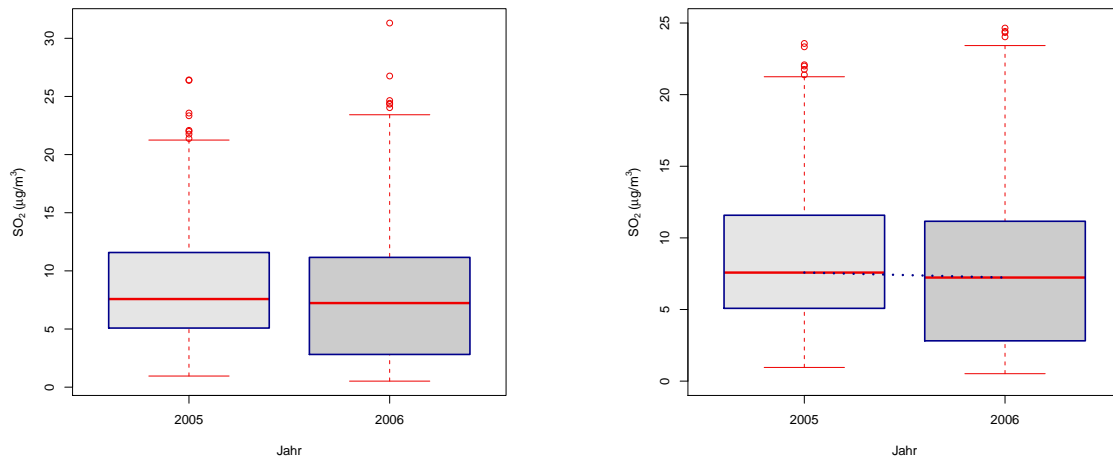


Abbildung 2.27: Boxplotserie der  $\text{SO}_2$ -Tagesmittelwerte bzgl. Jahr

Eine Aufteilung bezüglich des Jahres zeigt auch hier größtenteils keinen signifikanten Unterschied. So sind in etwa die Größen wie Median und Maximum ungefähr gleich hoch. Nur beim 1. Quartil trifft dies nicht zu.

### 2.2.3 Unterscheidung nach der Saison

Durchaus interessanter ist auch hier die Aufteilung der Daten nach der Saison. Es wird also unterschieden, ob die Werte in der Sommer- oder Wintersaison aufgezeichnet wurden. Die Tabelle gibt einen ersten Überblick über die Verteilungen der jeweiligen Datensätze.

Tabelle 2.10: Statistische Kenngrößen des Schadstoffes  $\text{SO}_2$  ( $\mu\text{g}/\text{m}^3$ ) bzgl. Saison

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
$\text{SO}_2$	6.95	8.31	0.003	58.72	6.14	3.93	11.31	0.29
- Sommer	4.95	5.49	0.003	27.50	3.89	2.27	7.66	0.17
- Winter	9.58	11.11	0.033	58.72	6.66	6.36	14.51	0.24

Auch hier zeigt sich ein signifikanter Unterschied in der Saison. Das Maximum des Winters ist mit  $58.72 \mu\text{g}/\text{m}^3$  ungefähr doppelt so groß wie das des Sommers mit  $27.50 \mu\text{g}/\text{m}^3$ . Auch bei den Größen wie Median und Mittelwert ist dieses Verhältnis gültig.

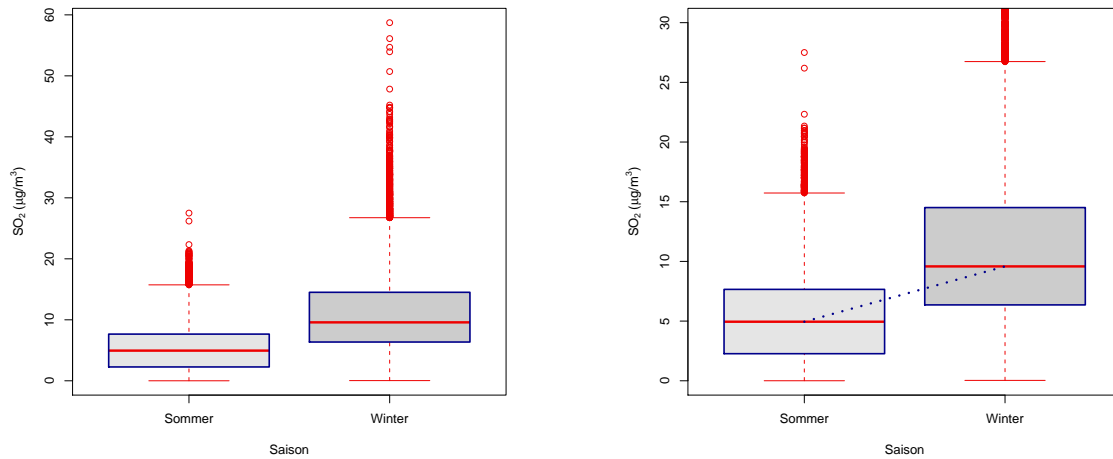


Abbildung 2.28: Boxplotserie der SO<sub>2</sub>-Halbstundenwerte bzgl. Saison

Im Sommer gibt es sehr wenige und vom Wert niedrige Ausreißer. Beide Saisonen zeigen ein paar auffallend hohe Werte. Obwohl anhand der Tabelle eine leichte Rechtsschiefe beider Verteilungen ablesbar ist, zeigt die Betrachtung der Boxen, dass die Daten der Sommersaison einer symmetrischen Verteilung sehr nahe sind.

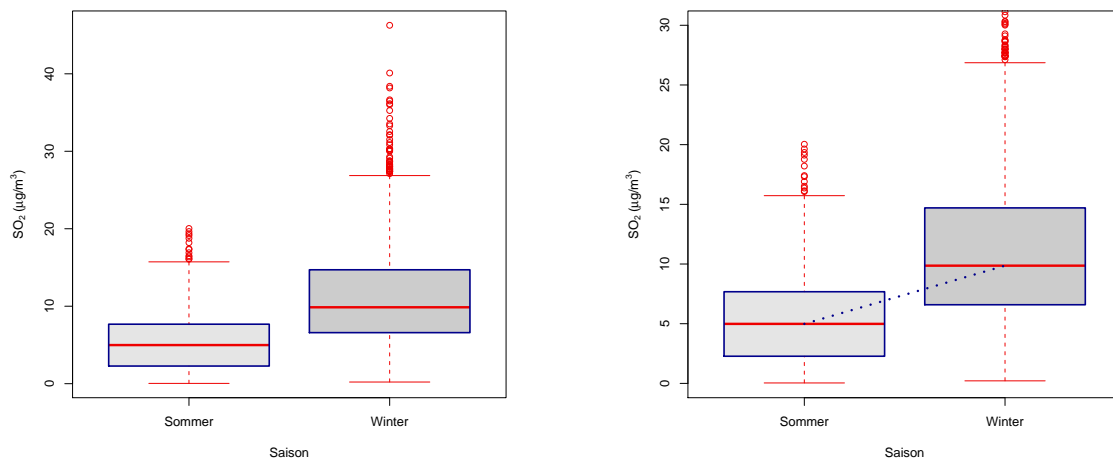


Abbildung 2.29: Boxplotserie der SO<sub>2</sub>-4-Stunden-Mittelwerte bzgl. Saison

Ganz klar zeigt sich in den obigen Abbildungen der Anstieg des Medians im Winter, so wie auch, dass die Streubreite im Winter größer ist. Das 1. Quartil des Winters und das 3. Quartil des Sommers liegen beinahe auf derselben Höhe.

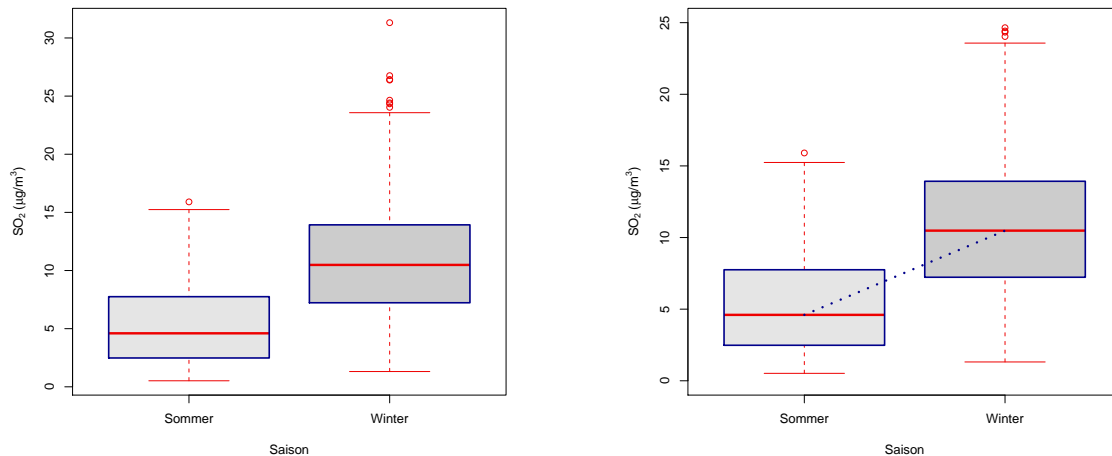


Abbildung 2.30: Boxplotserie der SO<sub>2</sub>-Tagesmittelwerte bzgl. Saison

Bei der Mittelung auf Tageswerte sieht man, dass es nur noch wenige Ausreißer im Winter gibt, im Sommer bleibt sogar bloß ein einzelner übrig. Wesentlich ist, dass auch bei den Schwefeldioxidaten ein starker Unterschied zwischen Sommer und Winter besteht.

## 2.2.4 Unterscheidung nach dem Monat

Obwohl die Aufteilung bezüglich der Saison keine ungewöhnlichen Ergebnisse darbrachte, wird die folgende Differenzierung nach den Monaten ein völlig neues Bild zeigen. Bereits an den statistischen Kenngrößen kann man einige Auffälligkeiten erkennen.

Tabelle 2.11: Statistische Kenngrößen des Schadstoffes SO<sub>2</sub> (µg/m<sup>3</sup>) bzgl. Monat

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
SO <sub>2</sub>	6.95	8.31	0.003	58.72	6.14	3.93	11.31	0.29
- Jänner	15.68	16.55	3.007	58.72	7.56	10.92	21.11	-0.03
- Februar	9.35	11.27	1.168	56.11	6.55	6.72	14.34	0.37
- März	8.93	10.16	3.222	44.16	5.03	6.47	12.56	0.35
- April	7.55	8.06	3.687	27.50	2.81	6.14	9.11	0.42
- Mai	9.01	8.47	0.344	21.34	3.56	5.77	11.11	-0.49
- Juni	3.00	5.46	0.056	22.33	4.99	2.07	7.22	0.90
- Juli	3.63	3.42	0.005	14.86	2.03	1.73	4.57	-0.29
- August	4.74	4.14	0.003	14.75	2.91	1.23	6.31	-0.33
- September	2.54	3.50	0.058	20.69	2.80	1.48	4.74	0.55
- Oktober	5.92	6.25	0.033	27.17	3.85	3.36	8.25	0.05
- November	8.74	9.22	0.966	32.83	4.58	6.08	11.91	-0.02
- Dezember	12.52	13.12	1.718	39.26	6.49	8.23	16.66	0.04

Ganz klar zu sehen ist, dass die Werte des Medians und Mittelwertes sehr unterschiedlich sind. Aber auch die Schiefekoeffizienten sind different. In den Monaten Jänner, Mai, Juli, August und November sind sie sogar negativ, das heißt es liegt eine Linksschiefe vor. Ansonsten sind die Daten rechtsschief verteilt.

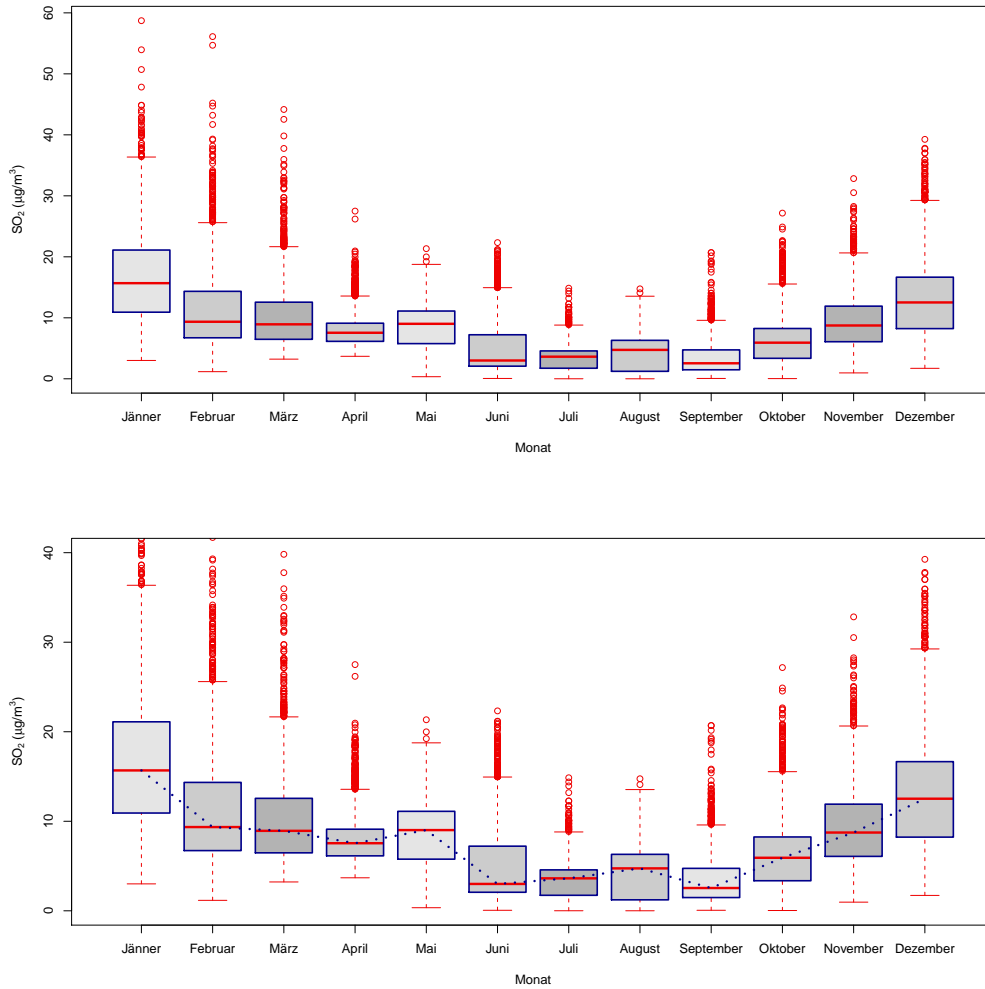


Abbildung 2.31: Boxplotserie der SO<sub>2</sub>-Halbstundenwerte bzgl. Monat

Die Unregelmäßigkeit zeigt sich sehr gut bei den beiden obigen Abbildungen. Von einem kontinuierlichen Abfall oder Anstieg kann auf keinen Fall gesprochen werden. Selbst die Verteilung der Ausreißer scheint keinem Muster zu folgen. Ein große Streubreite zeigen etwa die Daten zum Monat März. Im Februar gibt es ein paar sehr auffällig hohe Werte. Monate wie der Mai oder August zeigen beinahe keine Ausreißer.

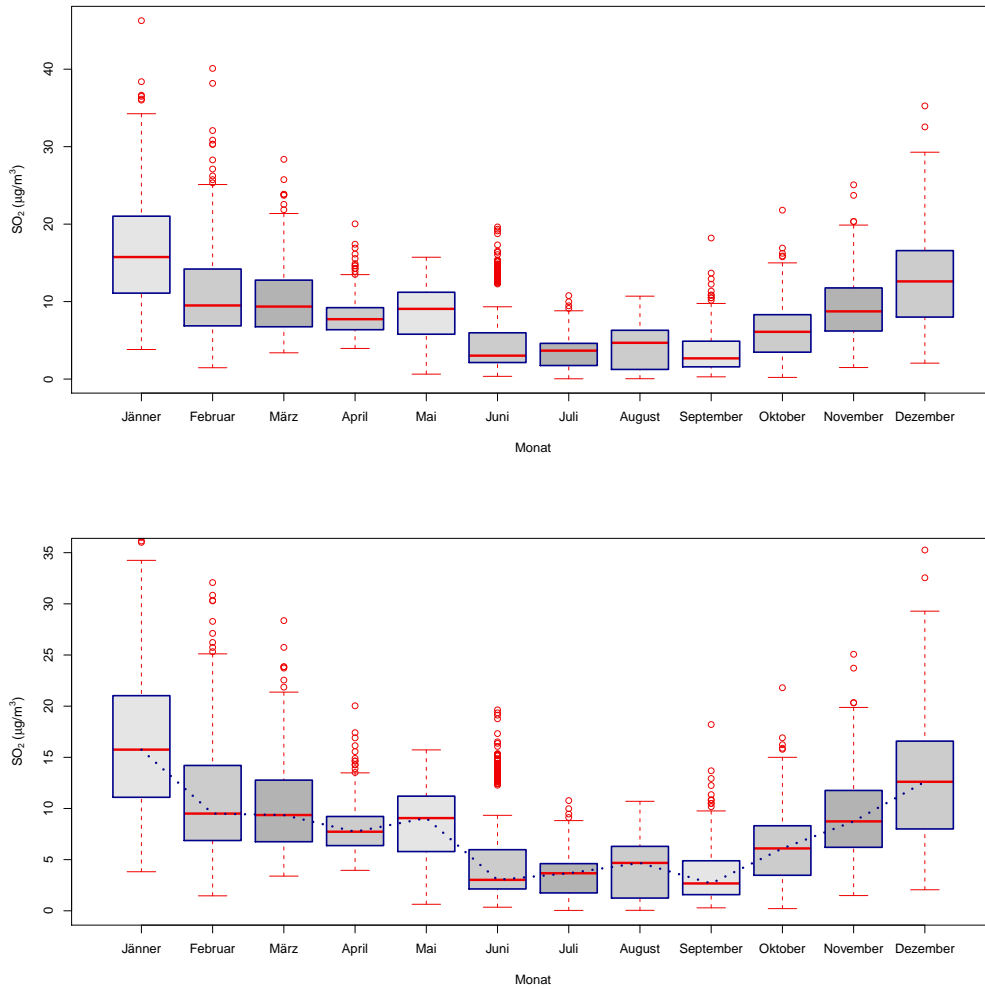


Abbildung 2.32: Boxplotserie der SO<sub>2</sub>-4-Stunden-Mittelwerte bzgl. Monat

Auch die Mediane sind stark variierend: Dieser fällt von Jänner zu Feber und bleibt schließlich bis zum Mai in der ungefähr gleichen Höhe. Danach gibt es einen starken Abfall. Nach einem etwas höheren Wert im August, steigt der Median ab September bis zum Dezember stetig an.

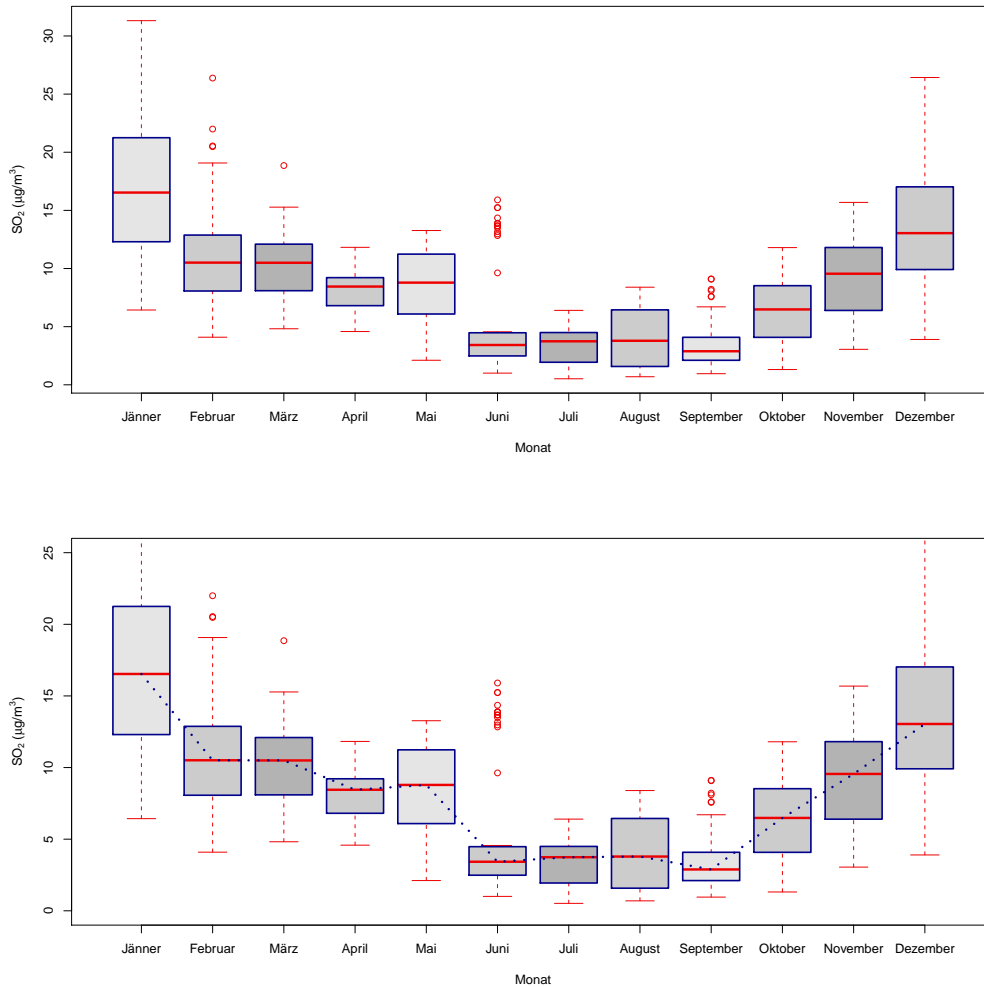


Abbildung 2.33: Boxplotserie der SO<sub>2</sub>-Tagesmittelwerte bzgl. Monat

Die Abbildungen der Tagesmittelwerte zeigen wiederholt die Auffälligkeiten dieser Aufteilung. Auch die Größen der interquartilen Bereiche sind nicht einheitlich. Interessant ist auch noch der Monat Juni, der bei den Tagesmittelwerten ungewöhnlich viele hohe Ausreißer hat.

### 2.2.5 Unterscheidung nach dem Tagestyp

Die als nächstes betrachtete Aufteilung ist die mittels der Tagestypen. Hier sind die Tage also in 3 Kategorien zusammengefasst sind. Die Kenngrößen zu dieser Betrachtung sind in der folgenden Tabelle aufgezeichnet.

Tabelle 2.12: Statistische Kenngrößen des Schadstoffes SO<sub>2</sub> (µg/m<sup>3</sup>) bzgl. Tagestyp

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
SO <sub>2</sub>	6.95	8.31	0.003	58.72	6.14	3.93	11.31	0.29
- Werktag	7.29	8.70	0.003	58.72	6.37	4.12	11.79	0.30
- Samstag	6.64	7.66	0.015	41.54	5.60	3.65	10.28	0.26
- Sonn-/Feiertag	6.06	7.35	0.006	37.64	5.51	3.38	10.20	0.29

Sowohl das Minimum als auch das Maximum des Gesamtdatensatzes kommen an einem Werktag vor. Trotzdem sind die Sonn- bzw. Feiertage die belastungsschwächsten Tage. Die Größen wie Median, Mittelwert und Standardabweichung nehmen von den Werktagen zu den Sonn- bzw. Feiertagen ab. Nur der Schiefekoeffizient ist an den Samstagen am kleinsten.

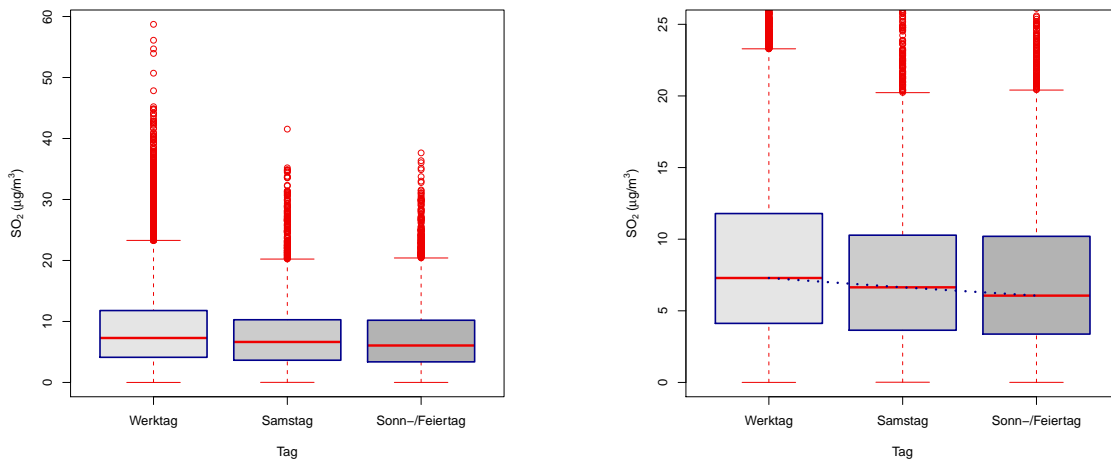


Abbildung 2.34: Boxplotserie der SO<sub>2</sub>-Halbstundenwerte bzgl. Tagestyp

Die meisten und größten Ausreißer gibt es an den Werktagen. Die Anzahl dieser nimmt aber ab einem Wert von circa 45 µg/m<sup>3</sup> stark ab. Der Samstagsdatensatz zeigt einen auffällig hohen Wert. Bei allen Boxplots gibt es eine nur leichte Rechtsschiefe.

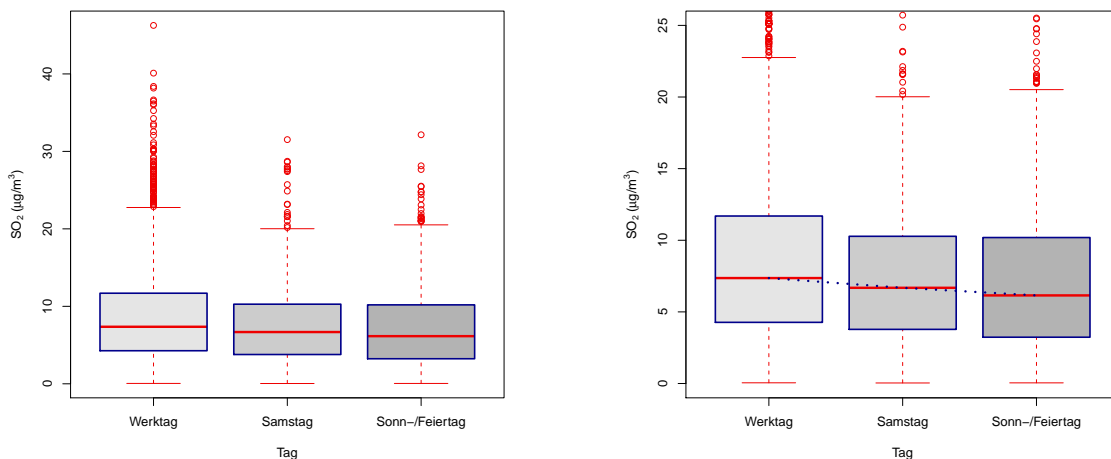


Abbildung 2.35: Boxplotserie der SO<sub>2</sub>-4-Stunden-Mittelwerte bzgl. Tagestyp

Die Mediane sinken zwar Richtung Sonntag, jedoch nicht allzu stark. Die Boxplots gehörend zu den Samstagen und Sonn- bzw. Feiertagen scheinen sehr ähnlich. So sind unter anderem die interquartilen Bereiche der Samstags- und Sonn- bzw. Feiertage ungefähr gleich groß.

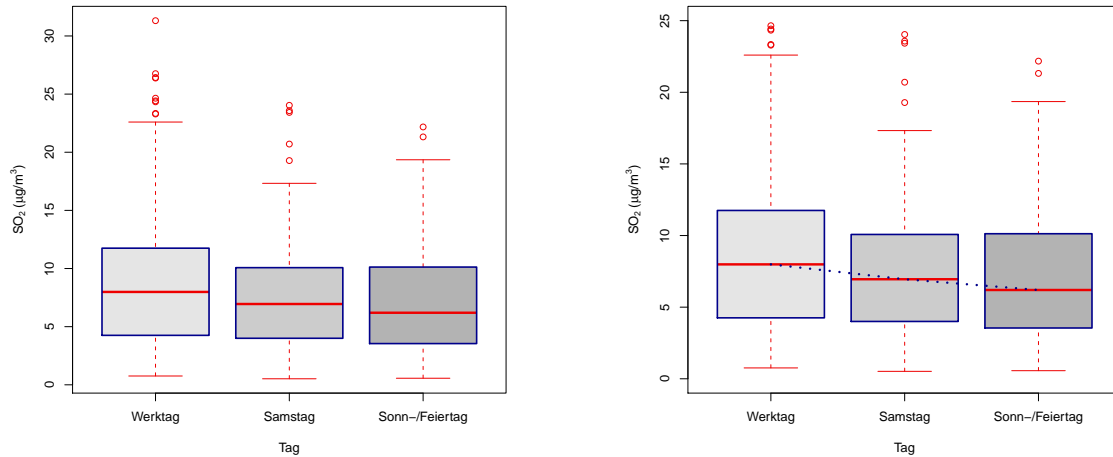


Abbildung 2.36: Boxplotserie der SO<sub>2</sub>-Tagesmittelwerte bzgl. Tagestyp

Es gibt zwar einen Unterschied zwischen den Daten zu den Werktagen und denen zu den restlichen Tagen. Dieser ist aber nicht so groß wie in dem zuvor betrachteten Schadstoff Stickstoffmonoxid.

### 2.2.6 Unterscheidung nach dem Wochentag

Da man bei der Aufteilung nach den Tagen keinen allzu wesentlichen Unterschied aufdecken konnte, möchte man sehen, ob es vielleicht an den Werktagen differente Verteilungen gibt. Man betrachtet also nun eine Aufteilung nach Wochentagen.

Tabelle 2.13: Statistische Kenngrößen des Schadstoffes SO<sub>2</sub> (µg/m<sup>3</sup>) bzgl. Wochentag

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
SO <sub>2</sub>	6.95	8.31	0.003	58.72	6.14	3.93	11.31	0.29
- Montag	6.68	8.14	0.010	38.26	5.67	4.23	11.05	0.26
- Dienstag	7.38	8.72	0.004	44.83	6.24	4.29	11.88	0.24
- Mittwoch	7.58	8.83	0.031	58.72	6.41	4.16	11.91	0.28
- Donnerstag	7.36	8.88	0.009	50.71	6.77	3.82	12.14	0.38
- Freitag	7.32	8.71	0.003	56.11	6.42	4.03	11.88	0.30
- Samstag	6.68	7.76	0.015	41.54	5.69	3.68	10.38	0.26
- Sonntag	5.89	7.16	0.006	36.36	5.49	3.18	9.81	0.35

Die Kenngrößen wie Median, Mittelwert, Standardabweichung und interquartiler Bereich sind an den Tagen Dienstag bis Freitag ungefähr gleich hoch. Belastungsschwächster Werktag ist der Montag. Hier ist auch das Maximum mit 38.26 µg/m<sup>3</sup> auffallend klein. Sämtliche Schiefeffizienten liegen ungefähr um den Wert 0.3.



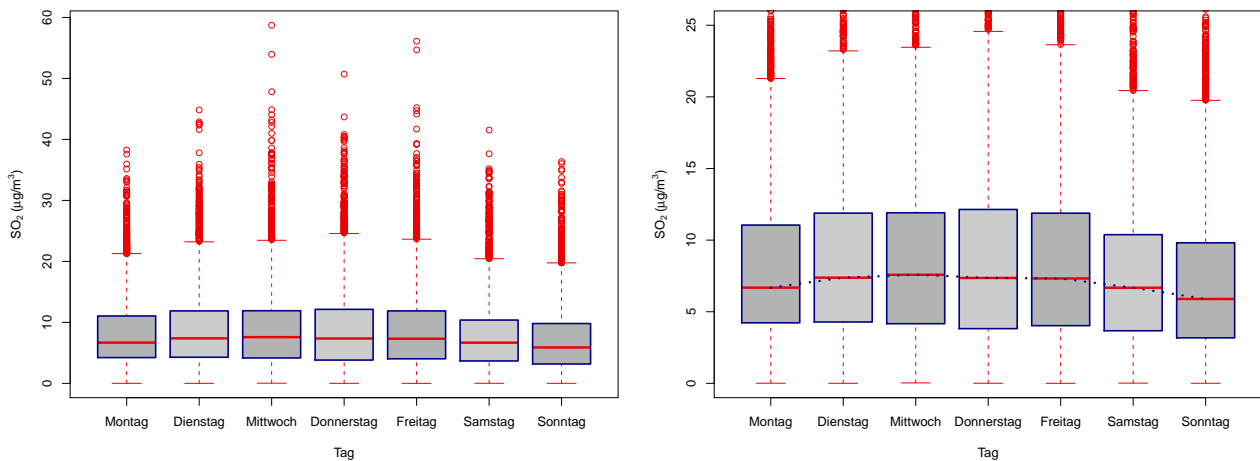


Abbildung 2.37: Boxplotserie der SO<sub>2</sub>-Halbstundenwerte bzgl. Wochentag

An den Tagen Mittwoch und Freitag gibt es die höchsten Ausreißer. Ansonsten ist die Streubreite ungefähr gleich groß. Die niedrigsten Ausreißer gibt es sonntags. Die oberen und unteren Boxgrenzen der Boxplots von Montag bis Freitag sind ungefähr auf der selben Höhe.

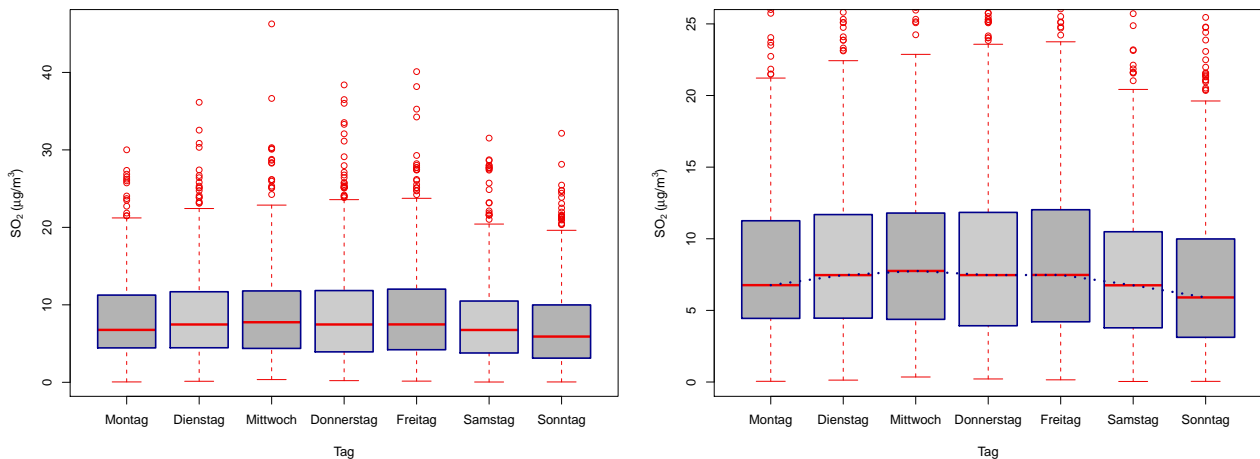


Abbildung 2.38: Boxplotserie der SO<sub>2</sub>-4-Stunden-Mittelwerte bzgl. Wochentag

Die Mediane steigen zur Wochenmitte von beiden Seiten an. Dennoch zeigen die Boxplots der Werkzeuge keine signifikanten Unterschiede. Die detailliertere Anschauung mittels Aufspaltung der Tage in Wochentage bringt somit keine neuen Ergebnisse.

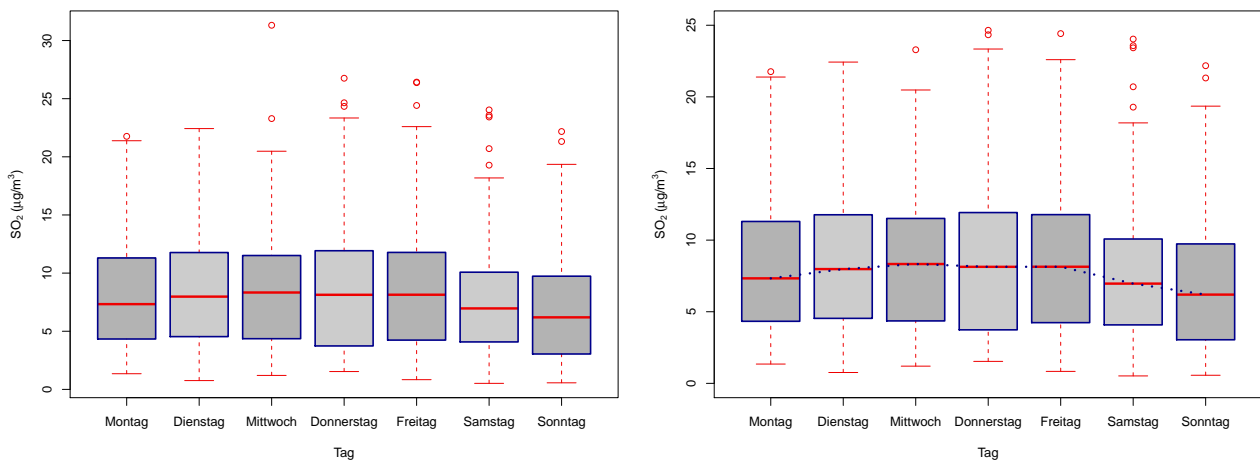


Abbildung 2.39: Boxplotserie der SO<sub>2</sub>-Tagesmittelwerte bzgl. Wochentag

Insgesamt kann man sagen, dass sich die Daten unterschieden nach den Tagen, egal ob nach 3 Kategorien zusammengefasst oder bezogen auf die Woche, kaum unterscheiden. Die Belastung variiert also bezüglich der Tage wenig.

### 2.2.7 Unterscheidung nach der Uhrzeit

Am Ende der explorativen Analyse der Messdaten zum Schadstoff Schwefeldioxid betrachtet man den Verlauf innerhalb eines Tages. Bei diesem Schadstoffdatensatz scheint eine Aufteilung des Tages in nur 3 Zeitspannen am sinnvollsten. Auch hier sind diese Zeitspannen von unterschiedlicher Länge (4h, 10h, 10h). Die Kenngrößen zu den jeweiligen Zeiten sehen so aus:

Tabelle 2.14: Statistische Kenngrößen des Schadstoffes SO<sub>2</sub> (µg/m<sup>3</sup>) bzgl. Uhrzeit

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
SO <sub>2</sub>	6.95	8.31	0.003	58.72	6.14	3.93	11.31	0.29
- 0-4 Uhr	5.96	6.55	0.003	30.59	4.43	3.08	9.05	0.05
- 4-14 Uhr	8.03	9.52	0.013	58.72	6.67	4.80	12.90	0.26
- 14-0 Uhr	6.61	7.88	0.005	40.52	6.01	3.47	10.69	0.37

Zwischen 4 und 14 Uhr ist die Schwefeldioxidbelastung am größten, am geringsten von 0 bis 4 Uhr. Die in der Zeitspanne 0-4 Uhr aufgezeichneten Werte sind nahezu symmetrisch verteilt. Hier liegt der Schiefekoeffizient mit 0.05 nämlich nahezu bei 0.

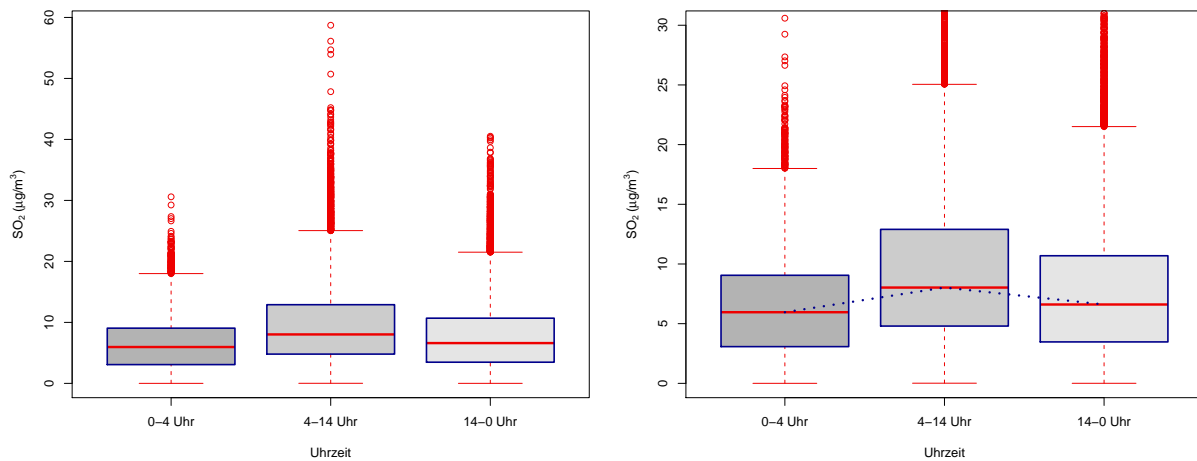


Abbildung 2.40: Boxplotserie der SO<sub>2</sub>-Halbstundenwerte bzgl. Uhrzeit

Die meisten Ausreißer gibt es in der 2. Zeitspanne. Wenige und sehr niedrige Ausreißer gibt es bei den Daten, aufgezeichnet von 0-4 Uhr. Die Belastung ist somit an der Tagesmitte am stärksten. Zu sehen ist auch, dass in jeder Zeitspanne auch sehr kleine Belastungswerte aufgezeichnet wurden.

Möglicherweise wird der Unterschied in den Zeitspannen bei einer aufgeteilten Ansicht noch deutlicher. Aus diesem Grund hat man die Boxplotserie nicht nur nach den Zeitspannen sondern auch nach der Saison bzw. nach dem Tagestyp aufgeteilt. Diese Anschauungen liefern folgende Ergebnisse.

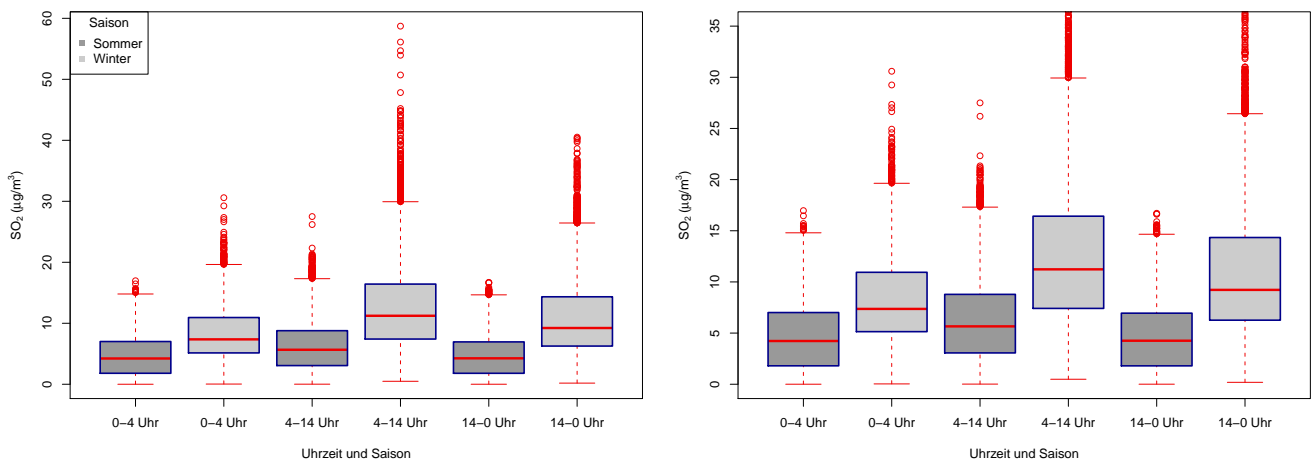


Abbildung 2.41: Boxplotserie der SO<sub>2</sub>-Halbstundenwerte bzgl. Uhrzeit und Saison

Alle drei Boxplotserien zu den jeweiligen Zeitspannen zeigen den starken Unterschied in der Saison. Am ausgeprägtesten ist dieser in der Zeitspanne 4-14 Uhr, in der auch das Belastungsmaximum dieser Serie liegt. Außerdem hat der Boxplot zu dieser Zeitspanne in der Wintersaison die höchsten Ausreißer und somit die größte Streuung. Der Unterschied im Tagesverlauf in der Sommersaison ist nicht so deutlich. Sowohl die Wertebereiche als auch die Streubreiten sind ähnlich. Signifikanter ist der Unterschied zwischen den Zeitspannen in der Wintersaison.

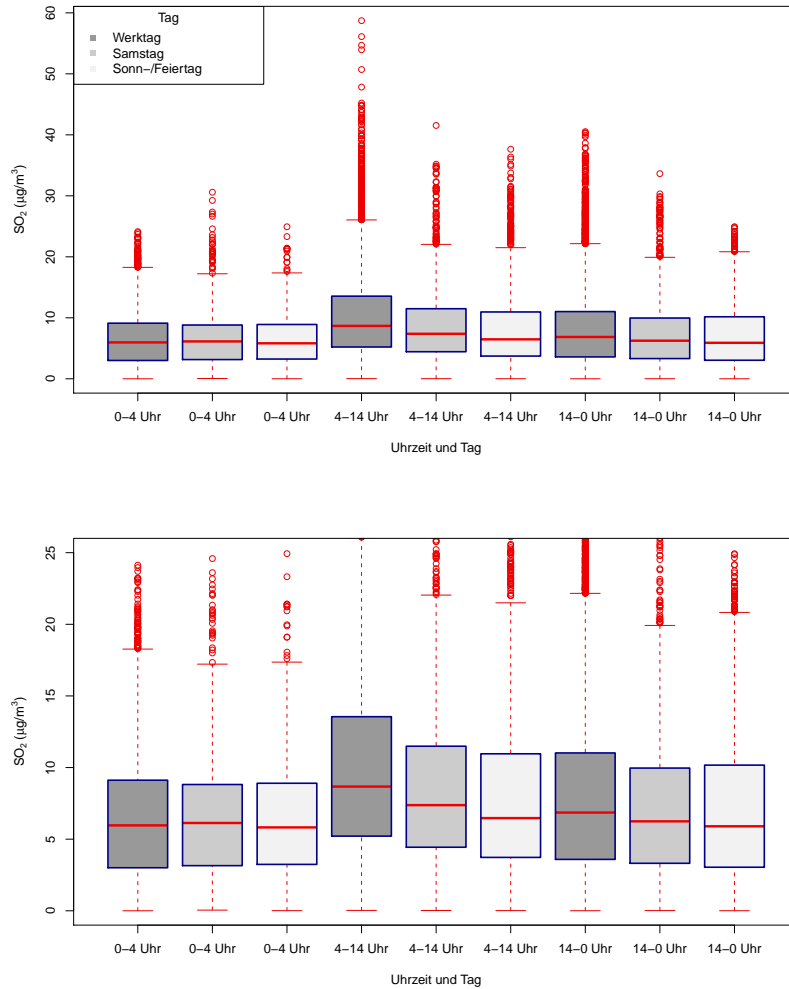


Abbildung 2.42: Boxplotserie der SO<sub>2</sub>-Halbstundenwerte bzgl. Uhrzeit und Tag

In der Zeitspanne 0-4 Uhr zeigt sich der Unterschied in den Tagen kaum bis gar nicht. Am deutlichsten wird er in der Zeitspanne 4-14 Uhr. Wiederum liegt das Maximum der Schwefeldioxidbelastung, und somit auch die höchsten Ausreißer, zwischen 4 und 14 Uhr. Genauer gesagt, bei dieser Aufspaltung in der Zeitspanne 4-14 Uhr und an den Werktagen. Es gibt mehrere ähnlich kleine Boxplots. Jedoch kann das Minimum, auch anhand der Größe der Ausreißer, in der Zeitspanne 0-4 Uhr an Sonn- bzw. Feiertagen bestimmt werden. Insgesamt zeigt diese Serie, bis auf die Höhe der Ausreißer, nicht allzu signifikante Unterschiede in den Boxplots.

## 2.3 Schadstoff: Kohlenmonoxid - CO

### 2.3.1 Vollständige Betrachtung

Zu Beginn der Analyse möchte man sehen, in welchem Größenbereich die Daten zum Schadstoff Kohlenmonoxid ungefähr liegen, um ein erstes Bild über die Verteilung zu erhalten. Die statistischen Kenngrößen dafür sehen folgendermaßen aus:

Tabelle 2.15: Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ )

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
CO	0.61	0.80	0.118	5.89	0.63	0.37	0.98	0.97

Das Maximum ist mit  $5.89 \mu\text{g}/\text{m}^3$  beinahe 50 mal so groß wie das Minimum. Der Wert des Maximums ( $5.89 \mu\text{g}/\text{m}^3$ ) verglichen mit dem des 3. Quartils ( $0.98 \mu\text{g}/\text{m}^3$ ) zeigt die hohe Streubreite der Daten nach oben. Der Schiefeffizient ist positiv, wodurch man auf eine Rechtsschiefe schließen kann.

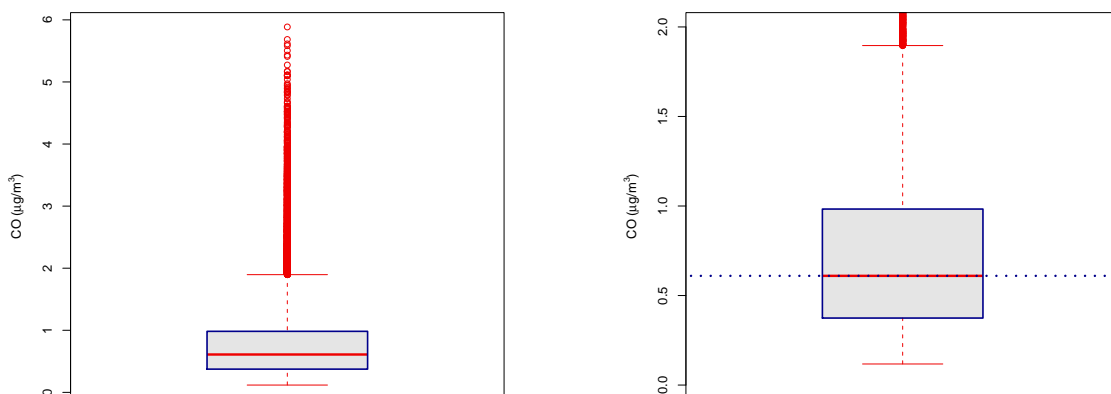


Abbildung 2.43: Boxplot der CO-Halbstundenwerte

Insgesamt ist die Anzahl der Ausreißer sehr groß, wenn auch die Dichte derer ab einem Wert von  $5 \mu\text{g}/\text{m}^3$  ein wenig abnimmt. Nach unten gibt es, wie auch bei den beiden Schadstoffendatensätzen zuvor, keine Ausreißer. Die linke Grafik zeigt auch deutlich das extreme Verhältnis Spannweite zu interquartiler Bereich.

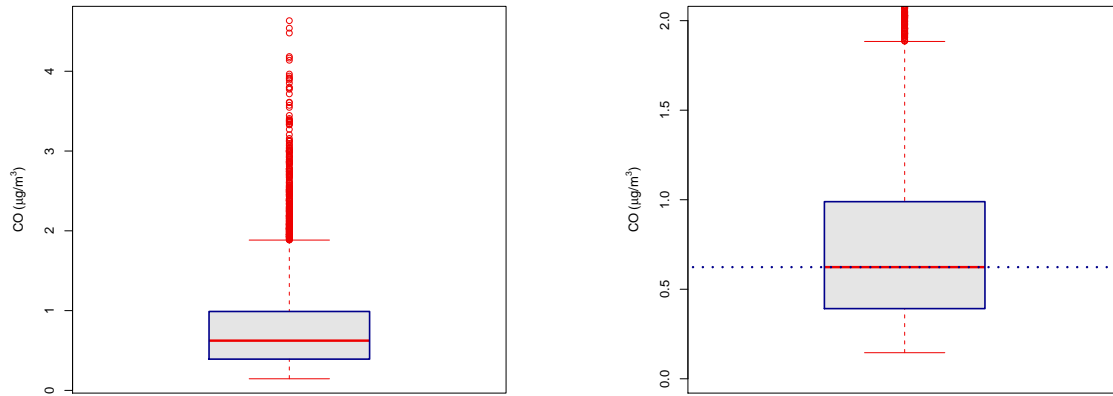


Abbildung 2.44: Boxplot der CO-4-Stunden-Mittelwerte

Der Median liegt etwas im unteren Bereich der Box. Die Annahme auf Rechtsschiefe wird also bestätigt. Das Minimum ist im Vergleich zu den beiden anderen Schadstoffanalysen eher groß. Es ist hier nämlich am weitesten von der 0  $\mu\text{g}/\text{m}^3$ -Grenze entfernt.

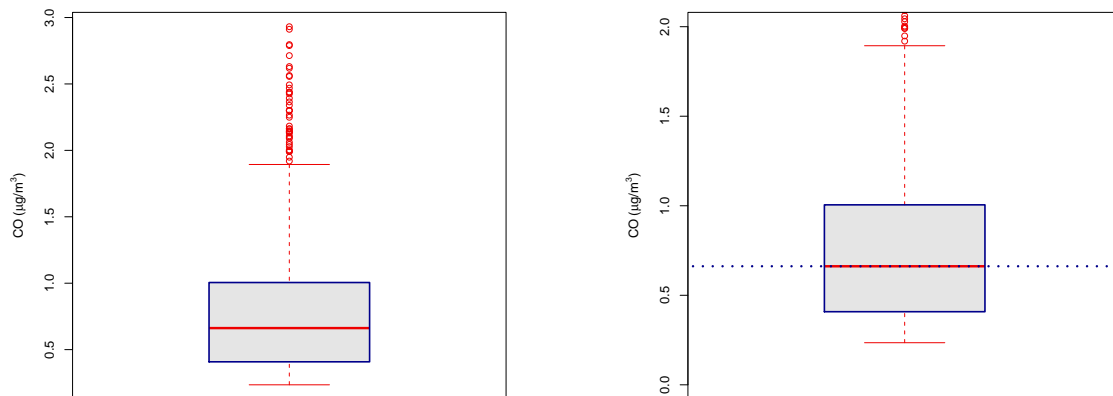


Abbildung 2.45: Boxplot der CO-Tagesmittelwerte

Die Abbildungen mit den Boxplots zu den Tagesmittelwerten des Schadstoffes Kohlenmonoxid zeigen nochmals die Ergebnisse der Analysen zuvor. Deutlich wird auch hier das etwas höhere Minimum.

### 2.3.2 Unterscheidung nach dem Jahr

Begonnen wird wieder mit der Aufteilung bezüglich der Jahre. Bei den Analysen der beiden Schadstoffe zuvor wurde kein signifikanter Unterschied erkannt. Die nachfolgende Tabelle zeigt, dass diese Annahme auch hier korrekt erscheint, jedoch am ehesten von einem Unterschied gesprochen werden kann.

Tabelle 2.16: Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Jahr

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
CO	0.61	0.80	0.118	5.89	0.63	0.37	0.98	0.97
- Jahr 2005	0.51	0.71	0.133	5.17	0.60	0.33	0.85	1.32
- Jahr 2006	0.69	0.89	0.118	5.89	0.66	0.45	1.11	0.73

Die Mediane der beiden Teildatensätze unterscheiden sich um fast  $0.2 \mu\text{g}/\text{m}^3$ . Die Rechtsschiefe der Daten zeigt sich im Jahr 2005 deutlicher als im Jahr 2006. Weiters unterscheiden sich das 1. und 3. Quartil.

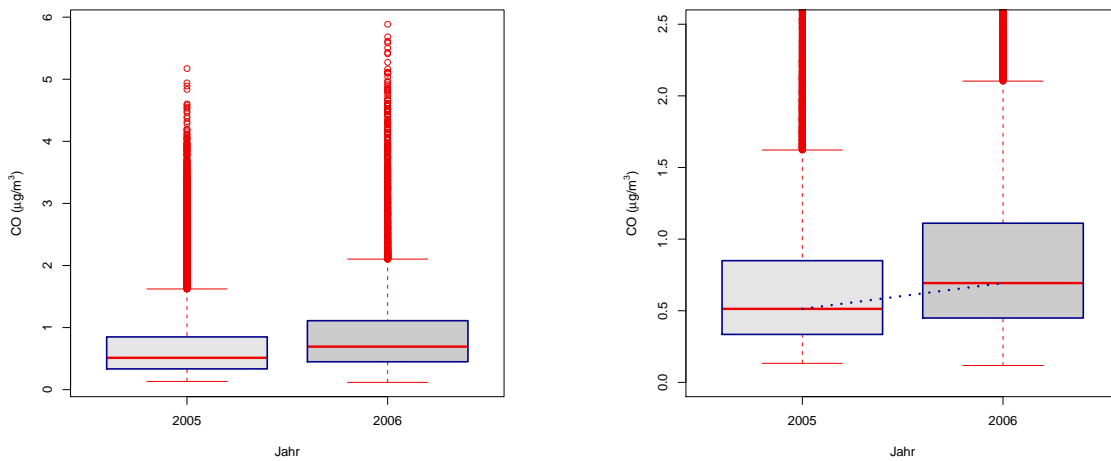


Abbildung 2.46: Boxplotserie der CO-Halbstundenwerte bzgl. Jahr

Im Jahr 2006 ist die Dichte der Ausreißer im höheren Bereich größer als im Jahr 2005. Außerdem ist das Maximum etwas höher im Jahr 2006. Bei beiden Boxplots zeigt sich auch hier das extreme Verhältnis zwischen Wertebereich und interquartilen Bereich.

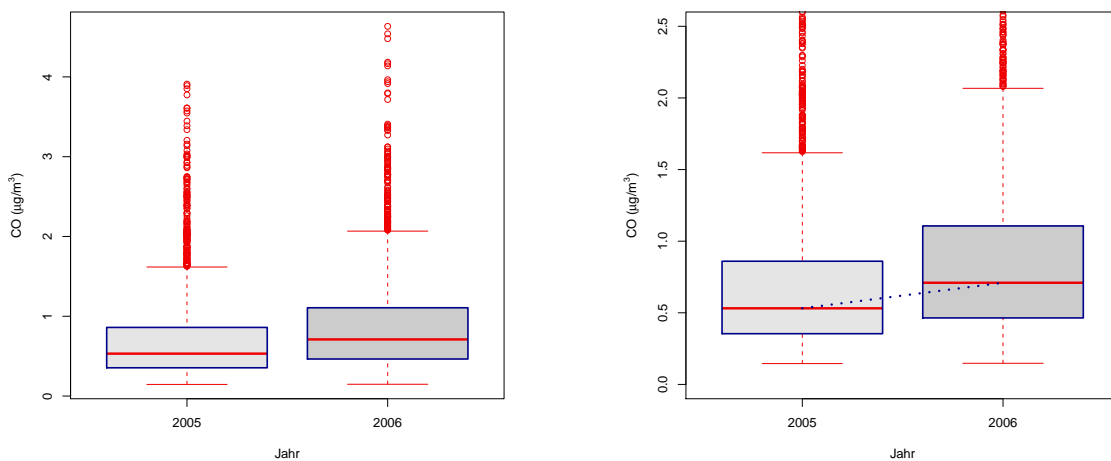


Abbildung 2.47: Boxplotserie der CO-4-Stunden-Mittelwerte bzgl. Jahr

Der Median des Datensatzes gehörend zum Jahr 2005 und das 1. Quartil fürs Jahr 2006 liegen beinahe auf derselben Höhe. Dies bedeutet, dass 50 % der Daten von 2005 ungefähr unter der Grenze liegen, unter der 25 % der Daten von 2006 liegen. Jedoch ist der interquartile Bereich so klein, dass diese Tatsache nicht auf einen signifikanten Unterschied hinweist.

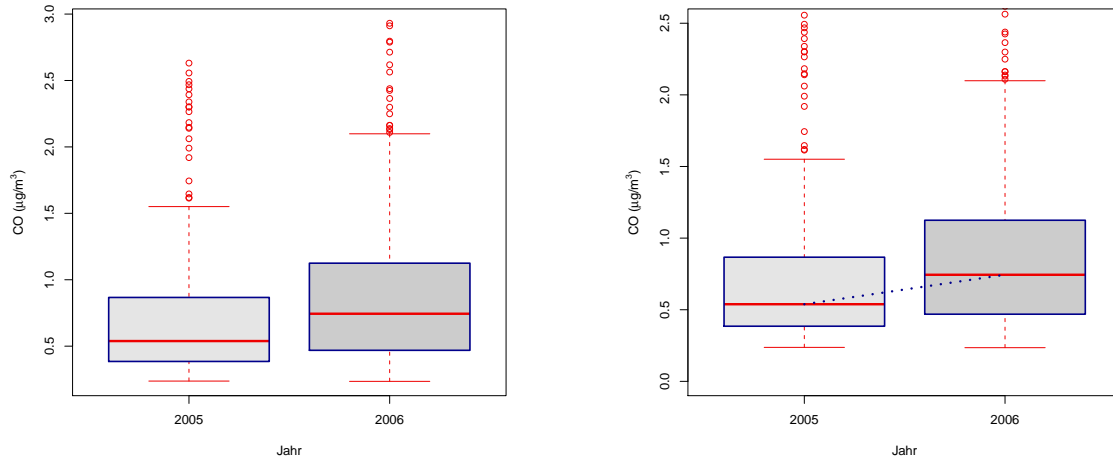


Abbildung 2.48: Boxplotserie der CO-Tagesmittelwerte bzgl. Jahr

Der Unterschied in den Jahren ist also beim Schadstoff Kohlenmonoxid am ehesten vorhanden. Im Jahr 2006 war die Schadstoffbelastung nämlich etwas stärker.

### 2.3.3 Unterscheidung nach der Saison

Die signifikanteste Unterscheidung ist die mittels der Saison. Diese Signifikanz wird, wie auch bei allen anderen Analysen, zu Beginn durch statistische Kenngrößen in einer Tabelle aufgezeigt.

Tabelle 2.17: Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Saison

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
CO	0.61	0.80	0.118	5.89	0.63	0.37	0.98	0.97
- Sommer	0.40	0.47	0.118	2.30	0.23	0.29	0.60	0.28
- Winter	0.94	1.14	0.172	5.89	0.73	0.62	1.43	0.58

Der Unterschied der Daten zu den beiden Saisonen zeigt sich in Größen wie Median und Mittelwert. Aber auch die Maxima bestätigen die Verschiedenheit der beiden Datensätze. Deren Unterschied beträgt nämlich sogar  $3.59 \mu\text{g}/\text{m}^3$ .



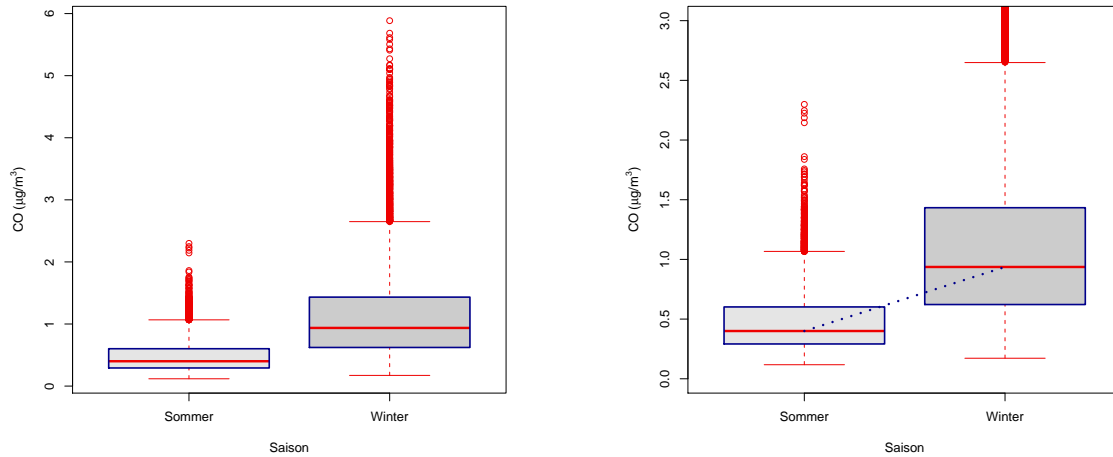


Abbildung 2.49: Boxplotserie der CO-Halbstundenwerte bzgl. Saison

Die Streubreite ist im Winter größer. Auffällig sind im Sommer ein paar Halbstundenwerte über  $2 \mu\text{g}/\text{m}^3$ . Die Relation zwischen den interquartilen Bereichen ist auch erwähnenswert. Dieser ist beim Winterdatensatz fast dreimal so groß wie der des Sommerdatensatzes.

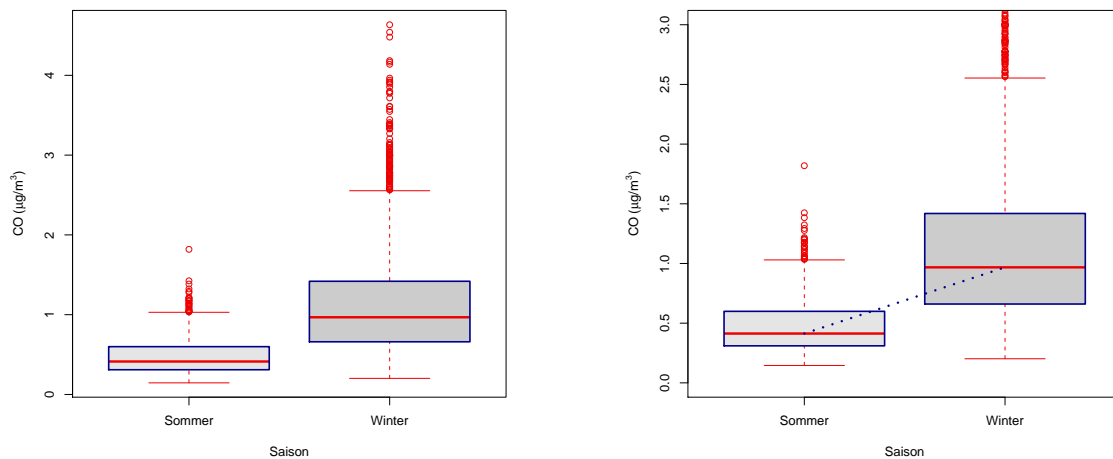


Abbildung 2.50: Boxplotserie der CO-4-Stunden-Mittelwerte bzgl. Saison

Wiederum erkennt man, dass das 1. Quartil des Winters und das 3. Quartil des Sommers beinahe auf der selben Höhe liegen. Dies bedeutet, 75 % der Sommerdaten in dem Wertebereich liegen, in dem 25 % der Winterdaten sind. Die beiden Mediane unterscheiden sich um mehr als  $0.5 \mu\text{g}/\text{m}^3$ . Man kann also von einem deutlichen Anstieg der Werte sprechen.

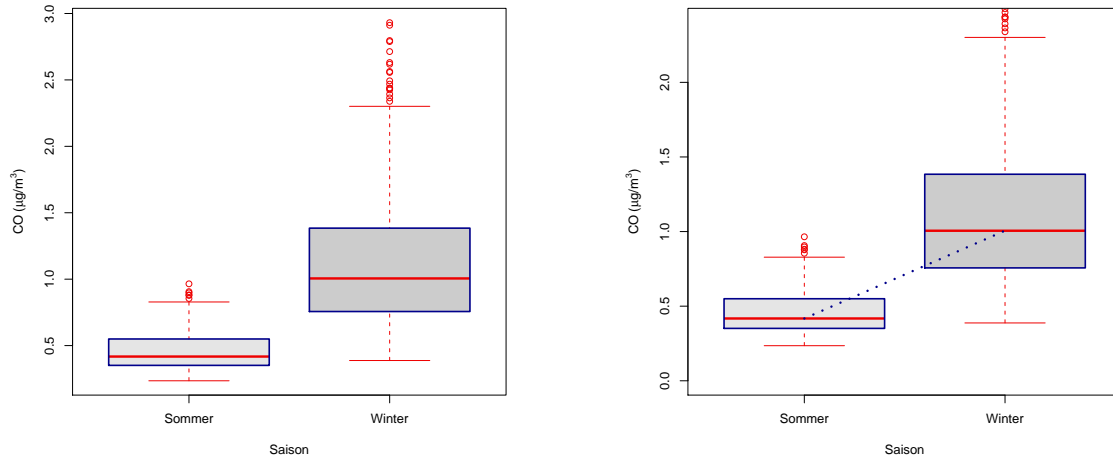


Abbildung 2.51: Boxplotserie der CO-Tagesmittelwerte bzgl. Saison

Die Tagesmittelwerte zeigen das extremste Bild. Der Median der Winterdaten liegt höher als das Maximum der Sommerdaten. Dies verdeutlicht den wesentlichen Unterschied in der Saison.

### 2.3.4 Unterscheidung nach dem Monat

Nach der Aufteilung bezüglich der Saison folgt die detailliertere Aufteilung, nämlich die bezüglich der Monate. Auch bei den Schadstoffdaten zum Kohlenmonoxid wurde ein klarer Unterschied zwischen Winter- und Sommerdaten aufgezeigt. Die nachfolgende Tabelle zeigt nun, auf welche Art der Anstieg bzw. Abfall erfolgt.

Tabelle 2.18: Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Monat

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
CO	0.61	0.80	0.118	5.89	0.63	0.37	0.98	0.97
- Jänner	1.26	1.45	0.173	5.89	0.90	0.77	1.91	0.35
- Februar	0.84	1.07	0.203	5.58	0.70	0.58	1.37	0.69
- März	0.63	0.76	0.186	2.86	0.40	0.47	0.93	0.53
- April	0.44	0.52	0.173	2.30	0.26	0.34	0.62	0.64
- Mai	0.35	0.39	0.128	1.25	0.15	0.28	0.45	0.40
- Juni	0.31	0.35	0.118	1.40	0.14	0.25	0.41	0.57
- Juli	0.33	0.39	0.142	1.42	0.18	0.26	0.48	0.37
- August	0.57	0.54	0.133	1.35	0.21	0.35	0.68	-0.32
- September	0.62	0.62	0.172	1.75	0.26	0.39	0.78	-0.12
- Oktober	0.79	0.85	0.179	2.89	0.37	0.60	1.06	0.07
- November	0.99	1.13	0.251	3.92	0.57	0.70	1.41	0.35
- Dezember	1.38	1.56	0.172	5.68	0.86	0.96	1.96	0.25

Auch hier zeigt sich wiederum eine deutliche Differenz zwischen den Monaten. Diese zeigt sich vor allem an den unterschiedlich großen Maxima. Weiters ist der Schiefekoeffizient nicht in jedem Monat positiv. In manchen Monaten ist dieser also negativ und die Daten somit linksschief verteilt.

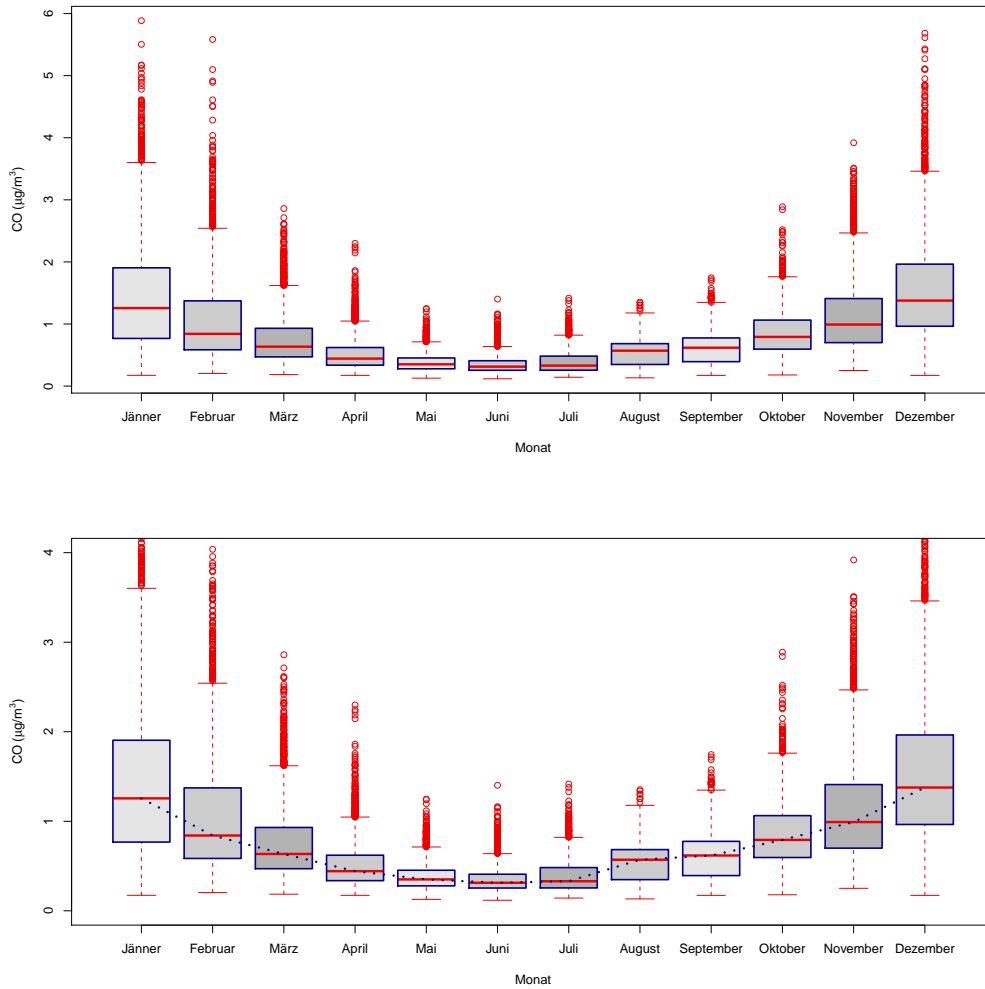


Abbildung 2.52: Boxplotserie der CO-Halbstundenwerte bzgl. Monat

Viele und auch hohe Ausreißer gibt es in den Monaten Jänner, Februar und Dezember. Die Boxplots zu den Monaten Juni bis September haben nur eine kleine Streubreite. Im Gegensatz zum Schadstoffdatensatz für Schwefeldioxid erfolgt hier der Übergang zwischen den Monaten kontinuierlich.

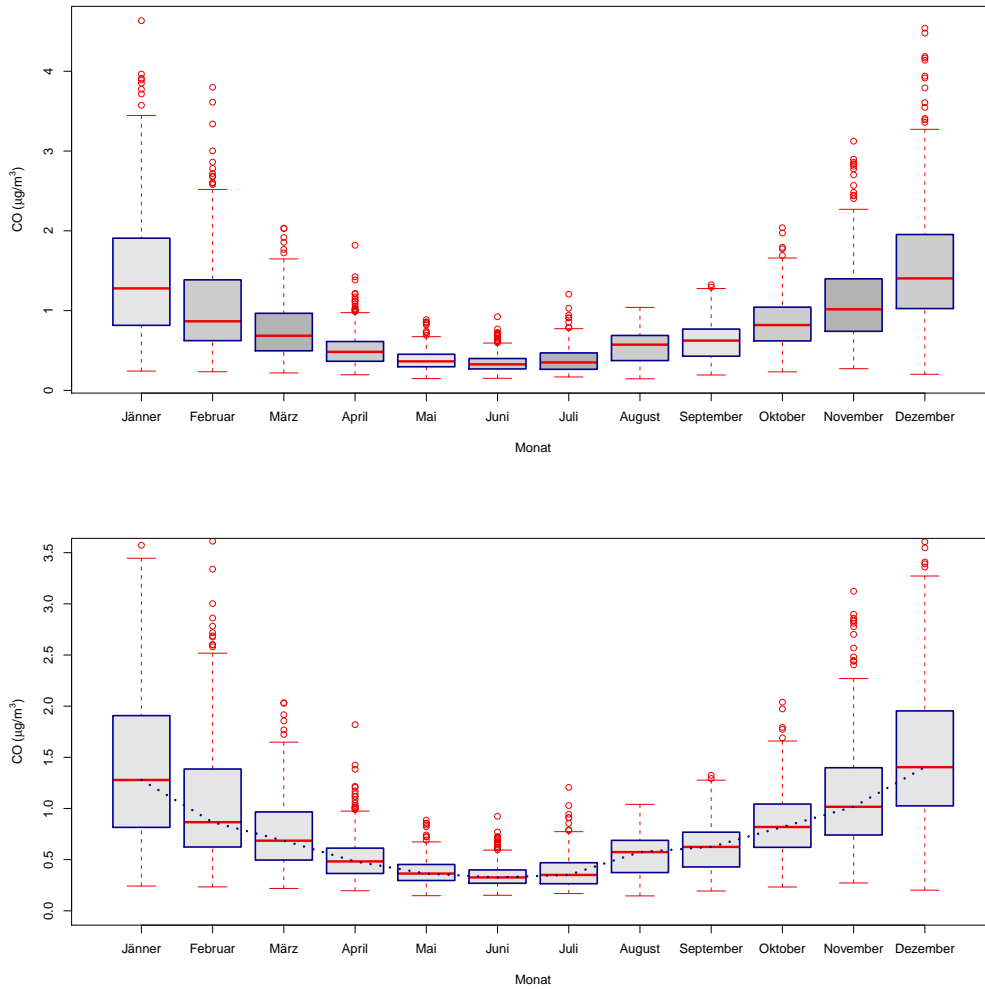


Abbildung 2.53: Boxplotserie der CO-4-Stunden-Mittelwerte bzgl. Monat

Als belastungsgeringster Monat erweist sich der Juni. In den Monaten Jänner und Dezember ist die Schwefeldioxidbelastung am größten. Einen deutlichen Anstieg des Medians gibt es beim Übergang von Juli zu August. Der August ist außerdem ein Beispiel für eine linksschiefe Verteilung der Daten.

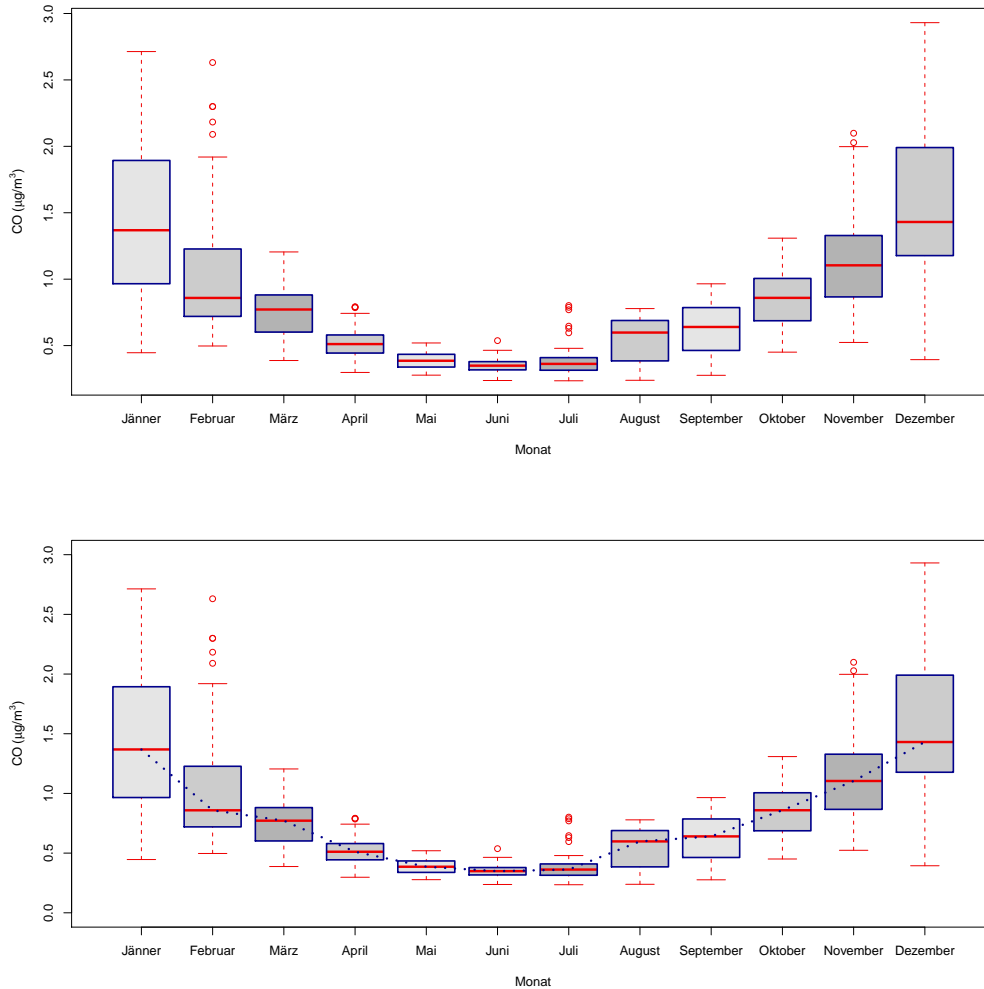


Abbildung 2.54: Boxplotserie der CO-Tagesmittelwerte bzgl. Monat

Die Tagesmittelwerte zeigen am deutlichsten den extremen Unterschied zwischen Monaten wie Dezember und Juni. Dieses Bild erklärt somit ganz ersichtlich den Unterschied, den man bei der Aufteilung bezüglich der Saison gesehen hat.

### 2.3.5 Unterscheidung nach dem Tagestyp

Die nächste Unterscheidung ist die bezüglich der in drei Kategorien zusammengefassten Tage. Der Datensatz wird zuerst mittels statistischer Kenngrößen auf mögliche Differenzen untersucht, die in der nachfolgenden Tabelle aufgezeichnet sind.

Tabelle 2.19: Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Tagestyp

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
CO	0.61	0.80	0.118	5.89	0.63	0.37	0.98	0.97
- Werktag	0.64	0.83	0.118	5.89	0.65	0.40	1.02	0.98
- Samstag	0.59	0.76	0.152	4.41	0.60	0.35	0.92	1.01
- Sonn-/Feiertag	0.54	0.71	0.133	5.61	0.57	0.32	0.90	0.93

Wie auch schon bei den Analysen der beiden anderen Schadstoffe zeigt sich, dass die Belastung an den Werktagen am größten ist. Belastungsgeringste Tage sind nach den Größen wie Median, Mittelwert und Standardabweichung die Sonn- bzw. Feiertage, wenn auch das Maximum der Samstage das kleinste ist.

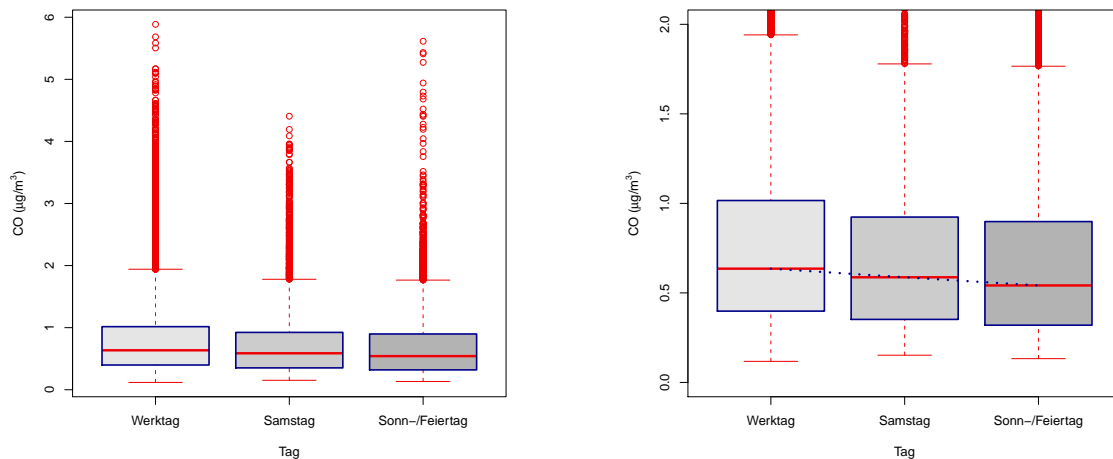


Abbildung 2.55: Boxplotserie der CO-Halbstundenwerte bzgl. Tagestyp

Freitags und sonntags erreichen die extremen Ausreißer fast die selbe Höhe. Jedoch unterscheidet sich die Struktur dieser in ihrer Anzahl. Die interquartilen Bereiche sind in allen 3 Kategorien fast gleich groß. Die oberen und unteren Grenzen derer sinken nur leicht Richtung Sonn- bzw. Feiertage.

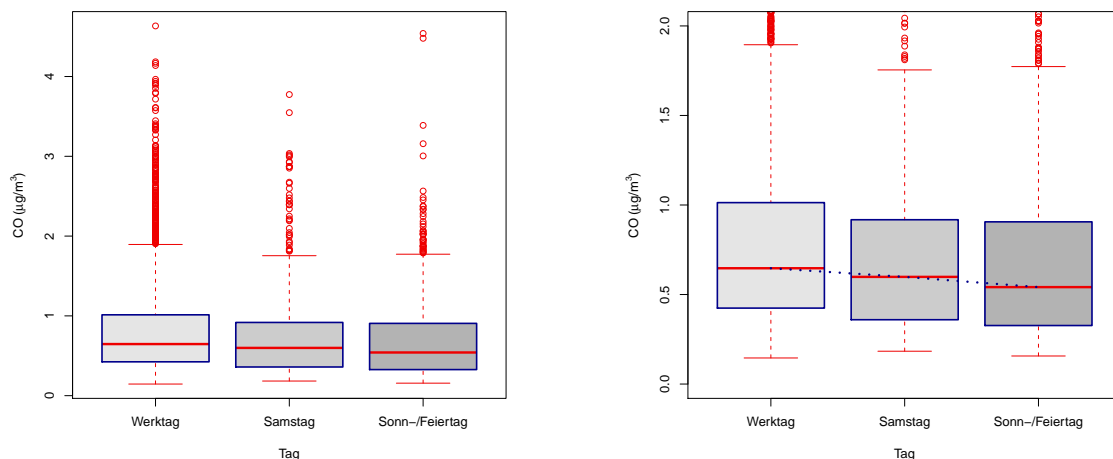


Abbildung 2.56: Boxplotserie der CO-4-Stunden-Mittelwerte bzgl. Tagestyp

Die Mediane liegen fast auf derselben Höhe bzw. fallen nur gering von den Werktagen zu den Sonn- bzw. Feiertagen. Da diese in allen drei Boxplots der Serie in der unteren Hälfte des interquartilen Bereichs liegen, kann man auf eine Rechtsschiefe der Daten schließen.

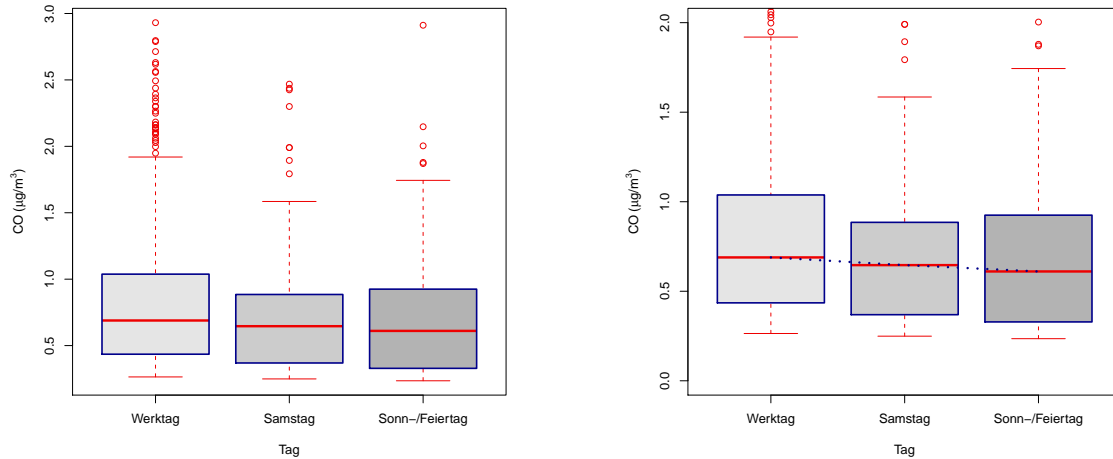


Abbildung 2.57: Boxplotserie der CO-Tagesmittelwerte bzgl. Tagestyp

Das leichte Absinken der Mediane und der Unterschied in den 3. Quartilen zeigen also insgesamt einen Unterscheid von den Werktagsdaten zu den anderen, wenn auch einen eher geringen.

### 2.3.6 Unterscheidung nach dem Wochentag

Nun kommt man zur Unterscheidung bezüglich der Wochentage. Dies dient nach der Unterscheidung in die 3 Tagestypen vor allem der Analyse der Werktage. Ob es an diesen Tagen unterschiedlich starke Belastungen gab, zeigt unter anderem die nachfolgende Tabelle der Kenngrößen.

Tabelle 2.20: Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Wochentag

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
CO	0.61	0.80	0.118	5.89	0.63	0.37	0.98	0.97
- Montag	0.59	0.76	0.119	4.59	0.56	0.37	0.96	0.75
- Dienstag	0.63	0.83	0.118	5.17	0.64	0.39	1.04	0.81
- Mittwoch	0.64	0.85	0.152	5.89	0.69	0.41	1.03	1.12
- Donnerstag	0.63	0.85	0.136	5.68	0.71	0.39	1.02	1.17
- Freitag	0.68	0.85	0.128	4.61	0.63	0.42	1.05	0.83
- Samstag	0.59	0.77	0.152	4.41	0.61	0.35	0.93	1.08
- Sonntag	0.54	0.70	0.133	5.61	0.57	0.31	0.84	1.07

Belastungsschwächster Werktag ist der Montag. Die Größen Median und Mittelwert liegen hier sogar unter denen vom Samstag. Allgemein belastungsschwächster Tag ist wiederum der Sonntag. Besonders hoch ist die Belastung an den Freitagen. Sämtliche Datensätze zeigen eine rechtsschiefe Verteilung.

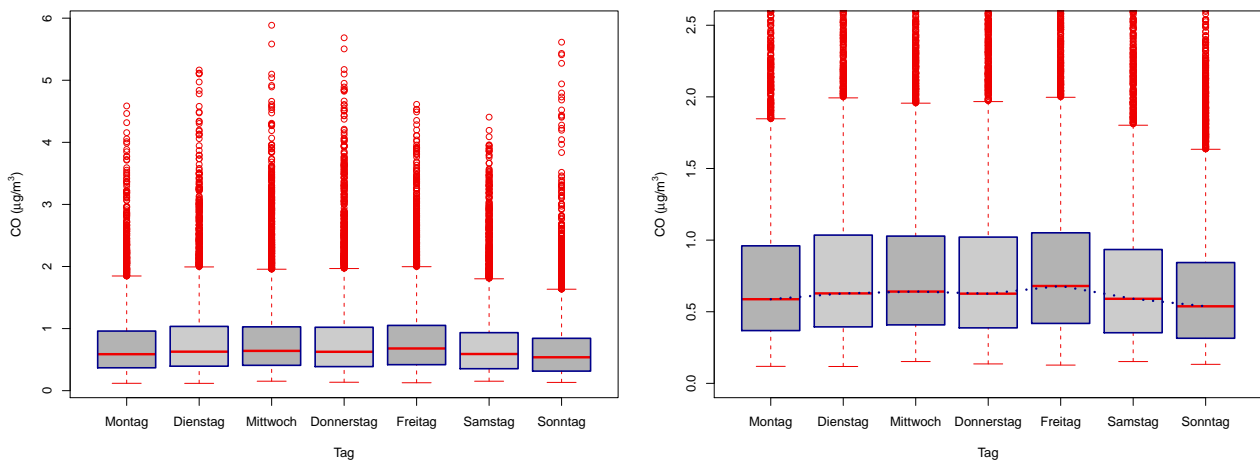


Abbildung 2.58: Boxplotserie der CO-Halbstundenwerte bzgl. Wochentag

Besonders hohe Ausreißer gibt es an den Tagen Mittwoch und Donnerstag. Aber auch die Sonntagsdaten zeigen einige auffällig hohe Werte. Die interquartilen Bereiche sind an allen Tagen ungefähr gleich groß, vor allem an den Werktagen.

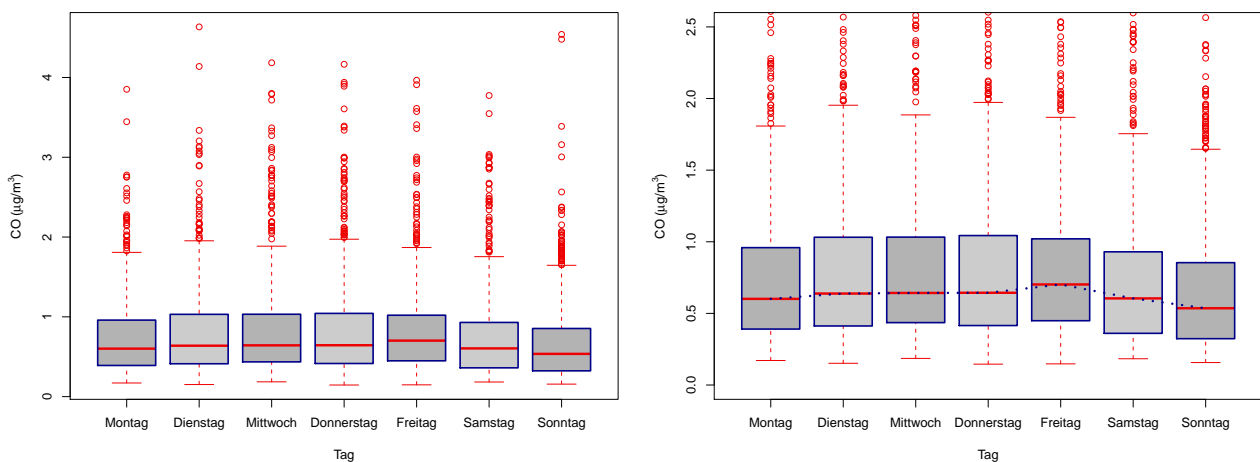


Abbildung 2.59: Boxplotserie der CO-4-Stunden-Mittelwerte bzgl. Wochentag

Der Median erreicht an den Freitagen sein Maximum. Er steigt dorthin von beiden Seiten an. Sämtliche Boxplots zeigen auf, dass die dazugehörigen Daten rechtsschief verteilt sind.



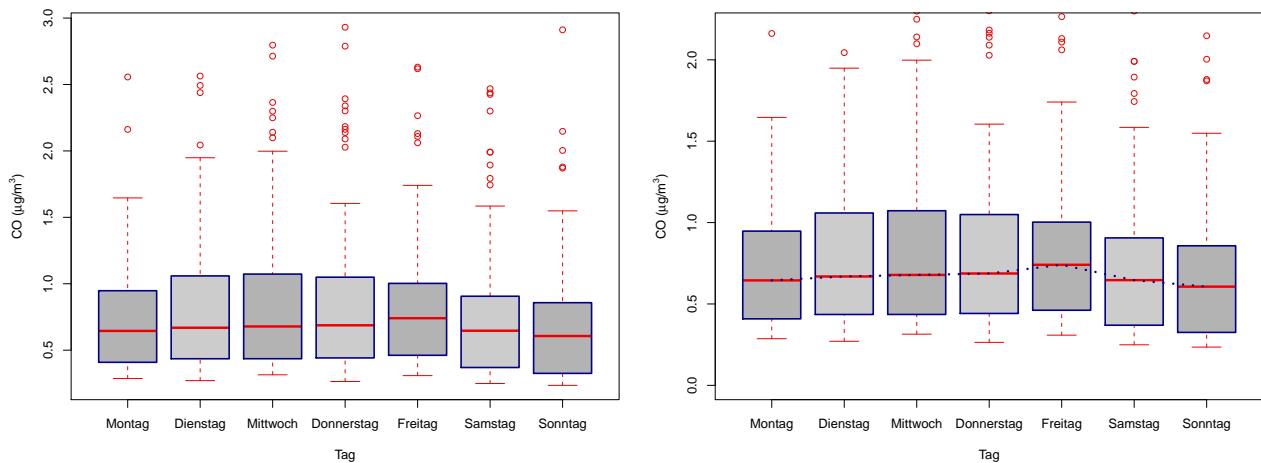


Abbildung 2.60: Boxplotserie der CO-Tagesmittelwerte bzgl. Wochentag

Die Boxplotserie der Tagesmittelwerte zeigt, dass auch an den belastungsschwächsten Tagen im Mittel bereits mindestens  $0.2 \mu\text{g}/\text{m}^3$  Kohlenmonoxid gemessen wurden. Somit ist Kohlenmonoxid der am stärksten in der Luft vorkommende, untersuchte Schadstoff.

### 2.3.7 Unterscheidung nach der Uhrzeit

Zum Schluss werden die Schadstoffdaten bezüglich der Uhrzeit analysiert. Beim Schadstoff Kohlenmonoxid wird eine Aufteilung in 4 Zeitspannen gewählt. Nachfolgende Tabelle zeigt die Zeitspannen, die von unterschiedlicher Länge sind, und die jeweils dazugehörigen Kenngrößen.

Tabelle 2.21: Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Uhrzeit

Daten	Median	Mittelwert	Min.	Max.	Std.-Abw.	1. Quartil	3. Quartil	Schiefe
CO	0.61	0.80	0.118	5.89	0.63	0.37	0.98	0.97
- 0-4 Uhr	0.50	0.69	0.118	3.99	0.59	0.29	0.84	1.25
- 4-11 Uhr	0.73	0.91	0.133	5.89	0.61	0.47	1.15	0.65
- 11-17 Uhr	0.52	0.62	0.154	3.39	0.39	0.33	0.74	0.80
- 17-0 Uhr	0.65	0.91	0.119	5.61	0.77	0.40	1.11	1.18

Von 0 bis 4 Uhr und von 11 bis 17 Uhr sind die Kenngrößen eher kleiner. Zu den zwei anderen Zeitspannen ergab die Berechnung eher größere Werte. Die Kenngrößen steigen und fallen also abwechselnd. Die Belastung ist zwischen 4 und 11 Uhr am stärksten.

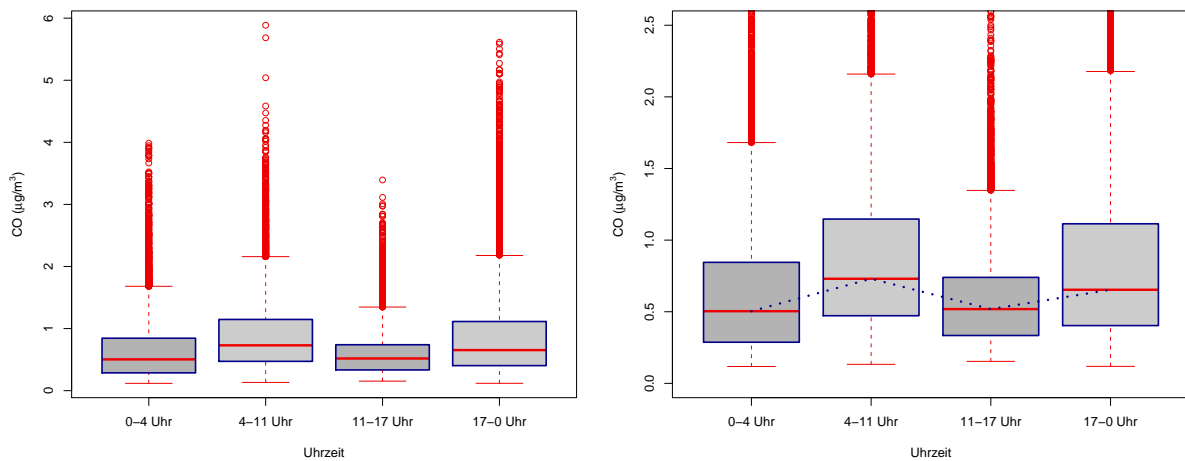


Abbildung 2.61: Boxplotserie der CO-Halbstundenwerte bzgl. Uhrzeit

Die meisten Ausreißer gibt es bei den Daten, die in der Zeitspanne 17-0 Uhr aufgezeichnet wurden. Zwischen 4-11 Uhr gibt es ein paar sehr hohe Werte. Bei den Boxplots mit höheren Medianen ist auch der interquartile Bereich größer. Die Kohlenmonoxidbelastung scheint von 11-17 Uhr am geringsten zu sein.

Auch hier möchte man zum Abschluss der Analyse detailliertere Boxplotserien betrachten und hat deswegen wiederum die Serien dargestellt, in denen die Daten nach der Uhrzeit und jeweils nach der Saison oder nach dem Tagestyp aufgespalten sind.

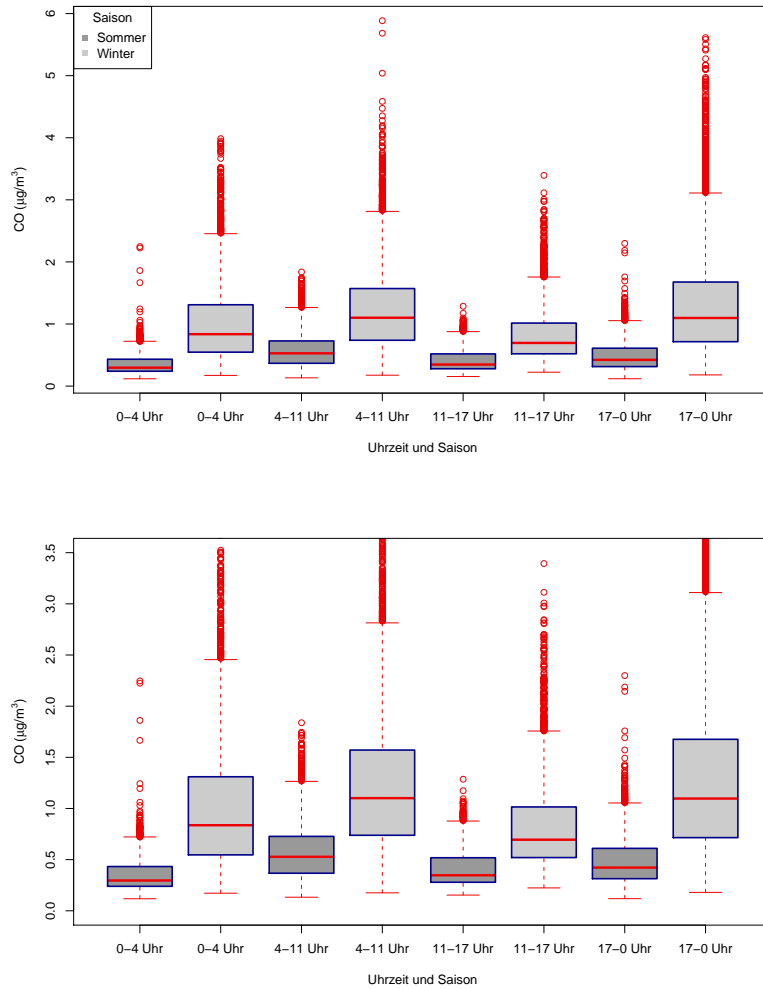


Abbildung 2.62: Boxplotserie der CO-Halbstundenwerte bzgl. Uhrzeit und Saison

Die Ergebnisse der Analyse der Daten, in der diese nach der Saison aufgeteilt waren, zeigen sich auch hier sehr stark und zwar in jeder der 4 Zeitspannen. Das bedeutet eine große Streubreite und hohe Werte bei den Boxplots gehörend zur Wintersaison und kleinere Werte mit weniger Streuung bei den Boxplots zur Sommersaison. Der Unterschied in den Zeitspannen wird bei den Daten gehörend zur Sommersaison nicht mehr so deutlich, ist aber dennoch vorhanden. Jedoch ist es dadurch schwierig ein Minimum in der Belastung zu bestimmen. Das Belastungsmaximum liegt ganz klar in der Wintersaison, jedoch sind sich hierbei die Boxplots zu den Zeitspannen 4-11 Uhr und 17-0 Uhr sehr ähnlich.

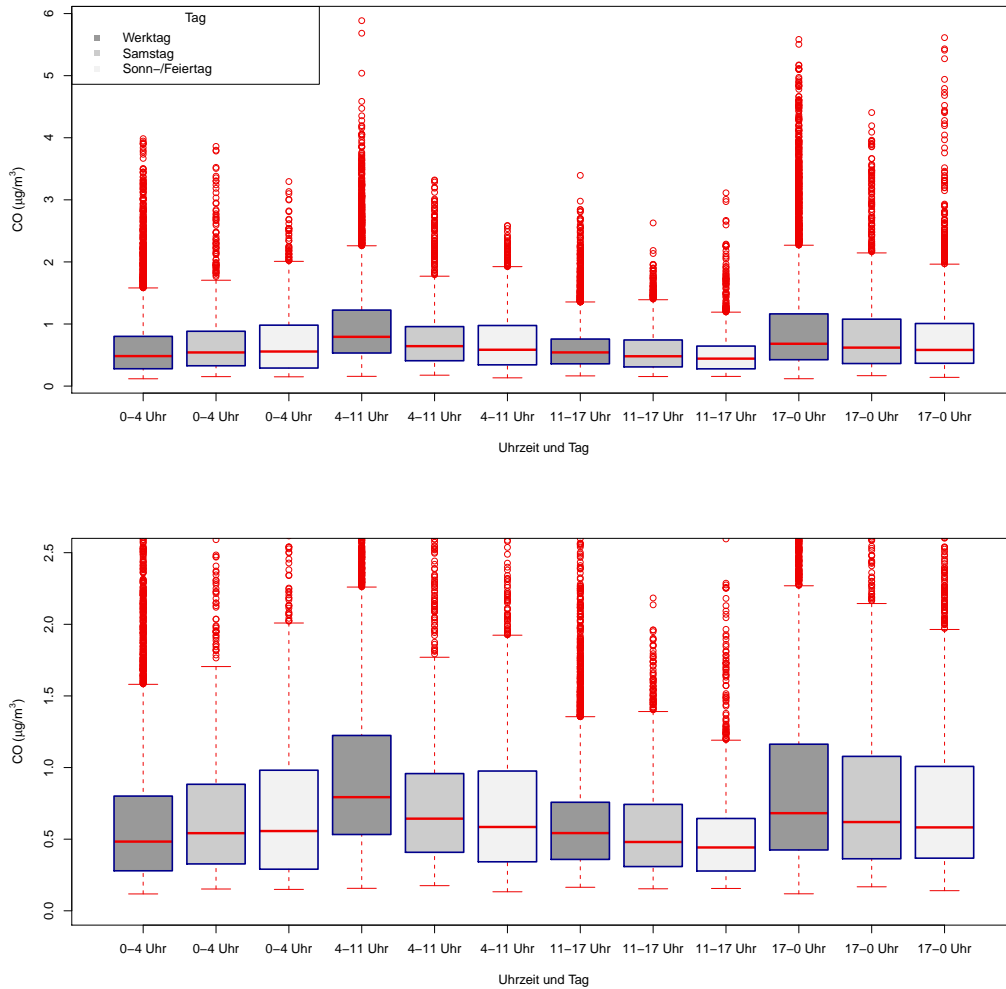


Abbildung 2.63: Boxplotserie der CO-Halbstundenwerte bzgl. Uhrzeit und Tag

In allen 4 Zeitspannen zeigt sich ein Unterschied bezüglich der Tagestypen, nur gibt es eine Auffälligkeit. Während bei den Zeitspannen 4-11 Uhr, 11-17 Uhr und 17-0 Uhr die Boxplotserie bezüglich der Tage dem zuvor erkannten Muster entspricht, nämlich einem Abfall in Richtung der Sonn- bzw. Feiertage, so ist dies in der Zeitspanne 0-4 Uhr nicht der Fall. In dieser Zeitspanne ist nämlich ein Anstieg des 3. Quartils zu bemerken, wenn auch die Höhe der Ausreißer wie gewohnt abfällt. Ansonsten ist der Unterschied in den Boxplots nicht allzu signifikant. Zu den belastungsstärksten Zeiten gehören die Zeitspannen 4-11 Uhr und 17-0 Uhr jeweils an den Werktagen.

## 3 Zusammenhangsanalyse

### 3.1 Zusammenhang der drei Schadstoffe

#### 3.1.1 Scatterplotmatrizen

Im Folgenden möchte man feststellen, ob es lineare Zusammenhänge zwischen den Verteilungen der einzelnen Schadstoffe gibt. Man betrachtet hierfür Scatterplotmatrizen und auch den Korrelationskoeffizienten von Pearson. Beides wird in den nächsten drei Abbildungen dargestellt bzw. angezeigt.

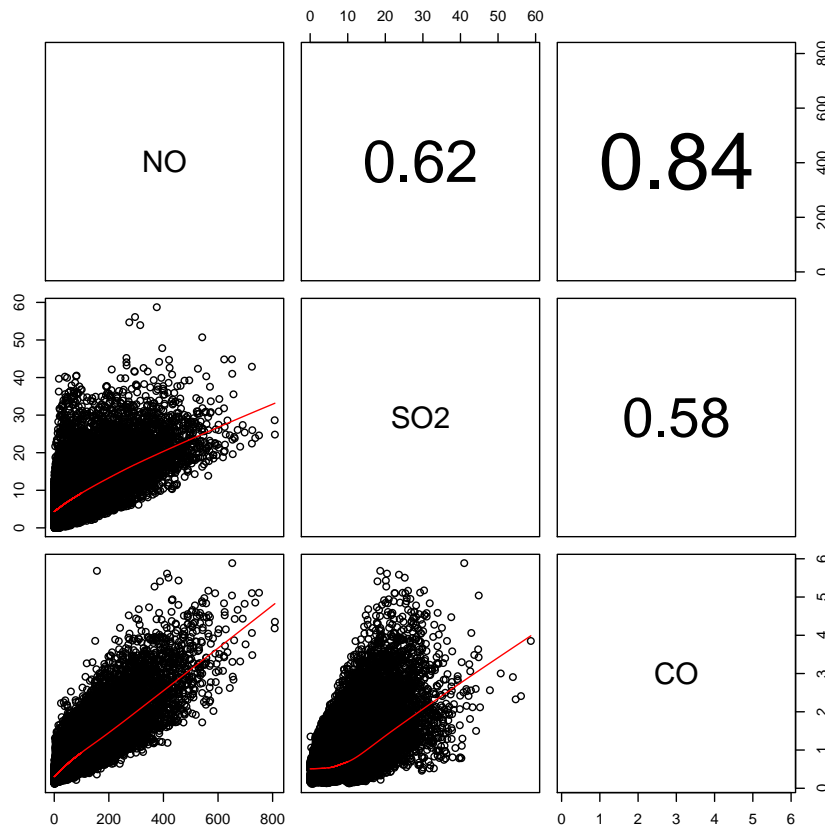


Abbildung 3.1: Scatterplotmatrix mit Kreuzkorrelationskoeffizienten der Halbstundenwerte

Der stärkste Zusammenhang besteht laut Korrelationskoeffizienten zwischen den Messdaten zu den Schadstoffen Stickstoffmonoxid (NO) und Kohlenmonoxid (CO), der schwächste zwischen Schwefeldioxid (SO<sub>2</sub>) und Kohlenmonoxid (CO). In allen drei Fällen kann man jedoch von einem eher starken Zusammenhang sprechen. Außerdem besteht ein positiver Zusammenhang, das heißt die Werte steigen jeweils gleichzeitig an.

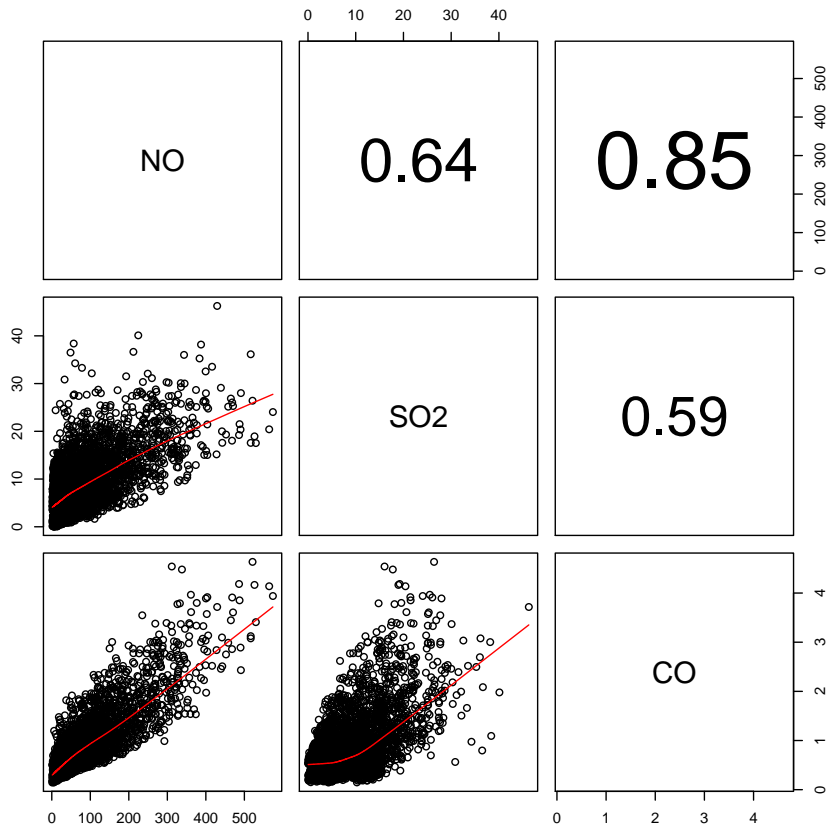


Abbildung 3.2: Scatterplotmatrix mit Kreuzkorrelationskoeffizienten der 4-Stunden-Mittelwerte

Auch die abgebildeten Scatterplotmatrizen zeigen diesen Sachverhalt. Während bei SO<sub>2</sub> und CO eine eher große „Punktwolke“ vorliegt, sammeln sich die Datenpunkte bei NO und CO in einem kleineren Gebiet an. Bei allen Scatterplots ist die Dichte der Punkte bei niedrigeren Werten größer.

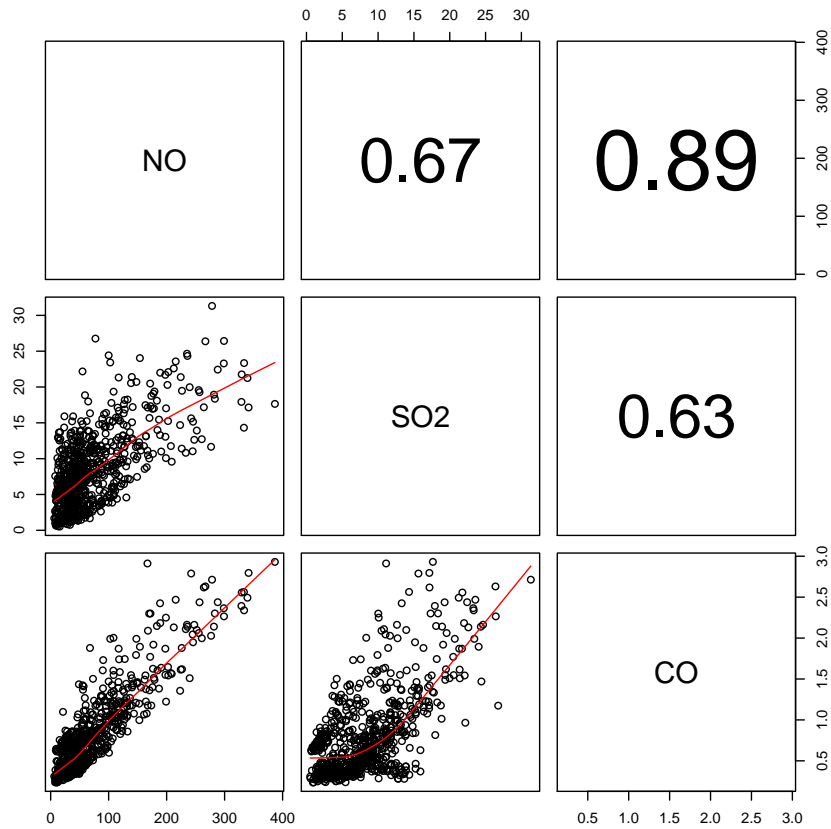


Abbildung 3.3: Scatterplotmatrix mit Kreuzkorrelationskoeffizienten der Tagesmittelwerte

Bei allen drei Scatterplots sind die Punkte in einer Art Trichter angeordnet. Je größer die Werte sind, desto größer ist die Streuung.

### 3.1.2 Scatterplotmatrizen bzgl. Tagestyp

Nach dem man nun einen ersten allgemeinen Eindruck über den linearen Zusammenhang der Daten erhalten hat, wirft man einen detaillierteren Blick auf die Sachlage. Dies wird zuerst mittels der Aufteilung der Daten nach dem Tagestyp, also nach den in 3 Kategorien zusammengefassten Tagen, erreicht.

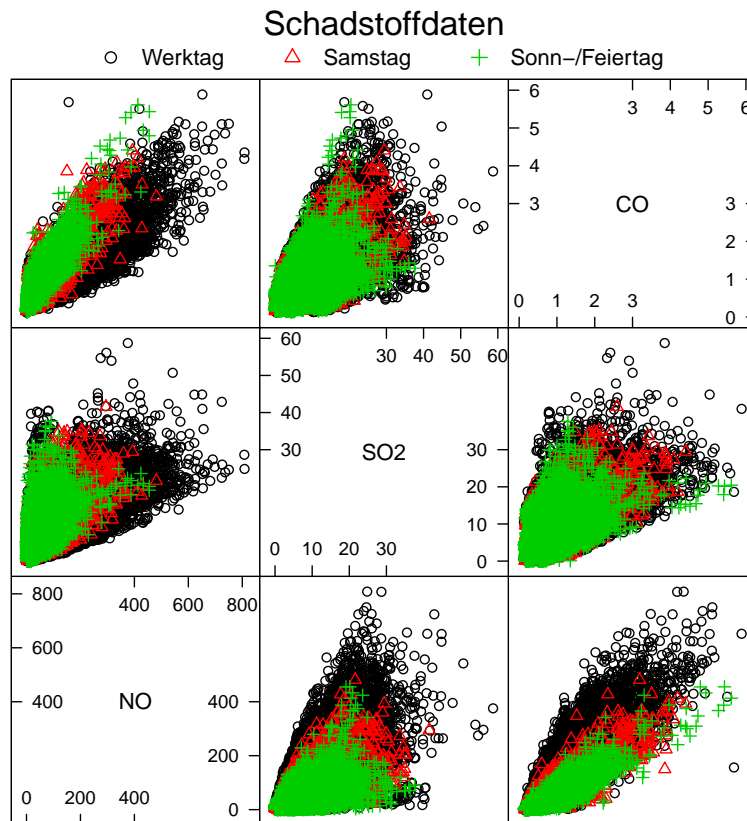


Abbildung 3.4: Scatterplotmatrix der Halbstundenwerte bzgl. Tagestyp

Alle drei Scatterplots zeigen, dass es an den Werktagen zu einer größeren „Punktwolke“ kommt als an den restlichen Tagen. Das bedeutet sowohl, dass die Daten der Werktage hohe Werte annehmen, als auch, dass die Streubreite dieser Daten groß ist. Daten mit niedrigeren Werten und einer kleinen Streubreite wurden eher an Sonn- bzw. Feiertagen aufgezeichnet. Weiters sind an Werktagen fünfmal mehr Daten als an Samstagen, Sonn- und Feiertagen vorhanden.



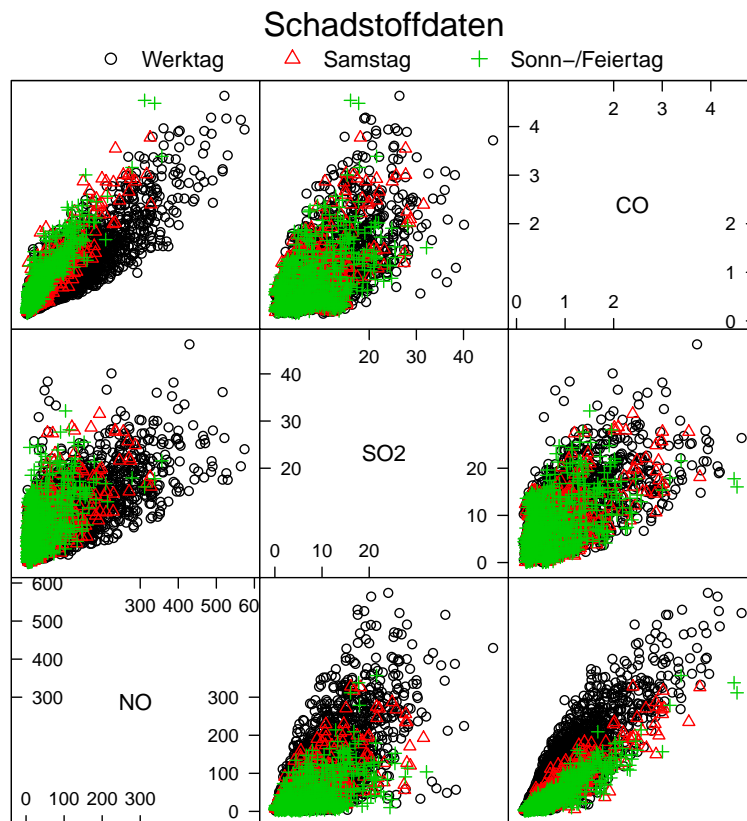


Abbildung 3.5: Scatterplotmatrix der 4-Stunden-Mittelwerte bzgl. Tagestyp

Speziell sieht man in den einzelnen Scatterplots folgendes: Der lineare Zusammenhang der Daten im Scatterplot von NO und CO zeigt sich auch in den aufgeteilten Datensätzen. Drei sehr große „Punktewolken“ sieht man bei der Analyse des Zusammenhangs von SO<sub>2</sub> und CO. Bei NO und SO<sub>2</sub> sammeln sich die Daten der Sonn- bzw. Feiertage in einem etwas kleineren Bereich an. Die Werktagsdaten zeigen hier eine große Streubreite.

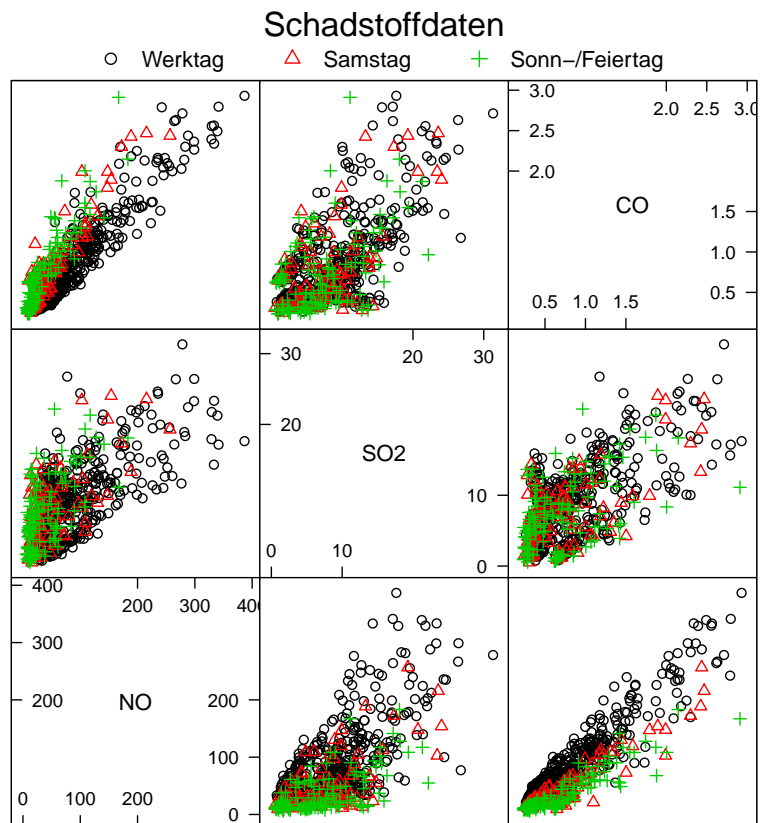


Abbildung 3.6: Scatterplotmatrix der Tagesmittelwerte bzgl. Tagestyp

Obige Abbildung zu den Tagesmittelwerten zeigt, dass auch die gemittelten Tageswerte hohe Ausreißer haben. Dies bedeutet, dass die Streubreite der Daten trotz Mittelung erhalten bleibt.

### 3.1.3 Scatterplotmatrizen bzgl. Saison

Neben der Unterscheidung der Daten nach dem Tag erwies sich auch die nach der Saison als interessant. Deswegen wählt man als weitere Anschauung die Scatterplotmatrix, bei der die Aufzeichnungen nach Sommer- bzw. Winterdaten unterschieden wurden.

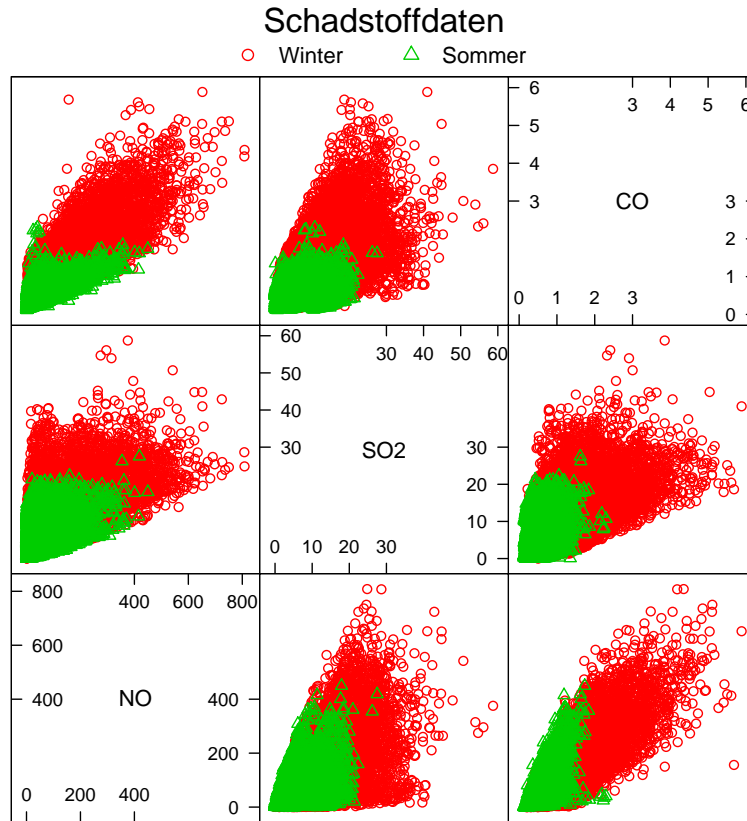


Abbildung 3.7: Scatterplotmatrix der Halbstundenwerte bzgl. Saison

Ganz klar zu sehen ist die Relation der beiden Teildatensätze zueinander. In allen drei Scatterplots ist die Ansammlung der Sommerwerte nur ungefähr halb so groß wie die der Winterwerte. Außerdem liegt die Datenwolke zur Sommersaison eher in der linken, unteren Hälfte der Datenwolke zur Wintersaison.

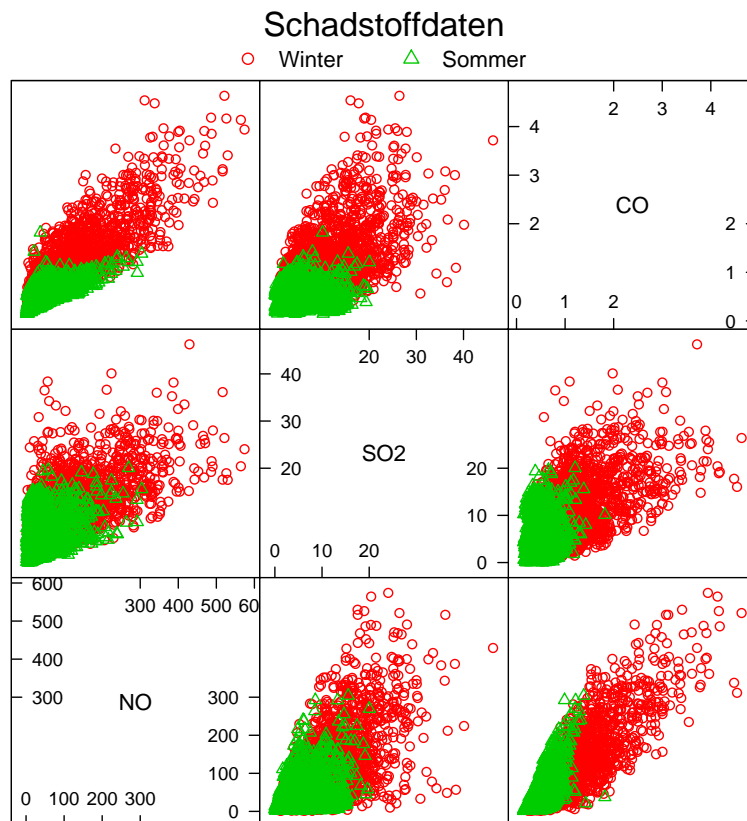


Abbildung 3.8: Scatterplotmatrix der 4-Stunden-Mittelwerte bzgl. Saison

Bei der Zusammenhangsanalyse von NO und SO<sub>2</sub> nehmen die Sommerdaten im Vergleich zu den anderen Scatterplots den größten Bereich ein. Während sich die Punkte beim Scatterplot von SO<sub>2</sub> und CO eher in einer runden Form ansammeln, sind sie beim Plot zu NO und CO eher entlang einer Linie aufgetragen.

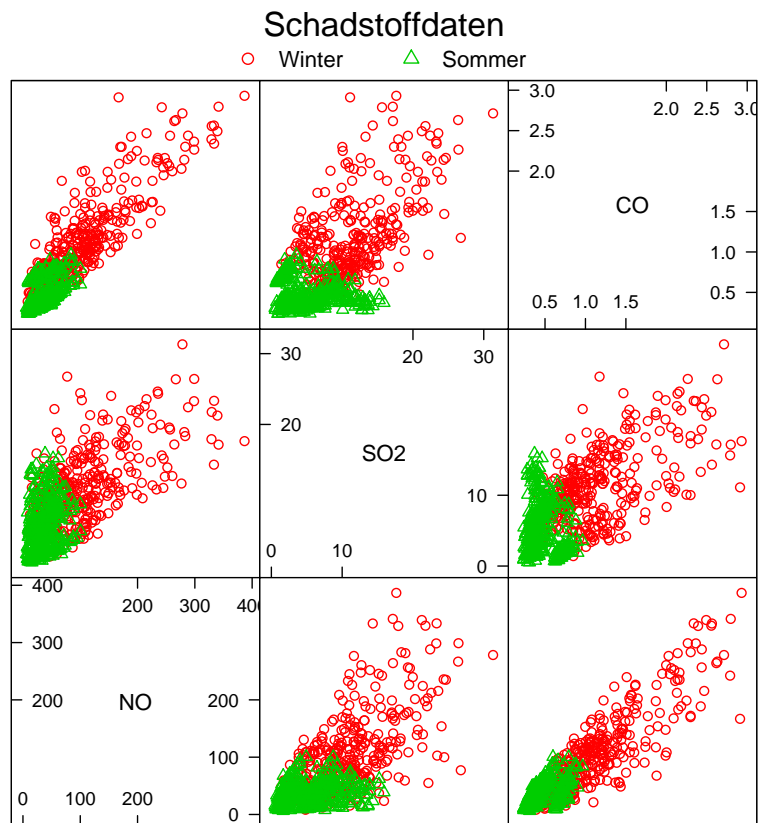


Abbildung 3.9: Scatterplotmatrix der Tagesmittelwerte bzgl. Saison

Die im Winter aufgezeichneten Daten zeigen trotz Tagesmittelung eine große Streubreite und hohe Werte. Die Sommerdaten sammeln sich bei allen Scatterplots in einem sehr kleinen Bereich an. Am größten ist dieser Zusammenhang zwischen den NO- und CO-Daten auch in der Aufteilung.

### 3.2 Spezialfall: Zusammenhang von NO und CO

Den stärksten Zusammenhang zeigten die Datensätze zu den Schadstoffen Stickstoffmonoxid (NO) und Kohlenmonoxid (CO). Aus diesem Grund wirft man noch einmal einen Blick auf die Sachlage und betrachtet deswegen die Scatterplots zu den Halbstunden-, 4-Stunden-Mittel- und Tagesmittelwerten:

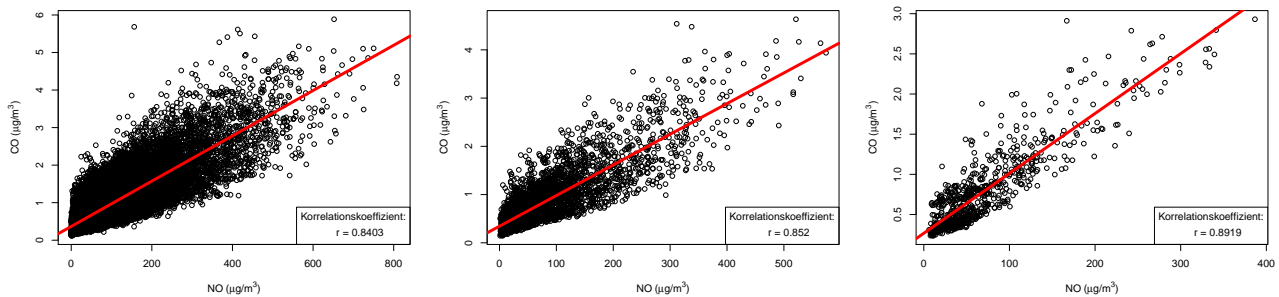


Abbildung 3.10: Scatterplots: NO und CO

Die Dichte der Daten ist bis zu einem Kohlenmonoxidwert von  $3 \mu\text{g}/\text{m}^3$  und einem Stickstoffmonoxidwert von  $400 \mu\text{g}/\text{m}^3$  am größten. Trotz Mittelung der Daten bleibt die Streuung in den höheren Wertebereich erhalten. Der Zusammenhang ist ganz klar positiv. Das heißt, wenn ein hoher Kohlenmonoxidwert aufgezeichnet wurde, wurde auch ein hoher Stickstoffmonoxidwert aufgezeichnet, und umgekehrt. Der berechnete Korrelationskoeffizient von Pearson bestätigt mit einem Wert über 0.8 die Stärke des Zusammenhangs.

## 4 Zeitreihenanalyse

### 4.1 Stickstoffmonoxid - NO

Zu Beginn der Zeitreihenanalyse des Schadstoffs Stickstoffmonoxid wirft man einen Blick auf die Darstellung der Zeitreihe an sich. Die folgenden drei Abbildungen zeigen den zeitlichen Verlauf der Stickstoffmonoxidbelastung vom 1.1.2005 bis zum 31.12.2006 mittels der Halbstunden-, 4-Stunden-Mittel- und Tagesmittelwerte.

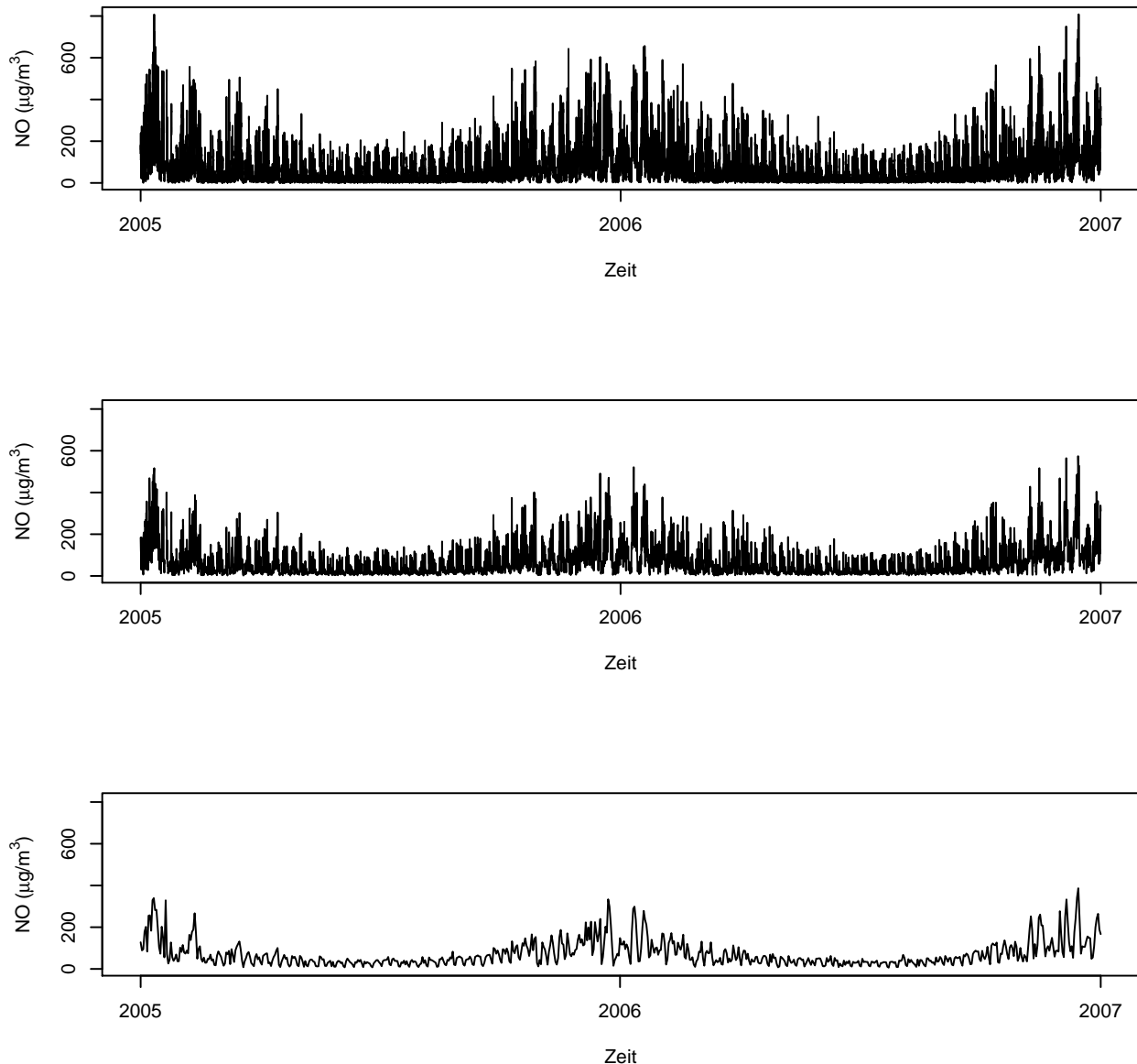


Abbildung 4.1: Zeitreihen des Schadstoffes NO

Betrachtet man die niedrigeren Werte so sieht man, dass um das Jahresende bzw. um den Jahresanfang herum größere Werte als in der Jahresmitte aufgezeichnet wurden. Nach oben zeigen die extrem großen Werte einen deutlichen Abfall Richtung Jahresmitte auf. Zwischen den beiden Jahren ist auch mittels Zeitreihenverlauf kein wesentlicher Unterschied erkennbar.

Die Tagesmittelung zeigt weiters ein Vorkommen eines regelmäßigen Schemas: Nach einigen hohen Werten kommen wenige kleinere Werte vor. Dieses Muster könnte die Abbildung der Entwicklung innerhalb einer Woche darstellen.

Detailliertere Aussagen über die zeitliche Entwicklung der Stickstoffmonoxidbelastung kann man mittels der Zerlegung der Zeitreihe machen. Folgende Grafik zeigt die Zeitreihe und die einzelnen Anteile der Zerlegung vom Trend bis hin zu den Residuen.

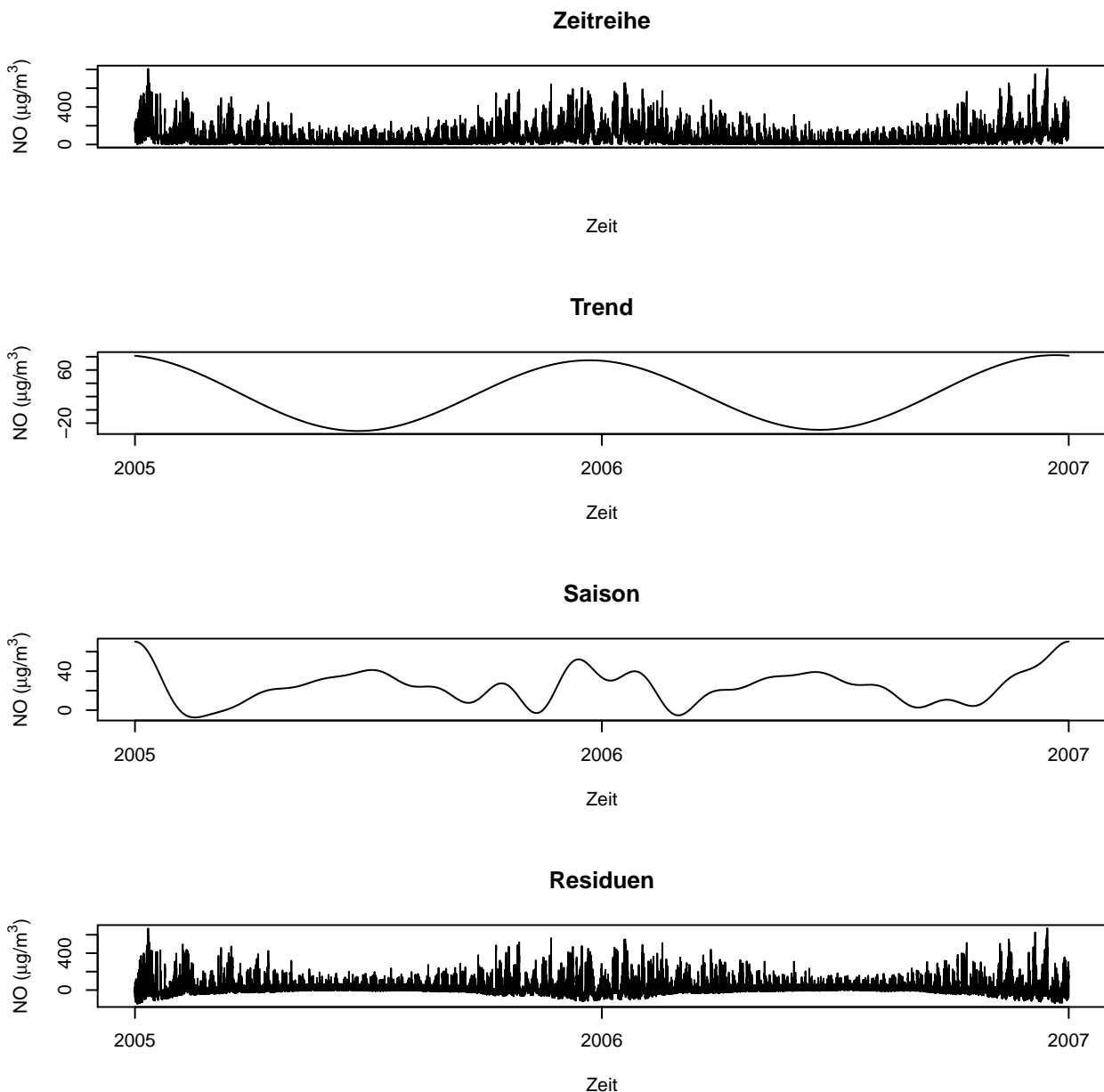


Abbildung 4.2: Zerlegung der Zeitreihe zu den NO-Halbstundenwerten

Der Trend zeigt ganz klar die Entwicklung die bereits bezüglich der Winter- bzw. Sommersaison aufgedeckt wurde. Das bedeutet einen Abfall der Schadstoffbelastung in Richtung Jahresmitte und einen Anstieg zum Jahreswechsel hin. Der saisonale Teil der Zerlegung zeigt kein besonderes Muster. Besonders hoch sind die Werte zu Beginn und am Ende der Aufzeichnungen. Die



Residuen machen jedoch den größten Teil der Zeitreihe aus. Teilweise erreichen diese eine Höhe von über  $600 \mu\text{g}/\text{m}^3$  gegenüber einer maximalen Schadstoffbelastung von ungefähr  $800 \mu\text{g}/\text{m}^3$ . Der größte Teil der Zeitreihe ist der zufällige Anteil, der mittels Residuen angegeben wird. Der Trend, der einen Wertebereich von ungefähr  $80 \mu\text{g}/\text{m}^3$  abdeckt, macht den zweitgrößten Anteil aus, in Relation zu den Residuen jedoch nur einen kleinen.

Um festzustellen, ob zwei Zeitpunkte der Zeitreihe miteinander korrelieren, betrachtet man die Autokorrelationsfunktion. Folgende Grafiken stellen diese jeweils für die Halbstundenwerte, 4-Stunden-Mittelwerte und Tagesmittelwerte dar.

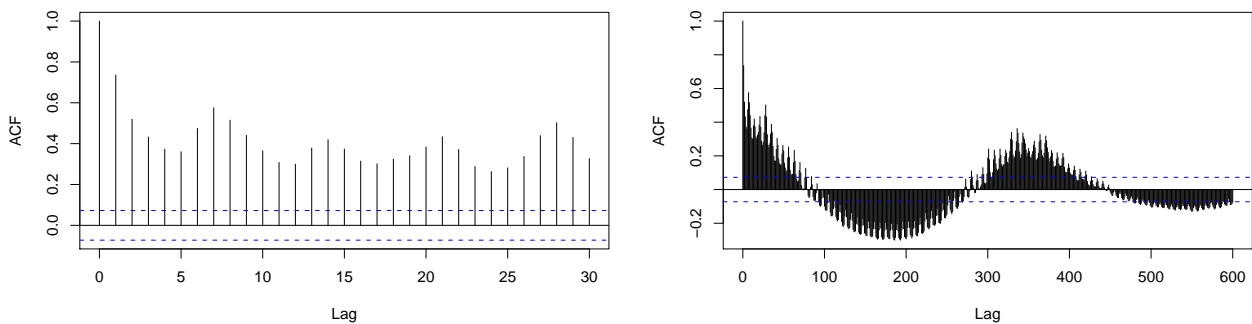


Abbildung 4.3: Autokorrelation der NO-Tagesmittelwerte

Die linke Abbildung zeigt die Autokorrelationsfunktion bis zu einem kleineren lag. Hier wird deutlich, dass der Korrelationskoeffizient beim lag 7 etwas höher als die Koeffizienten davor ist. Das bedeutet, dass zwei um 7 Tage, also um eine Woche, verschobene Werte stark korrelieren. Nicht nur beim lag 7, sondern bei allen lags, die ein Vielfaches von 7 sind, zeigt sich ein etwas höherer Korrelationskoeffizient. Sprich nicht nur der Wert eine Woche nach der Aufzeichnung, sondern auch der Wert nach 14 Tagen usw. ist dem aufgezeichneten Wert ähnlich. Neben dieser Beobachtung sieht man auch, dass sich das saisonale Verhalten der Zeitreihe in der Autokorrelationsfunktion widerspiegelt. So sind zum Beispiel Werte der Zeitreihe verschoben um etwa 180 Tage negativ zueinander korreliert, was eben ungefähr ein halbes Jahr ist. Beim lag 365 zeigt sich wiederum eine positive Korrelation. Dies spiegelt die Wiederholung bezüglich der Jahre wider.

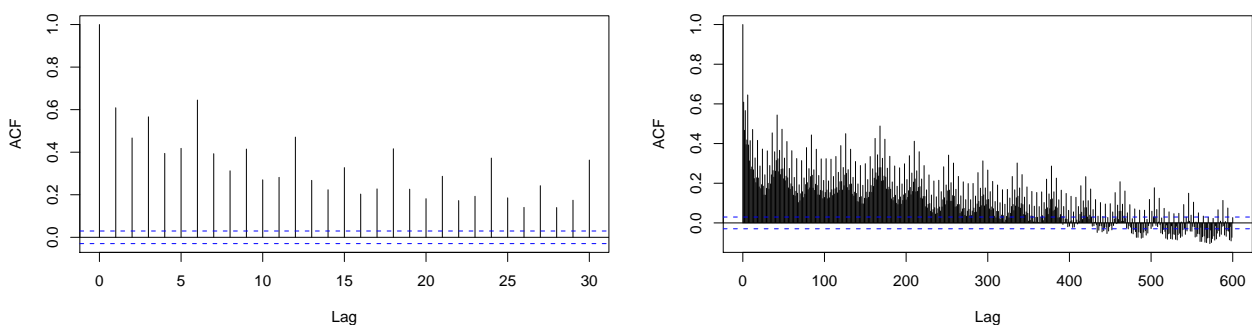


Abbildung 4.4: Autokorrelation der NO-4-Stunden-Mittelwerte

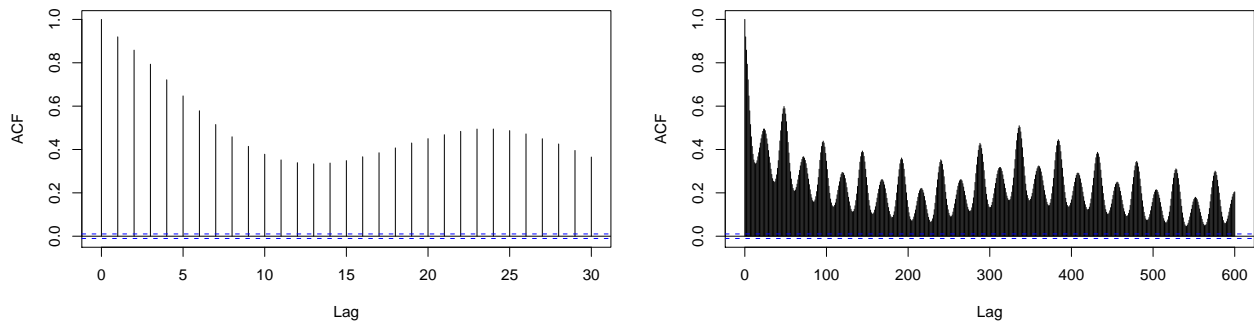


Abbildung 4.5: Autokorrelation der NO-Halbstundenwerte

Die obigen Abbildungen der Autokorrelation zu den Halbstunden- bzw. 4-Stunden-Mittelwerten zeigen grundsätzlich ungefähr dasselbe Bild wie das der Tagesmittelwerte, jedoch sind hier die Beobachtungen an einem anderen lag abzulesen. Dieser verschiebt sich nämlich von den Tages- zu den 4-Stunden-Mittelwerten um das 6-fache. Gegenüber den Halbstundenmittelwerten verschiebt es sich um das 48-fache. Zu erwähnen ist noch, dass sich durch die Mittelung der Wert des Autokorrelationskoeffizientes ein wenig ändert.

## 4.2 Schwefeldioxid - SO<sub>2</sub>

Einleitend betrachtet man die Darstellungen der in den Jahren 2005 und 2006 aufgenommenen Halbstundenwerte, 4-Stunden-Mittelwerte und Tagesmittelwerte der Schwefeldioxidbelastung als Zeitreihen. Diese sind in den folgenden Abbildungen dargestellt.

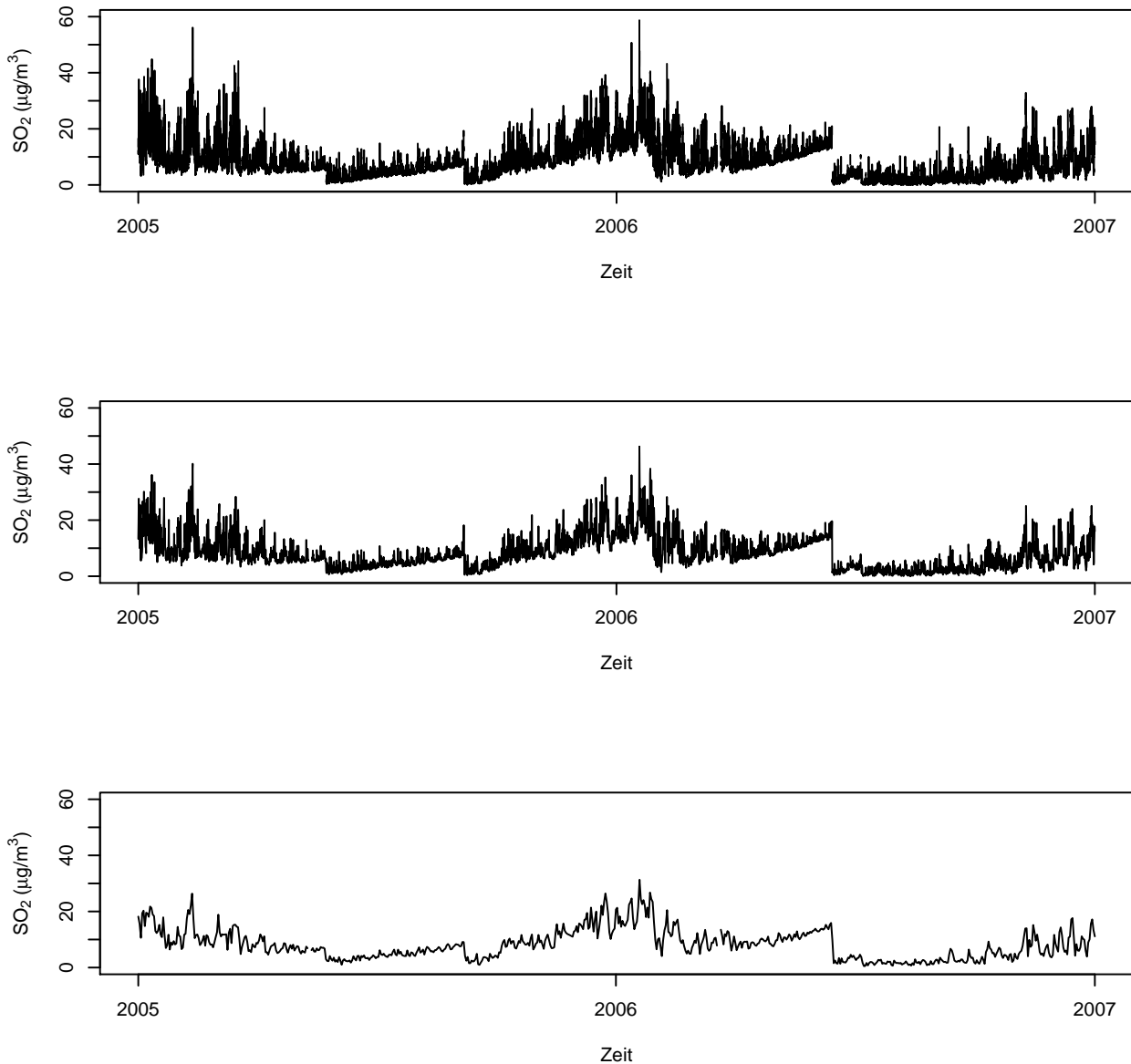


Abbildung 4.6: Zeitreihen des Schadstoffes SO<sub>2</sub>

Die dargestellten Zeitreihen zeigen ein sehr auffälliges Bild. Hier kann man kaum von einem monotonen Verhalten der Belastung sprechen. Ein Muster, das die Entwicklung innerhalb der Woche widerspiegelt, kann nicht aufgedeckt werden. Lediglich die Abnahme der Werte in Richtung Sommer kann bestätigt werden, wenn auch am Ende des Jahres 2006 die Belastung nicht signifikant stark ansteigt. Weiters ist einige Male ein schneller und starker Abfall der Schadstoffbelastung erkennbar. Besonders signifikant ist dieser etwa in der Mitte des Jahres 2006. Um diese Auffälligkeit genauer anzuschauen, betrachtet man die Zeitreihe für den Juni 2006.

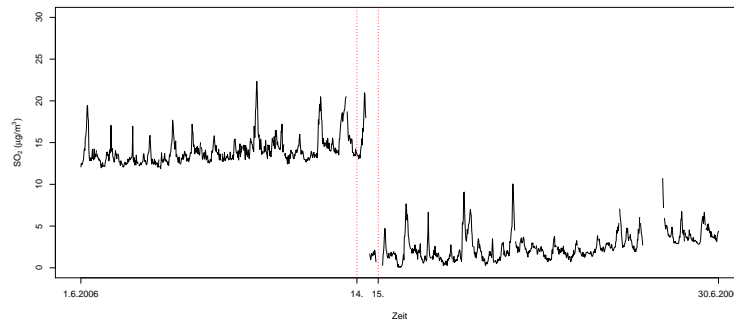


Abbildung 4.7: Auszug aus der Zeitreihe des Schadstoffes  $\text{SO}_2$  (Halbstundenwerte)

Der letzte große Werte wurde am 14.6.2006 zwischen 9:30-10:00 Uhr aufgenommen. Danach wurden einige Halbstundenwerte nicht aufgezeichnet. Der nächste in den Datensatz eingetragene Wert ist deutlich kleiner als der vom 14.6. Die beiden Werte unterscheiden sich nämlich um mehr als  $15 \mu\text{g}/\text{m}^3$ . Sieht man sich das Maximum der Belastung an den Abbildungen davor an, das ungefähr bei  $60 \mu\text{g}/\text{m}^3$  liegt, so wird die Signifikanz dieser Auffälligkeit deutlich.

Nach der Analyse der Zeitreihe an sich, betrachtet man die Zerlegung der Zeitreihe in gewisse Anteile, um so ein genaueres Bild über die zeitliche Entwicklung der Schwefeldioxidbelastung zu bekommen. Die Grafik dazu sieht wie folgt aus.

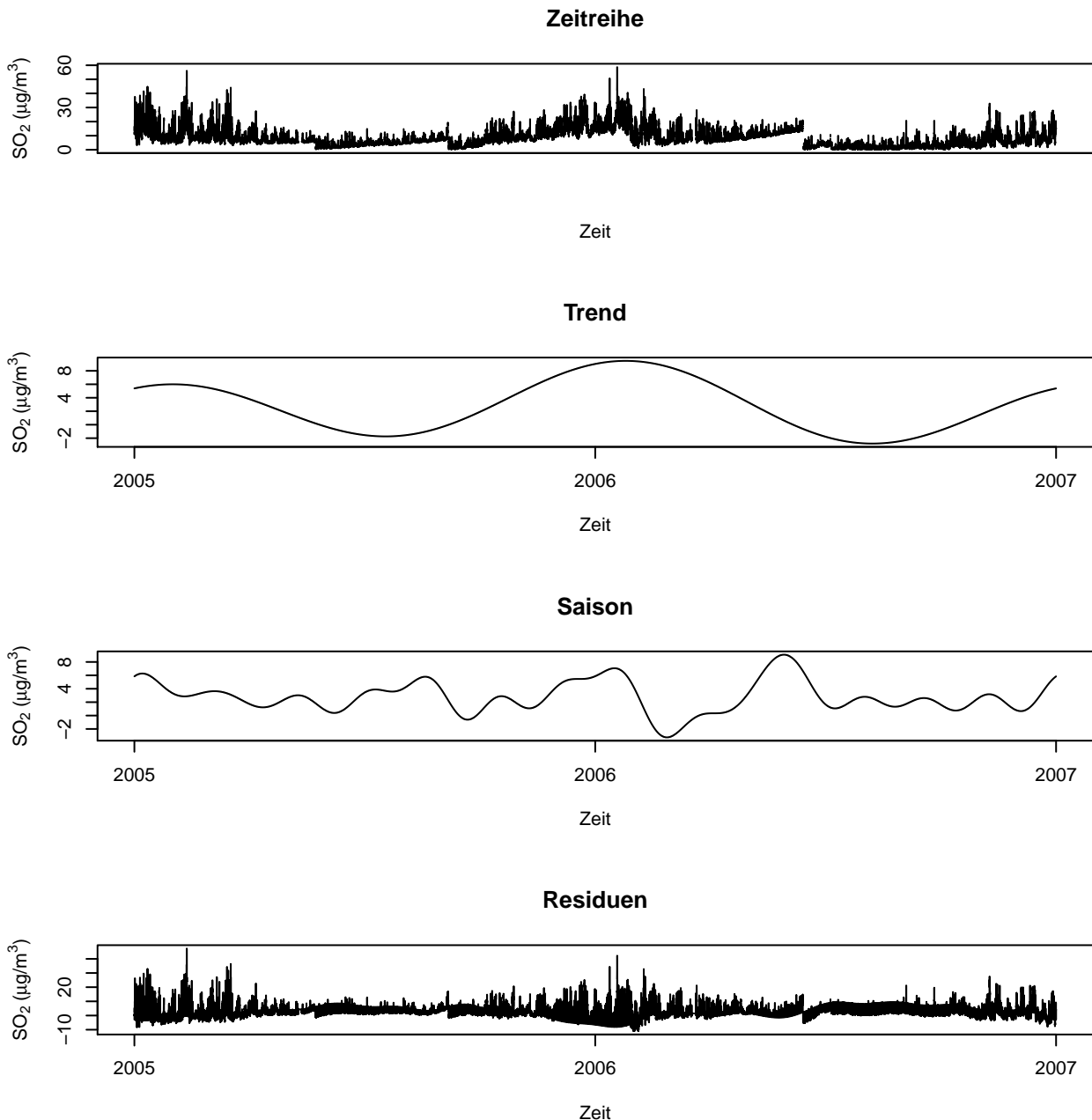


Abbildung 4.8: Zerlegung der Zeitreihe zu den SO<sub>2</sub>-Halbstundenwerten

Beim Schadstoff Schwefeldioxid zeigt der Trend-Anteil einen etwas anderen Verlauf als bei den restlichen Schadstoffen. Die Kurve nimmt nämlich ihr Maximum nicht zum Jahreswechsel hin an, sondern, leicht nach rechts verschoben, am Ende des Monats Jänner bzw. Anfang Februar. Weiters sind die Werte zum Aufzeichnungsanfang bzw. -ende nicht so groß. Dies zeigt insgesamt einen Unterschied in den Jahren 2005 und 2006. Deutlich ist jedoch aber auch hier das in Richtung Sommer die Werte abnehmen bzw. in Richtung Winter zunehmen. Weiters zeigt der saisonale Anteil kein Muster. Der Trend-Anteil erklärt ungefähr ein Sechstel der Zeitreihe. In der Kurve der Residuen zeichnen sich die Sprünge der eigentlichen Zeitreihe ab. Der Wertebereich der Residuen beträgt ungefähr  $50 \mu\text{g}/\text{m}^3$ . Ein sehr hoher Anteil, denn das Maximum der Zeitreihe liegt bei etwa  $60 \mu\text{g}/\text{m}^3$ .

Ob zwei verschiedene, also um eine gewisse Verzögerung verschobene Beobachtungen miteinander korrelieren, wird wiederum mit Hilfe der Autokorrelationsfunktion untersucht. Diese wird zum eigentlichen Datensatz und seinen beiden Mittelungen im Nachhinein grafisch dargestellt.

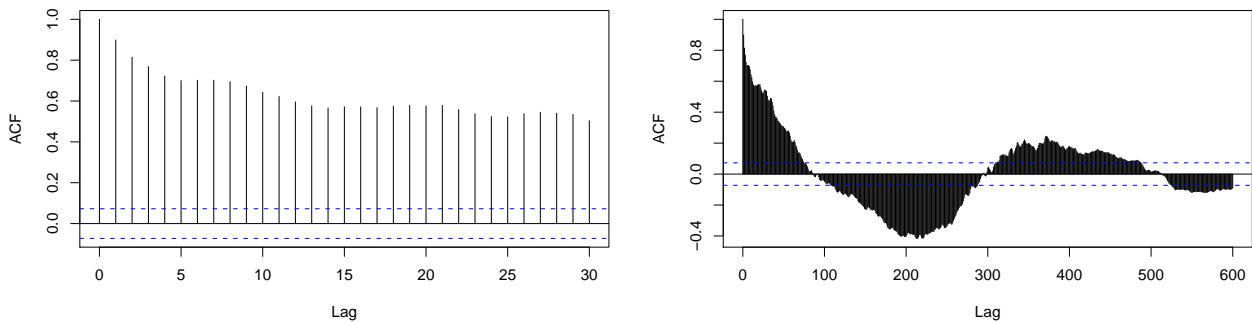


Abbildung 4.9: Autokorrelation der  $\text{SO}_2$ -Tagesmittelwerte

Hier können keine so deutliche Aussagen gemacht werden wie beim Schadstoff Stickstoffmonoxid. Anhand der linken Abbildung sieht man, dass bei kleineren lags von einer starken Korrelation gesprochen werden kann, das bedeutet ein Wert über 0,7. Jedoch gibt es keine lags bei denen sich ein extrem starker Zusammenhang zeigt. Weder wöchentliche noch jährliche Effekte können abgelesen werden. Die Autokorrelationsfunktion für die weitaus größeren lags zeigt eine Korrelation im negativen Bereich ab einem lag von etwa 100 bis zu einem von ungefähr 300. Dies bedeutet, dass die Werte, die um 100 bis 300 Tage verschoben sind, negativ miteinander korrelieren. Ab einem lag von circa 500 ist die Autokorrelation nicht mehr signifikant.

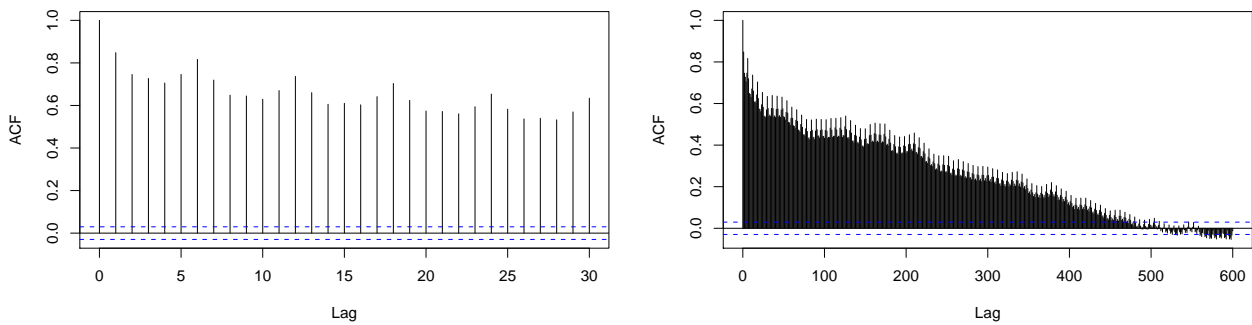


Abbildung 4.10: Autokorrelation der  $\text{SO}_2$ -4-Stunden-Mittelwerte

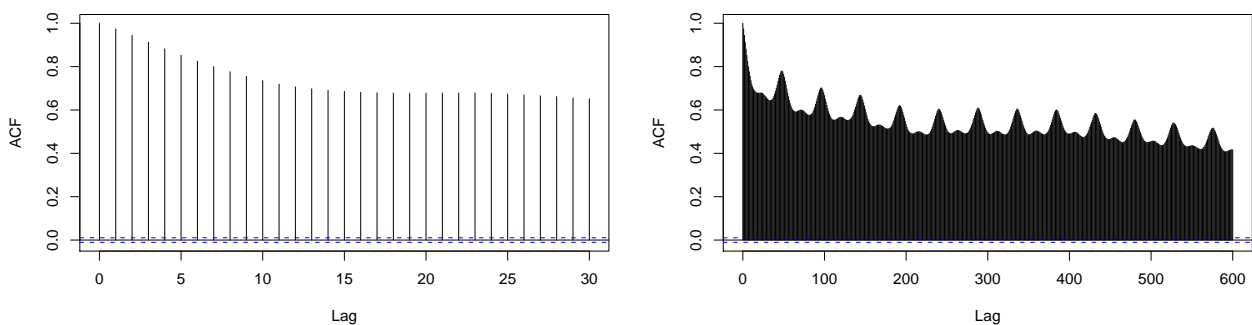


Abbildung 4.11: Autokorrelation der  $\text{SO}_2$ -Halbstundenwerte

Die obigen Abbildungen zeigen dieselben Ergebnisse wie die der Tagesmittelwerte, jedoch ist auch hier wiederum der lag anders abzulesen. So stellen die Abbildungen der Autokorrelation zu den Halbstunden- bzw. 4-Stunden-Mittelwerte einen kleineren Ausschnitt als die Abbildung der Autokorrelation zu den Tagesmittelwerten dar. Dieser geht bei den Halbstundenwerten bis zum lag 600, was bei den Tagesmittelwerten ungefähr dem lag 12 entspricht. Bei den 4-Stunden-Mittelwerten entspricht der maximale lag von 600 dem lag von 150 bei den Tagesmittelwerten.

### 4.3 Kohlenmonoxid - CO

Wiederum fängt man mit der einfachen Darstellung der Zeitreihe an. Anschließende Grafiken zeigen die zeitliche Entwicklung der Kohlenmonoxidbelastung mittels der aufgezeichneten Halbstundenwerte bzw. der berechneten 4-Stunden- und Tagesmittelwerte in den Jahren 2005 und 2006.

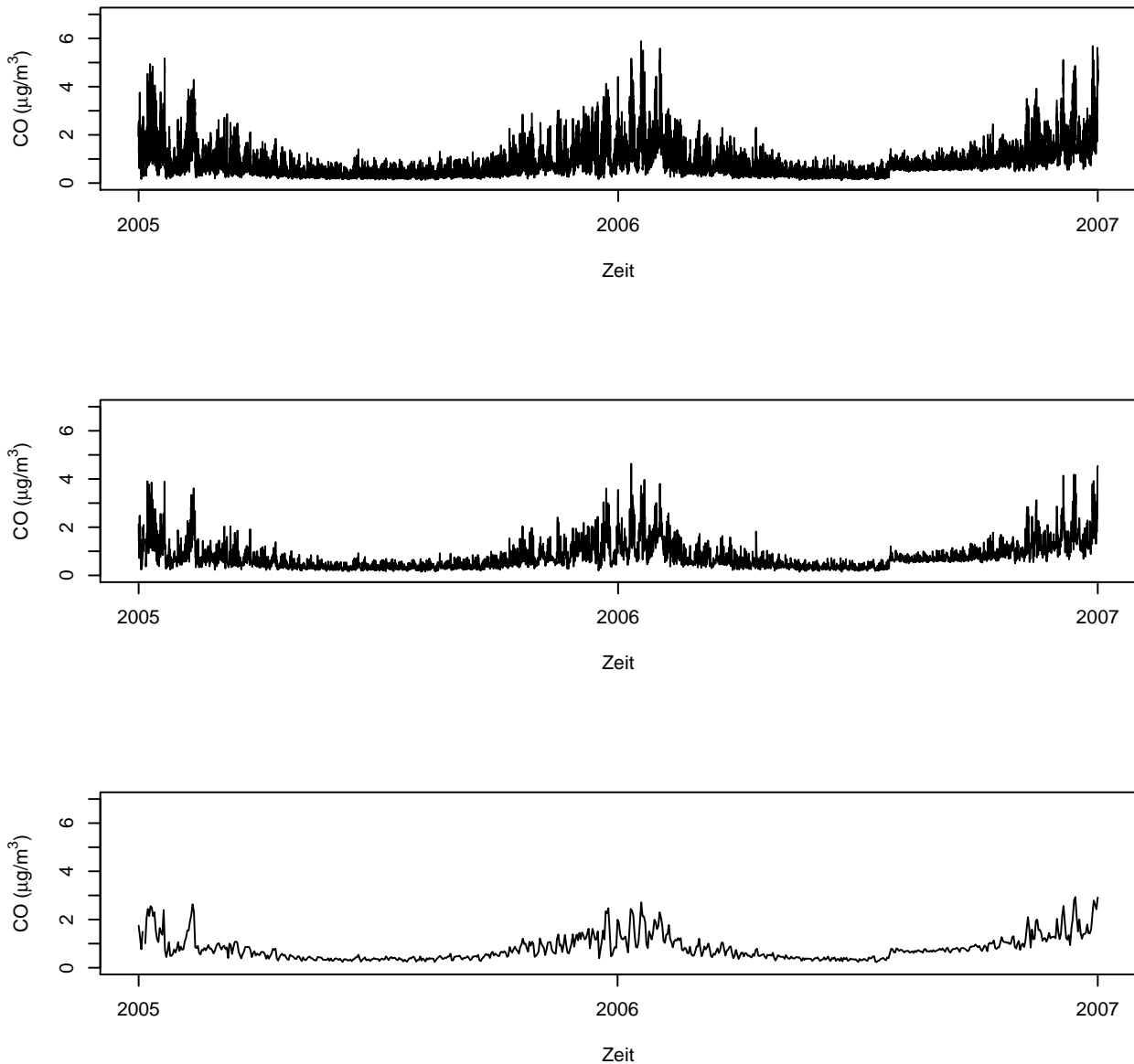


Abbildung 4.12: Zeitreihen des Schadstoffes CO

Die Zeitreihen zeigen einen eher kontinuierlichen Verlauf bezüglich der Saison. Es kann ein Abfall der Belastung Richtung Sommer und ein Anstieg Richtung Winter beobachtet werden. Jedoch zeigen die Tagesmittelwerte nicht das Muster, das der Entwicklung bezüglich der Wochentage entspricht. Auffällig ist ein belastungsarmes Intervall zu Beginn des Jahres 2005. Allgemein war die Belastung im Jahr 2006 etwas stärker als im Jahr 2005. Weiters ist hier, wie auch beim Schadstoff Schwefeldioxid, ein kleiner Sprung ungefähr in der Mitte des Jahres 2006 erkennbar. Ab diesem Sprung sind die Werte nach unten größer als die Aufzeichnungen zuvor.



Als Nächstes folgt die Analyse der Zerlegung der Zeitreihe in 3 Komponenten, um etwaige wiederkehrende Muster oder Trends aufzudecken und ein noch detaillierteres Bild über die Zeitreihe zu bekommen.

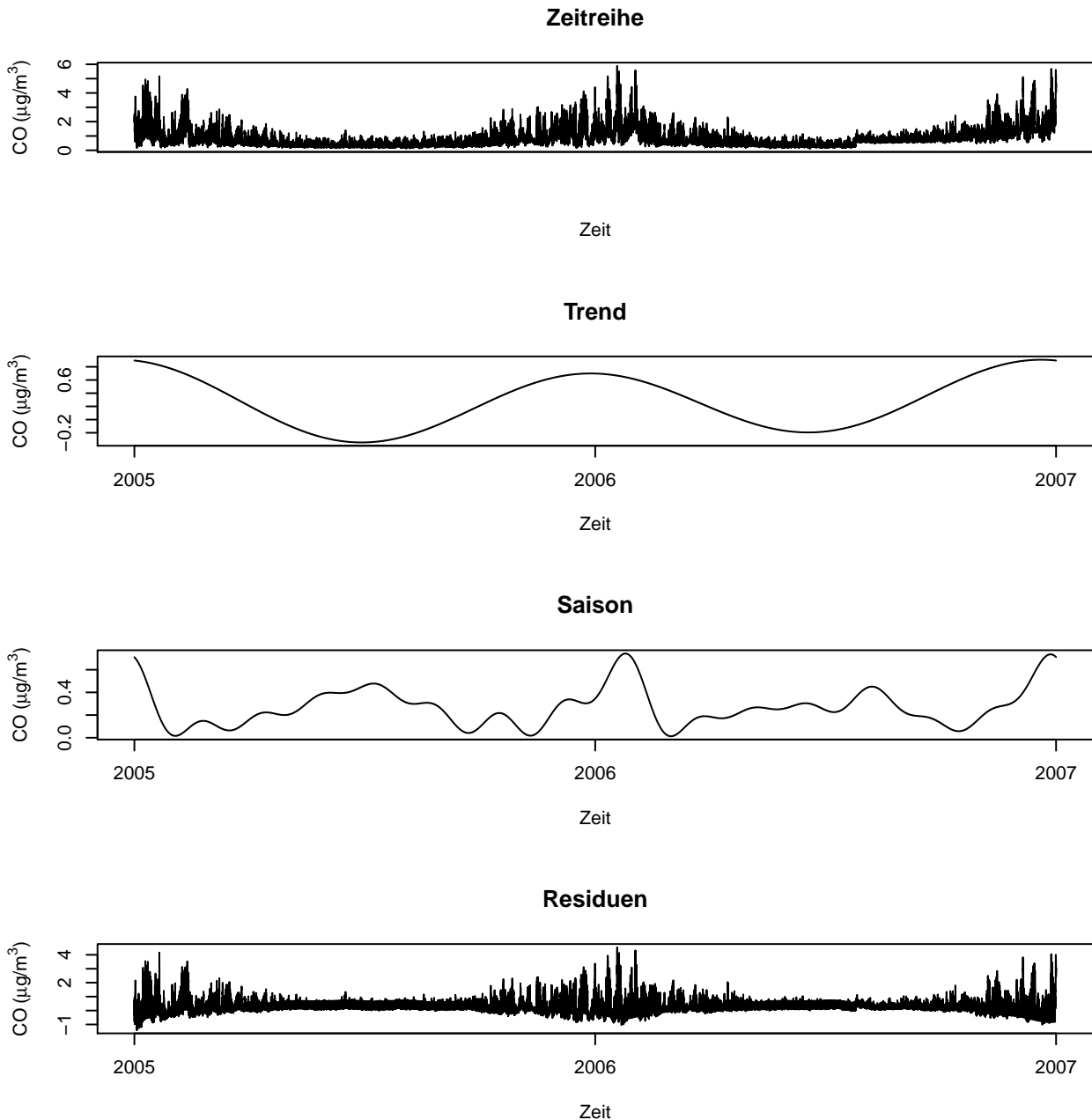


Abbildung 4.13: Zerlegung der Zeitreihe zu den CO-Halbstundenwerten

Die Aufteilung zeigt wiederum die erkannte Entwicklung in der Saison. Das bedeutet, dass es eine hohe Schadstoffbelastung in den Wintermonaten gab und eine niedrigere in den Sommermonaten. Dies zeigt hier der Trend, aber auch der saisonale Anteil der Zeitreihe zeigt höhere Werte um die Jahreswechsel herum an. Jedoch steckt aber auch bei der Zerlegung der Zeitreihe zu den Kohlenmonoxidzeichnungen der größte Anteil in den Residuen. Die Schadstoffbelastung nimmt innerhalb des Aufzeichnungszeitraums ein Maximum von ungefähr  $6 \mu\text{g}/\text{m}^3$  an.

Die Größenspanne der Residuen ist hier bis zu  $5 \mu\text{g}/\text{m}^3$  groß. Selbst wenn diese teilweise in den negativen Bereich fallen, um die Zerlegung deutlicher zu machen, ist dieser Anteil noch immer signifikant der größte. Dies bedeutet, dass auf Grund der hohen Unterschiede der Werte, besonders in den Wintermonaten an den einzelnen Tagen, nur ein sehr kleiner Anteil mittels saisonalen bzw. allgemeinen Trend erklärt werden kann.

Nach der Zerlegung der Zeitreihe kommt man nun ein weiteres Mal zur Untersuchung der Autokorrelation. Kommende 3 Abbildungen zeigen, ob und wie stark die Funktion der Kohlenmonoxidbelastung mit sich selbst korreliert.

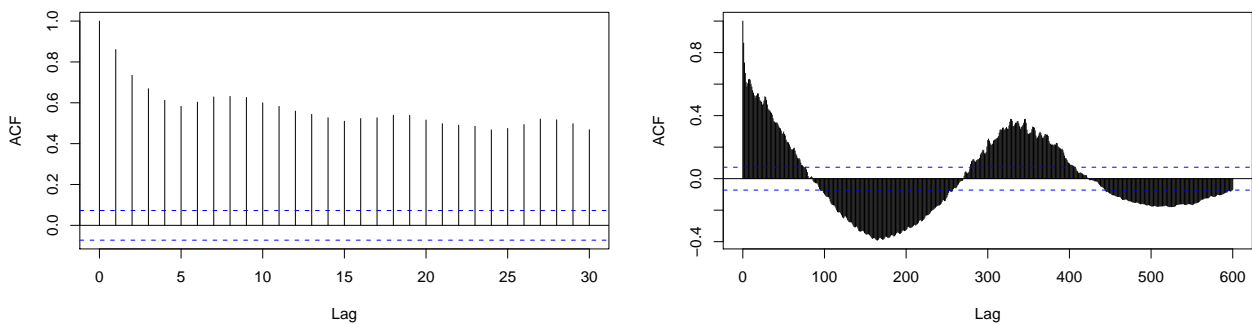


Abbildung 4.14: Autokorrelation der CO-Tagesmittelwerte

Bei dieser Autokorrelationsfunktion sind keine Analysen wie die bezüglich der Woche möglich. Beim lag 7 zeigt sich zwar ein leichter Anstieg, jedoch ist dieser gegenüber der benachbarten lags nicht signifikant. Allerdings ist hier ein saisonales Verhalten wieder aufzeigbar. Man erkennt klar, die negative Korrelation zwischen Werten die ungefähr ein halbes Jahr auseinander liegen. Dies entspricht in der Abbildung der Autokorrelation zu den Tagesmittelwerten ungefähr dem lag 163 und zeigt somit den Unterschied der Daten in der Saison auf. Auch die Autokorrelation bezüglich des Jahres, entspricht circa lag 365, ist erkennbar als deutlich positive Korrelation und zeigt somit die Ähnlichkeit der um ein Jahr verschobenen Daten.

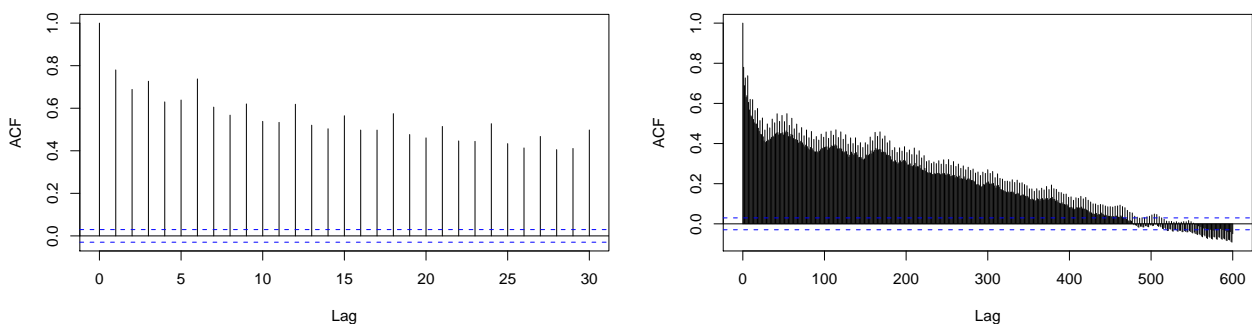


Abbildung 4.15: Autokorrelation der CO-4-Stunden-Mittelwerte

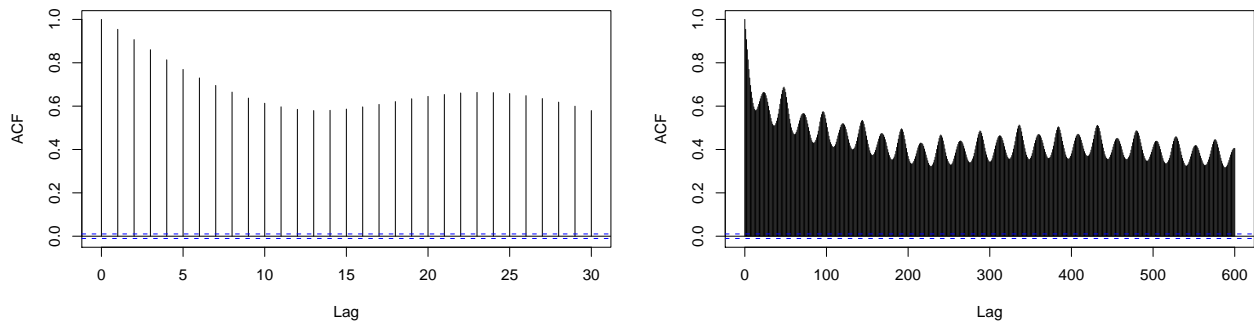


Abbildung 4.16: Autokorrelation der CO-Halbstundenwerte

Auch auf die Autokorrelationsfunktionen der Halbstundenwerte bzw. 4-Stunden-Mittelwerte zum Schadstoff Kohlenmonoxid können die Ergebnisse der Analyse der Autokorrelationsfunktion der Tagesmittelwerte der Kohlenmonoxidbelastung übertragen werden. So entspricht zum Beispiel der Wert der Autokorrelationsfunktion der Tagesmittelwerte bei lag 10, der ungefähr bei 0.6 liegt, dem Wert der Autokorrelationsfunktion der Halbstundenwerte bei lag 60, der eben auch ungefähr bei 0.6 liegt, wegen der Mittelung jedoch etwas kleiner ist. Bei den 4-Stunden-Mittelwerten würde dies dem lag 480 entsprechen. Man sieht, dass auch hier der Wert der Autokorrelation leicht unter 0.6 liegt.

## 4.4 Kreuzkorrelation

So wie man bei der explorativen Analyse den Zusammenhang zweier Verteilungen mittels Scatterplot aufgedeckt hat, kann man bei der Zeitreihenanalyse mittels der Kreuzkorrelationsfunktion den Zusammenhang zweier Zeitreihen deutlich machen. Nur hat man hier nicht einen Blick auf die Werte an sich, sondern auf die Größe, die angibt, wie stark die beiden Zeitreihen miteinander korrelieren, am selben Zeitpunkt und wenn die beiden Zeitpunkte um ein gewisses Intervall verschoben sind. Dies Verzögerung entspricht dem lag. Im Folgenden wurde diese Kreuzkorrelationsfunktion für sämtliche möglichen Konstellationen der Schadstoffe dargestellt.

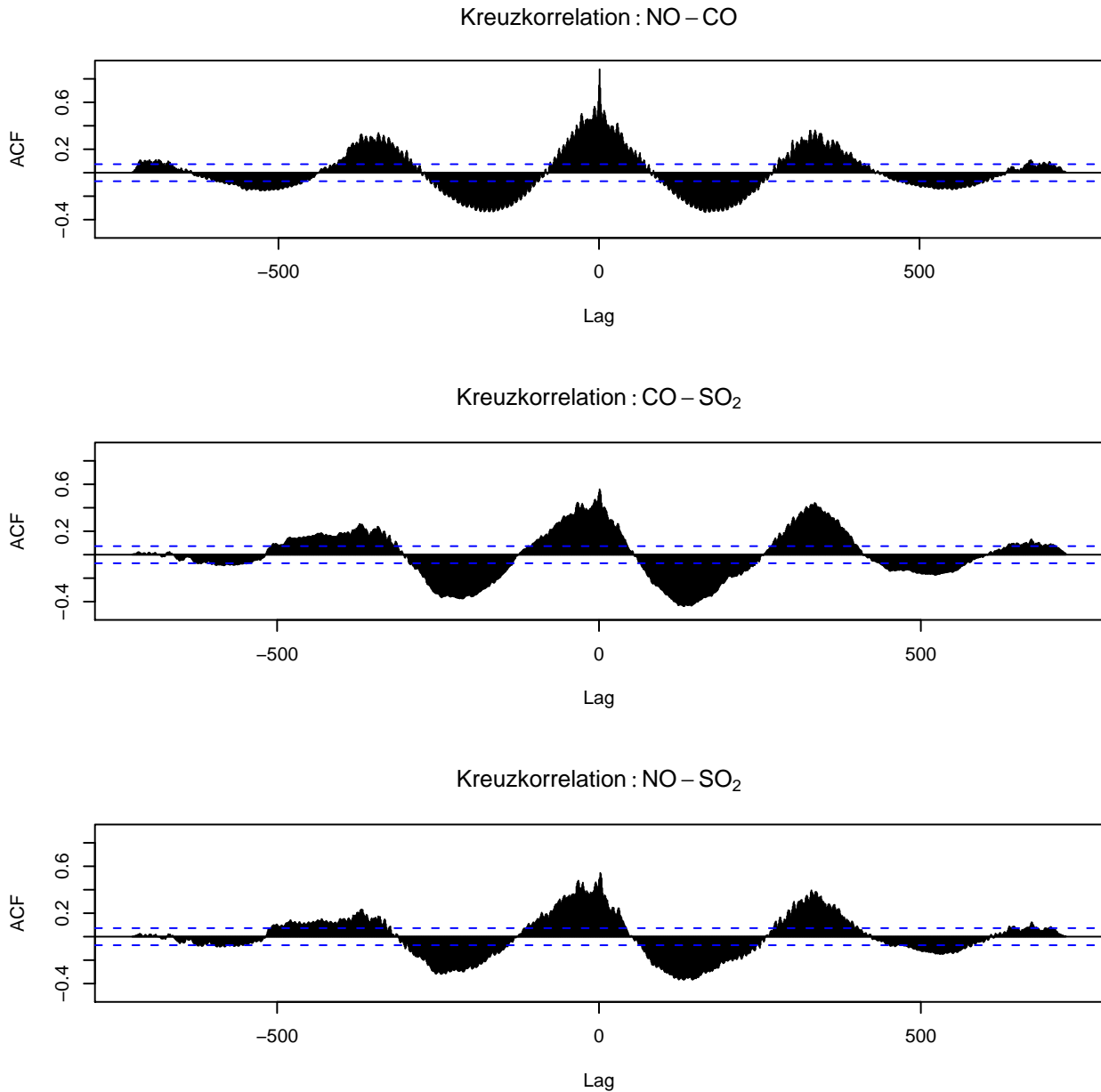


Abbildung 4.17: Kreuzkorrelation der Tagesmittelwerte

Am stärksten korrelieren die Zeitreihen zu den Schadstoffen Kohlenmonoxid und Stickstoffmonoxid. Hier ist u.a. der Koeffizient bei lag 0 am größten. Das bedeutet, dass die Halbstundenwerten zu den beiden Schadstoffen am selben Zeitpunkt sehr ähnlich sind. Betrachtet man die um ein Zeitintervall verschobenen Beobachtungen von Kohlenmonoxid und Stickstoffmonoxid, so nimmt die Korrelation mit der Größe des Intervalls, das dem lag entspricht, bis zum lag 160 stark ab bzw. fällt sogar ins Negative. Auch hier spiegelt der negative Korrelationskoeffizient beim lag 160 den großen Unterschied der Sommer- und Winterwerte wider. Ein anderes Bild zeigen die Kreuzkorrelationsfunktionen mit dem Schadstoff Schwefeldioxid. Die stärkste Korrelation zeigt sich zwar wiederum beim lag 0, jedoch ist die Abnahme bezüglich größerer oder kleinerer lags eine andere. Sie verläuft nämlich hier nicht, wie bei Kohlenmonoxid und Stickstoffmonoxid, symmetrisch in beide Richtungen. Betrachtet man den lag 100, sowohl positiv als auch negativ, so sieht man: Der Kohlenmonoxid-/Stickstoffmonoxidtagesmittelwert und der um 100 Tage zuvor aufgezeichnete Schwefeldioxidtagesmittelwerten korrelieren positiv miteinander. Ist jedoch der Schwefeldioxidbelastungswert 100 Tage nach dem Kohlenmonoxid-/Stickstoffmonoxidtagesmittelwert aufgezeichnet, korreliert er negativ damit. Der Korrelationskoeffizient liegt nämlich bereits im Negativen.

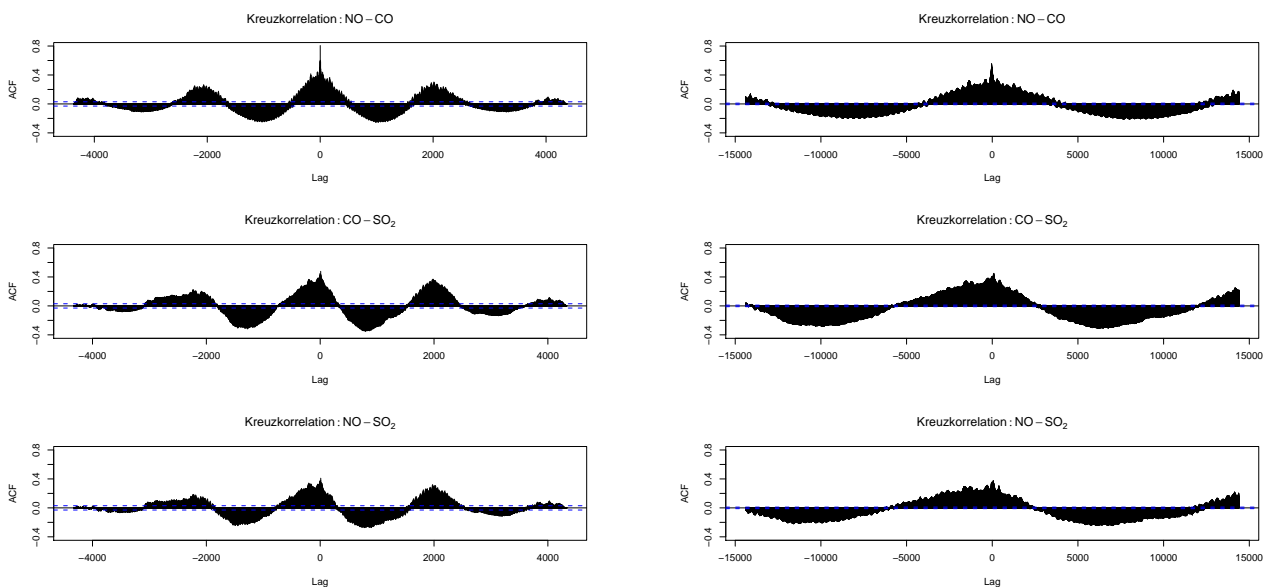


Abbildung 4.18: Kreuzkorrelation der 4-Stunden-Mittelwerte (links) und Halbstundenwerte (rechts)

Auch die Kreuzkorrelationsfunktionen zu den 4-Stunden-Mittelwerten und Halbstundenwerten zeigen das Bild und somit die Ergebnisse von zuvor. Die Verschiebung des lags kann hier wie schon zuvor bei den Autokorrelationsfunktionen interpretiert werden. Die Verzögerung wird also jeweils für die 4-Stunden-Mittelwerte um das 6-fache größer und bei den Halbstundenwerten um das 48-fache.

## 5 Conclusio

Die explorative Analyse und die Zeitreihenanalyse führen bei allen 3 Schadstoffen zu ähnlichen Ergebnissen. Es gibt kaum Unterschiede in der Belastung in den Jahren. Wesentliche Unterschiede findet man bei der Betrachtung bezüglich Saison und Tagestyp. Es gibt nämlich höhere Schadstoffbelastungen in den Wintermonaten und an den Werktagen. Bezüglich der Uhrzeit unterscheiden sich die Belastungsmaxima der Schadstoffe.

Die Verteilungen der 3 untersuchten Schadstoffe zeigen eine große Spannweite verglichen mit dem interquartilen Bereich. Dies zeigt, dass es eine große Streubreite und somit eine hohe Anzahl von Ausreißern gibt. Am deutlichsten wird diese Relation beim Schadstoff Stickstoffmonoxid (NO). 75 % der aufgezeichneten Daten sind hier unter einem Wert von  $90.60 \mu\text{g}/\text{m}^3$ , das Maximum der Daten liegt jedoch bei  $808.30 \mu\text{g}/\text{m}^3$ .

Aufgrund der ähnlichen Belastungsverläufe weist die Kreuzkorrelation hohe Abhängigkeiten auf. Den stärksten Zusammenhang zeigen die Schadstoffe Stickstoffmonoxid (NO) und Kohlenmonoxid (CO) mit einem Korrelationskoeffizienten von Pearson von über 0.8. Deren zeitliche Entwicklung ist also sehr ähnlich.

Als ein besonderer Schadstoff erweist sich Schwefeldioxid. Hier zeigen sowohl die Boxplotserien bezüglich des Monats als auch der Zeitreihenverlauf ein äußerst interessantes Bild. Es wurde nämlich ein auffälliger Abfall der Schwefeldioxidbelastung in der Mitte des Junis 2006 aufgedeckt.

# Literatur

- [1] Peng R.D. und Domenicini F.:  
*Statistical Methods for Environmental Epidemiology with R, A Case Study in Air Pollution and Health*, Springer, New York
- [2] Stadlober E. und Schauer J.:  
*Statistik - Skript*, Institut für Statistik, Technische Universität Graz,  
3. Auflage, September 2009, 226 Seiten
- [3] Homepage des Umweltbundesamts Österreich:  
<http://www.umweltbundesamt.at>
- [4] Homepage des Land Steiermark - Umweltinformation Steiermark:  
<http://www.umwelt.steiermark.at>
- [5] Homepage des Land Steiermark - Geographisches Informationssystem:  
<http://www.gis.steiermark.at>

# Abbildungsverzeichnis

1.1	Lage der Luftgütemessstation Graz-Don Bosco . . . . .	5
1.2	Lage der Luftgütemessstation Graz-Don Bosco . . . . .	5
2.1	Boxplot der NO-Halbstundenwerte . . . . .	8
2.2	Boxplot der NO-4-Stunden-Mittelwerte . . . . .	9
2.3	Boxplot der NO-Tagesmittelwerte . . . . .	9
2.4	Boxplotserie der NO-Halbstundenwerte bzgl. Jahr . . . . .	10
2.5	Boxplotserie der NO-4-Stunden-Mittelwerte bzgl. Jahr . . . . .	11
2.6	Boxplotserie der NO-Tagesmittelwerte bzgl. Jahr . . . . .	11
2.7	Boxplotserie der NO-Halbstundenwerte bzgl. Saison . . . . .	12
2.8	Boxplotserie der NO-4-Stunden-Mittelwerte bzgl. Saison . . . . .	12
2.9	Boxplotserie der NO-Tagesmittelwerte bzgl. Saison . . . . .	13
2.10	Boxplotserie der NO-Halbstundenwerte bzgl. Monat . . . . .	14
2.11	Boxplotserie der NO-4-Stunden-Mittelwerte bzgl. Monat . . . . .	15
2.12	Boxplotserie der NO-Tagesmittelwerte bzgl. Monat . . . . .	16
2.13	Boxplotserie der NO-Halbstundenwerte bzgl. Tagestyp . . . . .	17
2.14	Boxplotserie der NO-4-Stunden-Mittelwerte bzgl. Tagestyp . . . . .	17
2.15	Boxplotserie der NO-Tagesmittelwerte bzgl. Tagestyp . . . . .	18
2.16	Boxplotserie der NO-Halbstundenwerte bzgl. Wochentag . . . . .	19
2.17	Boxplotserie der NO-4-Stunden-Mittelwerte bzgl. Wochentag . . . . .	19
2.18	Boxplotserie der NO-Tagesmittelwerte bzgl. Wochentag . . . . .	20
2.19	Boxplotserie der NO-Halbstundenwerte bzgl. Uhrzeit . . . . .	21
2.20	Boxplotserie der NO-Halbstundenwerte bzgl. Uhrzeit und Saison . . . . .	21
2.21	Boxplotserie der NO-Halbstundenwerte bzgl. Uhrzeit und Tagestyp . . . . .	22
2.22	Boxplot der SO <sub>2</sub> -Halbstundenwerte . . . . .	23
2.23	Boxplot der SO <sub>2</sub> -4-Stunden-Mittelwerte . . . . .	24
2.24	Boxplot der SO <sub>2</sub> -Tagesmittelwerte . . . . .	24
2.25	Boxplotserie der SO <sub>2</sub> -Halbstundenwerte bzgl. Jahr . . . . .	25
2.26	Boxplotserie der SO <sub>2</sub> -4-Stunden-Mittelwerte bzgl. Jahr . . . . .	25
2.27	Boxplotserie der SO <sub>2</sub> -Tagesmittelwerte bzgl. Jahr . . . . .	26
2.28	Boxplotserie der SO <sub>2</sub> -Halbstundenwerte bzgl. Saison . . . . .	27
2.29	Boxplotserie der SO <sub>2</sub> -4-Stunden-Mittelwerte bzgl. Saison . . . . .	27
2.30	Boxplotserie der SO <sub>2</sub> -Tagesmittelwerte bzgl. Saison . . . . .	28
2.31	Boxplotserie der SO <sub>2</sub> -Halbstundenwerte bzgl. Monat . . . . .	29
2.32	Boxplotserie der SO <sub>2</sub> -4-Stunden-Mittelwerte bzgl. Monat . . . . .	30
2.33	Boxplotserie der SO <sub>2</sub> -Tagesmittelwerte bzgl. Monat . . . . .	31
2.34	Boxplotserie der SO <sub>2</sub> -Halbstundenwerte bzgl. Tagestyp . . . . .	32
2.35	Boxplotserie der SO <sub>2</sub> -4-Stunden-Mittelwerte bzgl. Tagestyp . . . . .	32
2.36	Boxplotserie der SO <sub>2</sub> -Tagesmittelwerte bzgl. Tagestyp . . . . .	33
2.37	Boxplotserie der SO <sub>2</sub> -Halbstundenwerte bzgl. Wochentag . . . . .	34
2.38	Boxplotserie der SO <sub>2</sub> -4-Stunden-Mittelwerte bzgl. Wochentag . . . . .	34
2.39	Boxplotserie der SO <sub>2</sub> -Tagesmittelwerte bzgl. Wochentag . . . . .	35
2.40	Boxplotserie der SO <sub>2</sub> -Halbstundenwerte bzgl. Uhrzeit . . . . .	36
2.41	Boxplotserie der SO <sub>2</sub> -Halbstundenwerte bzgl. Uhrzeit und Saison . . . . .	36
2.42	Boxplotserie der SO <sub>2</sub> -Halbstundenwerte bzgl. Uhrzeit und Tag . . . . .	37
2.43	Boxplot der CO-Halbstundenwerte . . . . .	38
2.44	Boxplot der CO-4-Stunden-Mittelwerte . . . . .	39
2.45	Boxplot der CO-Tagesmittelwerte . . . . .	39



2.46	Boxplotserie der CO-Halbstundenwerte bzgl. Jahr . . . . .	40
2.47	Boxplotserie der CO-4-Stunden-Mittelwerte bzgl. Jahr . . . . .	40
2.48	Boxplotserie der CO-Tagesmittelwerte bzgl. Jahr . . . . .	41
2.49	Boxplotserie der CO-Halbstundenwerte bzgl. Saison . . . . .	42
2.50	Boxplotserie der CO-4-Stunden-Mittelwerte bzgl. Saison . . . . .	42
2.51	Boxplotserie der CO-Tagesmittelwerte bzgl. Saison . . . . .	43
2.52	Boxplotserie der CO-Halbstundenwerte bzgl. Monat . . . . .	44
2.53	Boxplotserie der CO-4-Stunden-Mittelwerte bzgl. Monat . . . . .	45
2.54	Boxplotserie der CO-Tagesmittelwerte bzgl. Monat . . . . .	46
2.55	Boxplotserie der CO-Halbstundenwerte bzgl. Tagestyp . . . . .	47
2.56	Boxplotserie der CO-4-Stunden-Mittelwerte bzgl. Tagestyp . . . . .	47
2.57	Boxplotserie der CO-Tagesmittelwerte bzgl. Tagestyp . . . . .	48
2.58	Boxplotserie der CO-Halbstundenwerte bzgl. Wochentag . . . . .	49
2.59	Boxplotserie der CO-4-Stunden-Mittelwerte bzgl. Wochentag . . . . .	49
2.60	Boxplotserie der CO-Tagesmittelwerte bzgl. Wochentag . . . . .	50
2.61	Boxplotserie der CO-Halbstundenwerte bzgl. Uhrzeit . . . . .	51
2.62	Boxplotserie der CO-Halbstundenwerte bzgl. Uhrzeit und Saison . . . . .	52
2.63	Boxplotserie der CO-Halbstundenwerte bzgl. Uhrzeit und Tag . . . . .	53
3.1	Scatterplotmatrix mit Kreuzkorrelationskoeffizienten der Halbstundenwerte . . .	54
3.2	Scatterplotmatrix mit Kreuzkorrelationskoeffizienten der 4-Stunden-Mittelwerte	55
3.3	Scatterplotmatrix mit Kreuzkorrelationskoeffizienten der Tagesmittelwerte . . .	56
3.4	Scatterplotmatrix der Halbstundenwerte bzgl. Tagestyp . . . . .	57
3.5	Scatterplotmatrix der 4-Stunden-Mittelwerte bzgl. Tagestyp . . . . .	58
3.6	Scatterplotmatrix der Tagesmittelwerte bzgl. Tagestyp . . . . .	59
3.7	Scatterplotmatrix der Halbstundenwerte bzgl. Saison . . . . .	60
3.8	Scatterplotmatrix der 4-Stunden-Mittelwerte bzgl. Saison . . . . .	61
3.9	Scatterplotmatrix der Tagesmittelwerte bzgl. Saison . . . . .	62
3.10	Scatterplots: NO und CO . . . . .	63
4.1	Zeitreihen des Schadstoffes NO . . . . .	64
4.2	Zerlegung der Zeitreihe zu den NO-Halbstundenwerten . . . . .	65
4.3	Autokorrelation der NO-Tagesmittelwerte . . . . .	66
4.4	Autokorrelation der NO-4-Stunden-Mittelwerte . . . . .	66
4.5	Autokorrelation der NO-Halbstundenwerte . . . . .	67
4.6	Zeitreihen des Schadstoffes SO <sub>2</sub> . . . . .	68
4.7	Auszug aus der Zeitreihe des Schadstoffes SO <sub>2</sub> (Halbstundenwerte) . . . . .	69
4.8	Zerlegung der Zeitreihe zu den SO <sub>2</sub> -Halbstundenwerten . . . . .	70
4.9	Autokorrelation der SO <sub>2</sub> -Tagesmittelwerte . . . . .	71
4.10	Autokorrelation der SO <sub>2</sub> -4-Stunden-Mittelwerte . . . . .	71
4.11	Autokorrelation der SO <sub>2</sub> -Halbstundenwerte . . . . .	71
4.12	Zeitreihen des Schadstoffes CO . . . . .	73
4.13	Zerlegung der Zeitreihe zu den CO-Halbstundenwerten . . . . .	74
4.14	Autokorrelation der CO-Tagesmittelwerte . . . . .	75
4.15	Autokorrelation der CO-4-Stunden-Mittelwerte . . . . .	75
4.16	Autokorrelation der CO-Halbstundenwerte . . . . .	76
4.17	Kreuzkorrelation der Tagesmittelwerte . . . . .	77
4.18	Kreuzkorrelation der 4-Stunden-Mittelwerte (links) und Halbstundenwerte (rechts)	78

# Tabellenverzeichnis

2.1	Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) . . . . .	8
2.2	Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Jahr . . . . .	10
2.3	Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Saison . . . . .	12
2.4	Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Monat . . . . .	13
2.5	Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Tagestyp . . . . .	16
2.6	Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Wochentag . . . . .	18
2.7	Statistische Kenngrößen des Schadstoffes NO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Uhrzeit . . . . .	20
2.8	Statistische Kenngrößen des Schadstoffes SO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ ) . . . . .	23
2.9	Statistische Kenngrößen des Schadstoffes SO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ ) bzgl. Jahr . . . . .	25
2.10	Statistische Kenngrößen des Schadstoffes SO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ ) bzgl. Saison . . . . .	26
2.11	Statistische Kenngrößen des Schadstoffes SO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ ) bzgl. Monat . . . . .	28
2.12	Statistische Kenngrößen des Schadstoffes SO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ ) bzgl. Tagestyp . . . . .	31
2.13	Statistische Kenngrößen des Schadstoffes SO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ ) bzgl. Wochentag . . . . .	33
2.14	Statistische Kenngrößen des Schadstoffes SO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ ) bzgl. Uhrzeit . . . . .	35
2.15	Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) . . . . .	38
2.16	Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Jahr . . . . .	40
2.17	Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Saison . . . . .	41
2.18	Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Monat . . . . .	43
2.19	Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Tagestyp . . . . .	46
2.20	Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Wochentag . . . . .	48
2.21	Statistische Kenngrößen des Schadstoffes CO ( $\mu\text{g}/\text{m}^3$ ) bzgl. Uhrzeit . . . . .	50