

## Bakkalaureatsprojekt

# Zeitreihenanalyse von $PM_{10}$ und $NO_x$ für Graz-Nord, 2005/2006

verfasst im Rahmen der Lehrveranstaltung  
Bakkalaureatsprojekt TM (Statistik), PR 506.558

am

Institut für Statistik der TU Graz

unter Anleitung von

Univ.-Prof. DI Dr. Ernst Stadlober

durch

**Christof Peternell**

0530639

im Sommersemester 2010

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am .....

.....

# Danksagung

Hiermit möchte ich mich recht herzlich bei Herrn Univ.-Prof. DI Dr. Ernst Stadlober sowie Frau Brigitte Pfeiler bedanken, die mir diese Arbeit ermöglicht haben.

## Abstract

This paper deals with the analysis of air pollution of the northern part of Graz. The measurements include two pollutants,  $PM_{10}$  and  $NO_x$ , monitored over a time span of two years, from January 1<sup>st</sup> 2005 to December 31<sup>st</sup> 2006. The measurements were taken every 30 minutes. The goal is to show some characteristics of the data by the use of statistical parameters, boxplots and division of the data into the categorial attributes summer/winter, month and day of the week (weekday, saturday, sunday).

The analysis of the coherency of the time series from  $PM_{10}$  and  $NO_x$  is done by scatterplots. Again, the data is divided into categorial attributes. The data is analysed as four hour- and daily-means.

We used the models given in the sections 5.3 and 5.4 of the book „*Statistical Methods for Environmental Epidemiology with R*“, by R. D. Peng and F. Dominici: timescale decompositions and monitoring the autocorrelation functions of the time series  $PM_{10}$  and  $NO_x$  on several time scales. For this the statistical program R 2.11.0 was used.

One major result of the analysis is that the air pollution of  $PM_{10}$  and  $NO_x$  is much higher in the winter months, caused by lower temperatures and the bigger amount of fossil fuels used in that season.

## Zusammenfassung

In diesem Projekt wird eine Schadstoffanalyse für die Region Graz-Nord mithilfe von Zeitreihen des Zeitraums 01.01.2005 bis 31.12.2006 erstellt. Die betrachteten Schadstoffe sind  $PM_{10}$  und  $NO_x$ . Dabei werden sowohl Halbstunden-, als auch Vierstunden- und Tagesmittelwerte betrachtet. Als Charakteristika der Daten werden statistische Kenngrößen und Boxplotserien gruppiert nach kategorischen Merkmalen wie Saison (Sommer/Winter), Monat, Wochentag (Werktag/Samstag/Sonn- und Feiertag) dargestellt.

Die Zusammenhangsanalyse der  $PM_{10}$  - und  $NO_x$  - Zeitreihen erfolgt anhand von Scatterplots, erneut unter Einbeziehung kategorischer Merkmale. Dabei werden die im Buch „*Statistical Methods for Environmental Epidemiology with R*“ [1] von R. D. Peng und F. Dominici in den Abschnitten 5.3 und 5.4 beschriebenen Methoden herangezogen, um die Autokorrelation der Zeitreihen  $PM_{10}$  und  $NO_x$  auf verschiedenen Zeitskalen zu betrachten. Für die Arbeit wurde das Statistikprogramm R 2.11.0 verwendet.

Anhand der Analysen lässt sich zeigen, dass die Konzentrationen von  $PM_{10}$  und  $NO_x$  in den Wintermonaten höher ausfallen, als in den Sommermonaten. Dies lässt sich auf die geringeren Temperaturen und den höheren Verbrauch von fossilen Brennstoffen in den Wintermonaten zurückführen. Es lässt sich weiters eine positive Korrelation zwischen  $PM_{10}$  und  $NO_x$  feststellen.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Exploration der Zeitreihen für <math>PM_{10}</math> und <math>NO_x</math></b>	<b>3</b>
2.1	Lagemaße und Streuungsmaße . . . . .	6
2.2	Schiefemaße und Maße für die Wölbung . . . . .	8
2.3	Zusammenhangsanalyse . . . . .	41
2.4	Untersuchung der Auto- und Kreuzkorrelationsstruktur . . . . .	50
<b>3</b>	<b>Betrachtung der Zeitreihen <math>PM_{10}</math> und <math>NO_x</math> auf verschiedenen Zeitskalen</b>	<b>56</b>
3.1	Zerlegung der Zeitskala mit der Funktion <i>tsdecomp</i> . . . . .	64
3.2	Untersuchung von Störgrößen . . . . .	67
<b>4</b>	<b>Conclusio</b>	<b>70</b>
	<b>Quellenverzeichnis</b>	<b>71</b>

## Abbildungsverzeichnis

1	Darstellung der Zeitreihen $PM_{10}$ bzgl. 1/2h-Werte, 4h- und Tagesmittelwerte . . .	4
2	Darstellung der Zeitreihen $NO_x$ bzgl. 1/2h-Werte, 4h- und Tagesmittelwerte . . .	5
3	Dichteschätzer (Normalverteilung) für $NO_x$ und $PM_{10}$ anhand der Tagesmittelwerte	10
4	Dichteschätzer (Normalverteilung) für $PM_{10}$ bzgl. Sommer/Winter, Tagesmittelwerte . . . . .	12
5	Dichteschätzer (Normalverteilung) für $NO_x$ bzgl. Sommer/Winter, Tagesmittelwerte . . . . .	13
6	Boxplots von $PM_{10}$ und $NO_x$ bzgl. Sommer/Winter, Halbstundenwerte . . . . .	14
7	Boxplots von $PM_{10}$ und $NO_x$ bzgl. Sommer/Winter, Vierstundenwerte . . . . .	15
8	Darstellung der Boxplots $PM_{10}$ und $NO_x$ bzgl. Sommer/Winter, Tagesmittelwerte	16
9	Darstellung der Boxplots $PM_{10}$ und $NO_x$ bzgl. Monat, Halbstundenwerte . . . . .	18
10	Detailansicht Boxplots $PM_{10}$ und $NO_x$ bzgl. Monat, Halbstundenwerte . . . . .	19
11	Darstellung der Boxplots $PM_{10}$ und $NO_x$ bzgl. Monat, Vierstundenmittelwerte .	20
12	Detailansicht Boxplots $PM_{10}$ und $NO_x$ bzgl. Monat, Vierstundenmittelwerte . .	21
13	Darstellung der Boxplots $PM_{10}$ und $NO_x$ bzgl. Monat, Tagesmittelwerte . . . . .	22
14	Detailansicht Boxplots $PM_{10}$ und $NO_x$ bzgl. Monat, Tagesmittelwerte . . . . .	23
15	Darstellung des Dichteschätzers von $PM_{10}$ bzgl. Wochentage, Tagesmittelwerte .	25
16	Dichteschätzer (Normalverteilung) von $PM_{10}$ bzgl. Wochentage, Tagesmittelwerte	27
17	Darstellung der Boxplots $PM_{10}$ und $NO_x$ bzgl. Sommer, Halbstundenwerte . . .	28
18	Detailansicht der Boxplots $PM_{10}$ und $NO_x$ bzgl. Sommer, Halbstundenwerte . .	29
19	Darstellung der Boxplots $PM_{10}$ und $NO_x$ bzgl. Sommer, Vierstundenmittelwerte	30
20	Detailansicht der Boxplots $PM_{10}$ und $NO_x$ bzgl. Sommer, Vierstundenmittelwerte	31
21	Detailansicht der Boxplots $PM_{10}$ und $NO_x$ bzgl. Sommer, Tagesmittelwerte . . .	32
22	Dichteschätzer (Normalverteilung) von $PM_{10}$ bzgl. Wochentage, Tagesmittelwerte	34
23	Dichteschätzers (Normalverteilung) von $NO_x$ bzgl. Wochentage, Tagesmittelwerte	35
24	Darstellung der Boxplots $PM_{10}$ und $NO_x$ bzgl. Winter, Halbstundenwerte . . . . .	36
25	Detailansicht der Boxplots $PM_{10}$ und $NO_x$ bzgl. Winter, Halbstundenwerte . . .	37
26	Darstellung der Boxplots $PM_{10}$ und $NO_x$ bzgl. Winter, Vierstundenmittelwerte .	38
27	Detailansicht der Boxplots $PM_{10}$ und $NO_x$ bzgl. Winter, Vierstundenmittelwerte	39
28	Detailansicht der Boxplots $PM_{10}$ und $NO_x$ bzgl. Winter, Tagesmittelwerte . . . .	40
29	Darstellung der Scatterplots von $PM_{10}$ und $NO_x$ bzgl. 1/2h-Werte, 4h- und Tagesmittelwerte . . . . .	41
30	Darstellung der Scatterplots von $PM_{10}$ gegen $NO_x$ bezüglich Halbstundenwerten	42
31	Darstellung der Scatterplots von $PM_{10}$ gegen $NO_x$ bezüglich Vierstundenmittelwerten . . . . .	43
32	Darstellung der Scatterplots von $PM_{10}$ gegen $NO_x$ bezüglich Tagesmittelwerten .	43
33	Darstellung der Scatterplots von $PM_{10}$ und $NO_x$ mit Halbstundenwerten, Vierstunden- und Tagesmittelwerte bzgl. Wochentage im Sommer . . . . .	47
34	Scatterplots von $PM_{10}$ und $NO_x$ mit Halbstundenwerten, Vierstunden- und Tagesmittelwerte bzgl. Wochentage im Winter . . . . .	49
35	Darstellung der Autokorrelationsfunktionen von $NO_x$ bzgl. 1/2h-Werte, Vierstunden- und Tagesmittelwerte . . . . .	51
36	Autokorrelationsfunktionen von $PM_{10}$ bzgl. 1/2h-Werte, Vierstunden- und Tagesmittelwerte . . . . .	53

37	Darstellung der Kreuzkorrelationsfunktionen von $PM_{10}$ und $NO_x$ bzgl. 1/2h-Werte, Vierstunden- und Tagesmittelwerte . . . . .	55
38	Autokorrelationsfunktionen von $PM_{10}$ und $NO_x$ bzgl. jährlicher Zeitskala . . . . .	57
39	Autokorrelationsfunktionen von $PM_{10}$ und $NO_x$ bzgl. quartaler Zeitskala . . . . .	58
40	Autokorrelationsfunktionen von $PM_{10}$ und $NO_x$ bzgl. monatlicher Zeitskala . . . . .	59
41	Autokorrelationsfunktionen von $PM_{10}$ und $NO_x$ bzgl. wöchentlicher Zeitskala . . . . .	60
42	Autokorrelationsfunktionen von $PM_{10}$ und $NO_x$ bzgl. täglicher Zeitskala . . . . .	61
43	Zerlegung der Zeitreihen $PM_{10}$ und $NO_x$ in jährlichen, saisonalen und wöchentlichen Trend . . . . .	63
44	Zerlegung der Zeitreihen $PM_{10}$ und $NO_x$ mit der Funktion <i>tsdecomp</i> in langfristigen, saisonalen und kurzfristigen Trend . . . . .	64
45	Genauere Zerlegung der Zeitreihen $PM_{10}$ mit der Funktion <i>tsdecomp</i> . . . . .	65
46	Genauere Zerlegung der Zeitreihen $NO_x$ mit der Funktion <i>tsdecomp</i> . . . . .	66
47	Darstellung der Zeitreihe der Lufttemperatur bzgl. Halbstundenwerte . . . . .	67
48	Darstellung der Trends von $PM_{10}$ nach Entfernung des Einflusses der Lufttemperatur . . . . .	68
49	Darstellung der Trends von $NO_x$ nach Entfernung des Einflusses der Lufttemperatur . . . . .	68

# 1 Einleitung

## Messstandort Graz-Nord

Der Messstandort Graz-Nord liegt in der Josef Ornig-Straße 16, 8051 Graz Gösting. Der Betreiber ist die steirische Landesregierung. Laut Umweltbundesamt [4] liegt die Messstation auf 353m Seehöhe, in einem halboffenen Becken am Gebirgsrand und in der näheren Umgebung befinden sich Ackerland, Wohngebiete, sowie Industrie- und Gewerbegebiete. Straßenverkehrsemissionen üben hier nur wenig Einfluss aus, da die Station am Stadtrand liegt und sich in der unmittelbaren Umgebung nur Grünflächen, lockere Bebauung, sowie wenig befahrene Innerorts - Straßen befinden.





## Beschreibung der Schadstoffe

### Feinstaub PM<sub>10</sub>

PM<sub>10</sub> ist die Abkürzung für **P**articulate **M**atter **10**. Damit bezeichnet man jenen Feinstaub, der einen Durchmesser kleiner als 10  $\mu\text{m}$  (Mikrometer) besitzt. Durch seinen enorm kleinen Durchmesser (zum Vergleich: der Durchmesser eines Haares ist zehn mal größer) gelangt PM<sub>10</sub> über Nase und Mund in die Lunge, was zu gesundheitlichen Folgen führen kann: Laut Bayerischem Landesamt für Gesundheit und Lebensmittelsicherheit [5] erhöht sich bei einem Anstieg der PM<sub>10</sub> - Belastung um 10  $\mu\text{g}/\text{m}^3$  das Risiko einer Bronchitis bei Kindern um 30%, das allgemeine Sterblichkeitsrisiko bei Erwachsenen um 10%.

Laut Umweltbundesamt Österreich [4] durfte bis 2004 der Tagesmittelwert von 50 $\mu\text{g}/\text{m}^3$  nur 35 mal pro Kalenderjahr überschritten werden, von 2005 bis 2009 nur noch 30 mal. Ab 2010 dürfen maximal 25 Überschreitungen pro Jahr gemessen werden.

Der Standort Graz-Nord hatte laut der Behörde im Zeitraum vom 01.01.2010 bis 24.09.2010, das heißt innerhalb von 267 Tagen genau 21 Überschreitungen zu verzeichnen.

In dieser Arbeit wird gezeigt, dass es im Zweijahres - Zeitraum vom 01.01.2005 bis 31.12.2006 insgesamt 124 Überschreitungen des kritischen PM<sub>10</sub> - Tagesmittelwertes gab, davon 58 mal im Jahr 2005 und 66 mal im Jahr 2006. In beiden Jahren wurde also die Anzahl der erlaubten 30 Überschreitungen beträchtlich überschritten.

### Stickoxide NO<sub>x</sub>

Die wichtigsten **Stickoxide** sind *Stickstoffmonoxid NO* und *Stickstoffdioxid NO<sub>2</sub>*. Beide zusammen werden als *Stickoxide NO<sub>x</sub>* bezeichnet. Stickoxide entstehen überwiegend als Nebenprodukte bei der Verbrennung von Brenn- und Treibstoffen. Hauptverursacher ist hierbei der Verkehr. Laut Umweltbundesamt [4] entsteht „in der kalten Jahreszeit aus gasförmigen Stickoxiden und Ammoniak partikelförmiges Ammoniumnitrat, welches zu einer großräumigen Belastung durch Feinstaub beiträgt. Im Sommer führen Stickstoffoxide zusammen mit Kohlenwasserstoffen zur Bildung von Ozon.“

Gesundheitliche Folgen durch Stickoxid-Belastung sind zum Beispiel Beeinträchtigung der Lungenfunktion und Verstärkung von Allergien. Weiters nehmen bei länger andauernder hoher Konzentration Herzrhythmusstörungen und Herzinfarkte zu.

Die Richtlinien des Umweltbundesamtes für NO<sub>x</sub> besagen, dass der kritische Wert für den Jahresmittelwert von 30  $\mu\text{g}/\text{m}^3$  nicht überschritten werden darf. In dieser Arbeit wird gezeigt, dass dieser in den Jahren 2005 und 2006 jeweils eingehalten wurde.

## 2 Exploration der Zeitreihen für $PM_{10}$ und $NO_x$

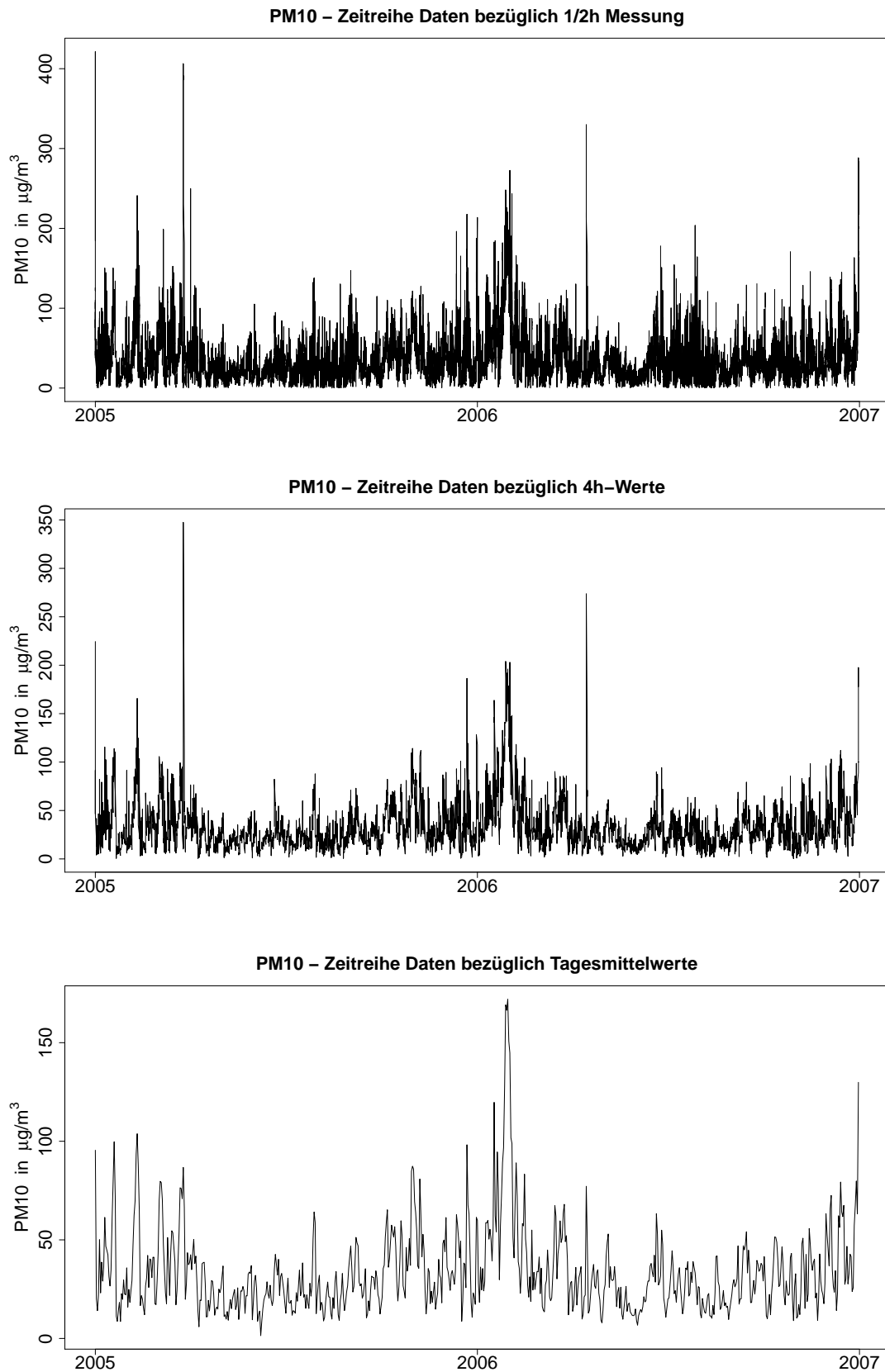
Die zugrunde liegende Messreihe besteht aus 35.040 Datenpunkten, gemessen im Halbstundentakt im Zweijahres - Zeitraum 01.01.2005 bis 31.12.2006. Dies entspricht acht Messungen alle vier Stunden, beziehungsweise 48 Messungen pro Tag.

Weiters werden auch Vierstunden- und Tagesmittelwerte (= 24-Stunden-Mittelwerte) betrachtet. Dann reduziert sich die Anzahl der Messungen auf 4.380 beziehungsweise 730 Werte.

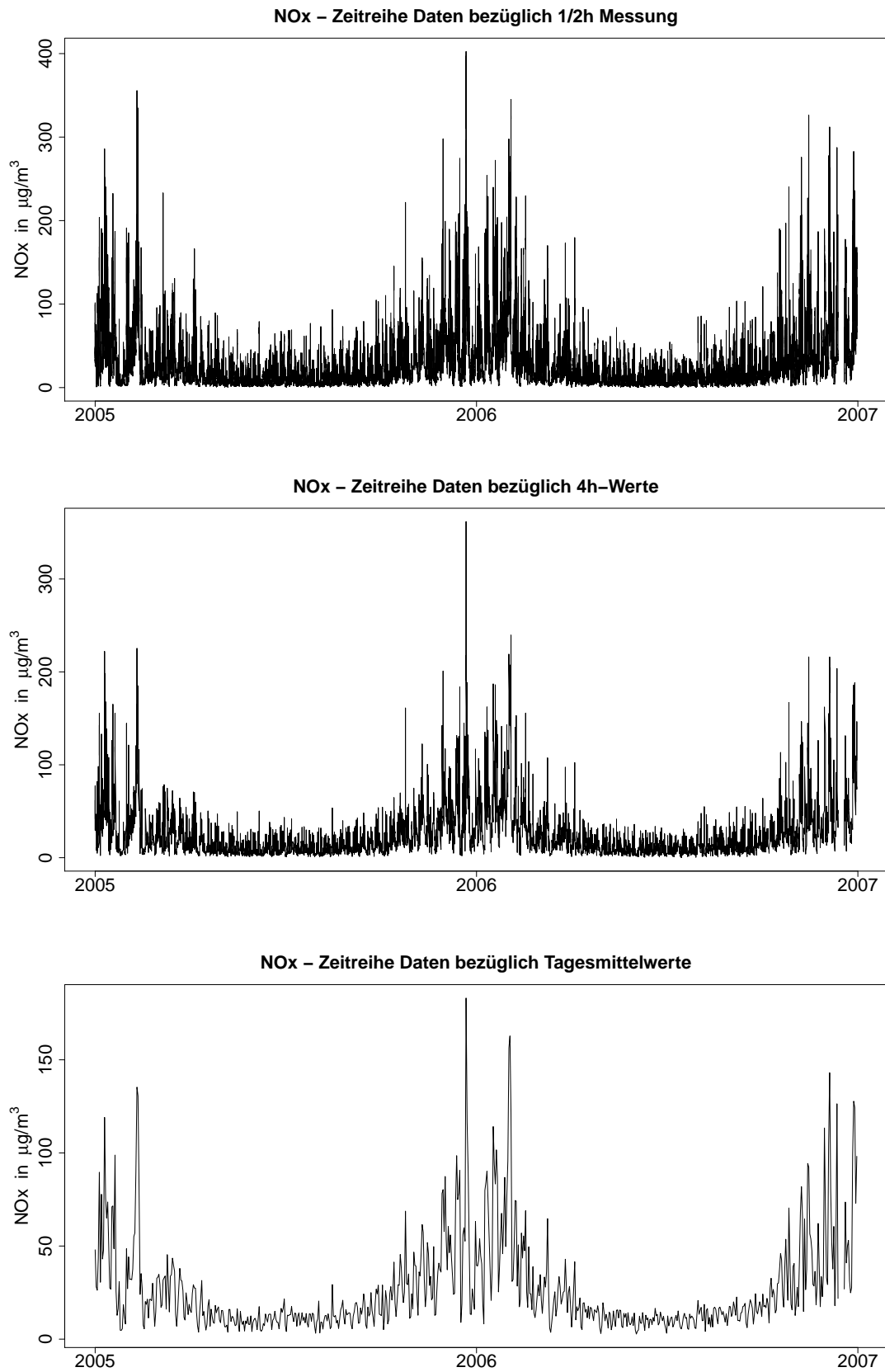
In Abbildung 1 erkennt man deutlich höhere  $PM_{10}$  - Werte im Bereich der Jahreswechsel. Zusätzlich kann man feststellen, dass, abgesehen von ein paar Ausreißern,  $PM_{10}$  seine Spitzenbelastungen um den Jahreswechsel 2005/2006 erreicht. Daraus lässt sich folgern, dass in den Wintermonaten die  $PM_{10}$  - Belastung höher ist, während sie in den Sommermonaten wieder sinkt. Auch für die Vierstunden- und Tagesmittelwerte kommt man auf die gleichen Ergebnisse. Durch die Mittelwertbildungen bei den Reihen der Vierstunden- und Tagesmittelwerte werden im Vergleich zu den Halbstundenwerten einige Spitzen herausgenommen und die Reihen erscheinen dadurch etwas glatter.

Analoge Erkenntnisse erhält man auch für die Zeitreihen von  $NO_x$  in Abbildung 2. Man sieht erneut, dass die Werte in den Wintermonaten tendenziell höher sind als in den Sommermonaten. Diese Tatsache tritt bei den Reihen von  $NO_x$  noch stärker auf, als bei jenen von  $PM_{10}$ , wie man in Abbildung 2 erkennen kann. Das führt dazu, dass man bei diesen Zeitreihen den sinusförmigen Trend noch viel deutlicher erkennen kann, als bei jenen der anderen Zielgröße.

Auch für die Reihen von  $NO_x$  kann man, abgesehen von ein paar Ausreißern, eine Spitzenbelastung um den Jahreswechsel 2005/2006 erkennen. Außerdem ist anhand der Abbildungen 1 und 2 gut ersichtlich, dass die Reihen von  $PM_{10}$  weniger hohe Ausreißer nach oben haben als jene von  $NO_x$ .



**Abbildung 1:** Darstellung der Zeitreihen  $PM_{10}$  bzgl. 1/2h-Werte, 4h- und Tagesmittelwerte



**Abbildung 2:** Darstellung der Zeitreihen  $NO_x$  bzgl. 1/2h-Werte, 4h- und Tagesmittelwerte

## 2.1 Lagemaße und Streuungsmaße

Nun sollen anhand der Messwerte statistische Kenngrößen der Daten berechnet werden. Vorab werden noch einige Definitionen wichtiger Kenngrößen angegeben, welche aus [2] entnommen wurden.

**Definition 1. Arithmetisches Mittel:**

Sei  $(x_1, \dots, x_n)$  eine realisierte Stichprobe.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{heißt arithmetisches Mittel oder Mittelwert.}$$

**Definition 2. Median:** Für geordnete Stichprobenwerte  $x_{(i)}$ ,  $i = 1, \dots, n$  heißt  $x_{med}$  Median, falls:

$$x_{med} = x_{(\frac{n+1}{2})} \quad \text{für } n \text{ ungerade,}$$
$$x_{med} = \frac{1}{2} \left( x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) \quad \text{für } n \text{ gerade.}$$

**Definition 3. Stichprobenvarianz:**

Die Stichprobenvarianz ist definiert als

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

**Definition 4. Standardabweichung:** Die Standardabweichung ist definiert als

$$s_L = \sqrt{s_L^2}.$$

**Definition 5. Variationskoeffizient:** Seien alle  $x_i \geq 0$  und  $\bar{x} > 0$ , dann ist der Variationskoeffizient

$$cv = \frac{s_L}{\bar{x}}$$

ein maßstabsunabhängiges Streuungsmaß, das die Abweichung relativ zu  $\bar{x}$  angibt.

**Definition 6. p-Quantil:** Das  $p$ -Quantil  $q_p$  ( $0 < p < 1$ ) der Stichprobe teilt die Daten so, dass mindestens  $p\%$  der Daten  $\leq q_p$  und mindestens  $(1-p)\%$  der Daten  $\geq q_p$  ist.

$q_{0.25}$  bezeichnet das **1. Quartil**,  $q_{0.5} = x_{med}$  das **2. Quartil** und  $q_{0.75}$  das **3. Quartil**.

**Definition 7. Interquartiler Bereich:**  $iqr = q_{0.75} - q_{0.25}$  heißt interquartiler Bereich.

**Definition 8. iqr-Standardabweichung:** Die  $iqr$ -Standardabweichung ist definiert als

$$s_q = \frac{iqr}{1,349} = \frac{q_{0.75} - q_{0.25}}{1,349},$$

wobei der Faktor von 1,349 durch die Normalverteilung motiviert ist.

Die Werte in der folgenden Tabelle sind in  $\mu\text{g}/\text{m}^3$  angegeben:

PM <sub>10</sub>	1/2h - Werte	4h - MW	TMW
Minimum	0,02	0,02	1,4
1st Quartil	17,00	17,00	19,00
Median	28,00	27,00	28,00
Mittelwert	35,00	34,00	34,00
3rd Quartil	44,00	42,00	41,00
Maximum	420,00	350,00	170,00
Standardabweichung	28,30	25,92	21,87
Stichprobenvarianz	801,12	671,79	477,67
Variationskoeffizient	22,90	19,90	14,16
intequartiler Bereich	27,68	25,25	21,92
iqr-Standardabweichung	20,52	18,72	16,25
Spannweite ( $x_{max} - x_{min}$ )	421,31	347,54	170,61
Anzahl der Zeitpunkte	35.040	4.380	730
Anzahl der bekannten Werte	33.828	4.370	730
Anzahl der fehlenden Werte	1.212	10	0

NO <sub>x</sub>	1/2h - Werte	4h - MW	TMW
Minimum	0,02	0,02	2,72
1st Quartil	7,30	7,90	11,00
Median	16,00	16,00	17,00
Mittelwert	27,00	26,00	26,00
3rd Quartil	33,00	33,00	32,00
Maximum	400,00	360,00	180,00
Standardabweichung	33,64	30,34	25,35
Stichprobenvarianz	1.131,57	920,03	641,88
Variationskoeffizient	41,79	35,04	24,48
intequartiler Bereich	25,92	24,59	20,57
iqr-Standardabweichung	19,21	18,23	15,25
Spannweite ( $x_{max} - x_{min}$ )	402,38	361,68	180,43
Anzahl der Zeitpunkte	35.040	4.380	730
Anzahl der bekannten Werte	33.985	4.345	725
Anzahl der fehlenden Werte	1.055	35	5

Die Werte der Tabelle sind auf 2 Dezimalstellen gerundet.

Bezüglich der Halbstundenwerte sind die Maxima und Minima der Zeitreihen NO<sub>x</sub> und PM<sub>10</sub> von der selben Größenordnung. Weiters sieht man, dass die Varianzen und Standardabweichungen für die Vierstundenmittelwerte kleiner sind als jene der Halbstundenwerte. Auch bei den Tagesmittelwerten werden diese Kenngrößen noch einmal verkleinert. Das bedeutet, dass die Verteilungen der Vierstunden- und Tagesmittelwerte glatter sind als jene der Halbstundenwerte. Dies ist auch in den Abbildungen 1 und 2 erkennbar.

Für die Zeitreihen der Halbstundenwerte von PM<sub>10</sub> und NO<sub>x</sub> sind die Stichprobenvarianzen

sehr hoch. Daher gilt, dass der Mittelwert für zukünftige Werte eine schlechte Vorhersage liefert. Geht man aber von den Halbstundenwerten zu den Vierstunden- und Tagesmittelwerten über, sieht man, dass die Varianzen immer kleiner werden, und daher der Mittelwert ein immer besseres Vorhersagewerkzeug darstellt.

Der Variationskoeffizient von NO<sub>x</sub> ist fast doppelt so hoch wie jener von PM<sub>10</sub>, was darauf schließen lässt, dass hier die Werte der einzelnen Zeitreihen viel stärker um ihre Mittelwerte schwanken als jene der Reihen von PM<sub>10</sub>. Auch der Variationskoeffizient wird bei den Reihen der Vierstunden- und Tagesmittelwerte kleiner. Das bedeutet, dass die Werte bei diesen beiden Zeitreihen weniger um den Mittelwert streuen.

Die hohen Spannweiten aller drei Zeitreihen sowohl von PM<sub>10</sub> als auch von NO<sub>x</sub> im Vergleich zu ihren relativ niedrigen dritten Quartilen lassen darauf schließen, dass es bei allen Reihen im oberen Bereich sehr hohe Ausreißer gibt, was in den Abbildungen 1 und 2 deutlich erkennbar ist.

## 2.2 Schiefemaße und Maße für die Wölbung

Im Folgenden sollen die Schiefemaße und Maße für die Wölbung der unterschiedlichen Zeitreihen berechnet werden. Zuerst müssen jedoch wieder ein paar grundlegende Kenngrößen definiert werden, welche aus [2] entnommen wurden.

### Definition 9. Momentkoeffizient und Quantilkoeffizient der Schiefe:

Der Momentkoeffizient der Schiefe ist definiert durch

$$g_1^m = \frac{m_3}{s^3} \quad \text{mit} \quad m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3.$$

Der Quantilkoeffizient der Schiefe ist definiert durch

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

Diese Maße geben die Abweichungen von der Symmetrie einer Verteilung an. Ist  $g_1^m = 0$ , so handelt es sich um eine symmetrische Verteilung. Ist der Wert positiv, so ist die Verteilung rechtsschief beziehungsweise linkssteil, das heißt, positive Abweichungen von der Symmetrie überwiegen. Bei einem negativen Momentkoeffizienten der Schiefe ist die Verteilung linksschief beziehungsweise rechtssteil, das heißt, negative Abweichungen von der Symmetrie überwiegen.

### Definition 10. Momentkoeffizient und Quantilkoeffizient der Wölbung:

Der Momentkoeffizient der Wölbung ist definiert durch

$$g_2^m = \frac{m_4}{s^4} - 3 \quad \text{mit} \quad m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4.$$

Der Quantilkoeffizient der Wölbung ist definiert durch

$$g_2^q = \frac{q_{0.975} - q_{0.025}}{q_{0.875} - q_{0.125}} - 1,75.$$

Diese Maße charakterisieren wie stark oder schwach die Randbereiche (Tails) von eingipfeligen Verteilungsfunktionen besetzt sind. Referenzverteilung ist hier die Normalverteilung (Maßzahl  $\approx 0$ ).

PM <sub>10</sub>	1/2h - Werte	4h - MW	TMW
Momentkoeffizient der Schiefe	2,63	2,59	2,29
Quantilkoeffizient der Schiefe	1,77	1,78	1,57
Momentkoeffizient der Wölbung	13,82	13,06	8,47
Quantilkoeffizient der Wölbung	0,28	0,28	0,20

NO <sub>x</sub>	1/2h - Werte	4h - MW	TMW
Momentkoeffizient der Schiefe	3,16	2,92	2,36
Quantilkoeffizient der Schiefe	4,95	4,30	4,70
Momentkoeffizient der Wölbung	14,37	12,32	6,79
Quantilkoeffizient der Wölbung	0,76	0,74	0,46

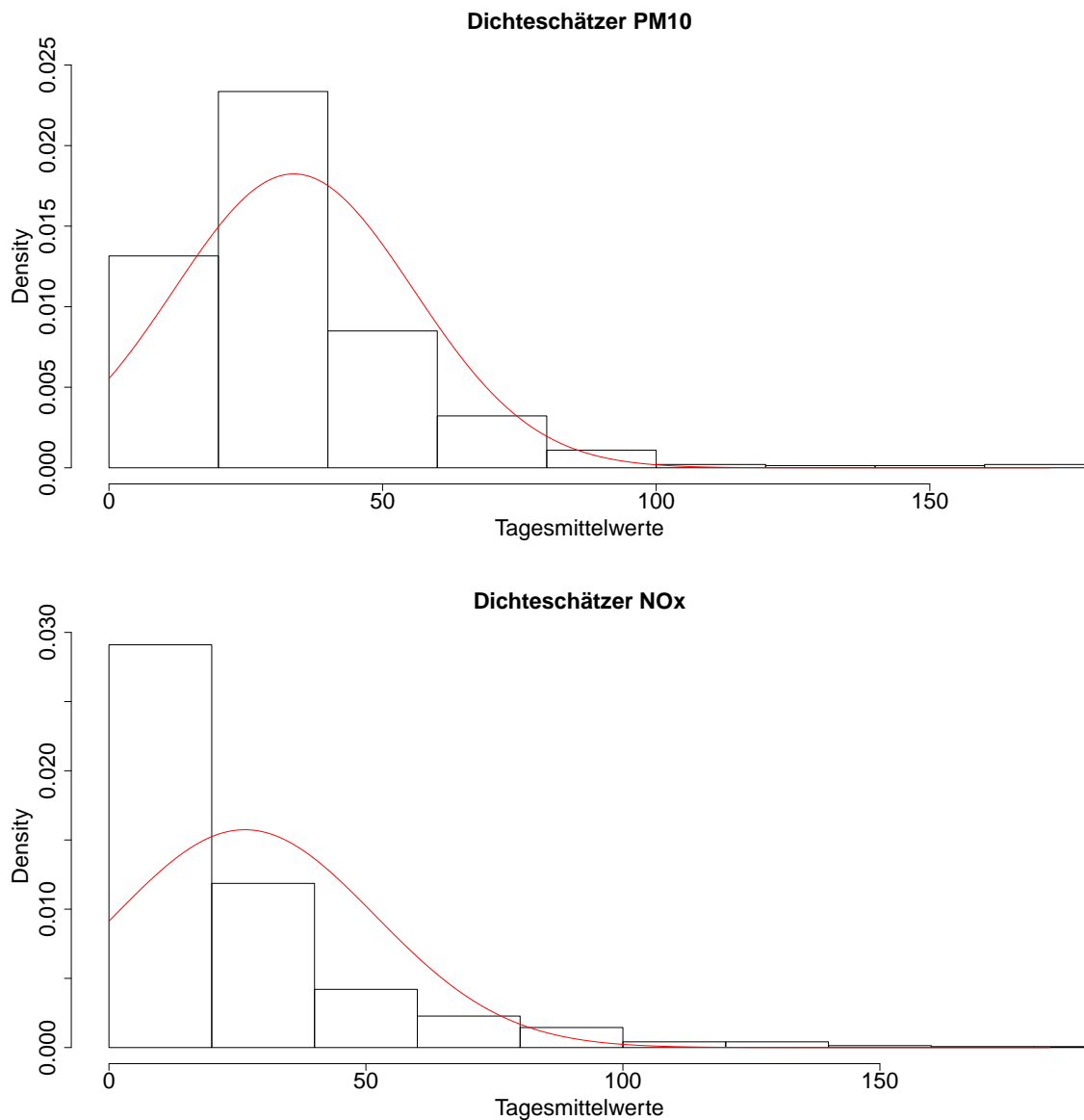
Da die Momentkoeffizienten der Schiefe für alle Zeitreihen der Schadstoffe PM<sub>10</sub> und NO<sub>x</sub> positiv sind, handelt es sich jeweils um rechtsschiefe beziehungsweise linkssteile Verteilungen der Datenpunkte, d.h. die positiven Abweichungen von der Symmetrie überwiegen.

Auch die Momentkoeffizienten der Wölbung sind sowohl bei PM<sub>10</sub> als auch bei NO<sub>x</sub> überall positiv. Das bedeutet, dass vor allem die Tails stärker besetzt sind. Dies resultiert erneut daraus, dass die Konzentrationen der beiden Schadstoffe in den Wintermonaten höher sind als jene im Sommer. Man spricht dann von einer „spitzen“ Verteilung der Daten.

Nachdem die Schiefemaße und die Maße für die Wölbung für die Vierstunden- und Tagesmittelwerte kleiner werden, wird bei diesen Werten die Verteilung der Daten immer symmetrischer und „flacher“.

Die Rechtsschiefe der Daten lässt sich auch anhand der Dichteschätzer der Normalverteilung  $N(\bar{x}, s^2)$  für die beiden Reihen der Tagesmittelwerte in Abbildung 3 deutlich erkennen. Die langen Tails bei den Dichteschätzern deuten darauf hin, dass es auch noch bei der Reihe der Tagesmittelwerte einige Ausreißer nach oben gibt.





**Abbildung 3:** Dichteschätzer (Normalverteilung) für  $NO_x$  und  $PM_{10}$  anhand der Tagesmittelwerte

## Boxplotserien

Nun werden die verschiedenen Merkmale (Halbstundenwerte, Vierstundenmittel- und Tagesmittelwerte) der drei Schadstoffe, aufgeteilt nach verschiedenen kategorischen Merkmalen anhand von Boxplotserien analysiert. Um einen besseren Überblick über die Verteilungen der Datenpunkte zu erhalten, wird bei den Boxplots auch eine Detailansicht des interquartilen Bereichs angegeben, falls das nötig ist. Für eine genauere Interpretation werden zusätzlich noch einige statistische Kenngrößen sowie die zu den kategorischen Merkmalen passenden Dichteschätzer für die Tagesmittelwerte angegeben.

**Kategorisches Merkmal: Sommer/Winter**

Anhand des kategorischen Merkmals „Sommer/Winter“ sollen nun einige statistische Kenngrößen betrachtet werden. Der betrachtete Zeitraum beinhaltet zwei Sommerperioden (01.04. bis 30.09. für 2005 und 2006) und zwei Winterperioden (jeweils 01.01. bis 30.03. und 1.10 bis 31.12.).

<b>PM<sub>10</sub> Winter</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	17.129	2.184	364
Mittelwert	43,52	42,67	42,67
Standardabweichung	33,43	31,05	25,98
Momentkoeffizient Schiefe	2,13	2,02	1,83
Momentkoeffizient Wölbung	9,22	7,97	5,14
Median	35,63	35,28	36,94
iqr-Bereich	36,26	34,51	30,31
iqr-Standardabweichung	26,88	25,58	22,47

<b>PM<sub>10</sub> Sommer</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	16.699	2.186	366
Mittelwert	26,22	25,04	24,92
Standardabweichung	18,04	14,93	11,25
Momentkoeffizient Schiefe	3,14	3,05	0,93
Momentkoeffizient Wölbung	29,33	36,74	1,26
Median	22,85	22,80	23,14
iqr-Bereich	19,69	17,82	15,38
iqr-Standardabweichung	14,59	13,21	11,40

Auch hier lassen sich bereits bekannte Ergebnisse feststellen:

Die Mittelwerte und die Standardabweichungen sind für alle drei Merkmale im Sommer viel niedriger als im Winter. Dies lässt erneut darauf schließen, dass die Konzentrationen der beiden Schadstoffe im allgemeinen in der kalten Jahreszeit höher sind als in den warmen Sommermonaten.

Der Momentkoeffizient der Wölbung zeigt für die Sommer-Messwerte eine erstaunlich hohe Ausprägungen in allen drei Merkmalen. Anhand des Schiefekoeffizienten für die Tagesmittelwerte erkennt man, dass die Tagesmittelwert - Verteilung im Sommer weniger schief ist als jene im Winter. Bei den Halbstunden- und Vierstundenmittelwerten ist dies jedoch genau umgekehrt.

Diese weniger schiefe Verteilung sieht man auch anhand des Dichteschätzers in Abbildung 4. Man erkennt weiters anhand der positiven Momentkoeffizienten der Schiefe eine Rechtsschiefe der Verteilungen der einzelnen Merkmale, die auch in Abbildung 4 für die Tagesmittelwerte gut sichtbar ist. Aufgrund der stärkeren Ausreißer der PM<sub>10</sub> - Messwerte im Winter ergibt sich für die Dichteschätzer eine unterschiedliche Skalierung.

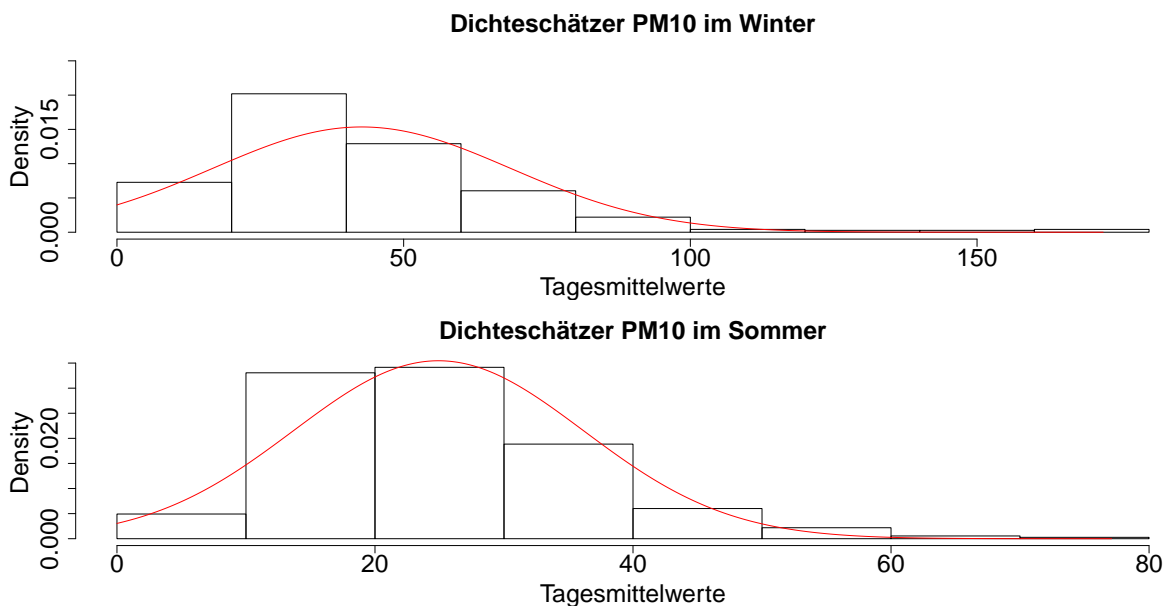


Abbildung 4: Dichteschätzer (Normalverteilung) für PM<sub>10</sub> bzgl. Sommer/Winter, Tagesmittelwerte

Auch für die Messdaten des Luftschadstoffs NO<sub>x</sub> sollen nun einige statistische Kenngrößen berechnet werden:

<b>NO<sub>x</sub> Winter</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	16.788,00	2.149,00	359,00
Mittelwert	41,89	40,91	40,81
Standardabweichung	41,35	36,83	29,33
Momentkoeffizient Schiefe	2,38	2,18	1,65
Momentkoeffizient Wölbung	8,10	6,86	3,20
Median	29,42	30,38	32,01
iqr-Bereich	37,03	34,83	30,38
iqr-Standardabweichung	27,45	25,82	22,52

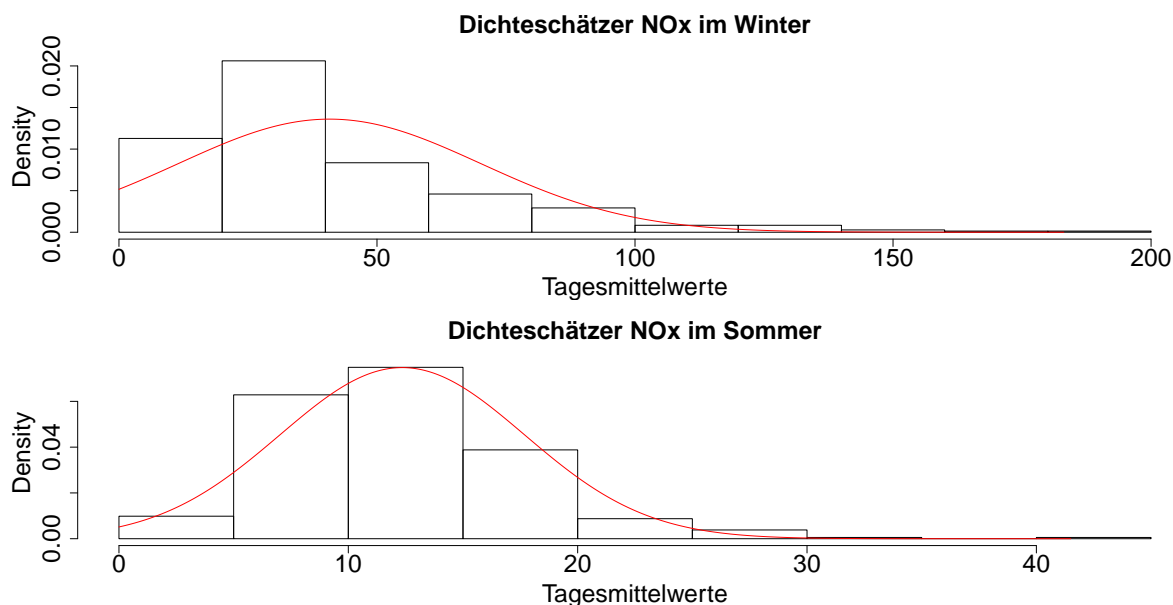
<b>NO<sub>x</sub> Sommer</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	17.197,00	2.196,00	366,00
Mittelwert	12,62	12,35	12,35
Standardabweichung	12,00	9,69	5,34
Momentkoeffizient Schiefe	2,86	1,94	1,07
Momentkoeffizient Wölbung	14,84	6,70	2,75
Median	9,26	9,88	12,07
iqr-Bereich	11,29	10,66	6,74
iqr-Standardabweichung	8,37	7,90	5,00

Auch für die Reihen von  $NO_x$  gilt, dass die Emissionen im Winter deutlich höher sind als im Sommer. Die Kurtosis ist für die Halbstundenwerte im Sommer höher als im Winter. Bei den Vierstunden- und Tagesmittelwerten verhält es sich genau umgekehrt: Hier sind die Verteilungen der Datenpunkte in den Wintermonaten spitzer als in den Sommermonaten. Der Unterschied zwischen der Kurtosis im Winter und im Sommer ist bei den Reihen von  $NO_x$  jedoch nicht so deutlich als bei jenen von  $PM_{10}$ .

Auch in Abbildung 5 sieht man, dass die Verteilung der Tagesmittelwerte im Sommer weniger schief ist als im Winter und auch weniger Ausreißer vorhanden sind. Auch die stärker besetzten Tails lassen sich in der Darstellung des Dichteschätzers für den Sommer gut erkennen.

Die Tatsache, dass sowohl für  $NO_x$  als auch für  $PM_{10}$  die Kurtosis für die Reihen der Tagesmittelwerte in den Sommermonaten höher ist, erkennt man an den Dichteschätzern in den Abbildungen 4 und 5 durch die spitzeren Verteilungsfunktionen. In den Wintermonaten ist konträr dazu eine eher flache Verteilungsfunktion für beide Schadstoffe zu erkennen.

Weiters erkennt man, dass die Dichte der Daten in den Sommermonaten höhere Werte annimmt als in den Wintermonaten, was darauf schließen lässt, dass die Messdaten im Sommer weniger streuen als im Winter.



**Abbildung 5:** Dichteschätzer (Normalverteilung) für  $NO_x$  bzgl. Sommer/Winter, Tagesmittelwerte

Nun werden die Boxplots anhand des kategorischen Merkmals „Sommer/Winter“ erzeugt, zunächst bezüglich der Halbstundenwerte:

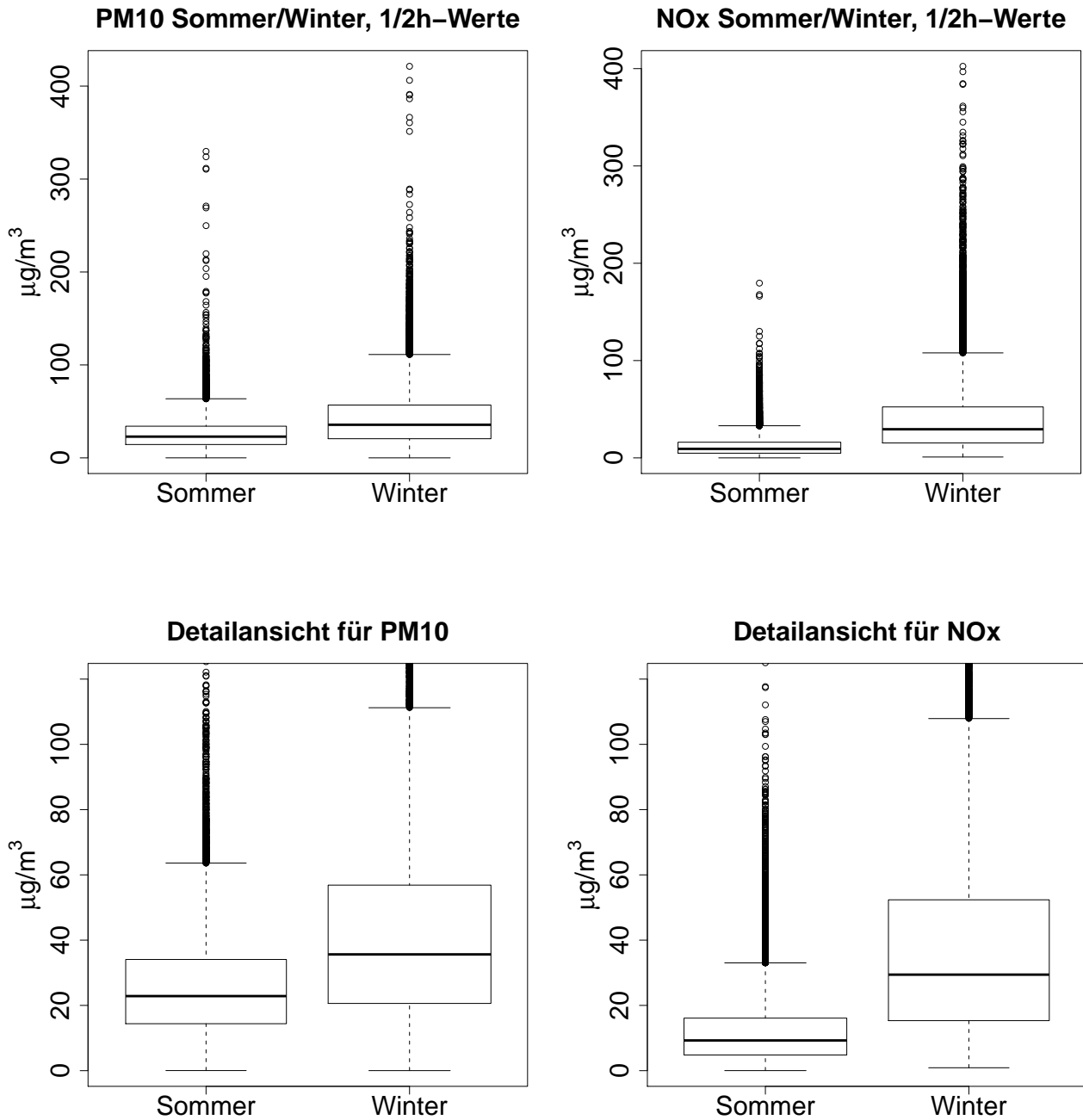


Abbildung 6: Boxplots von  $PM_{10}$  und  $NO_x$  bzgl. Sommer/Winter, Halbstundenwerte

Da durch die Ausreißer sehr wenig über die Verteilung der Datenpunkte im iqr-Bereich ausgesagt werden kann, wurde der Bereich eingeschränkt, um die Ansicht der zentralen 50% zu vergrößern.

Es ist wieder ersichtlich, dass die Messwerte im Winter höher ausfallen, als jene im Sommer. Die Werte der Zeitreihe von  $PM_{10}$  weisen im Sommer eine annähernd symmetrische Verteilung

innerhalb der zentralen 50% auf, da sich der Median ungefähr in der Mitte der Box befindet. Es existieren jedoch bei den Halbstundenwerten sehr starke Ausreißer nach oben. Da die Box der zentralen 50% ziemlich klein ist, bestehen die Daten zum Großteil aus Ausreißern.

Die Boxplots von  $PM_{10}$  und  $NO_x$  unterscheiden sich im Winter kaum voneinander, während die  $PM_{10}$ -Messwerte im Sommer ohne Berücksichtigung der Ausreißer deutlich höher ausfallen als jene von  $NO_x$ .

Für die Boxplots der Vierstundenmittelwerte erhält man folgende Graphiken:

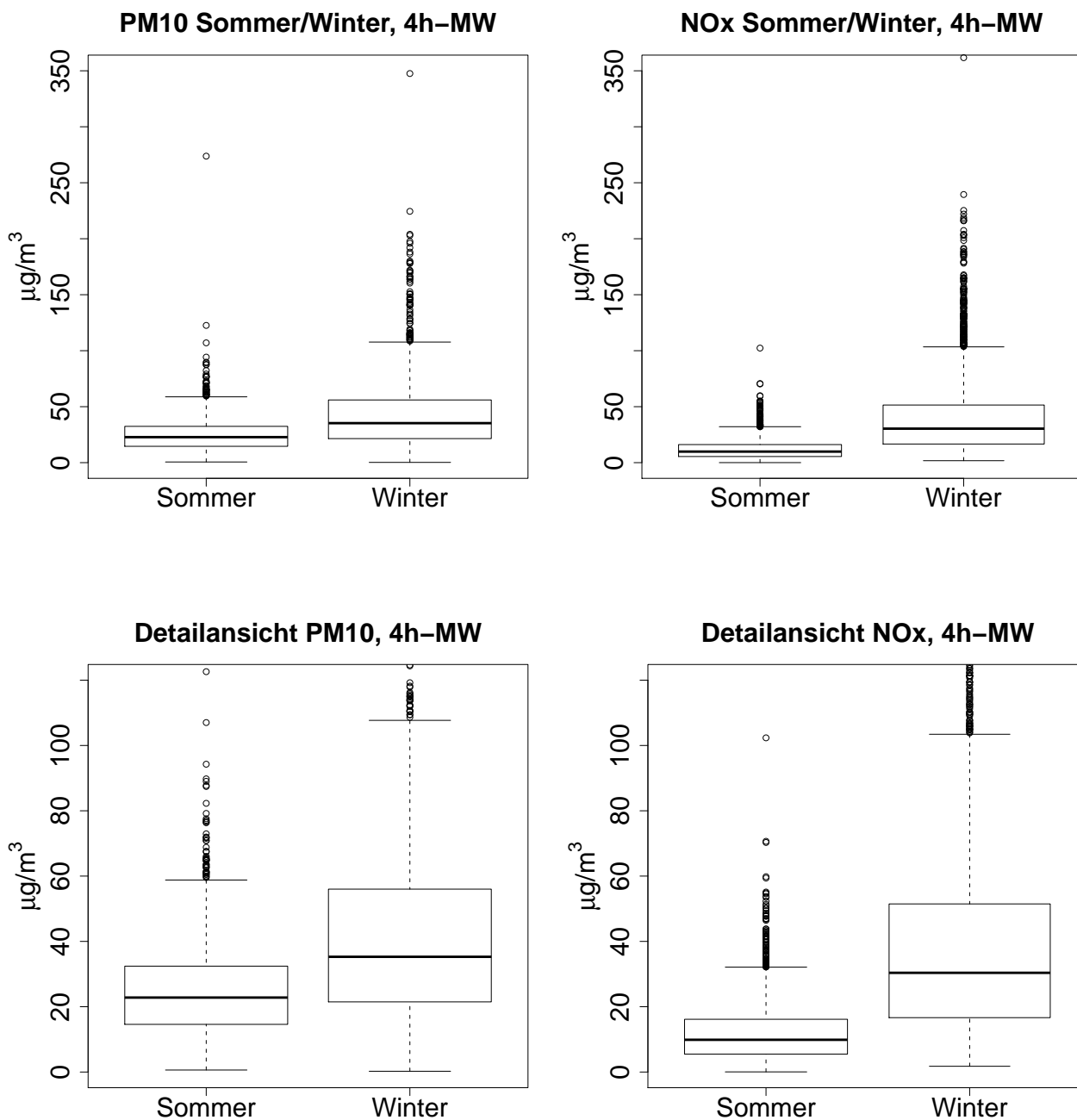


Abbildung 7: Boxplots von  $PM_{10}$  und  $NO_x$  bzgl. Sommer/Winter, Vierstundenwerte

Auch bei den Zeitreihen der Vierstundenmittelwerte kommt man zu ähnlichen Ergebnissen wie bei jenen der Halbstundenwerte, mit dem Unterschied, dass bei ihren Boxplots weniger Ausreißer zu sehen sind. Dies ist darauf zurückzuführen, dass die Messwerte der Reihen der Vierstundenmittelwerte durch Mittelwertbildung aus den Daten der Halbstundenwerte hervorgehen.

Die Boxplots für die Reihen der Tagesmittelwerte ergeben sich wie folgt:

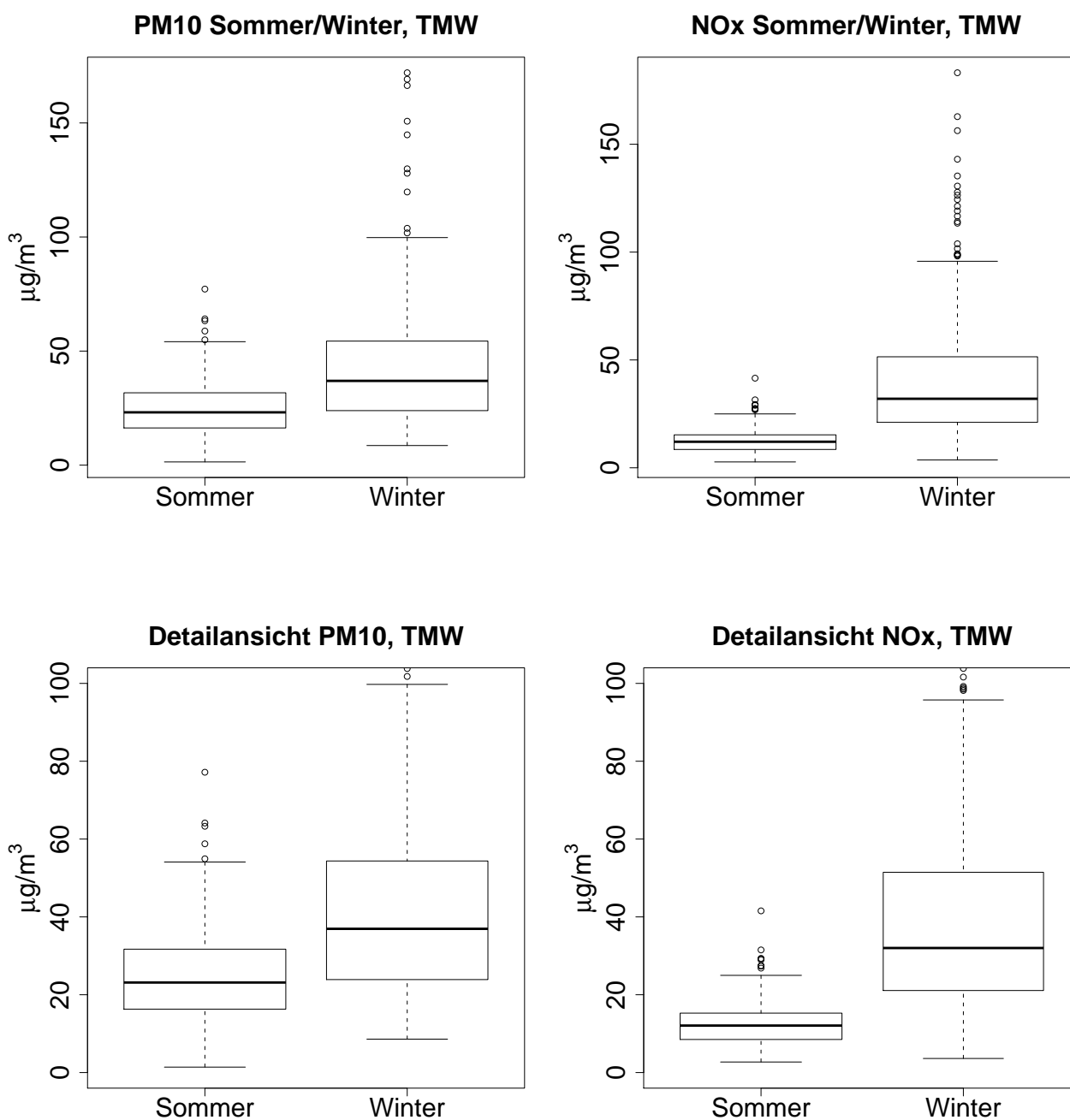


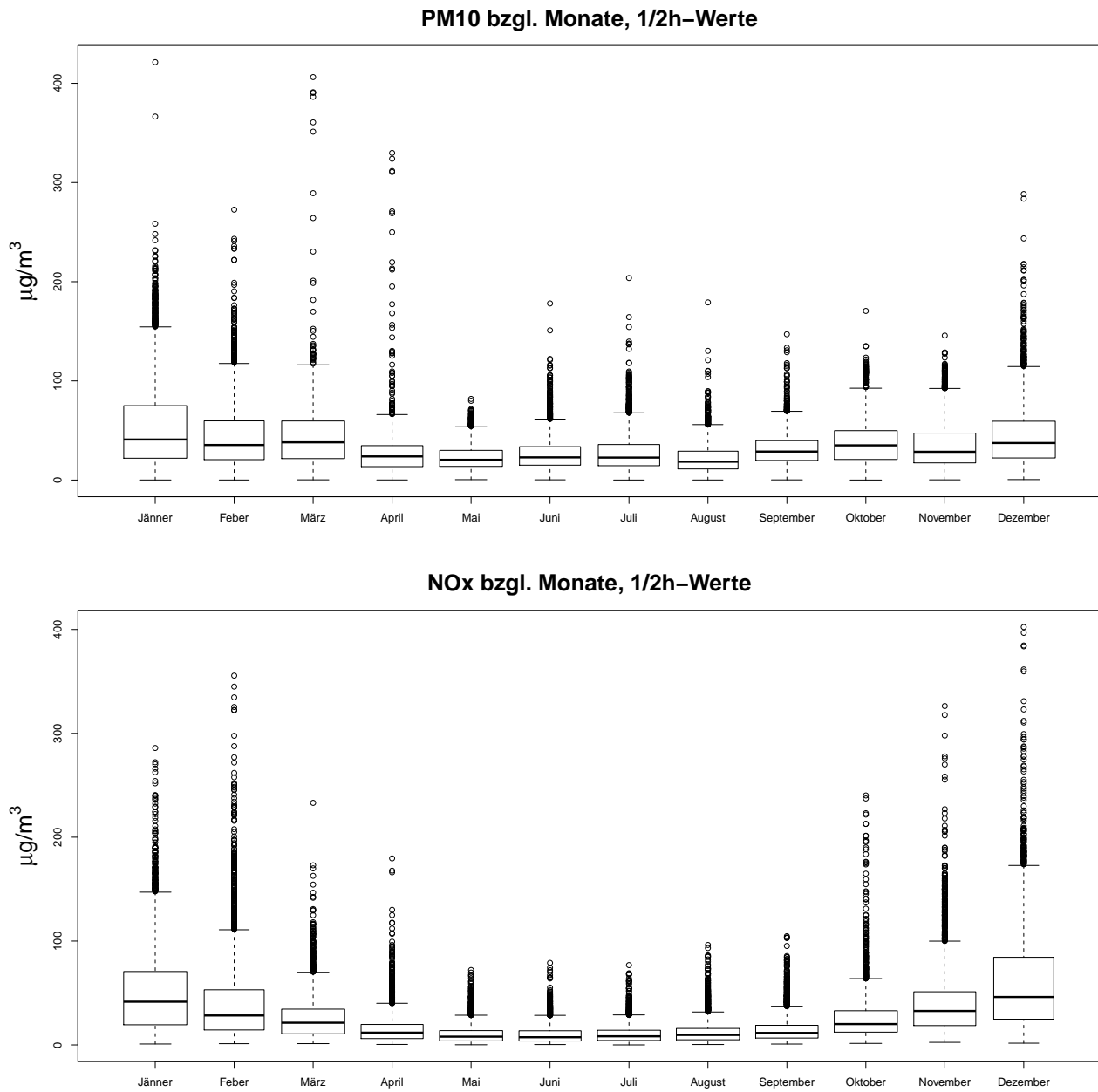
Abbildung 8: Darstellung der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Sommer/Winter, Tagesmittelwerte

Auch anhand von Abbildung 8 kann man ähnliche Schlüsse für die Tagesmittelwerte ziehen wie zuvor. Ausreißer sind hier beinahe zur Gänze verschwunden. Geht man von den Halbstundenwerten über die der Vierstundenmittelwerte zu jenen der Tagesmittelwerte über, sieht man, dass der 1.5-fache iqr-Bereich immer breiter wird. Das bedeutet wieder, dass bei den Tagesmittelwerte einige Spitzen weggenommen wurden und damit weniger hohe Ausreißer vorhanden sind. Anhand der Boxplots in den Abbildungen 6, 7 und 8 erkennt man erneut die Rechtsschiefe der Verteilungen der einzelnen Merkmale sowohl in den Sommer- als auch in den Wintermonaten.



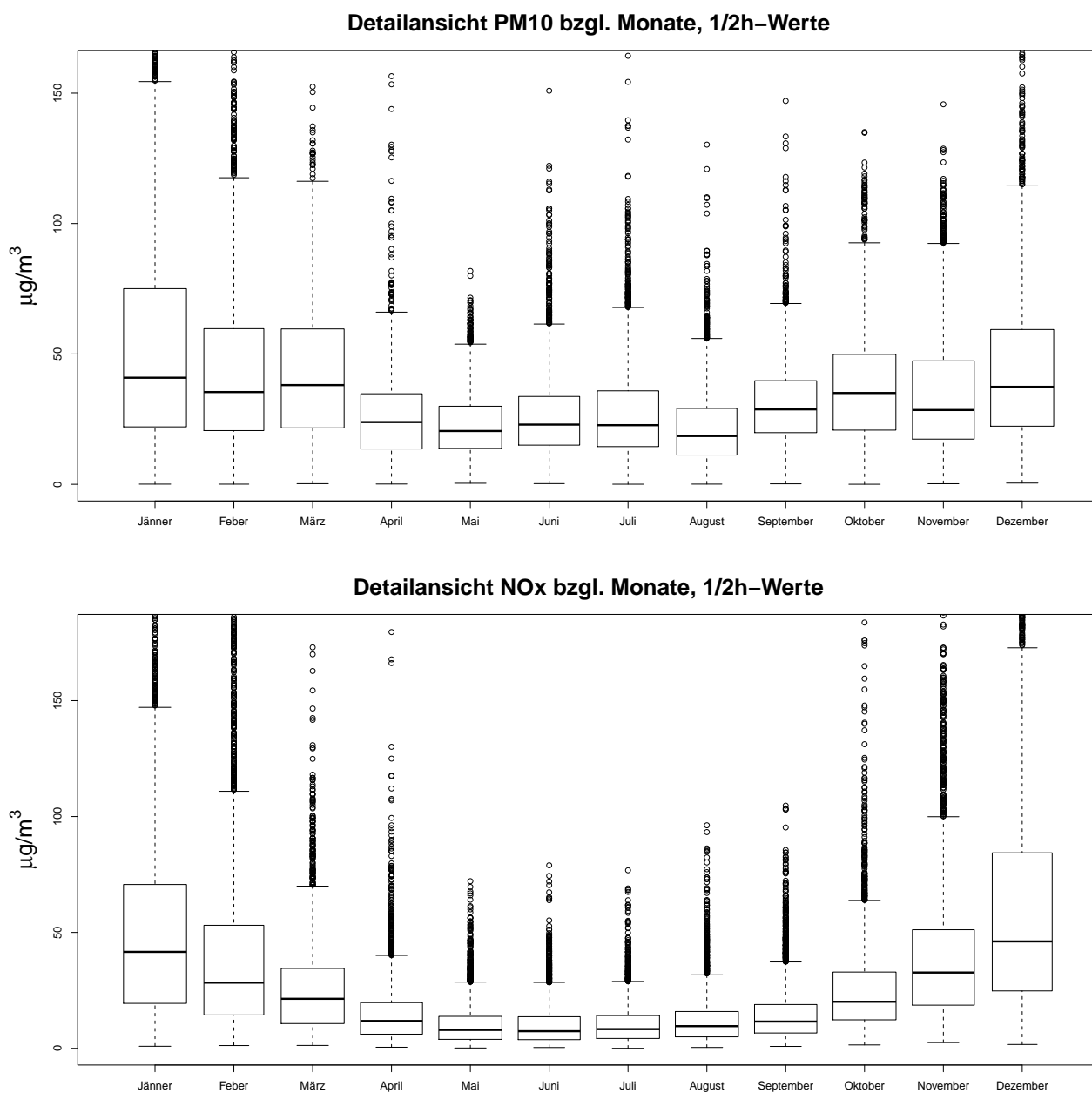
**Kategorisches Merkmal: Monat**

Nun werden Boxplotserien im Bezug auf die unterschiedlichen monatlichen Belastungen für  $PM_{10}$  und  $NO_x$  angegeben:



**Abbildung 9:** Darstellung der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Monat, Halbstundenwerte

Um die zentralen 50% der Daten besser darstellen zu können, werden nun wieder extreme Ausreißer entfernt:

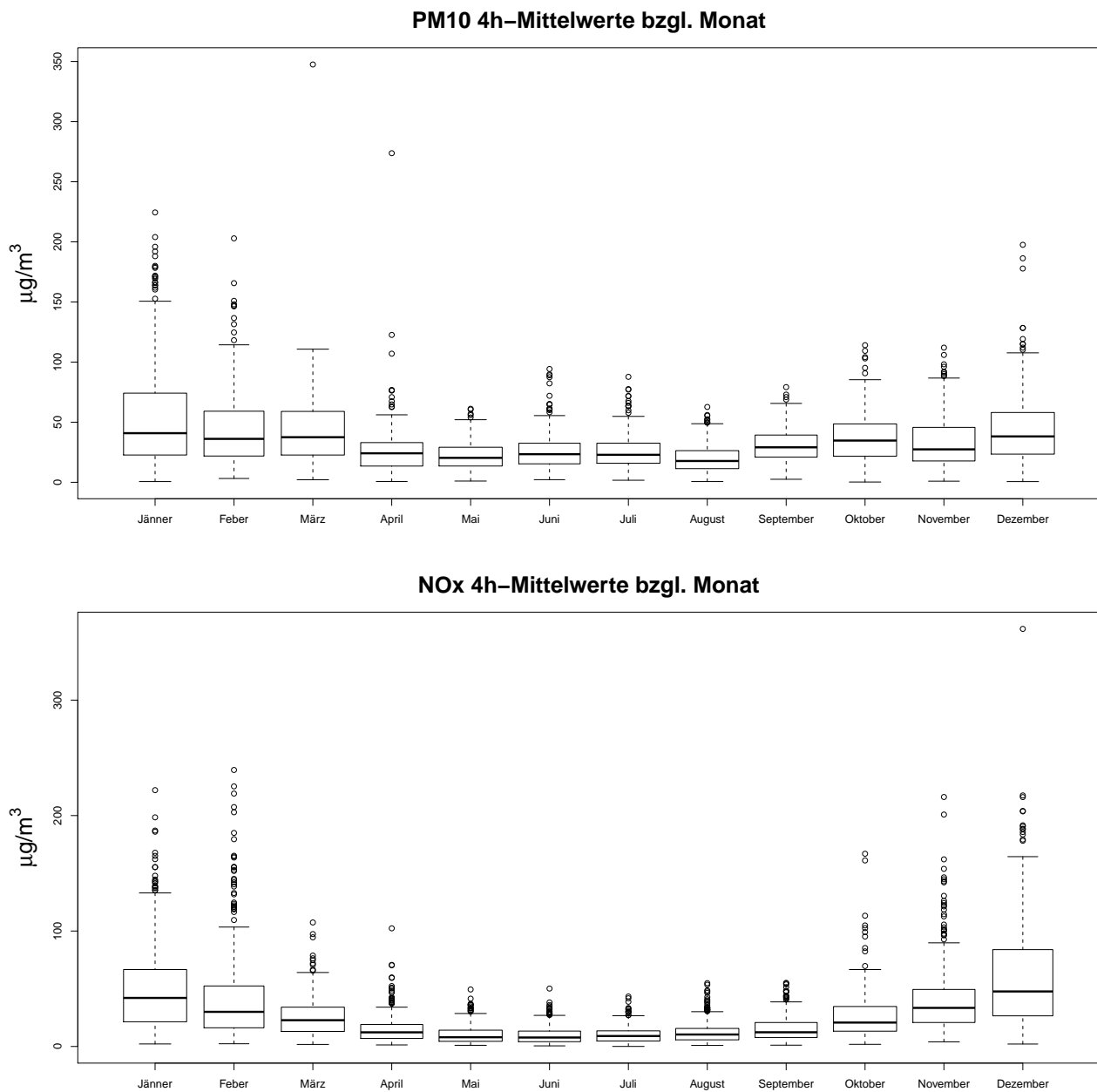


**Abbildung 10:** Detailansicht Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Monat, Halbstundenwerte

Abbildung 10 zeigt ein bereits bekanntes Ergebnis: Die Messwerte der Reihen von  $PM_{10}$  und  $NO_x$  nehmen in den Monaten Oktober bis März sehr hohe Werte an, mit sehr großen Ausreißern. Die höchsten Werte (ohne die Ausreißer zu berücksichtigen) fallen hierbei für  $PM_{10}$  im Jänner und für  $NO_x$  im Dezember an. In den Monaten Mai bis September sind die  $PM_{10}$ -Konzentrationen dafür wieder relativ niedrig, ehe sie im Oktober erneut zu steigen beginnen.

Generell kann man sagen, dass  $NO_x$  zwar in den Sommermonaten stärker absinkt als  $PM_{10}$ , dafür jedoch im Winter die  $NO_x$ -Belastung tendenziell höher ist als jene von  $PM_{10}$ . Mithilfe dieser Boxplots lässt sich wieder die bereits in der Einleitung angemerkte Abhängigkeit zwi-

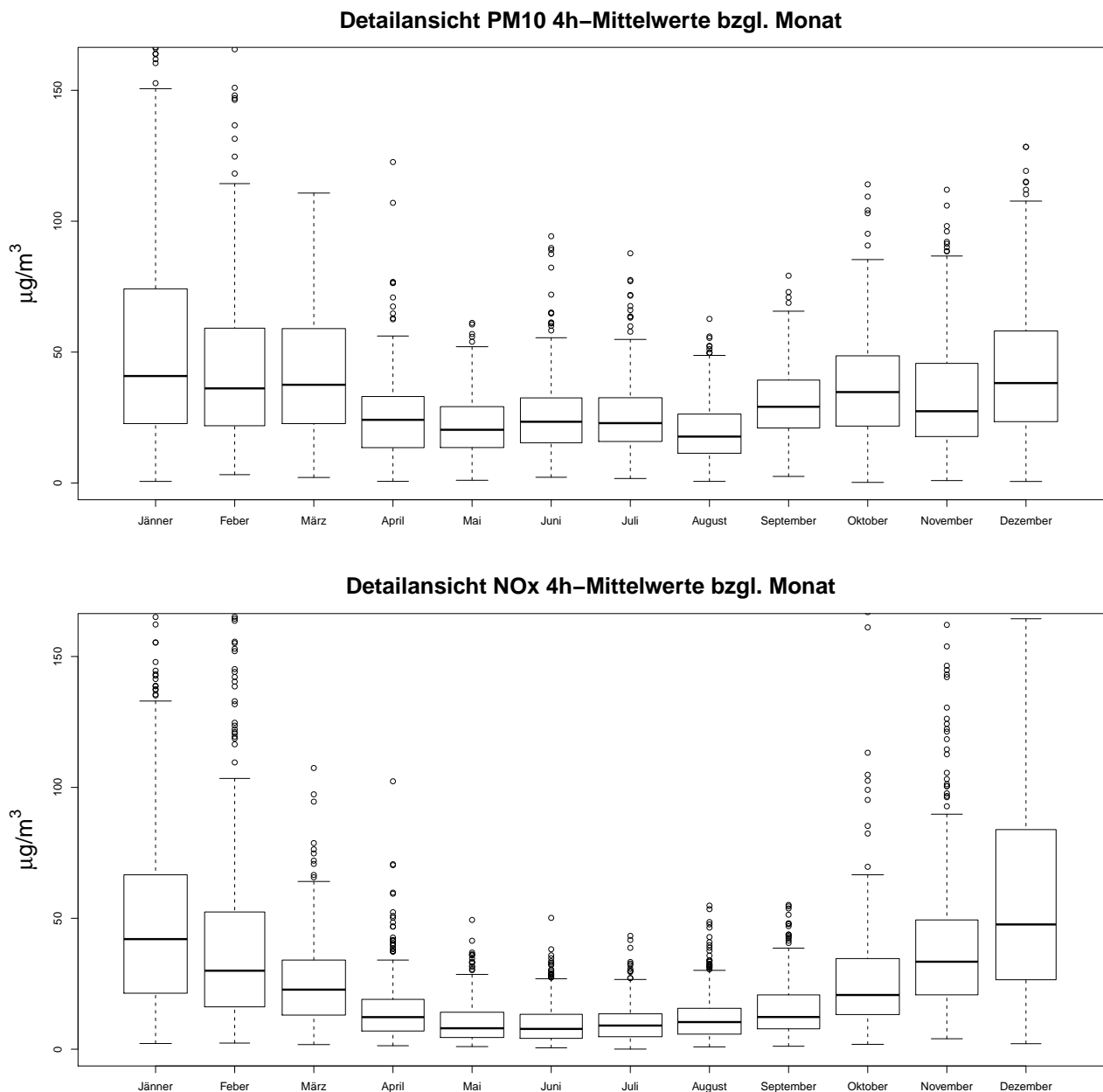
schen  $PM_{10}$  und  $NO_x$  vermuten. Allgemein fallen die Ausreißer für die  $NO_x$  - Daten stärker aus.



**Abbildung 11:** Darstellung der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Monat, Vierstundenmittelwerte

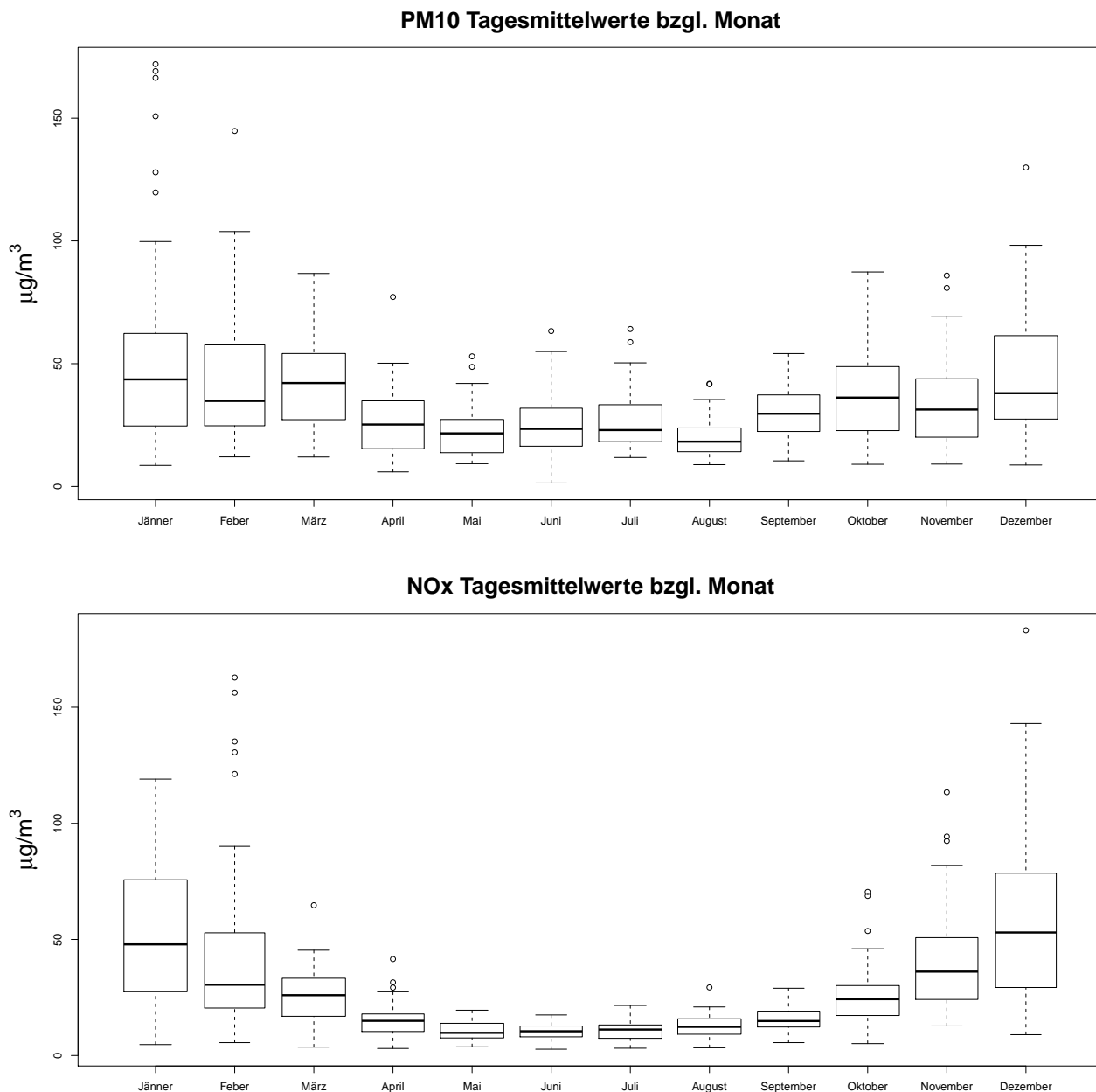
Man erkennt hier extreme Ausreißer in den Monaten März und April. Diese entsprechen dem 26.03.2005 beziehungsweise dem 15.04.2006, was jeweils der Ostersonntag war. In den Jahren 2005/2006 waren Osterfeuer in Graz noch nicht verboten, weshalb die extrem hohen Ausreißer wohl darauf zurückzuführen sind.

Auch in Abbildung 11 müssen extremen Ausreißer unberücksichtigt bleiben, um vernünftige Aussagen über die Verteilung der Daten im iqr-Bereich treffen zu können. Diese Detailansicht wird in Abbildung 12 dargestellt.



**Abbildung 12:** Detailansicht Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Monat, Vierstundenmittelwerte

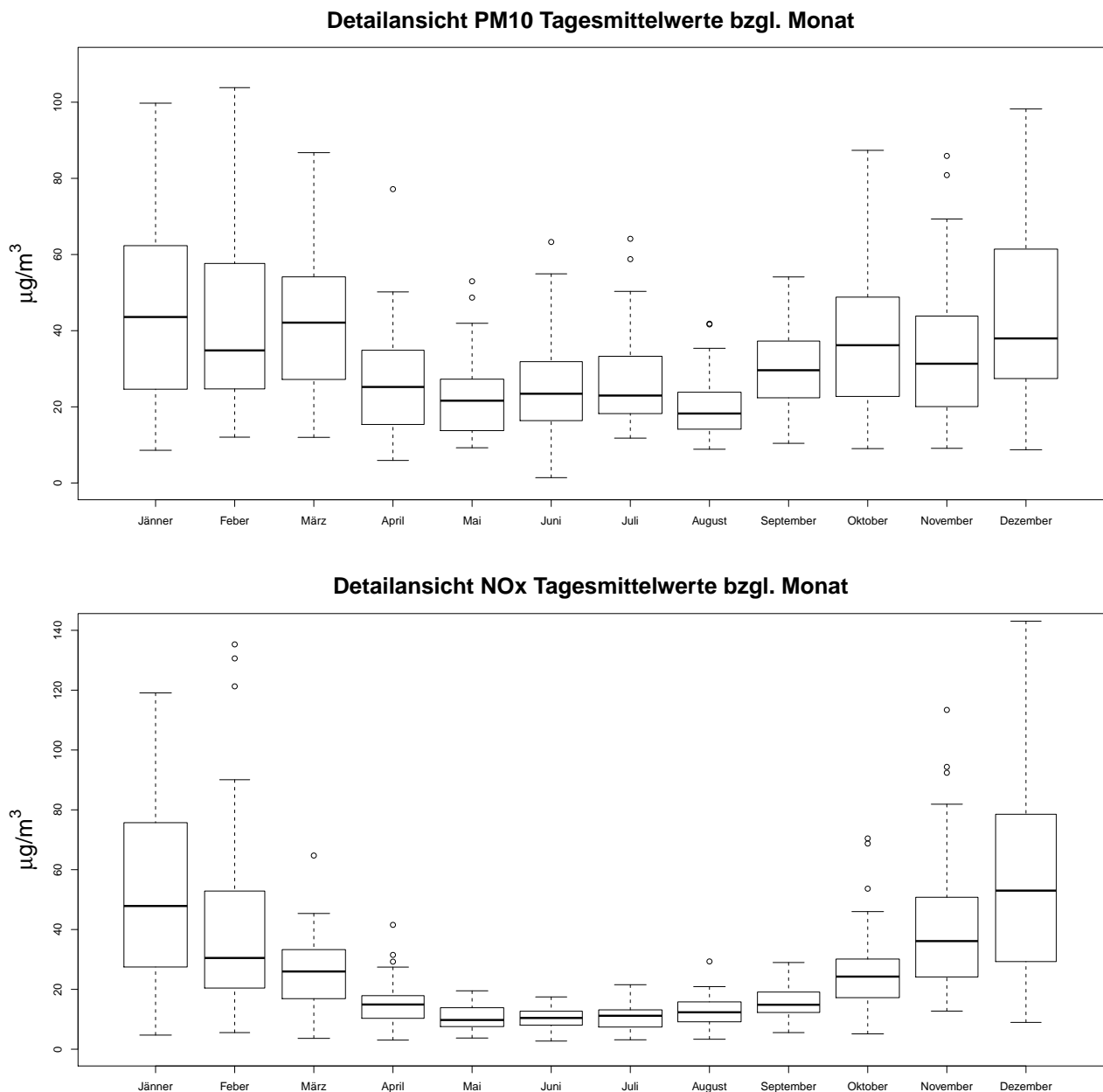
Auch hier spiegeln sich die Ergebnisse, die man bereits für die Halbstundenwerte erkennen konnte, wider: Die Werte beider Schadstoffe steigen in den Wintermonaten kontinuierlich an und man sieht erneut, dass die Messwerte zum Jahreswechsel am größten sind. Im Vergleich zu den Halbstundenwerten sind bei den Vierstundenmittelwerten wieder weniger Ausreißer zu erkennen.



**Abbildung 13:** Darstellung der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Monat, Tagesmittelwerte

Der Ausreißer für die  $PM_{10}$  - Messwerte der Tagesmittelwerte im Dezember entspricht dem 31.12.2006, woraus man schließen kann, dass eine erhöhte Feinstaubbelastung durch das Abschießen von Feuerwerkskörpern entsteht.

Auch für die Tagesmittelwerte werden nun Ausreißer in Abbildung 13 nicht berücksichtigt, um die Ansicht der zentralen 50% zu vergrößern und die daraus resultierenden Detailansichten werden in Abbildung 14 angegeben.



**Abbildung 14:** Detailansicht Boxplots PM<sub>10</sub> und NO<sub>x</sub> bzgl. Monat, Tagesmittelwerte

Die Boxplotserien der Reihen der Tagesmittelwerte für PM<sub>10</sub> und NO<sub>x</sub> verdeutlichen noch einmal die bereits für die Halbstunden- und Vierstundenmittelwerte gewonnenen Ergebnisse.

Bemerkenswert ist noch, dass bei allen drei Reihen von PM<sub>10</sub> die Werte im Juni und im Juli im Vergleich zu ihren unmittelbaren Nachbarn davor und danach höher sind.

**Kategorisches Merkmal: Wochentag**

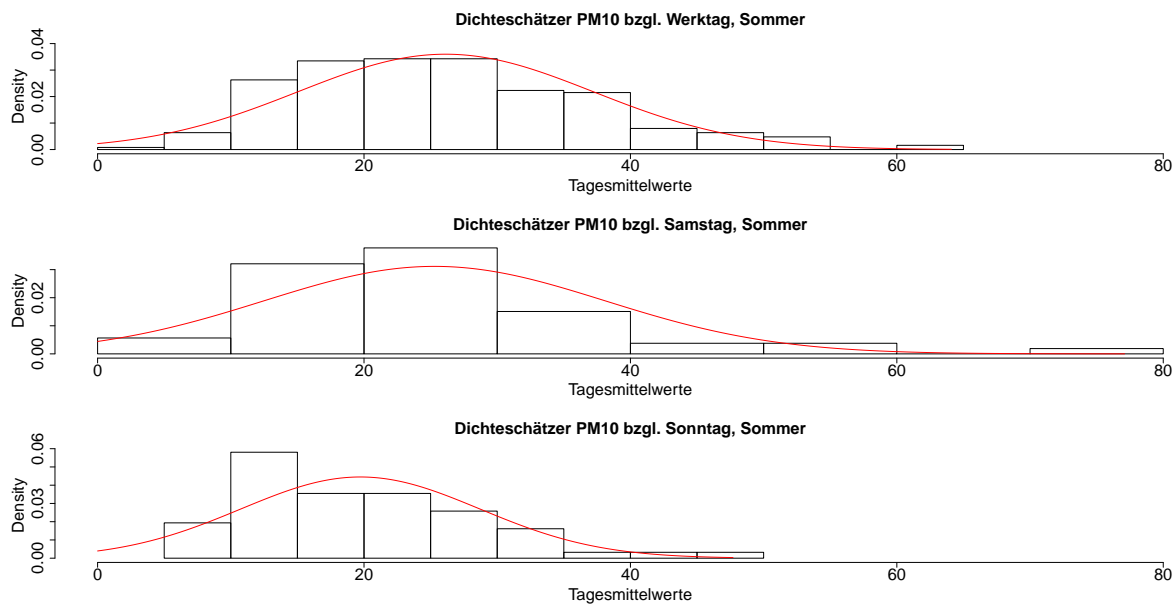
Nun wird eine Unterteilung nach den Wochentagen angestrebt: Werktag (Montag - Freitag), Samstag, Sonn- und Feiertag. Aufgrund der enormen Unterschiede in den Messwerten zwischen den Sommer- und Wintermonaten wird hierbei wieder eine zusätzliche Unterteilung durch das kategorische Merkmal *Sommer/Winter* vorgenommen.

Zunächst werden jedoch wieder einige statistische Kenngrößen für die Zeitreihen von PM<sub>10</sub> angegeben.

<b>PM<sub>10</sub> Werktag Sommer</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	11.480	1.496	251
Mittelwert	27,42	26,30	26,13
Standardabweichung	17,47	14,17	11,09
Momentkoeffizient Schiefe	1,55	0,91	0,66
Momentkoeffizient Wölbung	4,95	1,24	0,22
Median	23,89	24,43	24,47
iqr-Bereich	21,11	18,70	15,51
iqr-Standardabweichung	15,65	13,86	11,50

<b>PM<sub>10</sub> Samstag Sommer</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	2.432	318	53
Mittelwert	26,46	25,29	25,29
Standardabweichung	22,92	19,62	12,80
Momentkoeffizient Schiefe	6,27	6,86	1,62
Momentkoeffizient Wölbung	64,87	79,82	3,86
Median	23,44	22,91	23,27
iqr-Bereich	17,84	15,86	12,14
iqr-Standardabweichung	13,23	11,75	9,00

<b>PM<sub>10</sub> Sonntag Sommer</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	2.787	372	62
Mittelwert	21,06	19,72	19,72
Standardabweichung	14,16	11,78	8,96
Momentkoeffizient Schiefe	3,03	2,45	0,93
Momentkoeffizient Wölbung	21,98	15,65	0,58
Median	18,69	17,81	17,20
iqr-Bereich	14,99	12,18	12,64
iqr-Standardabweichung	11,11	9,03	9,37



**Abbildung 15:** Darstellung des Dichteschätzers von  $PM_{10}$  bzgl. Wochentage, Tagesmittelwerte

Anhand der Kenngrößen sieht man, dass es sich in allen drei Kategorien um rechtsschiefe Verteilungen der Messdaten handelt, deren Verteilung umso weniger symmetrisch ist, je größer die Zeitabstände zwischen den einzelnen Messpunkten werden.

Die Tatsache dass der Median in derselben Größenordnung liegt wie der Mittelwert, lässt eine nahezu symmetrische Verteilung erkennen.

Die Schiefekoeffizienten der einzelnen Merkmale sind in der Kategorie „Werktag“ am kleinsten, was darauf schließen lässt, dass die Verteilungsfunktion hier den niedrigsten Grad an Unsymmetrie aufweist.

Für Samstag und Sonntag erhält man bei den Zeitreihen der Halbstunden- und Vierstundenmittelwerte eine vergleichsweise hohe Kurtosis. Das bedeutet, dass es sich hier um sehr spitze Verteilungen handelt. Die flachste Verteilung erhält man Werktags für die Tagesmittelwerte. Dies sieht man auch sehr gut anhand der Dichteschätzer in Abbildung 15.

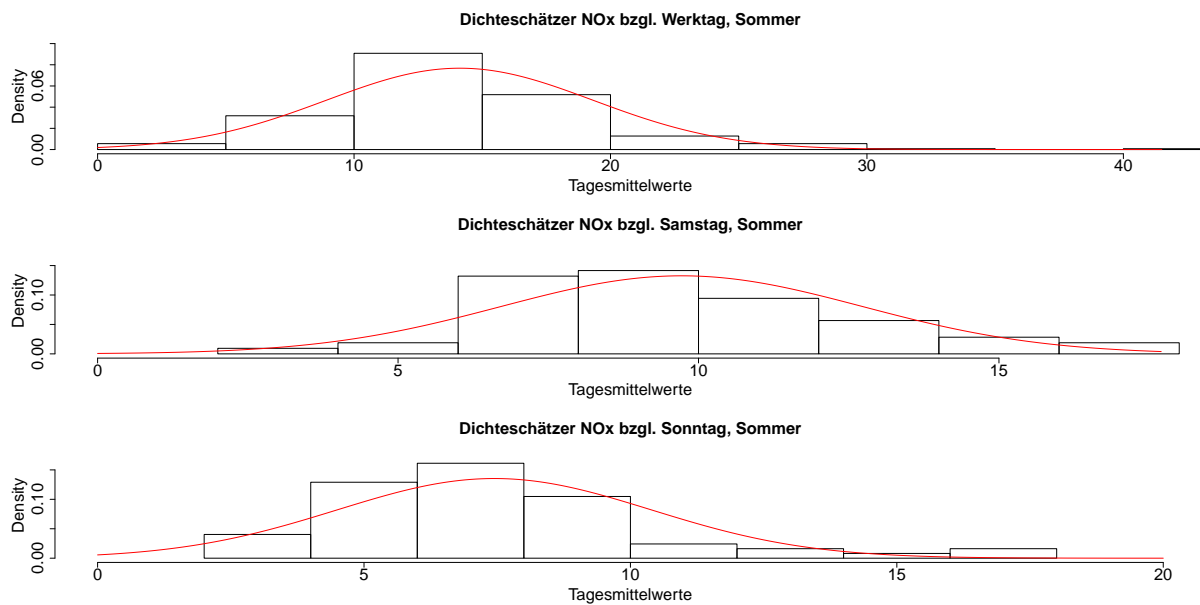


Für NO<sub>x</sub> erhält man im Sommer folgende statistische Kenngrößen:

<b>NO<sub>x</sub> Werktag Sommer</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	11.789	1.506	251
Mittelwert	14,43	14,12	14,12
Standardabweichung	13,36	10,57	5,20
Momentkoeffizient Schiefe	2,64	1,77	1,12
Momentkoeffizient Wölbung	12,24	5,46	3,42
Median	10,41	11,20	13,57
iqr-Bereich	12,74	12,13	5,44
iqr-Standardabweichung	9,44	8,99	4,03

<b>NO<sub>x</sub> Samstag Sommer</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	2.495	318	53
Mittelwert	9,91	9,72	9,72
Standardabweichung	7,22	6,09	3,01
Momentkoeffizient Schiefe	1,41	1,01	0,62
Momentkoeffizient Wölbung	2,92	1,08	0,38
Median	8,37	8,72	9,30
iqr-Bereich	9,04	8,09	4,04
iqr-Standardabweichung	6,70	6,00	2,99

<b>NO<sub>x</sub> Sonntag Sommer</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	2.913	372	62
Mittelwert	7,61	7,44	7,44
Standardabweichung	6,18	5,24	2,95
Momentkoeffizient Schiefe	1,52	1,13	1,18
Momentkoeffizient Wölbung	3,43	1,34	1,50
Median	5,99	6,43	6,81
iqr-Bereich	7,95	6,98	3,24
iqr-Standardabweichung	5,90	5,17	2,40



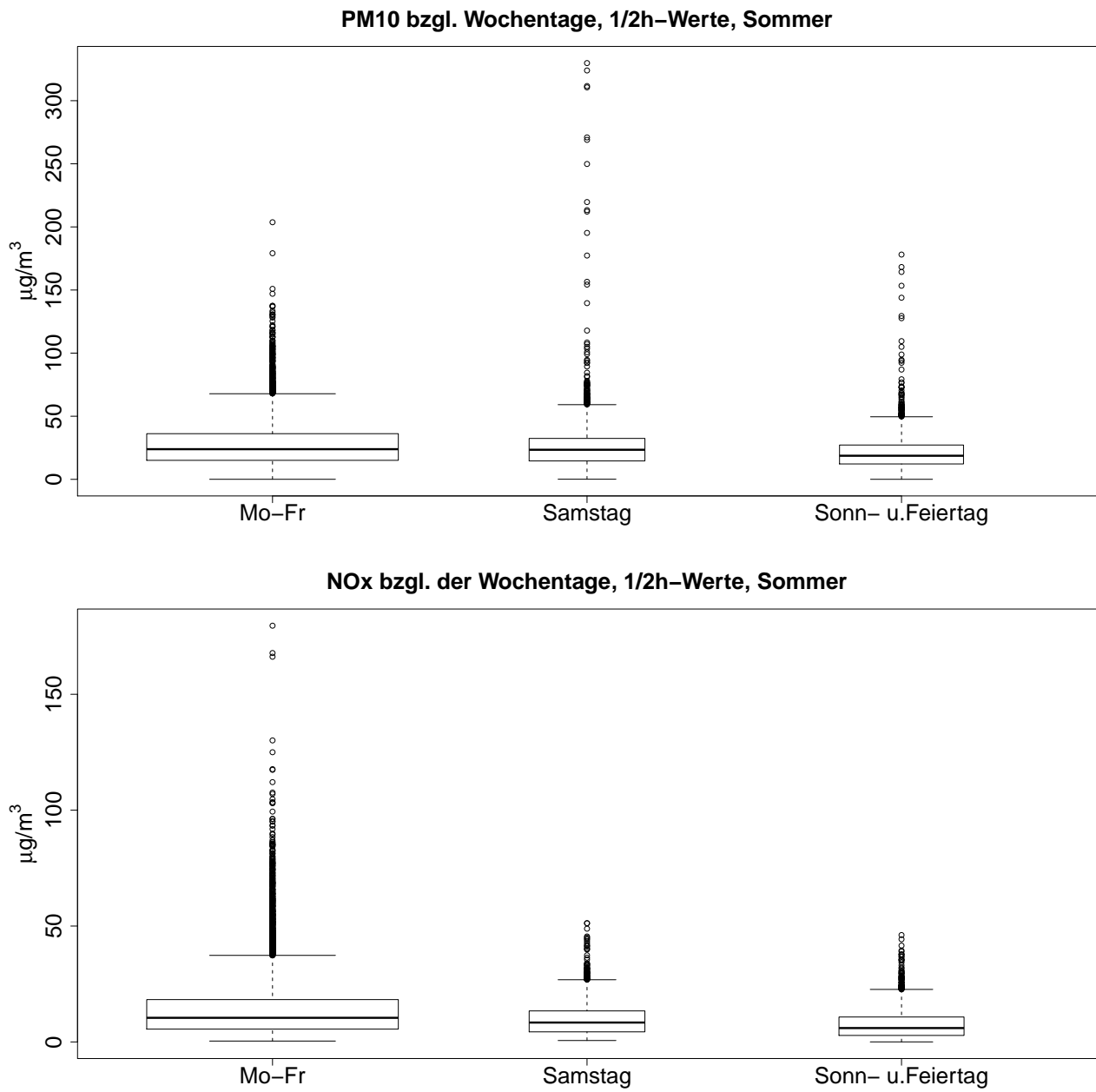
**Abbildung 16:** Dichteschätzer (Normalverteilung) von  $PM_{10}$  bzgl. Wochentage, Tagesmittelwerte

Bei den statistischen Kenngrößen für  $NO_x$  im Sommer erhält man relativ kleine Momentkoeffizienten der Schiefe und der Wölbung und daher annähernd symmetrische Verteilungen der Messpunkte, was auch der Dichteschätzer der Tagesmittelwerte in Abbildung 16 unterstreicht. Eine Ausnahme bildet hier die Kurtosis der Halbstundenwerte an den Werktagen.

Dadurch, dass die Schiefekoeffizienten alle positiv sind, sind die verschiedenen Verteilungen immer noch leicht rechtsschief.

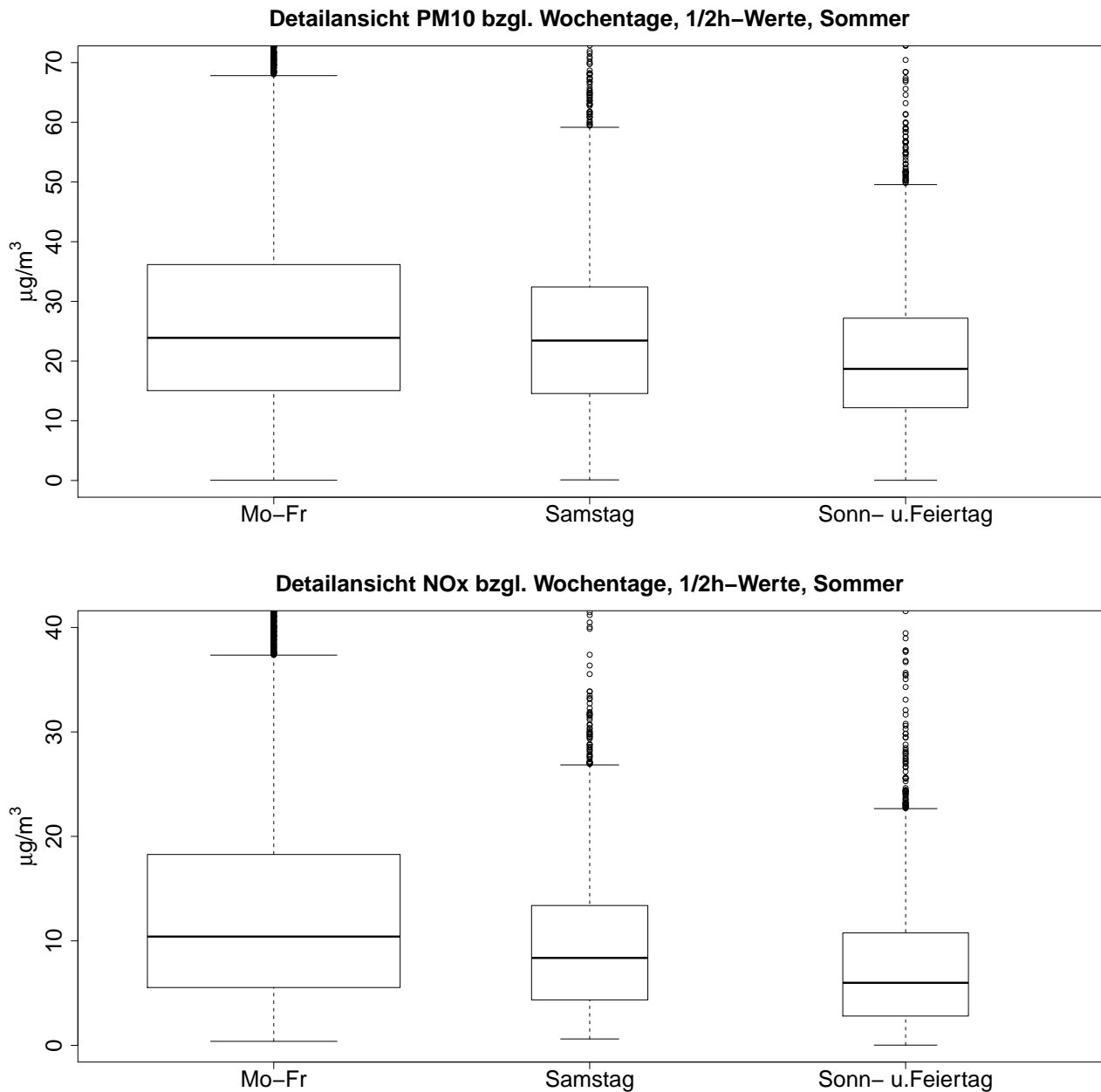
Die Mittelwerte der  $NO_x$  sind im Vergleich zu jenen von  $PM_{10}$  in allen drei Kategorien geringer und aufgrund der niedrigeren Wölbungskoeffizienten erkennt man Dichtefunktionen für  $NO_x$ , die in den Tails schwächer besetzt und damit „flacher“ sind als jene von  $PM_{10}$ .

Nun sollen die Boxplots der  $PM_{10}$  und  $NO_x$  bezüglich dem kategorischen Merkmal Sommer betrachtet werden.



**Abbildung 17:** Darstellung der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Sommer, Halbstundenwerte

Ein Entfernen der extremen Ausreißer und die dadurch entstehende Detailansicht des iqr-Bereichs ist in Abbildung 18 dargestellt:



**Abbildung 18:** Detailansicht der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Sommer, Halbstundenwerte

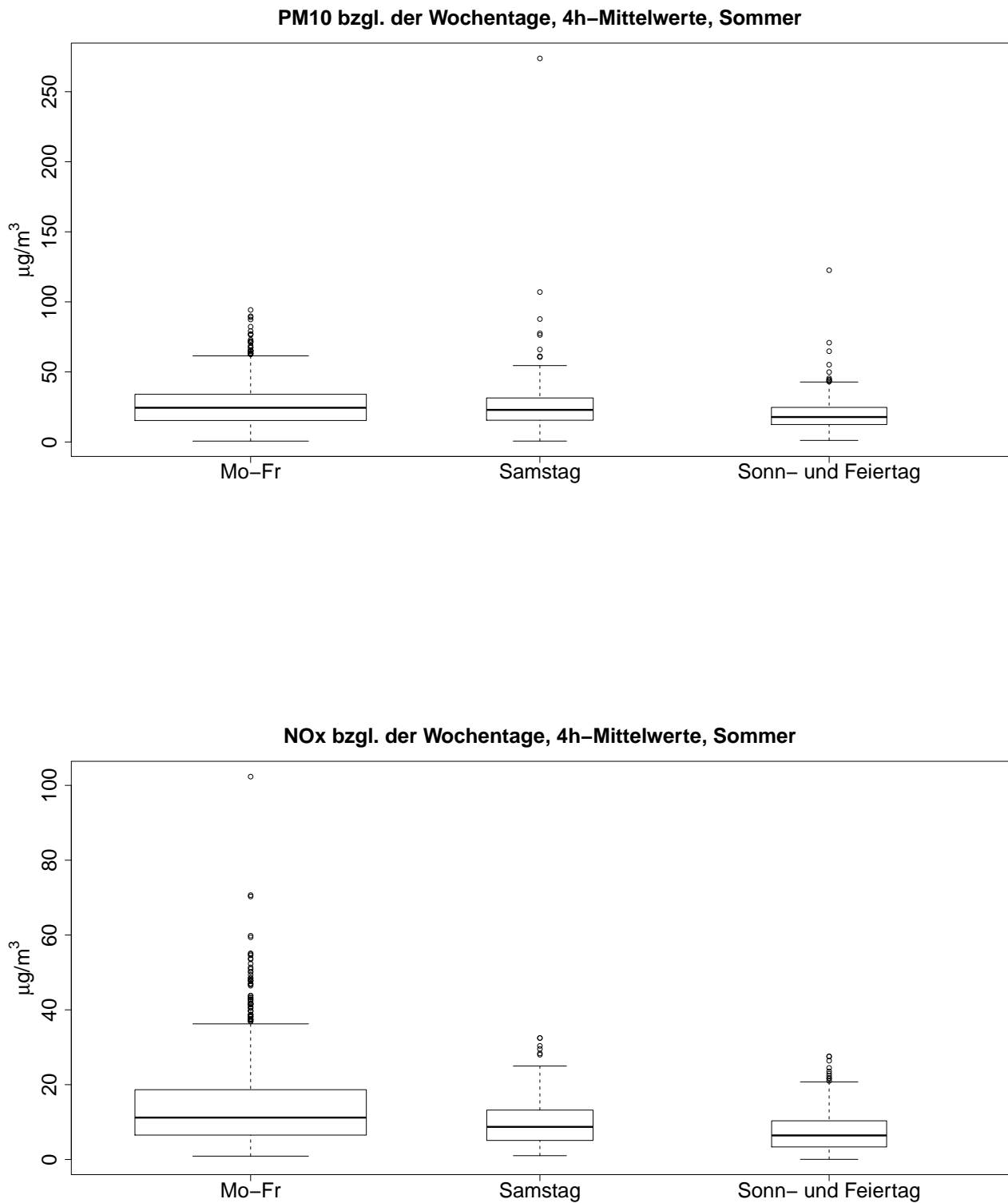
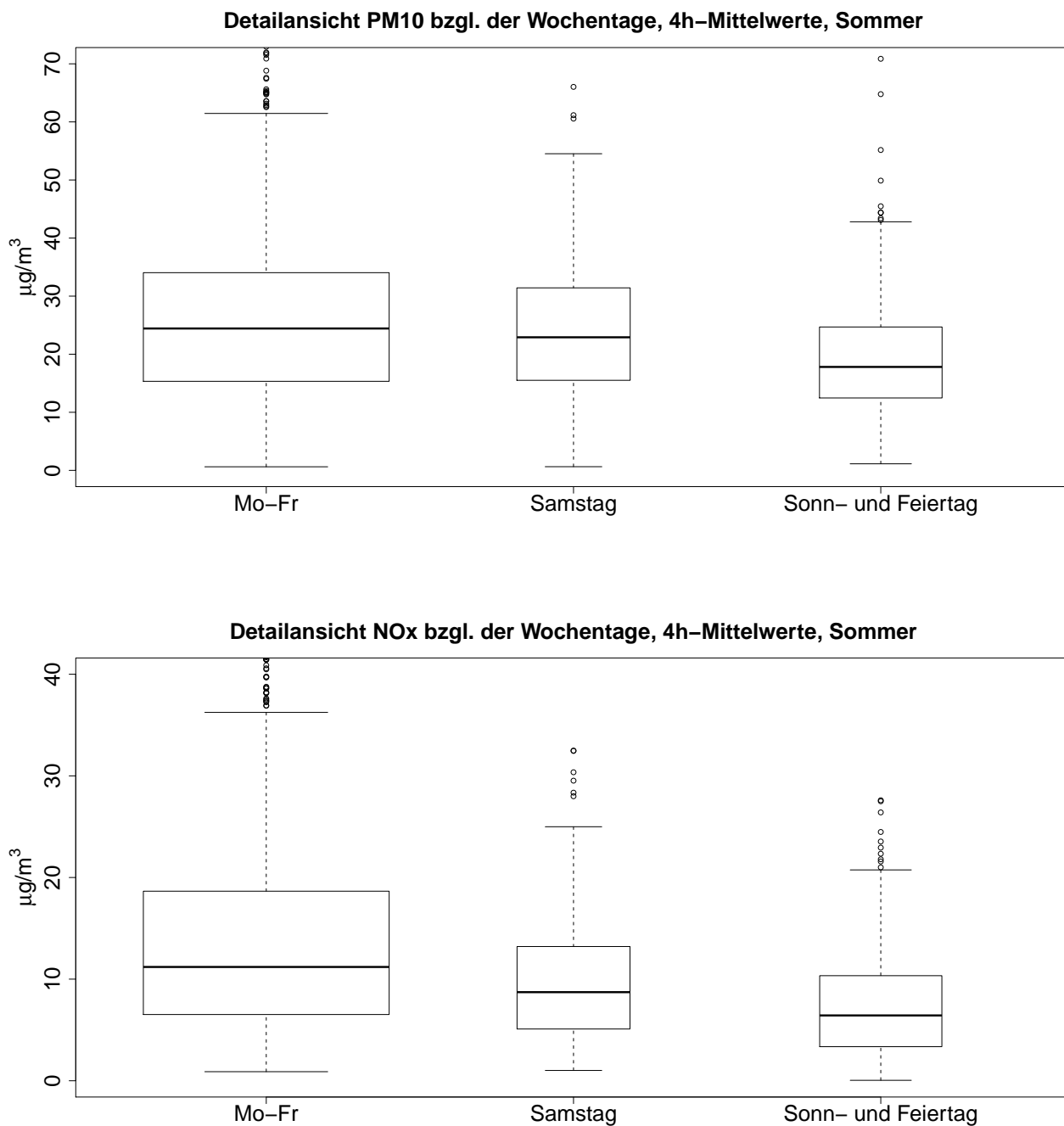


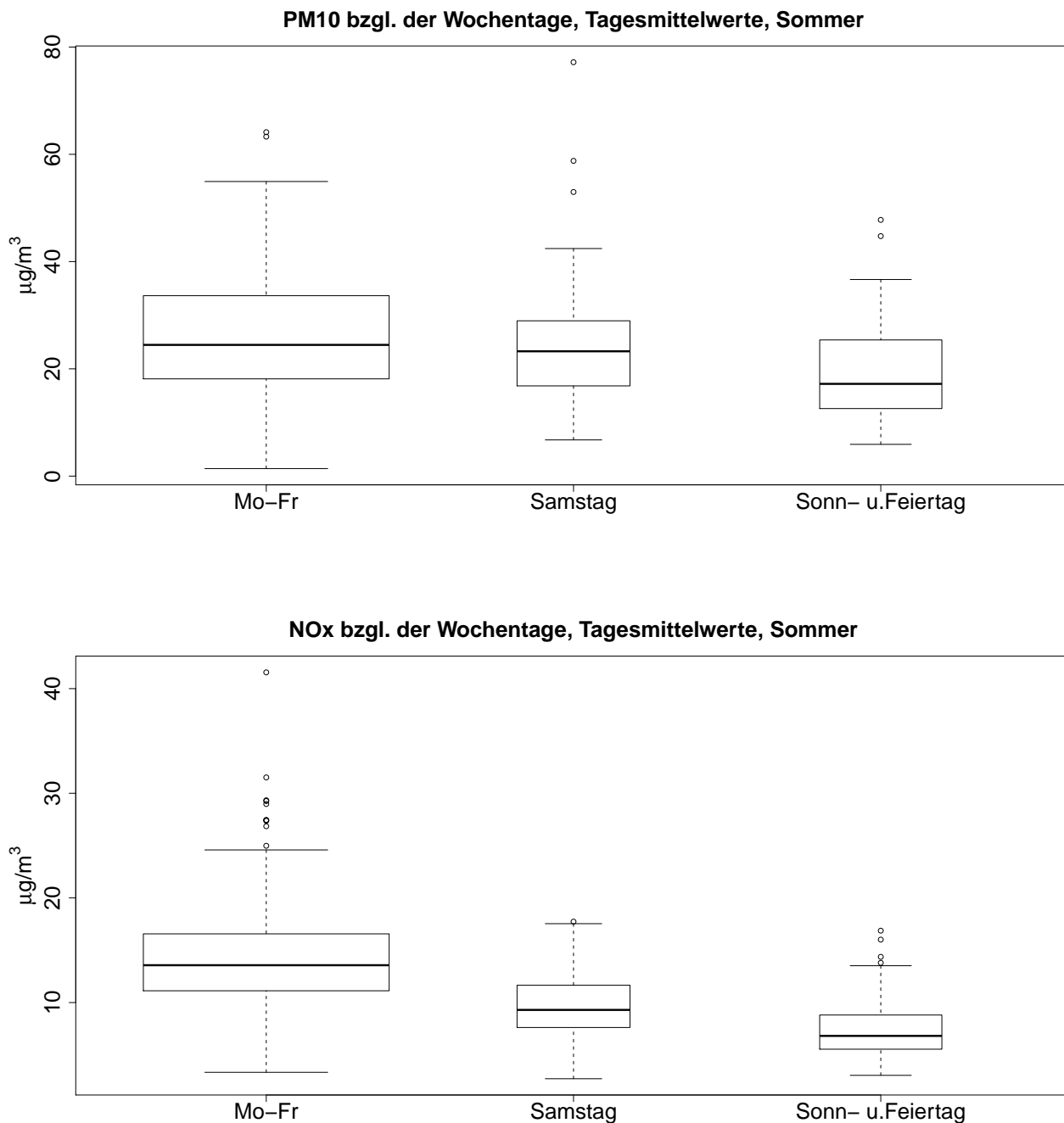
Abbildung 19: Darstellung der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Sommer, Vierstundenmittelwerte

Die beiden extremen Ausreißer von  $PM_{10}$  am Samstag beziehungsweise Sonn- und Feiertag entsprechen wieder dem 15.04.2006 beziehungsweise 16.04.2006, also dem Ostersonntag und Ostersonntag des Jahres 2006.

Auch bezüglich der Vierstundenmittelwerte wird für einen besseren Überblick über den iqr-Bereich extreme Ausreißer vernachlässigt.



**Abbildung 20:** Detailansicht der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Sommer, Vierstundenmittelwerte



**Abbildung 21:** Detailansicht der Boxplots PM<sub>10</sub> und NO<sub>x</sub> bzgl. Sommer, Tagesmittelwerte

Anhand der Boxplotserien in den Abbildungen 17 bis 21 sieht man, dass an den Werktagen mit Ausnahme einiger Ausreißer im Vergleich zu den beiden anderen Kategorien weitaus höhere Konzentration von PM<sub>10</sub> und NO<sub>x</sub> auftreten. Im allgemeinen nehmen die drei Reihen von PM<sub>10</sub> sowohl Werktags, als auch Samstags und Sonn- und Feiertags höhere Werte an als die von NO<sub>x</sub>.

Für die beiden Unterteilungen „Samstag“ und „Sonn- und Feiertag“ stellt man bei PM<sub>10</sub> extreme Ausreißer nach oben fest. Auch bei NO<sub>x</sub> gibt es sehr hohe Ausreißer, wobei sich diese vor allem an den Werktagen zeigen.

Betrachtet man die Vierstunden- und Tagesmittelwerte, kann man erneut feststellen, dass die Ausreißer mit zunehmendem Zeitunterschied zwischen den einzelnen Datenpunkten abnehmen. Der stärkste Ausreißer für die Sommerwerte von PM<sub>10</sub> am Samstag entspricht wieder dem Ostersonntag des Jahres 2006. Die anderen extremen Ausreißer von PM<sub>10</sub> und NO<sub>x</sub> besitzen kein signifikantes Datum, weshalb darauf auch nicht näher eingegangen wird.

Für die Wintermonate erhält man für PM<sub>10</sub> folgende statistische Kenngrößen:

<b>PM<sub>10</sub> Werktag Winter</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	11.650	1.482	247
Mittelwert	45,64	44,84	44,84
Standardabweichung	32,97	30,44	25,52
Momentkoeffizient Schiefe	1,42	1,35	1,49
Momentkoeffizient Wölbung	2,84	2,57	3,72
Median	38,28	37,71	40,31
iqr-Bereich	40,19	37,44	31,17
iqr-Standardabweichung	29,79	27,76	23,11

<b>PM<sub>10</sub> Samstag Winter</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	2.402	306	51
Mittelwert	40,72	39,95	39,95
Standardabweichung	35,72	33,59	26,00
Momentkoeffizient Schiefe	3,93	3,89	2,54
Momentkoeffizient Wölbung	26,29	25,75	9,60
Median	33,60	32,33	34,35
iqr-Bereich	29,03	27,03	23,45
iqr-Standardabweichung	21,52	20,04	17,38

<b>PM<sub>10</sub> Sonntag Winter</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	3.077	396	66
Mittelwert	37,72	36,63	36,63
Standardabweichung	32,44	30,37	26,90
Momentkoeffizient Schiefe	3,23	2,79	2,62
Momentkoeffizient Wölbung	18,43	10,61	8,26
Median	30,17	29,48	29,69
iqr-Bereich	26,47	24,42	19,65
iqr-Standardabweichung	19,62	18,10	14,57



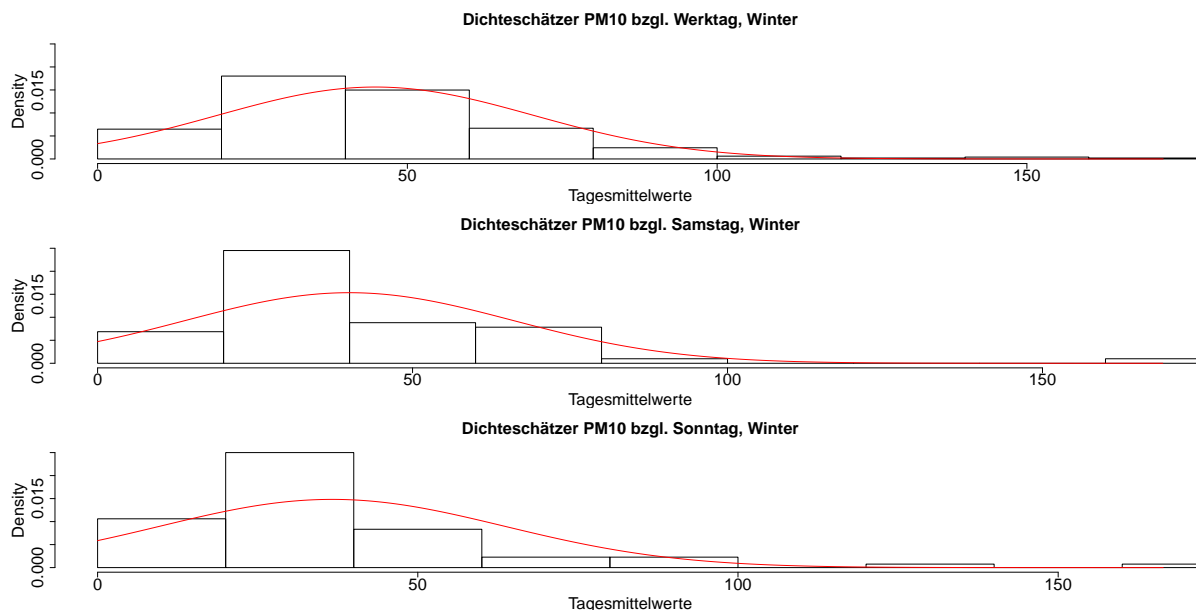


Abbildung 22: Dichteschätzer (Normalverteilung) von PM<sub>10</sub> bzgl. Wochentage, Tagesmittelwerte

Man bemerkt, dass der Mittelwert und die Standardabweichung für die Reihen des Schadstoffs PM<sub>10</sub> im Winter signifikant höher sind als die vergleichbaren Kenngrößen im Sommer. Das deutet wieder darauf hin, dass die Messwerte im Winter allgemein höher sind als im Sommer. Die Wölbung zeigt erneut, dass es sich um „spitze“ Verteilungen handelt, da die Kurtosis in allen Kategorien positiv ist. Weiters sieht man, dass die Verteilungen der einzelnen Reihen in allen Unterteilungen wieder eine leichte Rechtsschiefe aufzeigen.

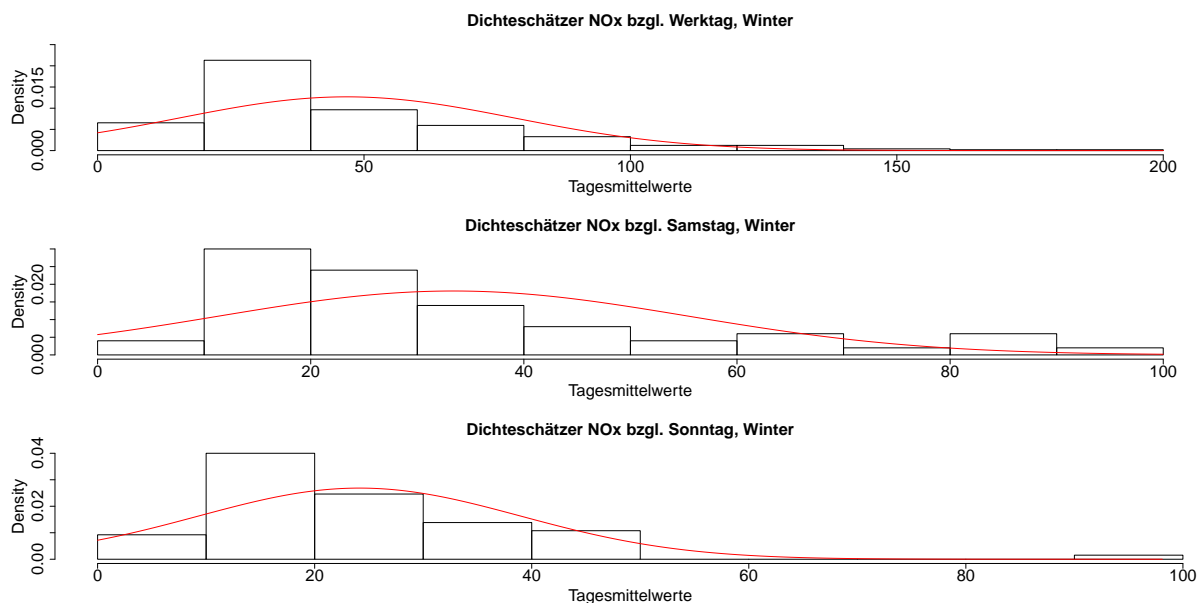
Dies sieht man auch sehr gut anhand der Dichteschätzer für die Tagesmittelwerte.

Für die Wintermonate erhält man für NO<sub>x</sub> folgende statistische Kenngrößen:

NO <sub>x</sub> Werktag Winter	1/2h-Werte	4h-MW	TMW
Anzahl Werte	11.385	1.459	244
Mittelwert	48,12	46,94	46,78
Standardabweichung	45,92	40,61	31,46
Momentkoeffizient Schiefe	2,12	1,95	1,46
Momentkoeffizient Wölbung	6,22	5,31	2,32
Median	34,00	35,87	36,43
iqr-Bereich	44,02	38,67	34,47
iqr-Standardabweichung	32,63	28,67	25,55

<b>NO<sub>x</sub> Samstag Winter</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	2.351	300	50
Mittelwert	34,09	33,39	33,39
Standardabweichung	29,09	26,12	22,05
Momentkoeffizient Schiefe	1,92	1,54	1,18
Momentkoeffizient Wölbung	4,71	2,12	0,40
Median	26,01	25,19	25,98
iqr-Bereich	29,06	27,60	22,32
iqr-Standardabweichung	21,54	20,46	16,55

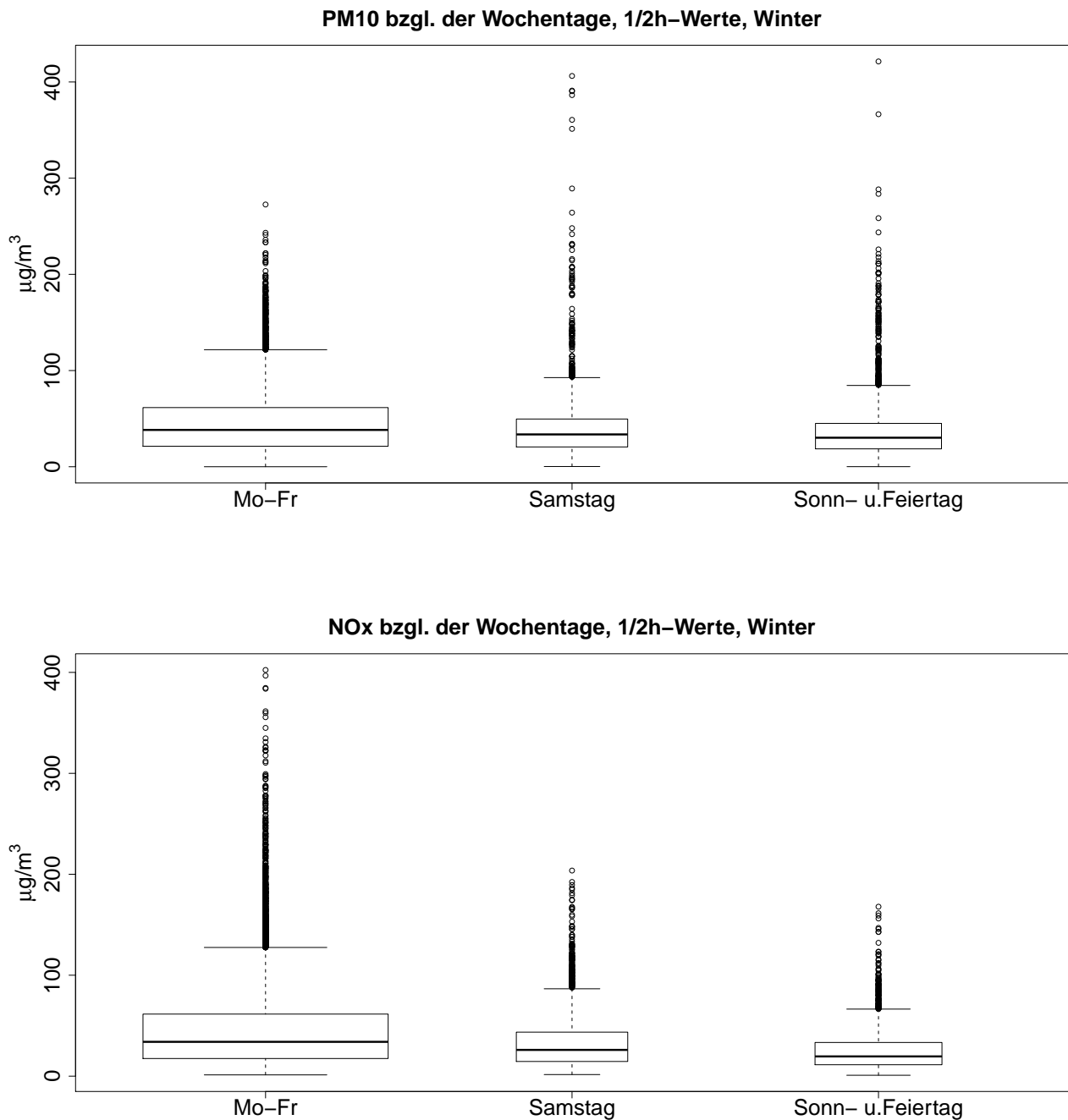
<b>NO<sub>x</sub> Sonntag Winter</b>	1/2h-Werte	4h-MW	TMW
Anzahl Werte	3.052	390	65
Mittelwert	24,67	24,13	24,13
Standardabweichung	20,01	17,98	14,85
Momentkoeffizient Schiefe	2,06	1,93	2,09
Momentkoeffizient Wölbung	6,90	6,86	7,37
Median	19,56	19,10	20,82
iqr-Bereich	22,05	20,50	16,26
iqr-Standardabweichung	16,35	15,20	12,05



**Abbildung 23:** Dichteschätzers (Normalverteilung) von NO<sub>x</sub> bzgl. Wochentage, Tagesmittelwerte

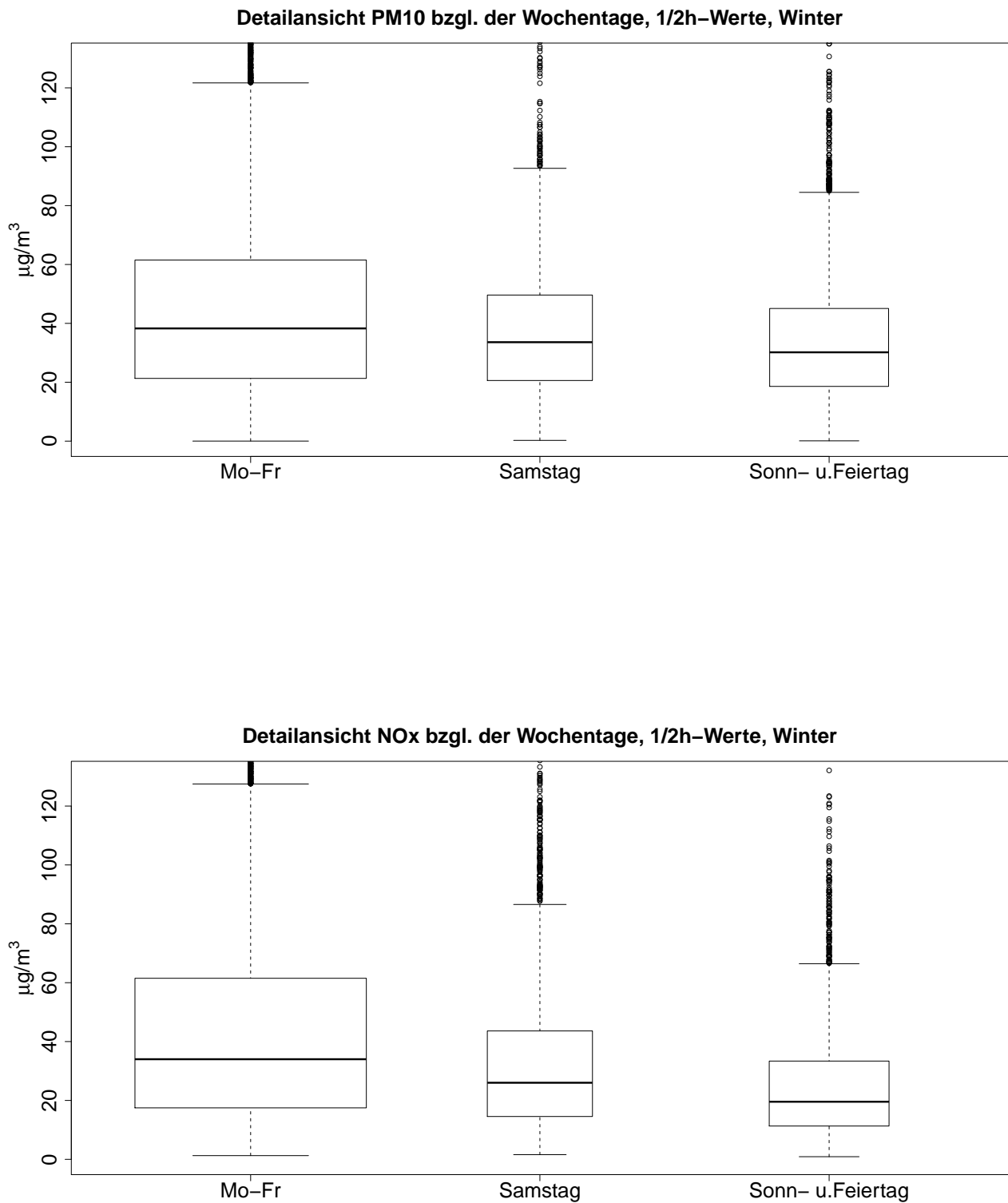
Auch für die Reihen von NO<sub>x</sub> lassen sich erneut spitze und rechtsschiefe Verteilungen feststellen. Man kann auch hier anhand der statistischen Kenngrößen sagen, dass die Konzentrationen von NO<sub>x</sub> im Winter jene im Sommer übersteigen.

Weiters sollen nun die Boxplotserien für das kategoriale Merkmal Winter erzeugt werden. Für



**Abbildung 24:** Darstellung der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Winter, Halbstundenwerte

eine bessere Betrachtung des iqr-Bereichs werden nun wieder die in Abbildung 24 auftretenden extremen Ausreißer eliminiert und die so entstehende Detailansicht wird in Abbildung 25 dargestellt.



**Abbildung 25:** Detailansicht der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Winter, Halbstundenwerte

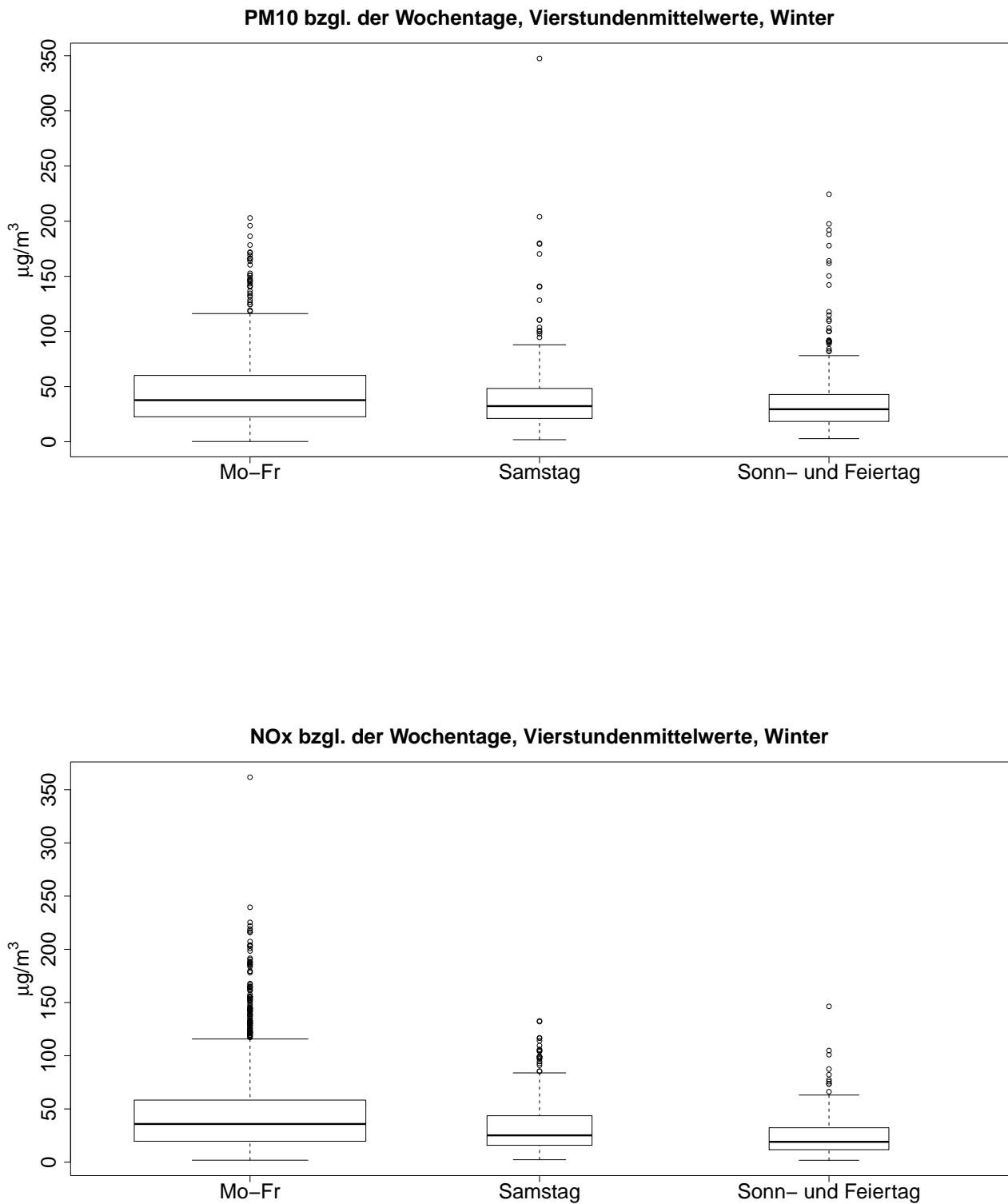


Abbildung 26: Darstellung der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Winter, Vierstundenmittelwerte

Der hohe Ausreißer der Vierstundenwerte für NO<sub>x</sub> im Sommer für die Kategorie *Samstag* entspricht hier wieder einem Ostersonntag, dem 26.03.2005.

Auch für die Vierstundenmittelwerte in Abbildung 26 muß der Wertebereich zur Veranschaulichung der Verteilung innerhalb der zentralen 50% eingegrenzt werden. Dies wird dann in Abbildung 27 geplottet.

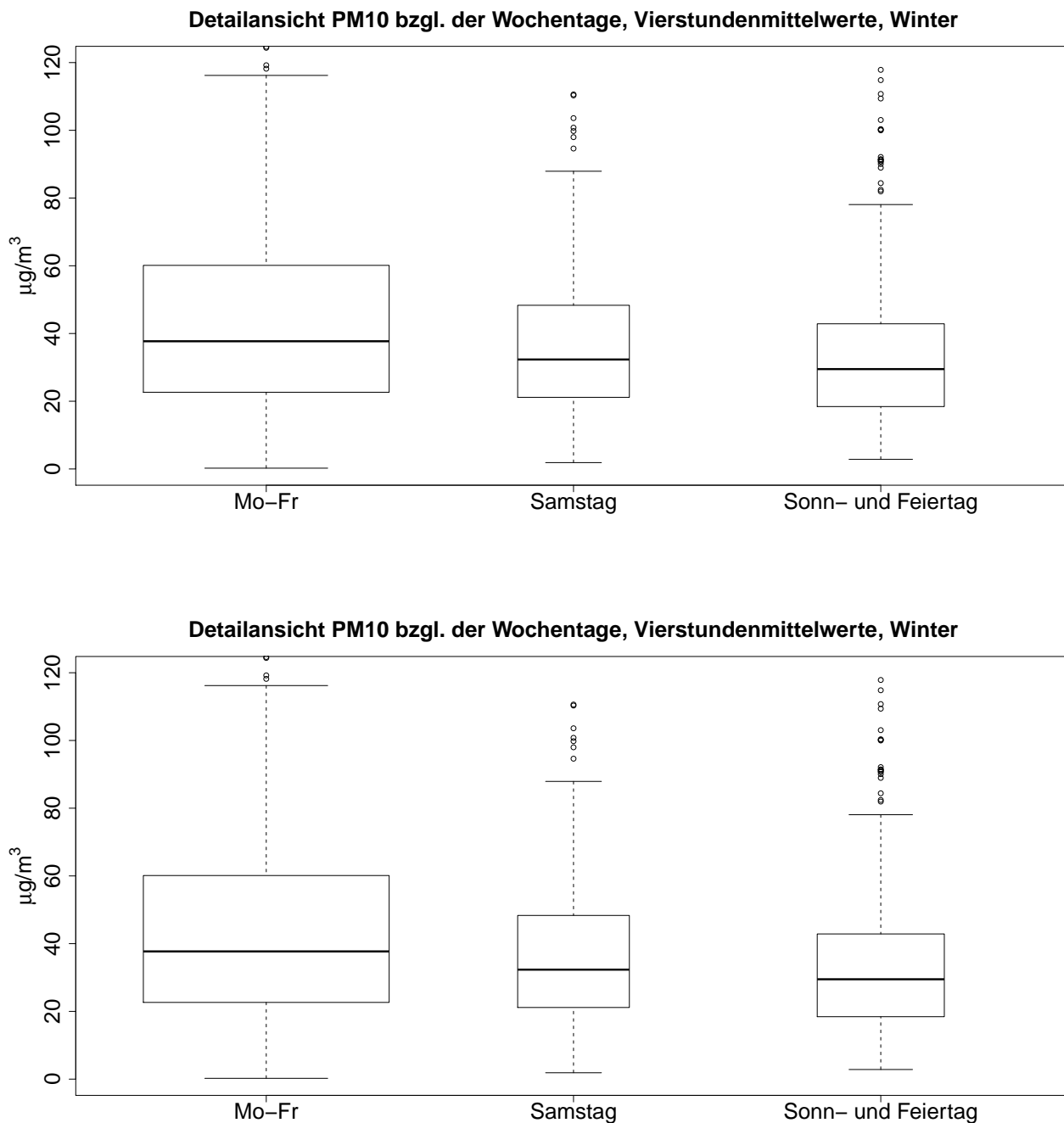
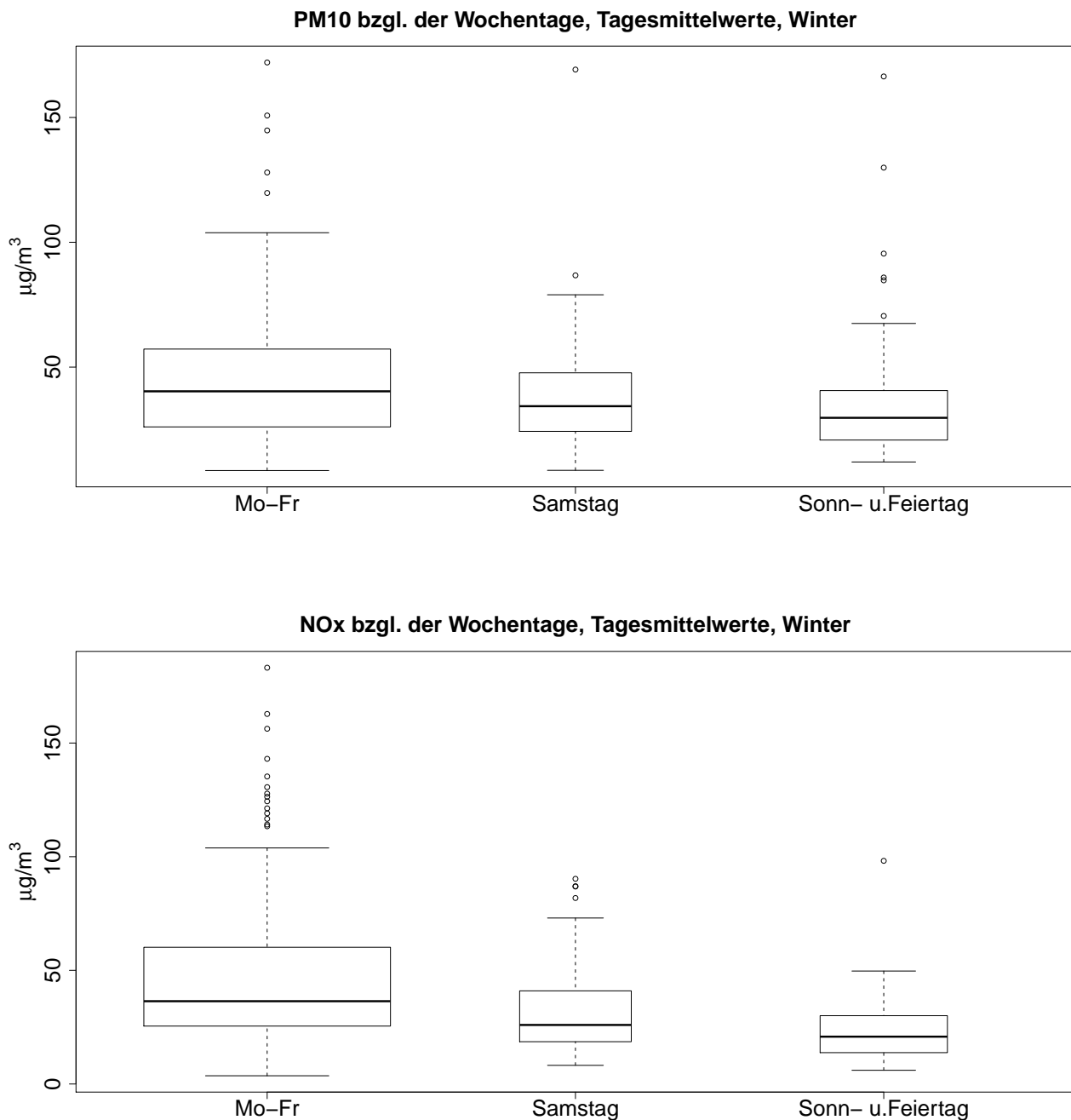


Abbildung 27: Detailansicht der Boxplots PM<sub>10</sub> und NO<sub>x</sub> bzgl. Winter, Vierstundenmittelwerte



**Abbildung 28:** Detailansicht der Boxplots  $PM_{10}$  und  $NO_x$  bzgl. Winter, Tagesmittelwerte

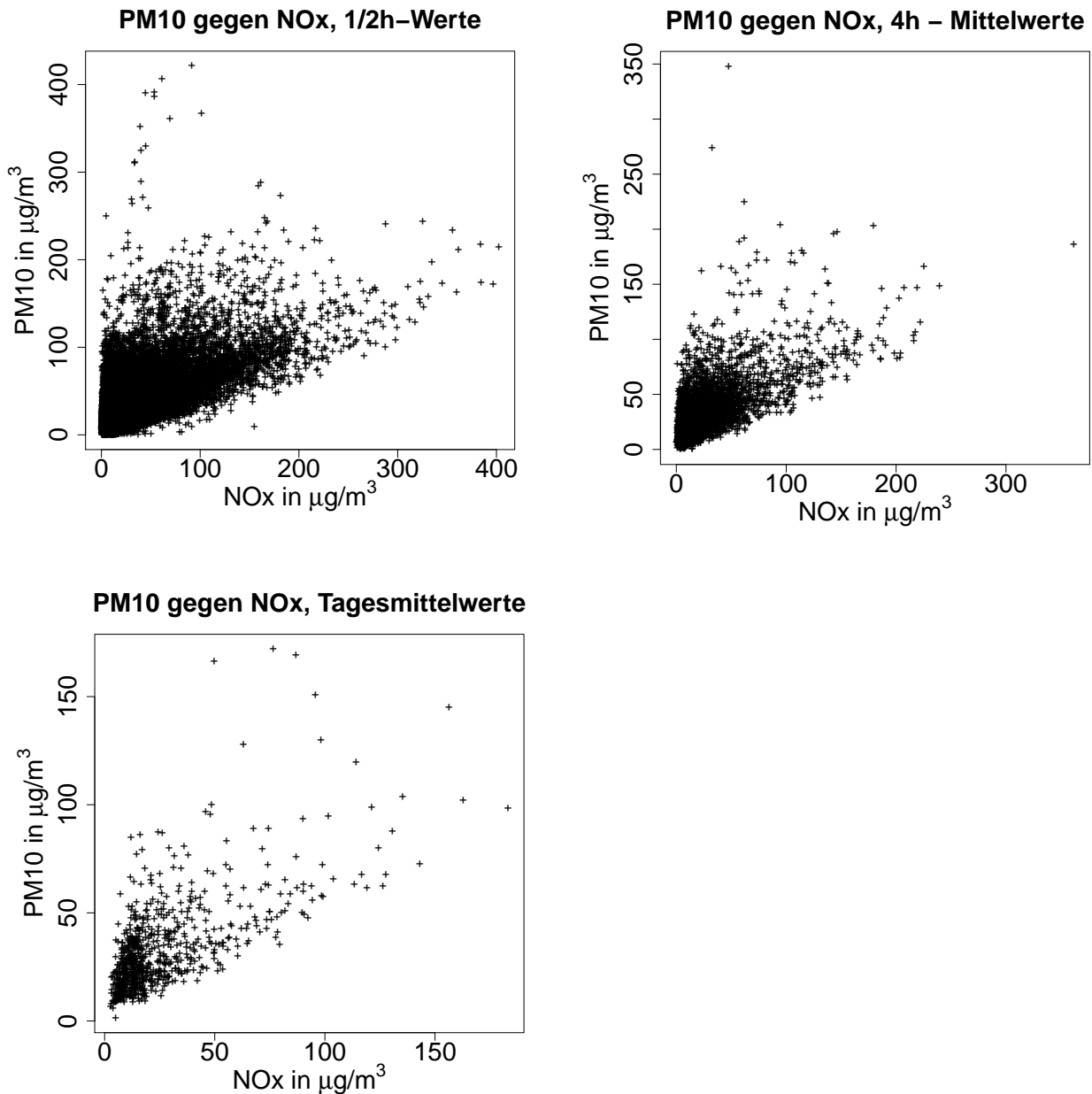
Auch hier werden bereits bekannte Ergebnisse erneut zum Ausdruck gebracht. Die stärksten Ausprägungen sind sowohl bei  $PM_{10}$  als auch bei  $NO_x$  mit Ausnahme einiger Ausreißer weiterhin an den Werktagen festzustellen. Die bereits besprochene Rechtsschiefe der Verteilungen der Daten innerhalb der Box der zentralen 50% wird anhand der oben gezeigten Boxplotserien noch verdeutlicht. Der  $NO_x$  - Ausreißer an Sonn- und Feiertagen im Winter für die Tagesmittelwerte entspricht dem Silvestertag 2006.

Geht man erneut von den Halbstundenwerten über die Vierstunden- zu den Tagesmittelwerten

über, sieht man, dass immer mehr Ausreißer geglättet werden. Weiters kann man auch erkennen, dass die Verteilungen der Tagesmittelwerte den kleinsten Grad an Unsymmetrie vorweisen.

### 2.3 Zusammenhangsanalyse

Im folgenden sollen die Zusammenhänge der einzelnen Reihen der beiden Schadstoffe anhand von Scatterplots und Korrelationskoeffizienten analysiert werden.



**Abbildung 29:** Darstellung der Scatterplots von  $PM_{10}$  und  $NO_x$  bzgl. 1/2h-Werte, 4h- und Tagesmittelwerte



Wenn man die Ausreißer vernachlässigt, ist in Abbildung 29 durchaus eine positive Korrelation zwischen PM<sub>10</sub> und NO<sub>x</sub> zu sehen. Ausreißer sind jedoch einige vorhanden.

Bei den Zeitreihen der Vierstunden- und Tagesmittelwerte ist die Korrelation sogar noch stärker sichtbar, da sich die Anzahl der Daten drastisch verringert hat. Die Streubreite wird ebenfalls geringer, da viele Ausreißer im Vergleich zu den Reihen der Halbstundenwerte geglättet wurden.

Dies lässt sich auch anhand des *Korrelationskoeffizienten von Pearson* feststellen (die Werte der folgenden Tabellen sind auf zwei Dezimalstellen gerundet):

		PM <sub>10</sub>		
		1/2h-Werte	4h-MW	TMW
NO <sub>x</sub>	1/2h-Werte	0,62		
	4h-MW	0,65		
	TMW	0,65		

Es lässt sich also bezüglich eines linearen Zusammenhangs zwischen PM<sub>10</sub> und NO<sub>x</sub> eine mittlere positive Korrelation feststellen. Man sieht, dass die Korrelation beim Übergang von den Halbstundenwerten über die Vierstunden- hin zu den Tagesmittelwerten immer etwas stärker wird.

### Kategorisches Merkmal: Sommer/Winter

Nun sollen die Daten wieder aufgeteilt werden in die kategorischen Merkmale Sommer beziehungsweise Winter, dargestellt jeweils mit Halbstundenwerten, Vierstundenmittel- und Tagesmittelwerten:

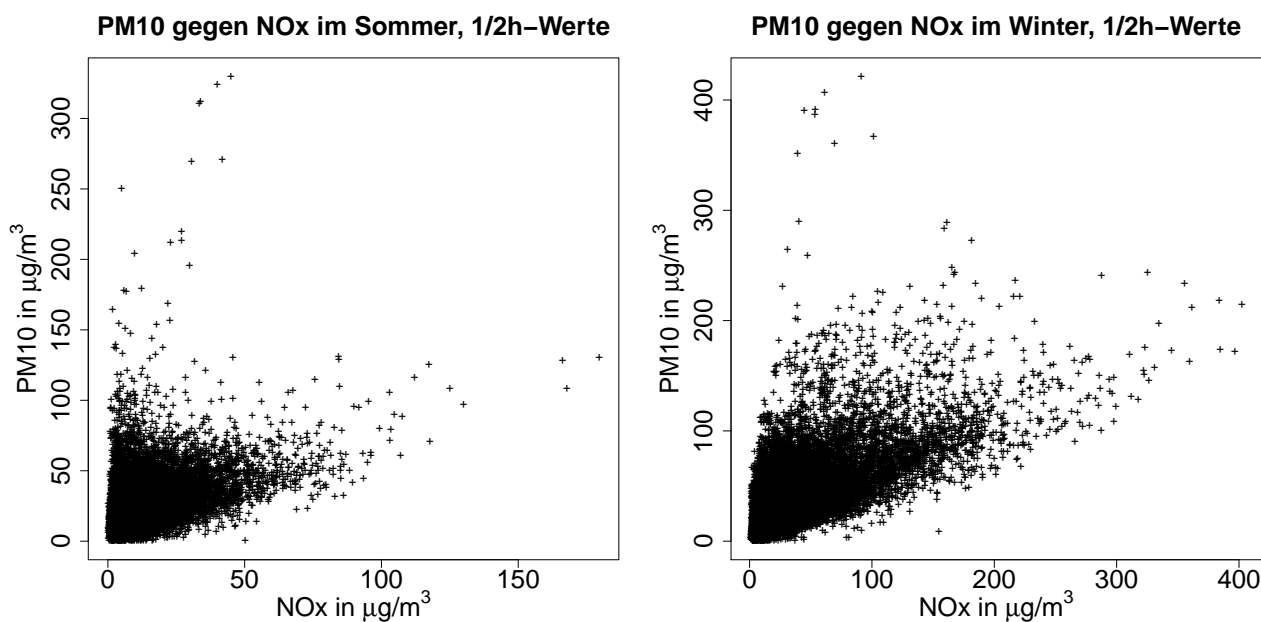


Abbildung 30: Darstellung der Scatterplots von PM<sub>10</sub> gegen NO<sub>x</sub> bezüglich Halbstundenwerten

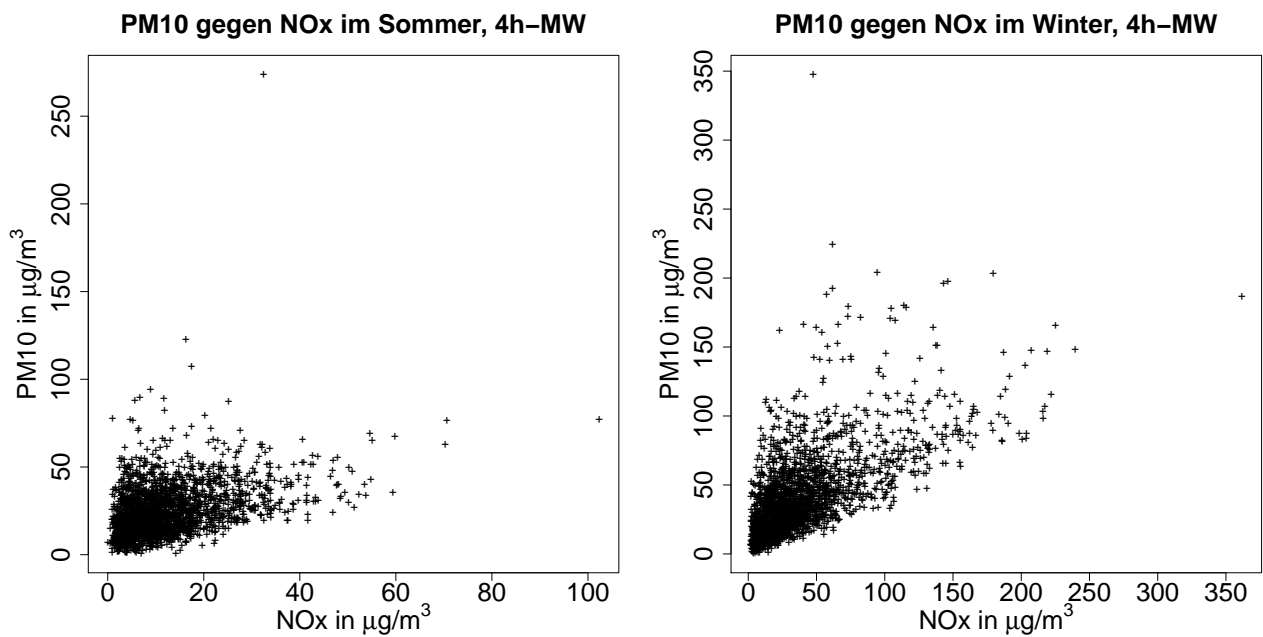


Abbildung 31: Darstellung der Scatterplots von  $PM_{10}$  gegen  $NO_x$  bezüglich Vierstundenmittelwerten

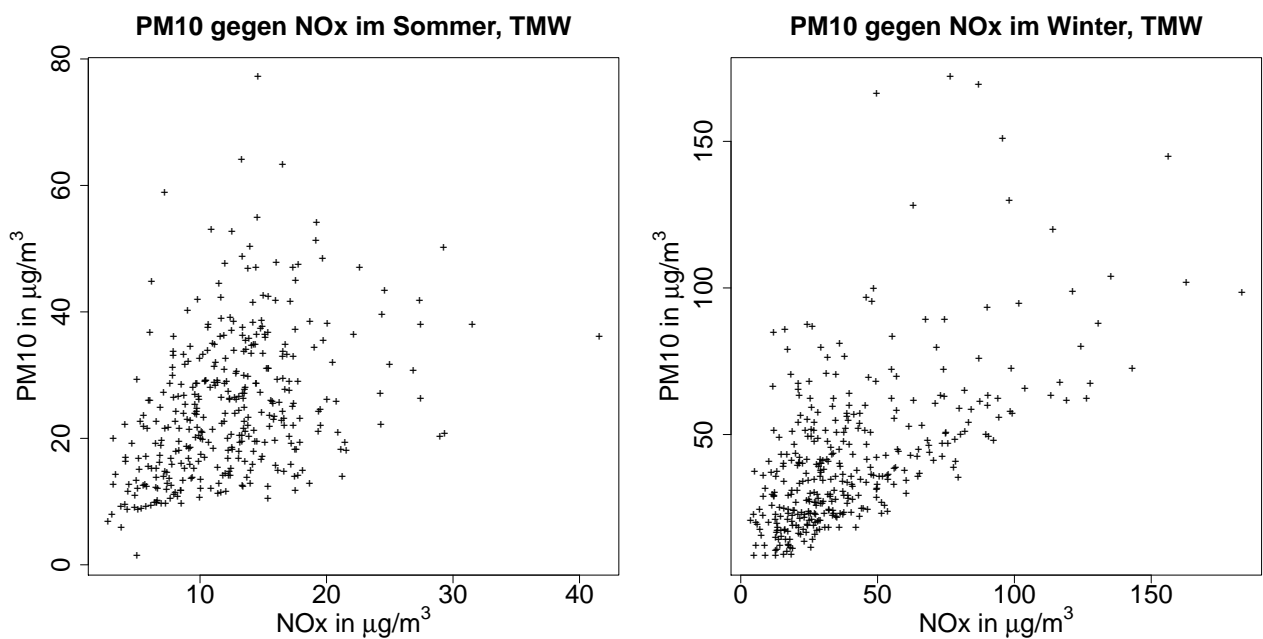


Abbildung 32: Darstellung der Scatterplots von  $PM_{10}$  gegen  $NO_x$  bezüglich Tagesmittelwerten

Bezüglich **Sommer** erkennt man wieder eine schwach positive Korrelation zwischen den drei Merkmalen von PM<sub>10</sub> und NO<sub>x</sub>, was auch die Korrelationskoeffizienten von Pearson zeigen (auf zwei Dezimalstellen gerundet):

<b>Sommer</b>		<b>PM<sub>10</sub></b>		
		1/2h-Werte	4h-MW	TMW
<b>NO<sub>x</sub></b>	1/2h-Werte	0,38		
	4h-MW		0,38	
	TMW			0,39

Für das kategorische Merkmal **Winter** erkennt man in den Abbildungen 30 bis 32 ebenfalls eine positive, jedoch viel höhere Korrelation als in den Sommermonaten.

Die Korrelationskoeffizienten von Pearson liefern folgende Werte:

<b>Winter</b>		<b>PM<sub>10</sub></b>		
		1/2h-Werte	4h-MW	TMW
<b>NO<sub>x</sub></b>	1/2h-Werte	0,63		
	4h-MW		0,63	
	TMW			0,59

Die Korrelationskoeffizienten in der Tabelle zeigen, dass in der kalten Jahreszeit bereits mittlere Zusammenhänge zwischen den beiden Schadstoffen vorliegen.

Bei den Tagesmittelwerten ist er am schwächsten, was bedeutet, dass bezüglich der Vorhersage von PM<sub>10</sub> durch NO<sub>x</sub> (oder umgekehrt) im Winter die Halbstundenwerte beziehungsweise Vierstundenmittelwerte zu bevorzugen sind. Weiters bedeuten die stets positiven Zusammenhänge zwischen diesen beiden Schadstoffen, dass bei einer hohen Konzentration von PM<sub>10</sub> auch die NO<sub>x</sub> - Emissionen hoch sind.

**Kategorisches Merkmal: Wochentag**

Aufgrund der unterschiedlichen Luftbelastungen durch PM<sub>10</sub> und NO<sub>x</sub> in Sommer/Winter, muss auch bei der Betrachtung des kategorischen Merkmals „Wochentag“ zwischen Sommer und Winter unterschieden werden. Die Korrelationskoeffizienten von Pearson bezüglich Werktag, Samstag und Sonntag ergeben sich wie folgt:

**Sommer:**

Werktag		PM <sub>10</sub>		
		1/2h-Werte	4h-MW	TMW
NO <sub>x</sub>	1/2h-Werte	0,40		
	4h-MW		0,38	
	TMW			0,33

Samstag		PM <sub>10</sub>		
		1/2h-Werte	4h-MW	TMW
NO <sub>x</sub>	1/2h-Werte	0,33		
	4h-MW		0,35	
	TMW			0,41

Sonn- und Feiertag		PM <sub>10</sub>		
		1/2h-Werte	4h-MW	TMW
NO <sub>x</sub>	1/2h-Werte	0,31		
	4h-MW		0,40	
	TMW			0,55

Die Korrelationskoeffizienten zeigen für die Halbstundenwerte und die Vierstundenmittelwerte in allen Kategorien eine schwach positive Korrelation an. Bei den Reihen der Tagesmittelwerte ergibt sich Sonntags bereits ein mittlerer positiver Zusammenhang, während in den beiden übrigen Kategorien die Korrelationen ebenfalls schwach ausgeprägt sind.

**Winter:**

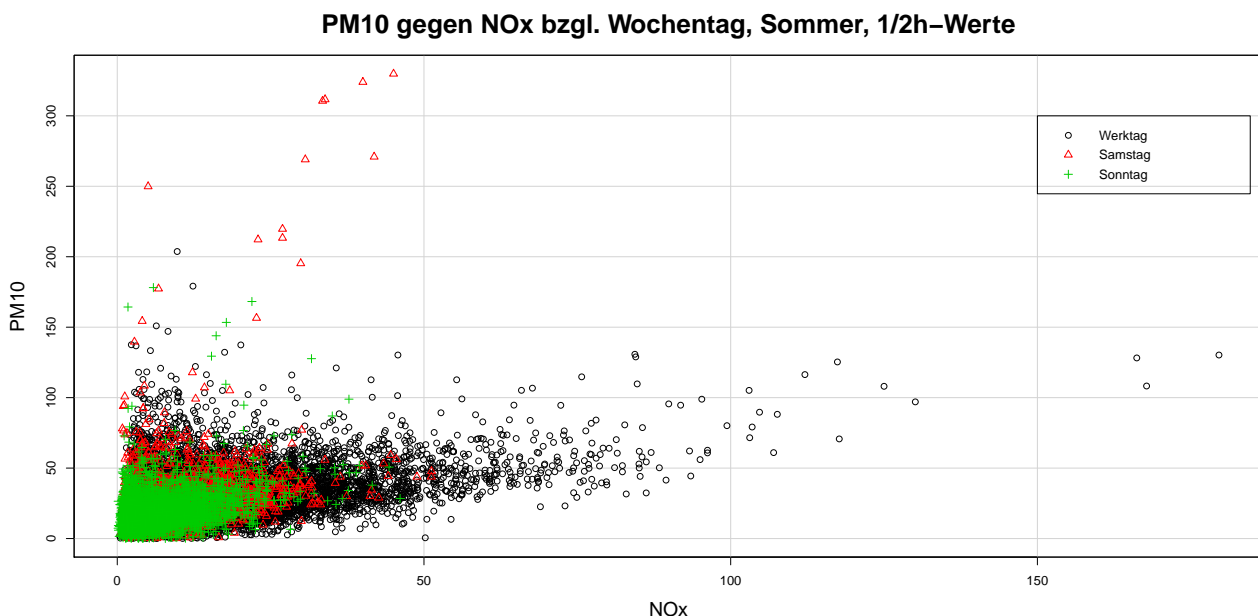
Werktag		PM <sub>10</sub>		
		1/2h-Werte	4h-MW	TMW
NO <sub>x</sub>	1/2h-Werte	0,68		
	4h-MW		0,66	
	TMW			0,62

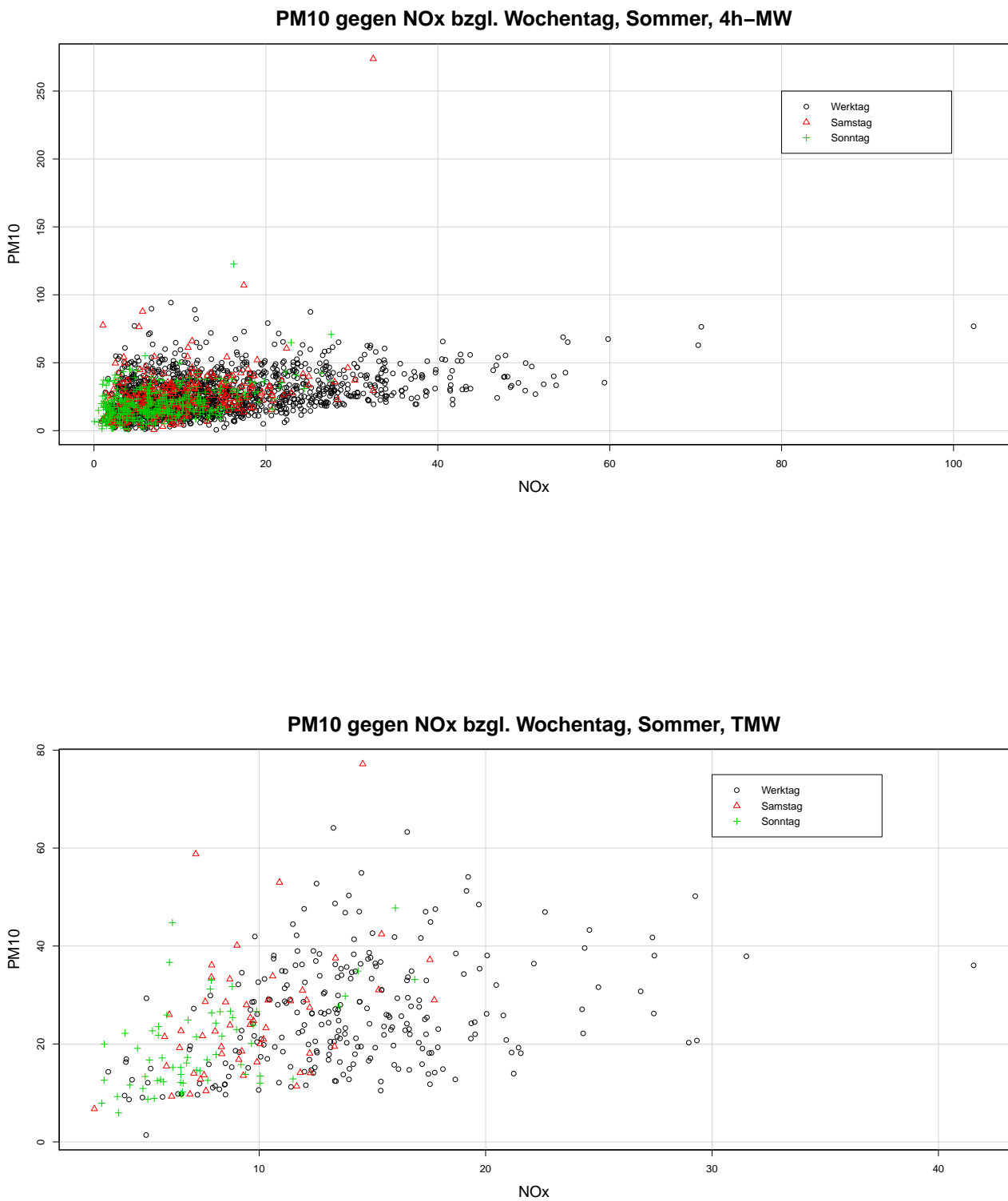
Samstag		$PM_{10}$		
		1/2h-Werte	4h-MW	TMW
$NO_x$	1/2h-Werte	0,58		
	4h-MW	0,58		
	TMW	0,60		

Sonn- und Feiertag		$PM_{10}$		
		1/2h-Werte	4h-MW	TMW
$NO_x$	1/2h-Werte	0,57		
	4h-MW	0,58		
	TMW	0,55		

Im Winter sind die beiden Schadstoffe in allen drei Kategorien bereits mittel positiv korreliert. An den Werktagen ist die Korrelation zwischen den einzelnen Reihen der beiden Zielgrößen am höchsten. Die Korrelation zwischen  $PM_{10}$  und  $NO_x$  ist also im Winter stärker als im Sommer. Die positiven Koeffizienten zeigen erneut einen direkten Zusammenhang: Sind die Werte von  $PM_{10}$  hoch, so auch jene von  $NO_x$  und umgekehrt.

Diese Ergebnisse kann man auch in den dazugehörigen Scatterplots in Abbildung 33 wiedererkennen.



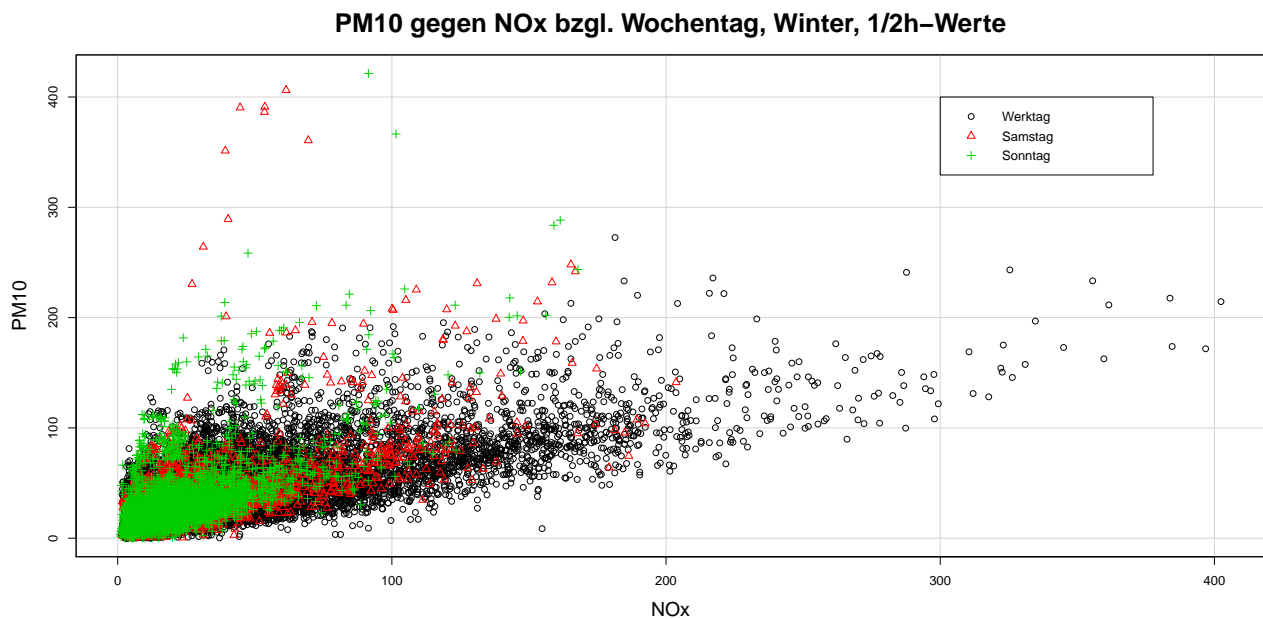


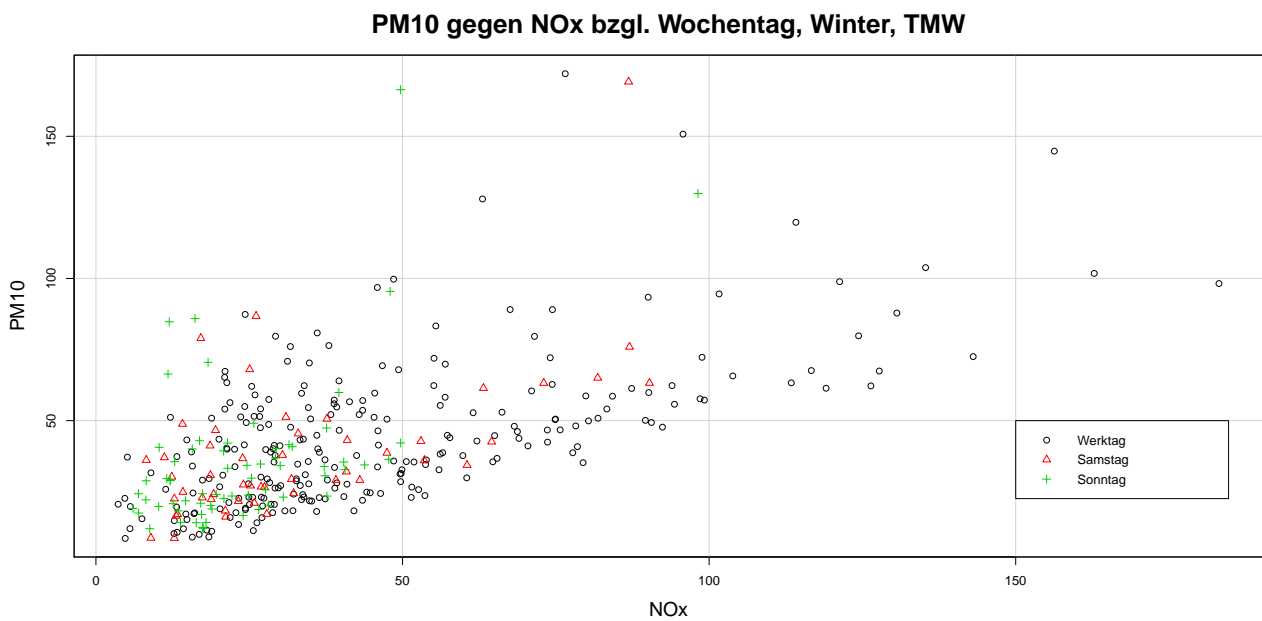
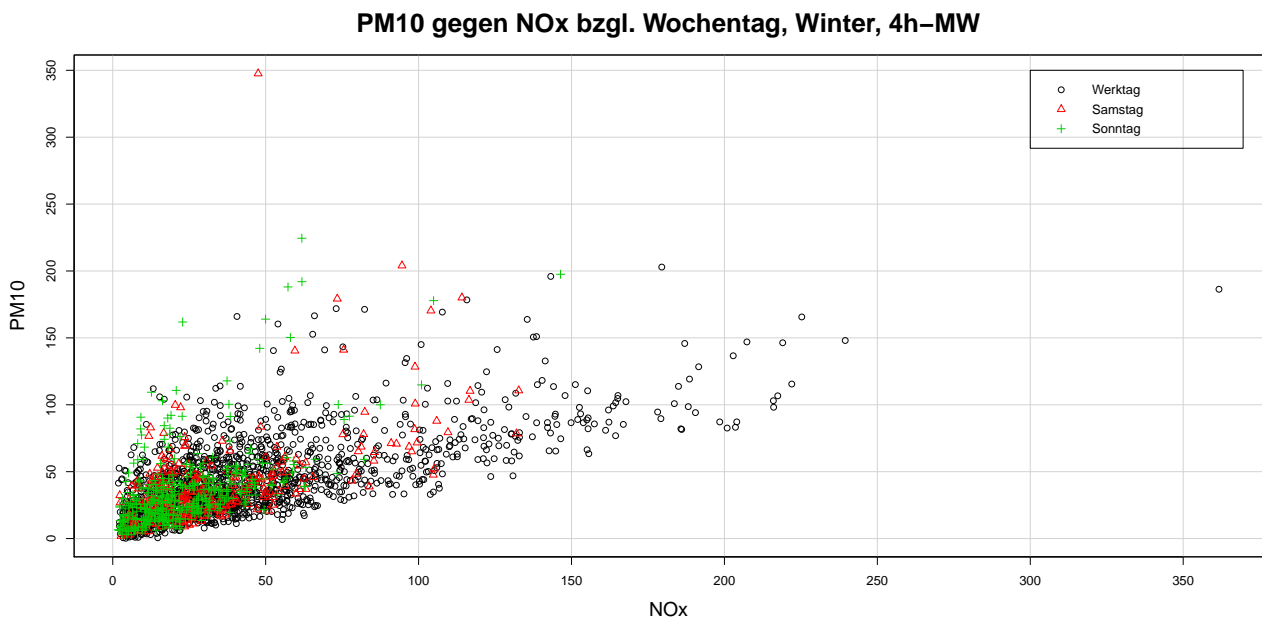
**Abbildung 33:** Darstellung der Scatterplots von  $PM_{10}$  und  $NO_x$  mit Halbstundenwerten, Vierstunden- und Tagesmittelwerte bzgl. Wochentage im Sommer

Die Scatterplotmatrizen zeigen noch einmal die zuvor erhaltenen Ergebnisse der Korrelationskoeffizienten: Aufgrund der geringeren Dichte der Daten ist bei den Tagesmittelwerten die mittlere Korrelation am besten zu erkennen.

Für die Messdaten der Wintermonate ergeben sich die Scatterplotmatrizen bezüglich dem kategorischen Merkmal Wochentag wie in Abbildung 34 dargestellt.

Die Scatterplotmatrizen der Messdaten im Winter zeigen die Verbesserung der Korrelation im Winter: Die Streubreite der Daten ist geringer als jene im Sommer.





**Abbildung 34:** Scatterplots von  $PM_{10}$  und  $NO_x$  mit Halbstundenwerten, Vierstunden- und Tagesmittelwerte bzgl. Wochentage im Winter



## 2.4 Untersuchung der Auto- und Kreuzkorrelationsstruktur

### Die Autokorrelationsfunktionen von PM<sub>10</sub> und NO<sub>x</sub>

Folgende Definition entstammt dem Buch „*Time Series Analysis and Its Applications (With R Examples)*“ [3].

#### Definition 11. Autokorrelationsfunktion

Die Autokorrelationsfunktion (ACF) ist definiert als

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}.$$

Der ACF misst die lineare Vorhersehbarkeit von  $x_t$  zu einem Zeitpunkt  $t$ , indem man nur den Wert  $x_s$  verwendet.

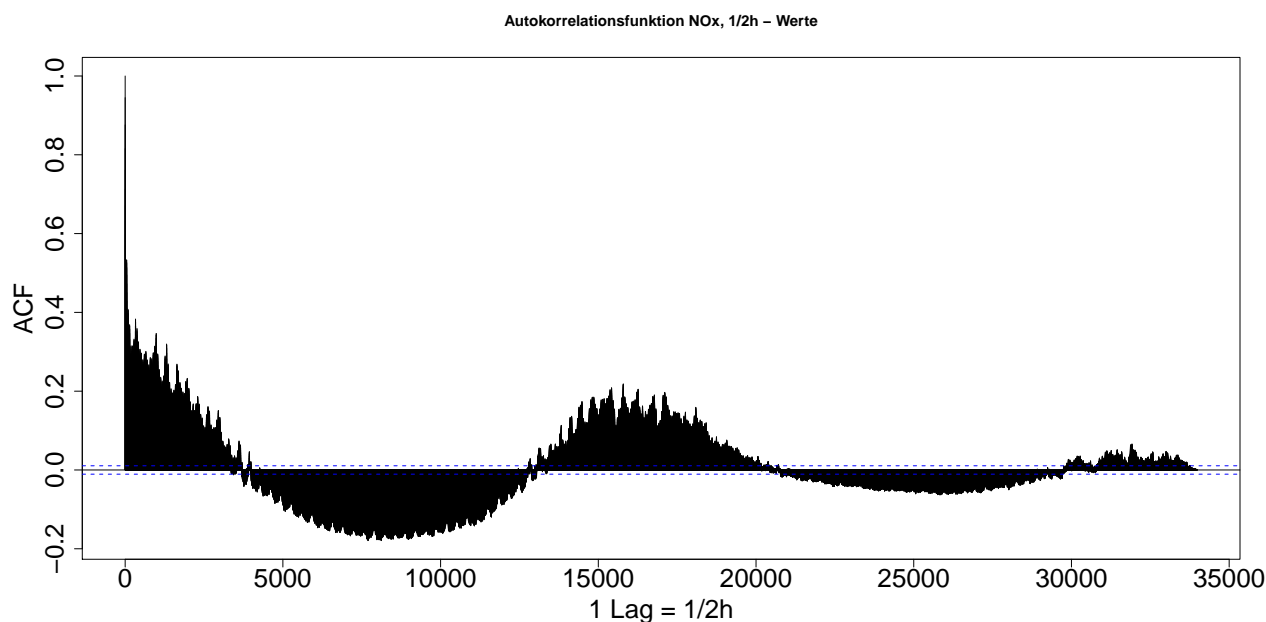
Der Wert  $\rho(s, t)$  liegt zwischen -1 und 1.

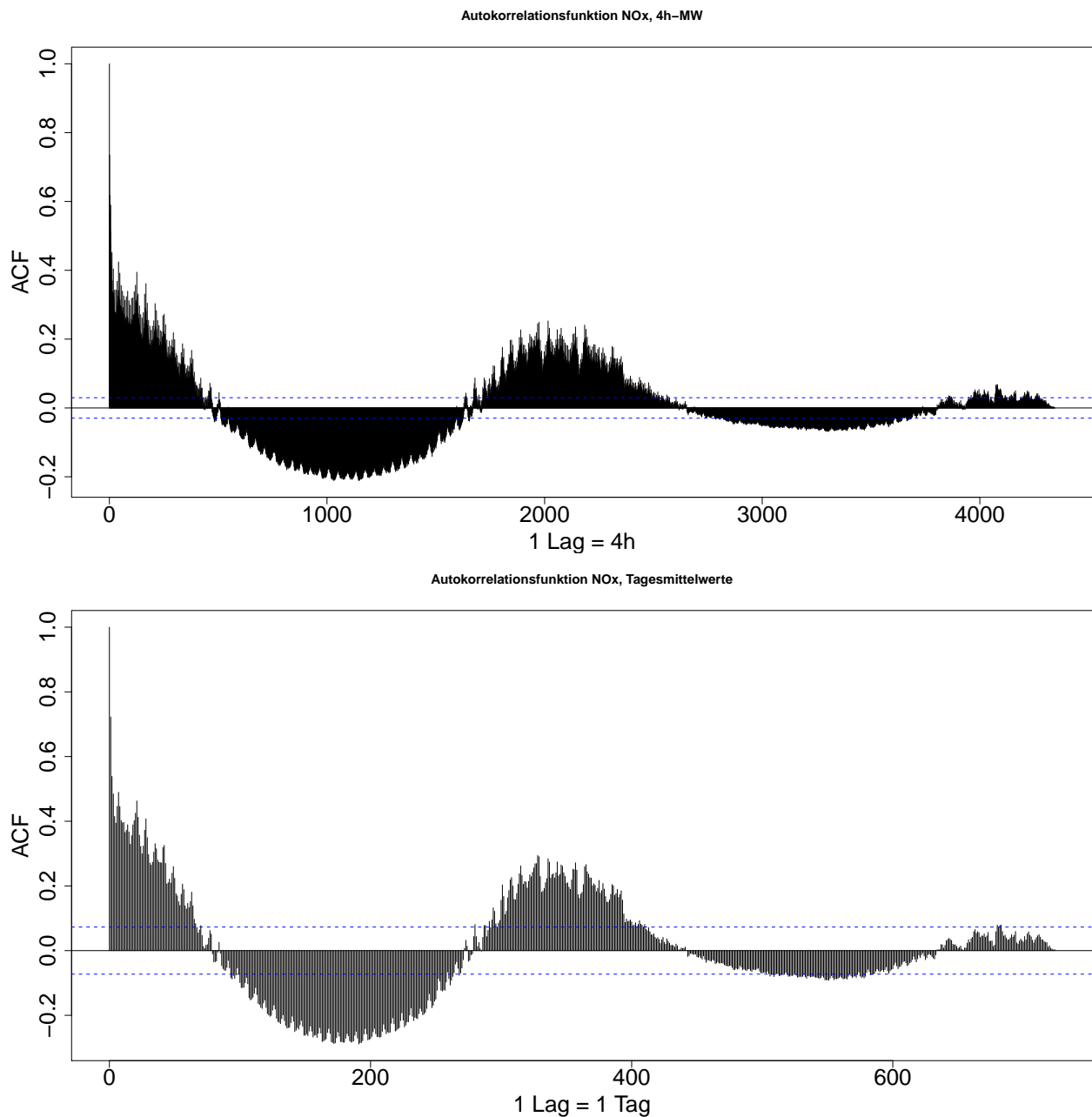
Es gilt:

$$-1 \leq \rho(s, t) \leq 1.$$

Es sollen nun die Autokorrelationsfunktionen der drei Zeitreihen von PM<sub>10</sub> und NO<sub>x</sub> analysiert werden.

In den folgenden Beispielen werden die theoretischen ACF - Werte durch deren Schätzung aus den Daten repräsentiert.





**Abbildung 35:** Darstellung der Autokorrelationsfunktionen von  $NO_x$  bzgl. 1/2h-Werte, Vierstunden- und Tagesmittelwerte

Die Autokorrelationsfunktionen in Abbildung 35 besitzen alle eine sinusförmige Schwingung. Es sind optisch kaum Unterschiede zwischen den Autokorrelationsfunktion der Halbstunden-, Vierstundenmittel- und Tagesmittelwerte erkennbar. Man kann erneut feststellen, dass die Korrelationen bei der Reihe der Tagesmittelwerte am stärksten ausgeprägt sind.

Bei der ACF der Halbstundenwerte erkennt man, dass Werte, die ca. 7.500 Datenpunkte entfernt sind (das entspricht einem Zeitraum von sechs Monaten), schwach negativ korreliert sind. Werte, die ca. 15.000 Datenpunkte voneinander entfernt sind (das entspricht demnach einem Zeitraum von einem Jahr), sind schwach positiv korreliert.

Dies bedeutet, dass Werte in den Sommermonaten mit den darauffolgenden Wintermonaten

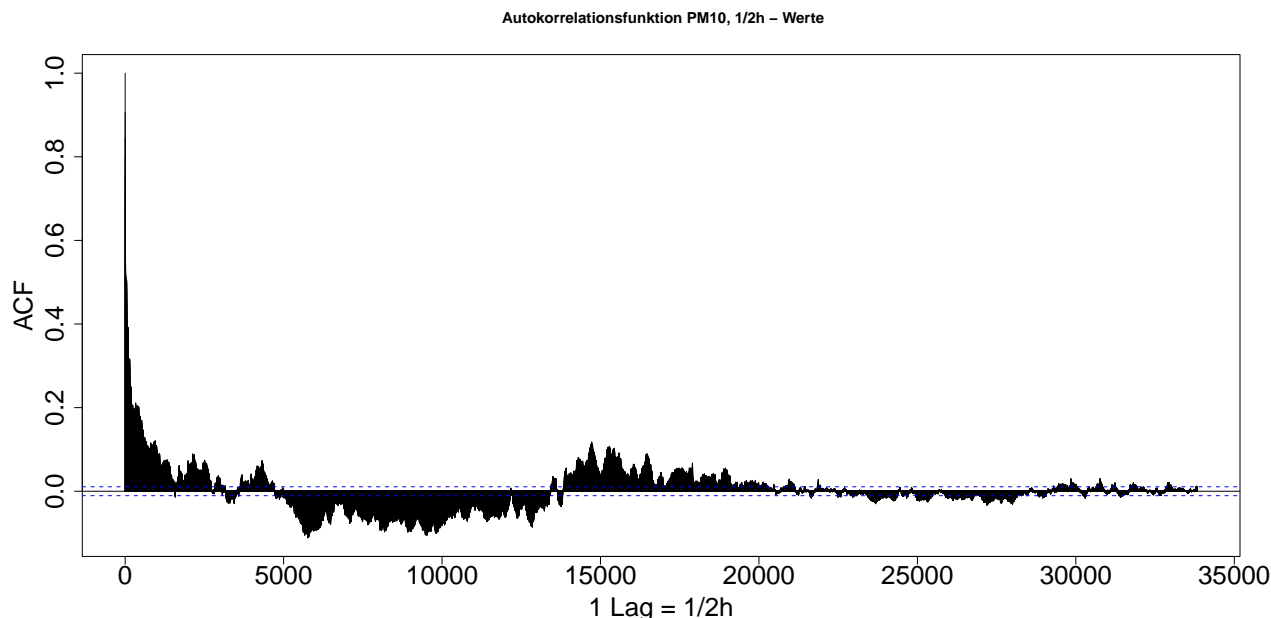
einen schwach negativen Zusammenhang aufweisen, während die Werte der Wintermonaten 2005 und 2006 beziehungsweise der Sommermonate 2005 und 2006 eine schwach positive Korrelation haben. Das heißt, wenn die Werte zum Zeitpunkt  $t = 0$  niedrig sind, jene zum Zeitpunkt  $t + h \approx 7.500$  hoch sein sollten und umgekehrt. Die positive Korrelation zum Lag  $h \approx 15.000$  bedeutet, dass Messpunkte, die ein Jahr auseinander liegen, in einem direkten Verhältnis zueinander stehen.

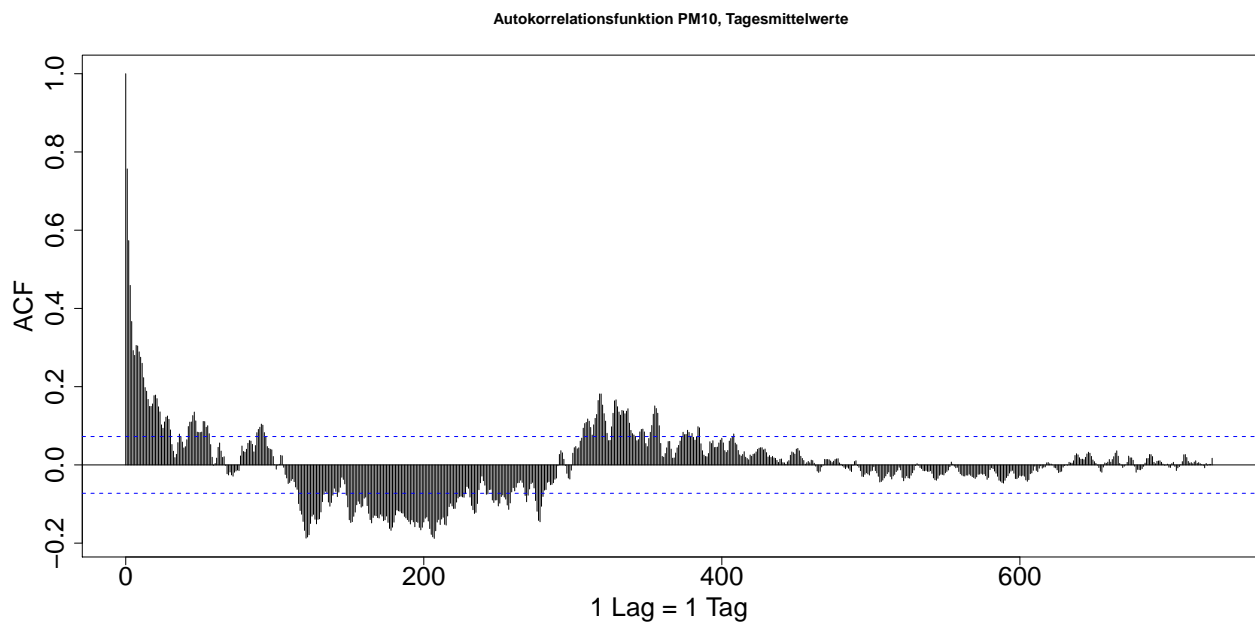
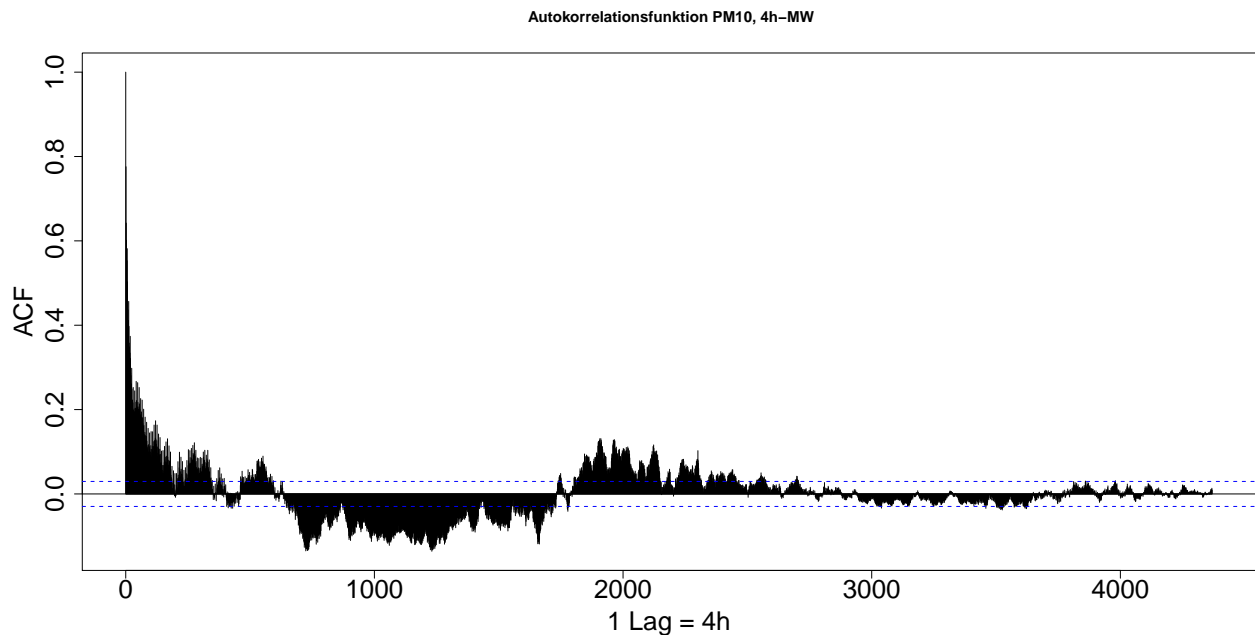
Für die ACF der Vierstunden- und Tagesmittelwerte erhält man analoge Erkenntnisse:

Bei den Vierstundenmittelwerten existiert bei Lag  $h \approx 1.000$  (das entspricht einem Zeitraum von ca. sechs Monaten) die stärkste negative Korrelation. Der größte positive Ausschlag der ACF bei den Vierstundenmittelwerten liegt bei Lag  $h \approx 2.000$  (das entspricht einem Zeitraum von ca. zwölf Monaten).

Für die Reihe der Tagesmittelwerte liegt der größte negative Ausschlag der Autokorrelationsfunktion bei Lag  $h \approx 160$  (was wieder ca. sechs Monaten entspricht) und der größte positive Ausschlag bei Lag  $h \approx 320$  (dies entspricht wieder einem Zeitraum von ca. einem Jahr).

In Abbildung 36 kann man für die Autokorrelationsfunktionen von  $PM_{10}$  ähnliche Ergebnisse wie in Abbildung 35 für die von  $NO_x$  erkennen: Die Zusammenhänge sind bei  $PM_{10}$  jedoch schwächer ausgeprägt und die sinusförmige Schwingung wird etwas verzerrt. Es lassen sich aber erneut die positiven und negativen Korrelationen zu den annähernd gleichen Zeitpunkten feststellen.





**Abbildung 36:** Autokorrelationsfunktionen von  $PM_{10}$  bzgl. 1/2h-Werte, Vierstunden- und Tagesmittelwerte

## Die Kreuzkorrelationsfunktionen von PM<sub>10</sub> und NO<sub>x</sub>

Nun werden anhand der Kreuzkorrelationsfunktionen der Reihen von PM<sub>10</sub> und NO<sub>x</sub> erneut Zusammenhänge zwischen diesen beiden Zielgrößen ausgearbeitet. Zu Beginn wird näher auf die Definition dieses Analysewerkzeugs eingegangen [3].

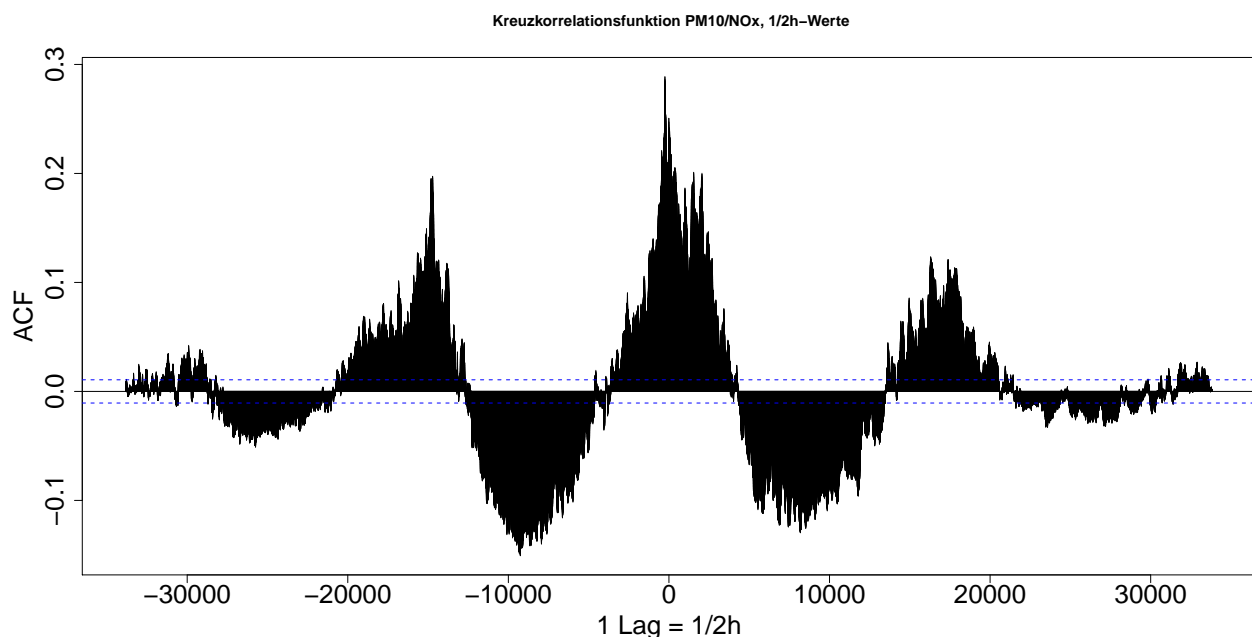
### Definition 12. Kreuzkorrelationsfunktion

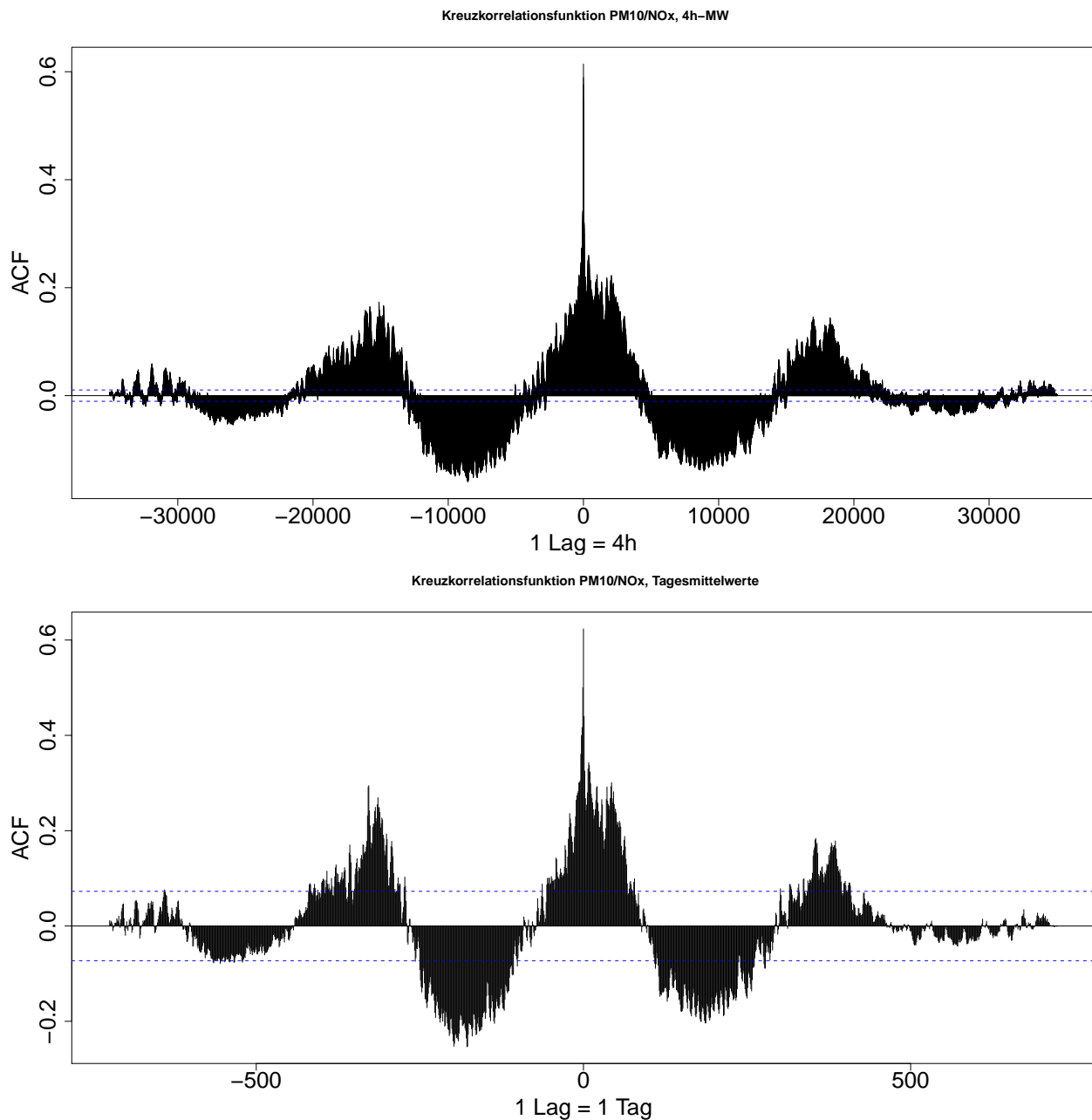
Die Kreuzkorrelationsfunktion (CCF) ist definiert als

$$\rho_{xy}(s, t) = \frac{\gamma_{xy}(s, t)}{\sqrt{\gamma_x(s, s)\gamma_y(t, t)}}.$$

Sie ist ein normiertes Maß für die lineare Vorhersehbarkeit von dem Wert  $x_s$  einer Zeitreihe  $x_t$  zu einem Zeitpunkt  $s$  durch die Kenntnis des Wertes  $y_t$  einer Zeitreihe  $y_t$  zu einem Zeitpunkt  $t$ .

Es sollen nun die Kreuzkorrelationsfunktionen der Reihen von PM<sub>10</sub> und NO<sub>x</sub> betrachtet werden.





**Abbildung 37:** Darstellung der Kreuzkorrelationsfunktionen von  $PM_{10}$  und  $NO_x$  bzgl. 1/2h-Werte, Vierstunden- und Tagesmittelwerte

Abbildung 37 lässt eine sehr sinusförmige Schwingung erkennen. Es ergibt sich eine Periodendauer von ca. 7.500 beziehungsweise ca. 15.000 Datenpunkten, was erneut einem Zeitraum von ca. sechs Monaten beziehungsweise einem Jahr entspricht. Das bedeutet, dass wenn die  $PM_{10}$ -Werte im Winter hoch sind, die  $NO_x$ -Werte im darauffolgenden Sommer niedrig und im nächsten Winter wieder dementsprechend hoch sind (und umgekehrt).

Die größte (positive) Korrelation zwischen  $PM_{10}$  und  $NO_x$  erhält man bei einem Lag  $h = 0$ . Hier entspricht der Wert der Kreuzkorrelationsfunktion genau dem Korrelationskoeffizienten von Pearson ( $r = 0,62$  für Halbstundenwerte,  $r = 0,65$  für Vierstunden- und Tagesmittelwerte).

Die Werte der Kreuzkorrelationsfunktionen für  $PM_{10}$  und  $NO_x$  werden bei den Vierstunden- und Tagesmittelwerten im Vergleich zu jenen der Halbstundenwerte minimal höher.

### 3 Betrachtung der Zeitreihen $PM_{10}$ und $NO_x$ auf verschiedenen Zeitskalen

Wie im Buch „*Statistical Methods for Environmental Epidemiology with R*“ [1] beschrieben wird, sollen nun die Reihen von  $PM_{10}$  und  $NO_x$  auf verschiedenen Zeitskalen analysiert werden.

Hier wird eine Zeitreihe  $Y_t$  folgendermaßen dargestellt:

$$Y_t = \text{Trend}_t + \text{Saisonalität}_t + \text{Residuum}_t,$$

wobei die Residuen hier eben für die zufälligen zeitlichen Einheiten stehen, die man der Einfachheit halber nicht betrachten möchte.

Um den jährlichen Trend berechnen zu können, wird ein *Moving Average* mit Länge 17.520 (für jeden Tag existieren 48 Messungen,  $48 \times 365 = 17.520$ ) bestimmt.

Den saisonalen Trend erhalten wir folgendermaßen:

$$\text{Saisonalität}_t + \text{Residuum}_t = Y_t - \text{Trend}_t,$$

wobei nun ein *Moving Average* der Länge 4.320 gebildet werden muss (für den Saisonalen Trend wird das Jahr in vier 90-tägige Zeitintervalle aufgeteilt).

Die Residuen ergeben sich dann aus dem Rest:

$$\text{Residuum}_t = Y_t - \text{Trend}_t - \text{Saisonalität}_t.$$

Auf diese Weise lässt sich die Zeitskala auf jede beliebige Genauigkeit verfeinern.

## Autokorrelationsfunktionen auf unterschiedlichen Zeitskalen

Jährliche Zeitskala:

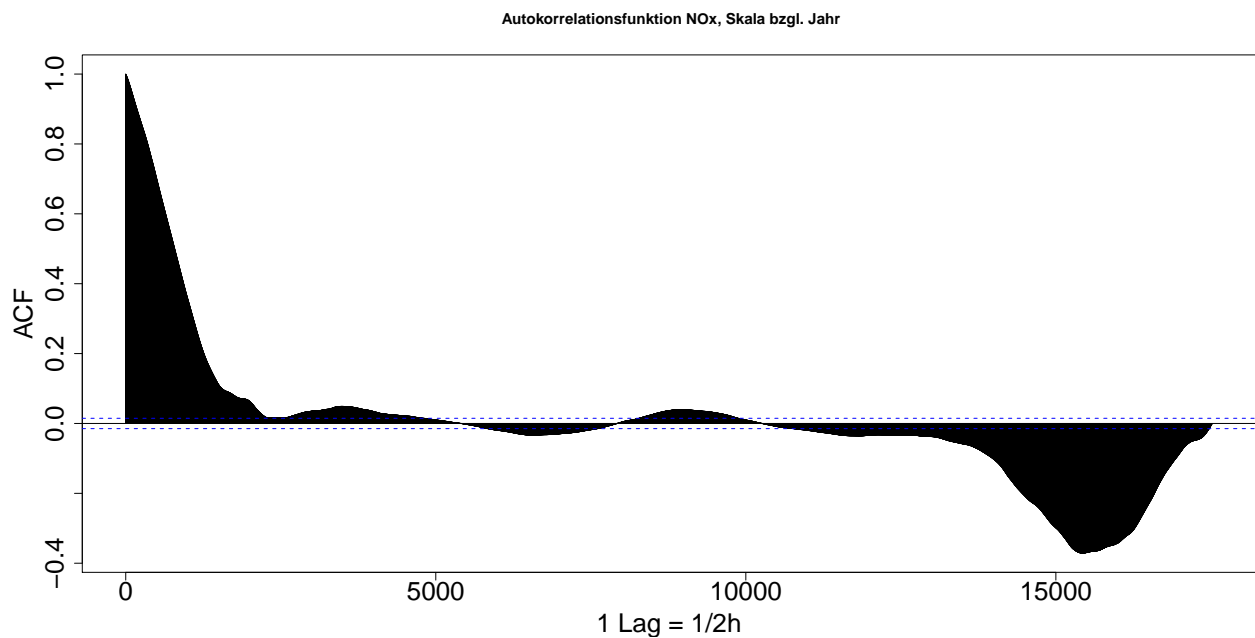
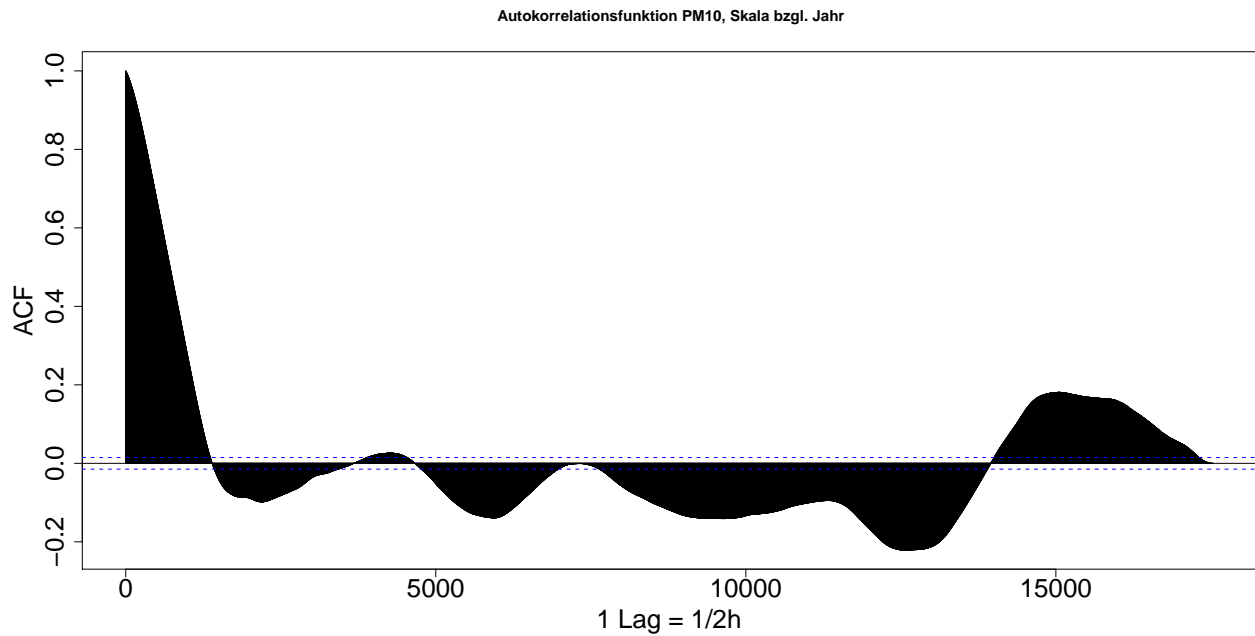
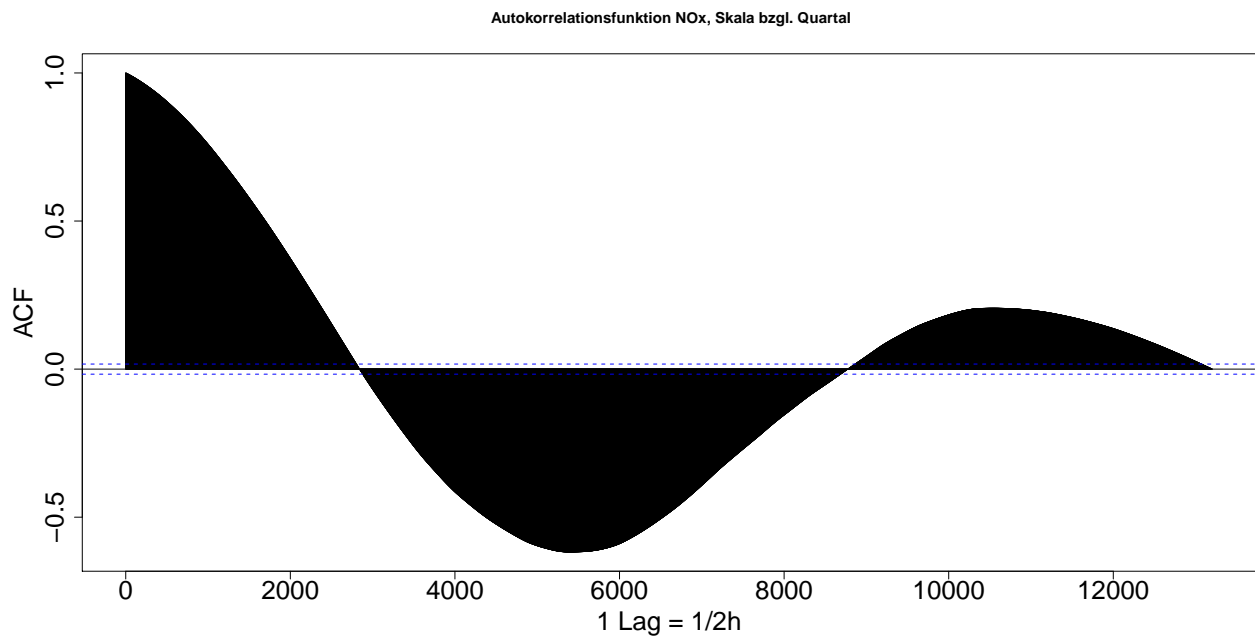
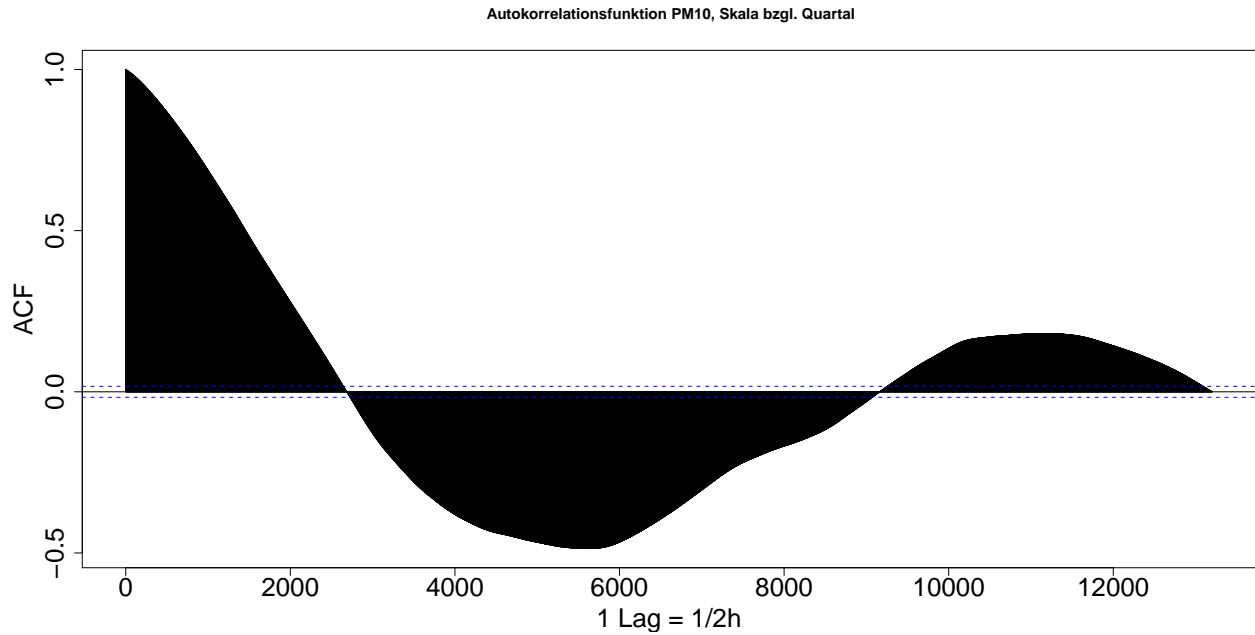


Abbildung 38: Autokorrelationsfunktionen von  $PM_{10}$  und  $NO_x$  bzgl. jährlicher Zeitskala



### 3 BETRACHTUNG DER ZEITREIHEN $PM_{10}$ UND $NO_x$ AUF VERSCHIEDENEN ZEITSKALEN

Saisonale Zeitskala (eine Saison entspricht ca. 3 Monaten):



**Abbildung 39:** Autokorrelationsfunktionen von  $PM_{10}$  und  $NO_x$  bzgl. quartaler Zeitskala

Monatliche Zeitskala:

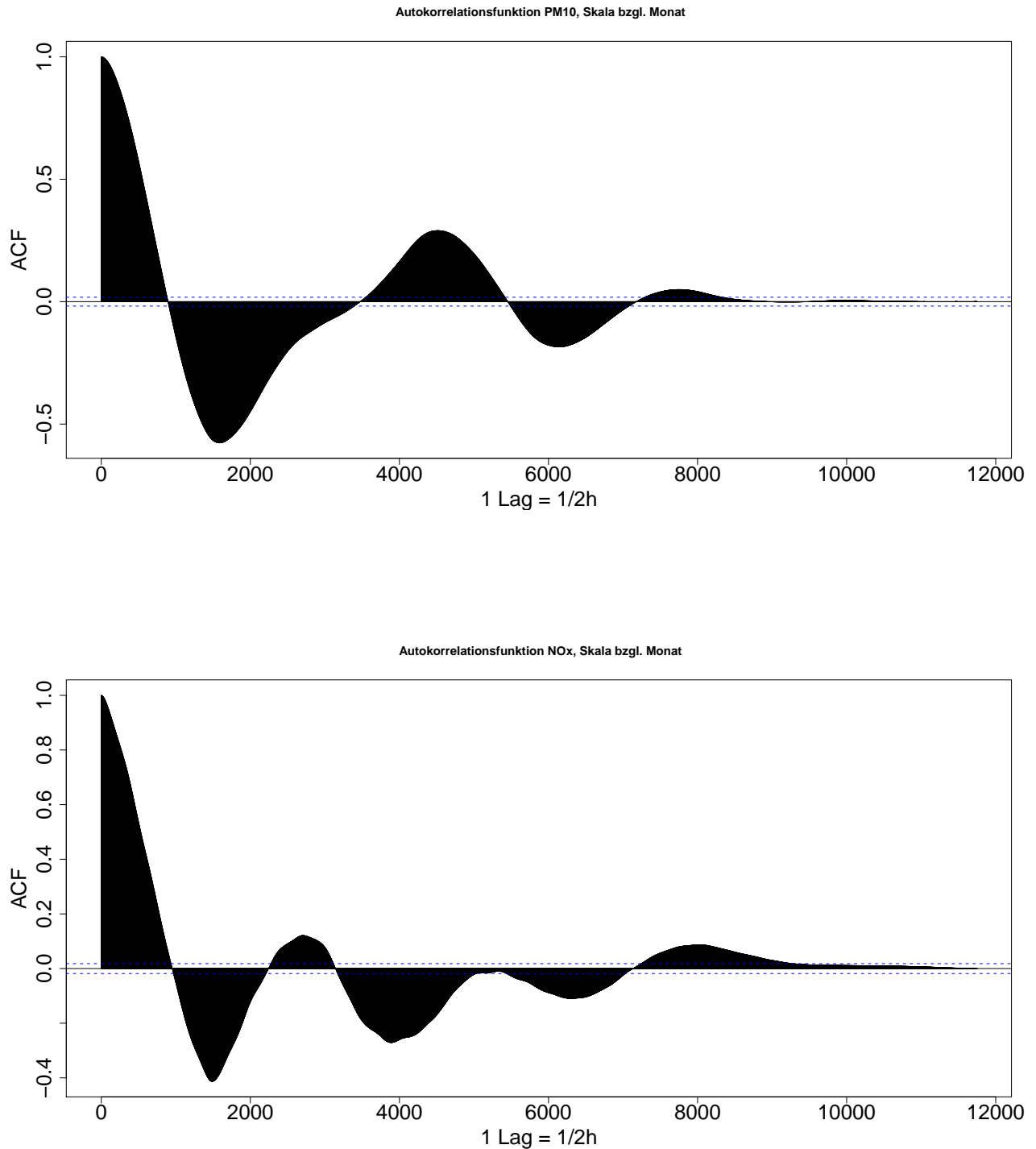
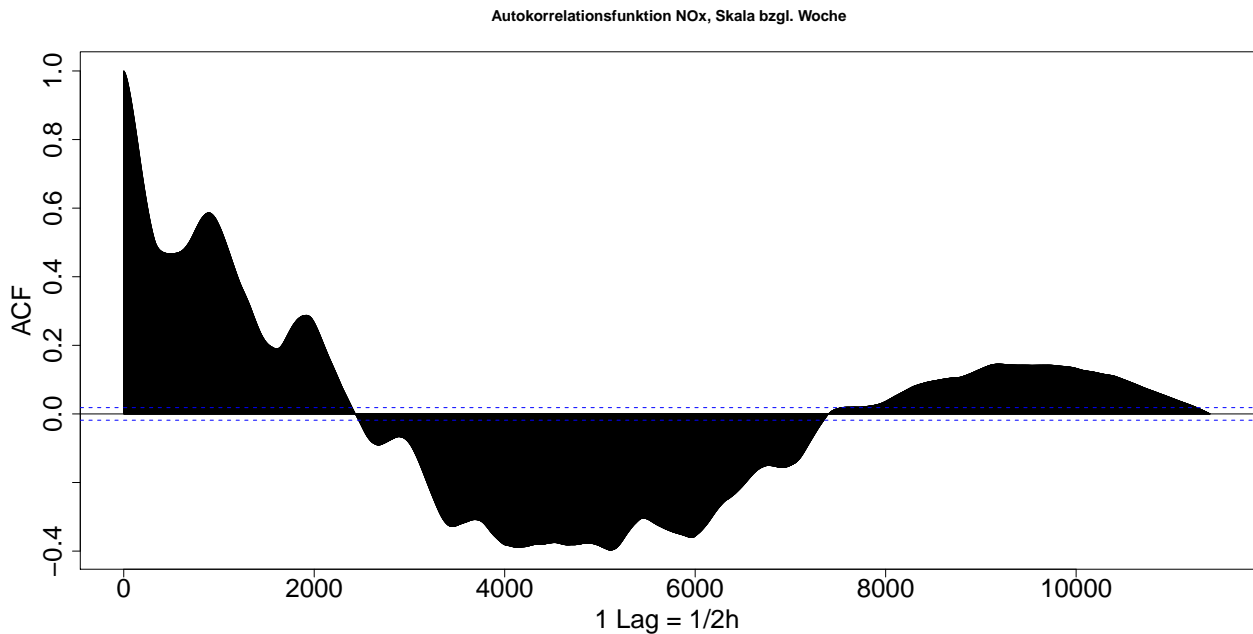
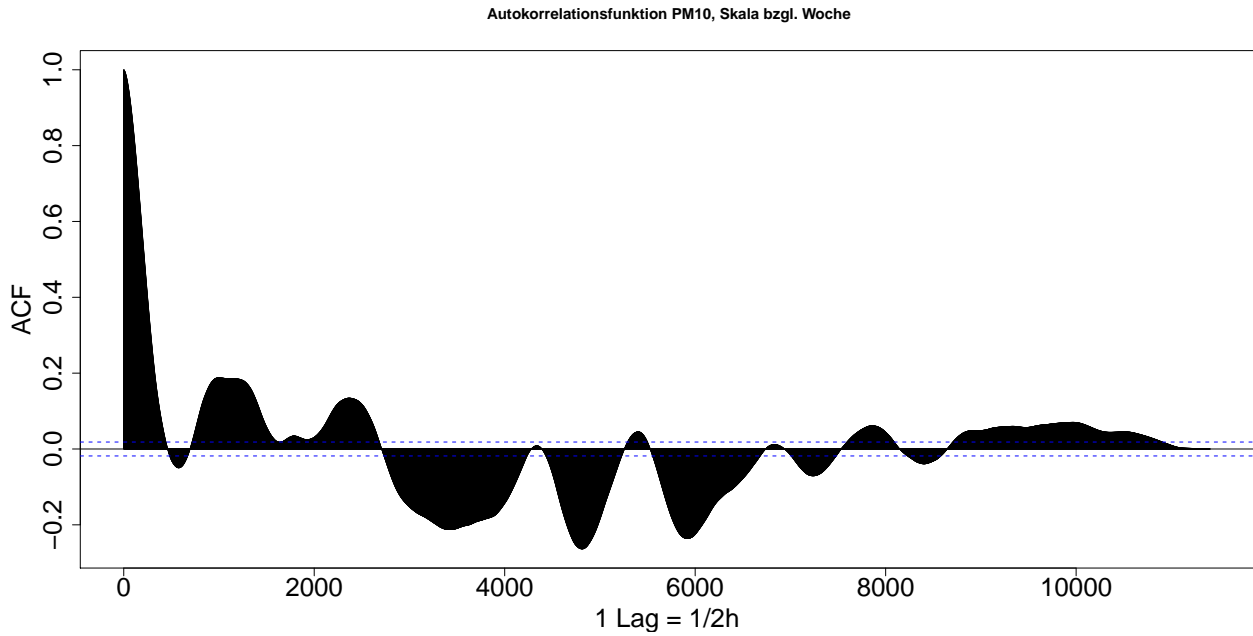


Abbildung 40: Autokorrelationsfunktionen von  $PM_{10}$  und  $NO_x$  bzgl. monatlicher Zeitskala

Wöchentliche Zeitskala:



**Abbildung 41:** Autokorrelationsfunktionen von  $PM_{10}$  und  $NO_x$  bzgl. wöchentlicher Zeitskala

Tägliche Zeitskala:

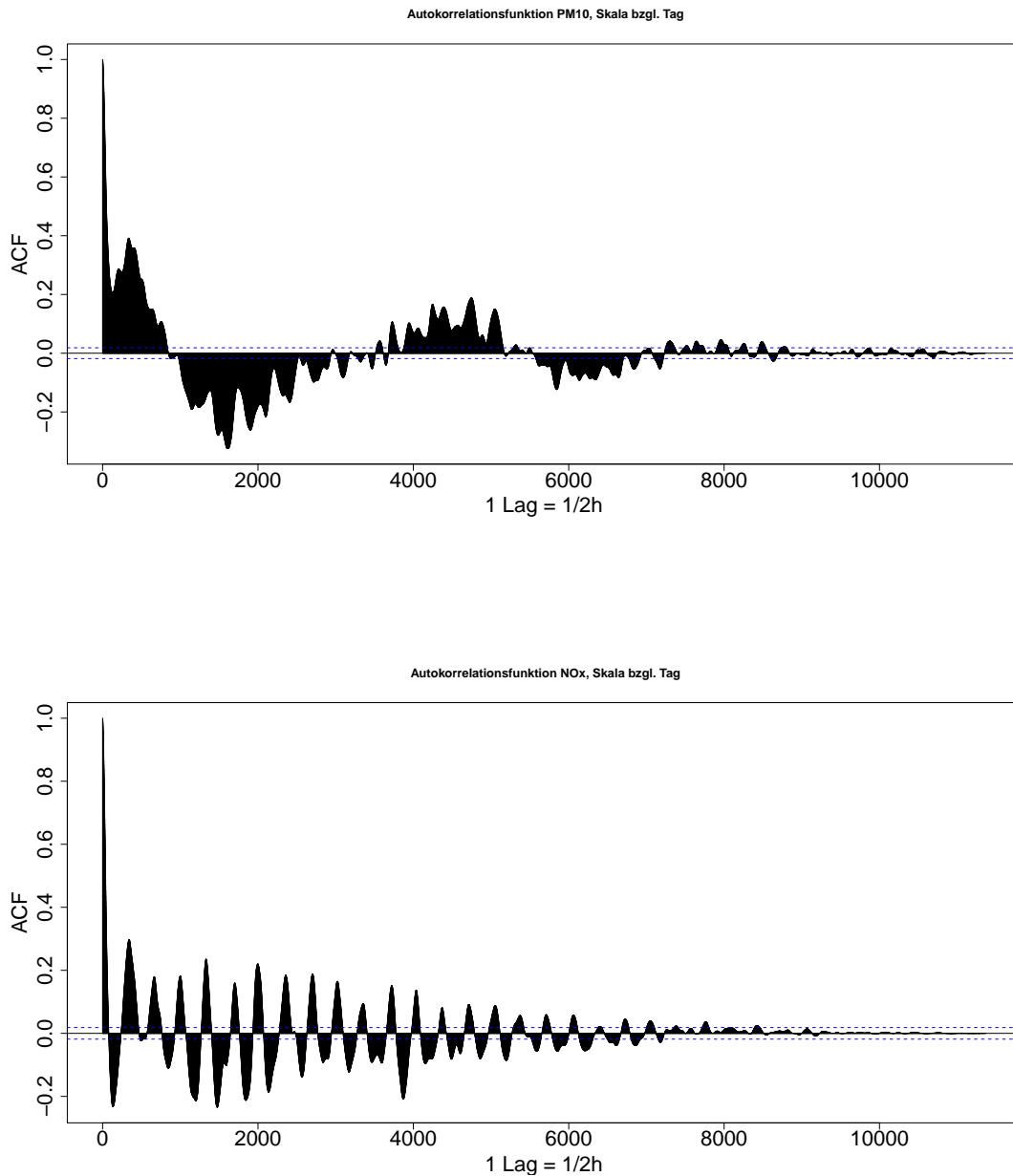


Abbildung 42: Autokorrelationsfunktionen von  $PM_{10}$  und  $NO_x$  bzgl. täglicher Zeitskala

Mit einer saisonalen Zeitskala wie in Abbildung 39 erhält man die höchste Korrelation, das heißt, es lassen sich hier am besten zukünftige Schadstoffwerte aus vorhandenen Daten voraussagen. Deshalb ist hier diese Zeitskala zu bevorzugen. Auch besitzt die Autokorrelationsfunktion hier am meisten Glätte.

## Zusammenhang von $PM_{10}$ und $NO_x$ auf unterschiedlichen Zeitskalen

Zunächst betrachtet man ein lineares Regressionsmodell zwischen  $PM_{10}$  und  $NO_x$ . Da hier die Funktion *filter* verwendet wird, fehlen in der Abbildung 43 jeweils die Anfangswerte.

In diesem Fall soll  $PM_{10}$  mit Hilfe eines linearen Zusammenhangs durch  $NO_x$  modelliert werden (alle Werte sind auf zwei Dezimalstellen gerundet). Hier werden nur die Reihen der Halbstundenwerte dieser beiden Zielgrößen verwendet.

	Estimate	Std. Error	t value	$Pr(>  t )$
(Intercept)	20,917	0,151	138,800	$< 2e - 16 * **$
$NO_x$	0,519	$< 3e - 2$	147,400	$< 2e - 16 * **$

Man erkennt also hier bereits einen positiven linearen Zusammenhang. Im betrachteten Fall ergibt sich folgendes Modell:

$$PM_{10} = 20,917 + 0,519 \cdot NO_x, \quad \text{für Halbstundenwerte}$$

Im weiteren soll auch der Zusammenhang zwischen einzelnen zeitlichen Komponenten der beiden Schadstoffe untersucht werden.

### Zerlegung der Zeitskala der Zeitreihen $PM_{10}$ und $NO_x$

Es sollen nun in diesem Abschnitt zunächst die zugrunde liegenden Zeitreihen  $PM_{10}$  und  $NO_x$  in ihre lang-, mittel- und kurzfristigen Komponenten aufgeteilt werden. Als Grundlage dient dazu Kapitel 5 des Buches [1].

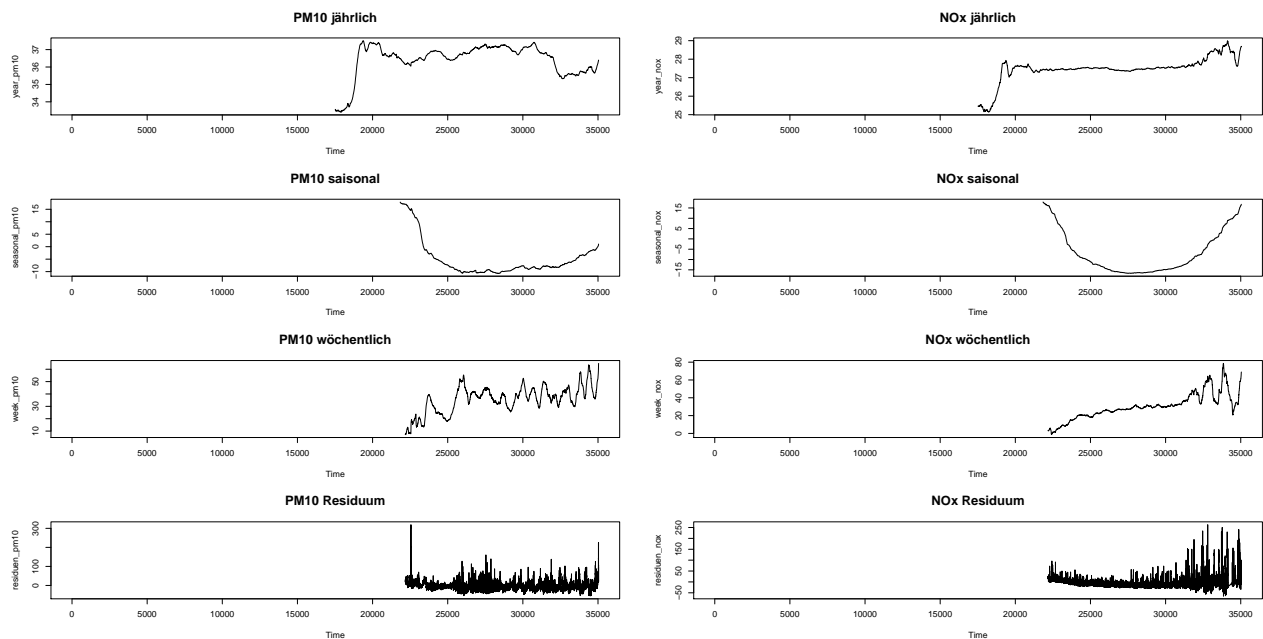
Im folgenden wird wieder eine lineare Regression zwischen  $PM_{10}$  und den unterschiedlichen Komponenten von  $NO_x$  erzeugt.

Anhand von Abbildung 43 lässt sich eine starke Ähnlichkeit der jeweiligen Trends erkennen.

Durch den Wertebereich erkennt man, dass  $PM_{10}$  bis auf den jährlichen Trend eine geringere Fluktuation aufweist, als  $NO_x$ . Die unterschiedlichen Trends der Zeitreihe  $NO_x$  sind hingegen glatter als jene von  $PM_{10}$ . Vor allem bei den jährlichen und wöchentlichen Trends fällt ein Unterschied bezüglich der Glätte auf.

Die stärksten Ähnlichkeiten lassen sich bei der saisonalen Komponente erkennen, wobei hier  $NO_x$  das Merkmal  $PM_{10}$  führt. Die Residuen von  $PM_{10}$  zeigen eine durchgehend ansatzweise gleiche Fluktuation auf, während die Residuen von  $NO_x$  gegen Ende signifikant ansteigen.

### 3 BETRACHTUNG DER ZEITREIHEN $PM_{10}$ UND $NO_x$ AUF VERSCHIEDENEN ZEITSKALEN



**Abbildung 43:** Zerlegung der Zeitreihen  $PM_{10}$  und  $NO_x$  in jährlichen, saisonalen und wöchentlichen Trend

Nun wird der Zusammenhang von  $PM_{10}$  und den unterschiedlichen  $NO_x$  - Trends betrachtet, das heißt, es soll  $PM_{10}$  durch den jährlichen, saisonalen und wöchentlichen Trend von  $NO_x$  dargestellt werden. Als Grundlage werden hierfür die Halbstundenwerte herangezogen.

	Estimate	Std. Error	t value	$Pr(>  t )$
(Intercept)	-21,54	27,33	-0,79	0,43
$NO_x$ jährlicher Trend	1,66	1,00	1,66	0,10
$NO_x$ saisonaler Trend	-0,02	0,03	-0,81	0,42
$NO_x$ wöchentlicher Trend	0,30	0,02	13,63	$< 2e - 16$ * **
$NO_x$ Residuen	0,42	0,01	62,40	$< 2e - 16$ * **

Bis auf den Intercept und den Koeffizienten des saisonalen Trends (der aber nicht signifikant ist) sind alle Koeffizienten positiv, was in einem positiven linearen Zusammenhang zwischen den Werten von  $PM_{10}$  und den jährlichen und wöchentlichen  $NO_x$  - Komponenten resultiert. Das Modell sieht hier folgend aus:

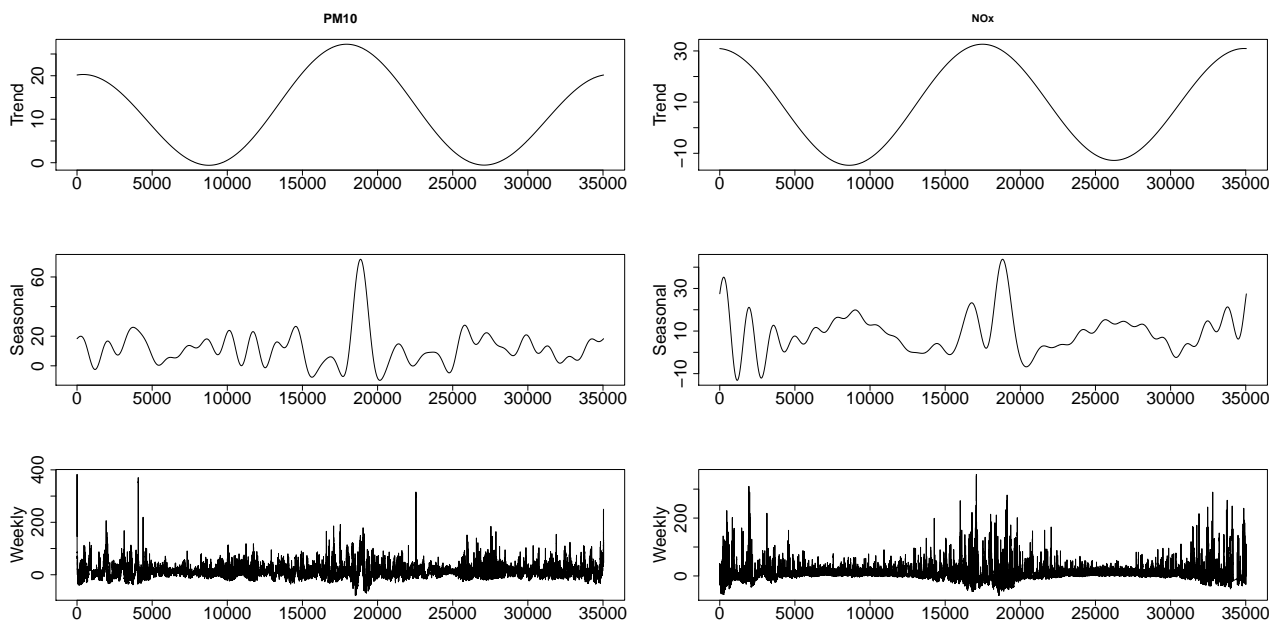
$$PM_{10} = 0,30 \cdot NO_{x,W} + 0,42 \cdot NO_{x,R}.$$

### 3.1 Zerlegung der Zeitskala mit der Funktion *tsdecomp*

Es wird in dem Buch [1] mit der R-Funktion *tsdecomp* aus dem Paket *tsModel* eine weitere Möglichkeit dargestellt, eine Zeitreihe auf verschiedenen Zeitskalen (exakt) zu zerlegen. Der ganze Ansatz beruht auf der Fourier-Transformation. Die Anzahl der Zyklen der jeweiligen Komponente müssen hier jeweils explizit angegeben werden.

Nun werden die Zeitreihen zunächst, ähnlich dem bereits betrachteten Fall, in drei Teile zerlegt, den langfristigen, saisonalen und kurzfristigen Trend, wobei hier unter dem saisonalen Trend ein monatlicher Trend verstanden wird.

Für den langfristigen Trend ergibt sich ein Zyklus über die gesamte Messdauer von zwei Jahren, für den saisonalen Trend erhält man 2-24 Zyklen, dem kurzzeitigen Trend entsprechen 25 oder mehr Zyklen. Es werden hierfür wieder die Halbstundenwerte herangezogen.



**Abbildung 44:** Zerlegung der Zeitreihen  $PM_{10}$  und  $NO_x$  mit der Funktion *tsdecomp* in langfristigen, saisonalen und kurzfristigen Trend

Für die lineare Regression zwischen  $PM_{10}$  und den Komponenten von  $NO_x$  ergibt sich folgendes:

	Estimate	Std. Error	t value	$Pr(>  t )$
(Intercept)	19,03	0,18	106,12	$< 2e - 16$ * **
Langzeit	0,53	$< 7e - 2$	73,25	$< 2e - 16$ * **
Saison	0,75	0,01	59,38	$< 2e - 16$ * **
Kurzzeit	0,49	$< 4e - 2$	116,00	$< 2e - 16$ * **

Der Unterschied zum vorigen Fall ist, dass man nun für alle drei Komponenten einen positiven Zusammenhang zu  $PM_{10}$  erhält.

Mit der Funktion *tsdecomp* lässt sich auch eine feinere Unterteilung der betrachteten Komponenten erstellen. Im Unterschied zu vorher soll die kurzfristige Komponente der Zeitreihen  $PM_{10}$  und  $NO_x$  noch weiter zerlegt werden. Der saisonale Trend entspricht wieder einem monatlichen Trend.

Die betrachteten Komponenten sind:

- Langzeittrend
- saisonaler Trend
- zwei bis vier Wochen
- ein bis zwei Wochen
- ein bis sieben Tage und
- < 1 Tag.

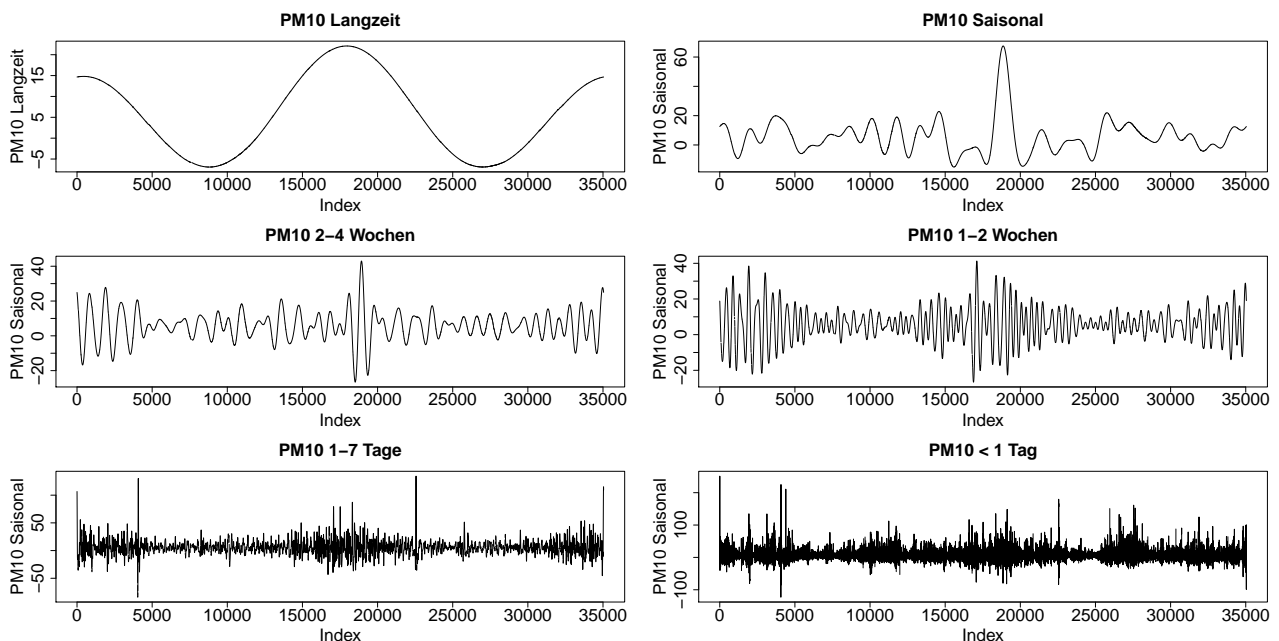
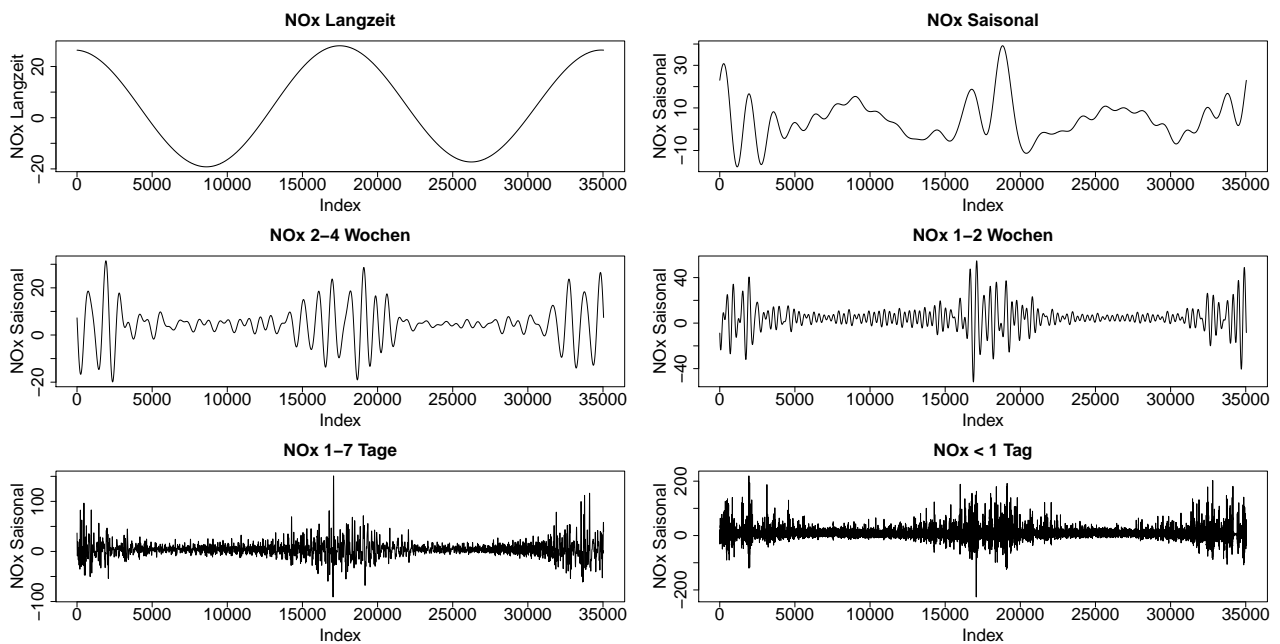


Abbildung 45: Genauere Zerlegung der Zeitreihen  $PM_{10}$  mit der Funktion *tsdecomp*

Bezüglich  $PM_{10}$  bemerkt man in Abbildung 45, dass der Höchstwert der Zeitreihe in den verschiedenen Zeitskalen ca. im Mittelpunkt der betrachteten Zeitpunkte angenommen wird. Dies entspricht dem Jahreswechsel von 2005/2006. Auch zu Beginn und gegen Ende der betrachteten Zeitreihe steigen die  $PM_{10}$  - Werte stark an. Daraus lassen sich maximale  $PM_{10}$  - Konzentrationen in den Winterperioden feststellen.



### 3 BETRACHTUNG DER ZEITREIHEN $PM_{10}$ UND $NO_x$ AUF VERSCHIEDENEN ZEITSKALEN



**Abbildung 46:** Genauere Zerlegung der Zeitreihen  $NO_x$  mit der Funktion *tsdecomp*

In Abbildung 46 sind die Maxima von  $NO_x$  in den Wintermonaten sogar noch besser erkennbar.

Nun soll analog wie zuvor eine lineare Regression zwischen  $PM_{10}$  und den verschiedenen Komponenten von  $NO_x$  erzeugt werden. Dazu dienen wieder die Halbstundenwerte als Basis.

	Estimate	Std. Error	t value	$Pr(>  t )$
(Intercept)	20,30	0,16	123,76	$< 2e - 16$ * **
Langzeit	0,53	$< 7e - 2$	73,25	$< 2e - 16$ * **
Saisonal	0,75	0,01	59,38	$< 2e - 16$ * **
2-4 Wochen	0,49	0,02	31,86	$< 2e - 16$ * **
1-2 Wochen	0,51	0,01	46,86	$< 2e - 16$ * **
1-7 Tage	0,49	0,01	67,74	$< 2e - 16$ * **
< 1 Tag	0,48	0,01	75,25	$< 2e - 16$ * **

Man erkennt eine sehr starke positive Abhängigkeit von  $PM_{10}$  und den Komponenten von  $NO_x$ , vor allem der saisonalen Komponente. Die Standardabweichung ist verschwindend klein. Aufgrund der hohen t-Werte kann man darauf schließen, dass  $PM_{10}$  signifikant von den Komponenten von  $NO_x$  abhängt.

Zur Erinnerung: Wenn der t-Wert stark von 0 abweicht beziehungsweise  $Pr(> |t|)$ , dann ist der entsprechende Koeffizient signifikant von 0 verschieden und die zugehörige Variable hat einen signifikanten Einfluss auf die Responsevariable  $PM_{10}$ .

### 3.2 Untersuchung von Störgrößen

In Abschnitt 5.4 [1] wird eine Möglichkeit aufgezeigt, wie eine dritte Variable den Zusammenhang zwischen  $PM_{10}$  und  $NO_x$  stören kann. Da die Werte von  $PM_{10}$  und  $NO_x$  sehr jahreszeitabhängig sind, wird vermutet, dass die Temperatur ein solcher Störfaktor sein könnte.

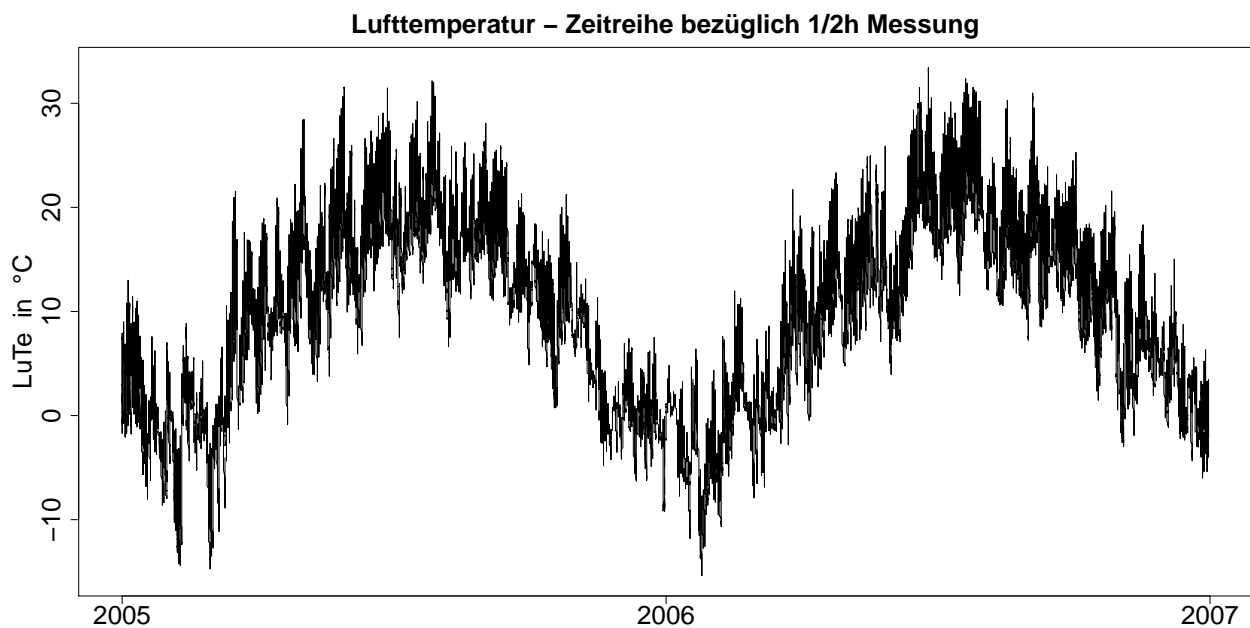


Abbildung 47: Darstellung der Zeitreihe der Lufttemperatur bzgl. Halbstundenwerte

Man kann nun den Einfluss der Lufttemperatur eliminieren, indem man eine lineare Regression der Lufttemperatur mit  $PM_{10}$  und  $NO_x$  durchführt und sich dann auf die Residuen beschränkt.

Die Abbildungen 48 und 49 zeigen die Trends von  $PM_{10}$  und  $NO_x$ , befreit vom Einfluss der Lufttemperatur.

### 3 BETRACHTUNG DER ZEITREIHEN $PM_{10}$ UND $NO_x$ AUF VERSCHIEDENEN ZEITSKALEN

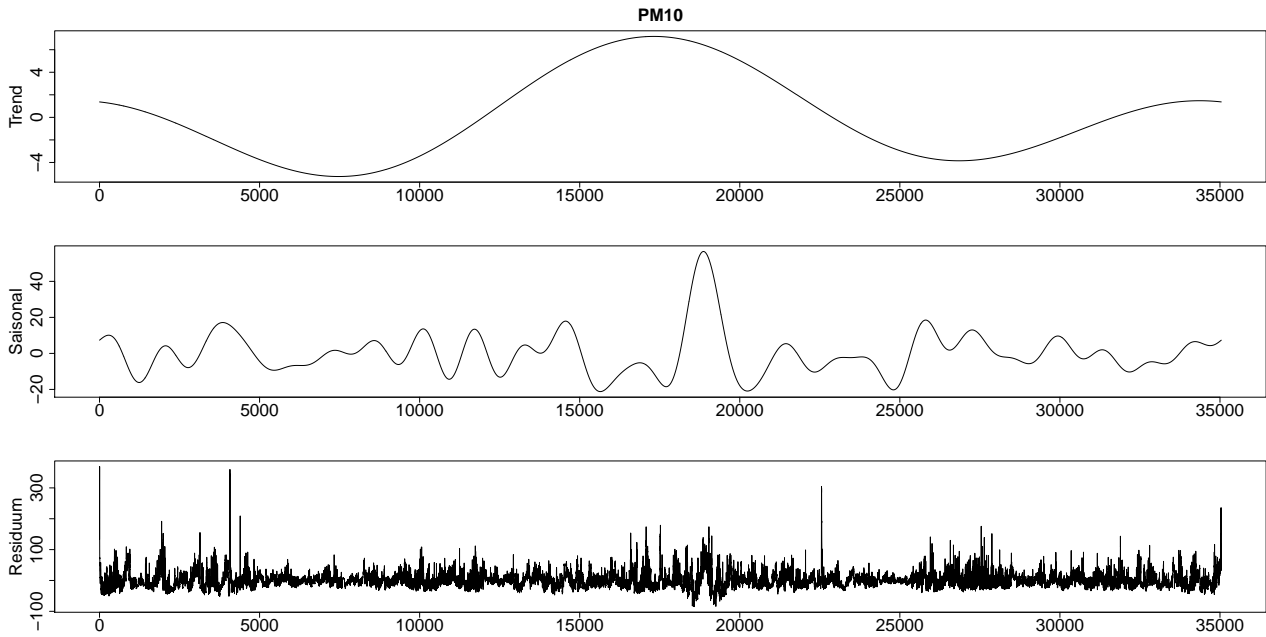


Abbildung 48: Darstellung der Trends von  $PM_{10}$  nach Entfernung des Einflusses der Lufttemperatur

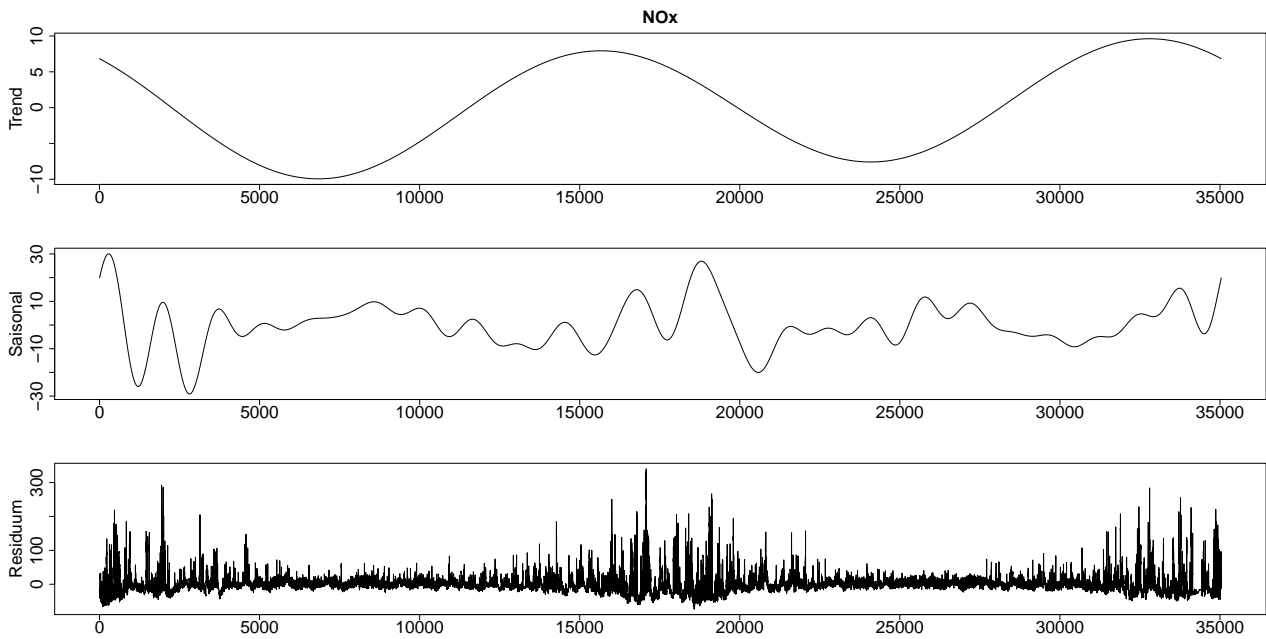


Abbildung 49: Darstellung der Trends von  $NO_x$  nach Entfernung des Einflusses der Lufttemperatur

### 3 BETRACHTUNG DER ZEITREIHEN $PM_{10}$ UND $NO_x$ AUF VERSCHIEDENEN ZEITSKALEN

---

Man erkennt, dass nun die Auslenkungen viel geringer ausfallen. Die Trends an sich bleiben jedoch ähnlich. Für die von der Lufttemperatur bereinigten Messwerte wird hier nun wieder eine Lineare Regression betrachtet, d.h.  $PM_{10}$  durch  $NO_x$  auf den verschiedenen Zeitskalen dargestellt:

	Estimate	Std. Error	t value	$Pr(>  t )$
(Intercept)	34,98	0,12	288,82	$< 2e - 16$ * **
Langzeit	0,84	0,02	43,52	$< 2e - 16$ * **
Saisonal	0,70	0,01	55,33	$< 2e - 16$ * **
Kurzzeit	0,49	$< 4e - 2$	113,01	$< 2e - 16$ * **

Es herrscht also auf allen Zeitskalen ein positiver Zusammenhang zwischen  $PM_{10}$  und  $NO_x$ , wobei er für die Langzeitkomponente am stärksten ausfällt. Die Abbildungen 48 und 49 bestätigen diese Erkenntnisse.

## 4 Conclusio

Aufgrund der Ergebnisse kann man feststellen, dass die  $PM_{10}$  - und  $NO_x$  - Werte in den Wintermonaten am höchsten sind, wodurch anzunehmen ist, dass die Belastung mit einem erhöhten Verbrauch von fossilen Brennstoffen in Verbindung steht: Das Auto wird häufiger als Fortbewegungsmittel genutzt und viele Haushalte besitzen immer noch umweltunfreundliche, altmodische Heizanlagen. Hier sollte zur Verringerung der Luftschadstoffbelastung ein Umdenken stattfinden und vermehrt auf erneuerbare Energiequellen zurückgegriffen werden, beziehungsweise die öffentlichen Verkehrsmittel genutzt werden, um die allgemeine Verkehrsbelastung zu senken.

Der kritische Tagesmittelwert für  $PM_{10}$  von  $50 \mu\text{g}/\text{m}^3$  wurde im Zeitraum 01.01.2005 bis 31.12.2006 exakt 124 mal überschritten, davon 58 mal im Jahr 2005 und 66 mal im Jahr 2006. Beides verstößt gegen die im Zeitraum 2005 - 2009 erlaubte Anzahl an Überschreitungen des Grenzwertes bei weitem: Er hätte maximal 30 mal pro Jahr überschritten werden dürfen.

Die Jahresmittelwerte für  $PM_{10}$  betragen 2005 ca.  $32,35 \mu\text{g}/\text{m}^3$  und 2006 ca.  $35,19 \mu\text{g}/\text{m}^3$ , das heißt, sie sind unter dem Grenzwert von  $40 \mu\text{g}/\text{m}^3$  geblieben.

Für  $NO_x$  ist kein Grenzwert für den Tagesmittelwert bekannt, jedoch wird der Jahresmittelwert von  $30 \mu\text{g}/\text{m}^3$  im betrachteten Zeitraum eingehalten:

Der Jahresmittelwert betrug 2005 ca.  $24,90 \mu\text{g}/\text{m}^3$  und 2006 wurden ca.  $27,63 \mu\text{g}/\text{m}^3$  gemessen.

Auch bezüglich des Störfaktors *Lufttemperatur* lässt sich ein gewisser Einfluss auf die Messwerte von  $PM_{10}$  und  $NO_x$  feststellen.

Die Korrelationsanalyse bestätigt, dass  $PM_{10}$  und  $NO_x$  in gewisser Weise zueinander in Verbindung stehen: Laut Umweltbundesamt [4] kann bekanntlich aus Stickoxiden Feinstaub entstehen. Somit tragen vermehrte Stickoxid - Konzentrationen auch zu einer erhöhten Belastung durch  $PM_{10}$  bei.

Über den Zeitraum 01.01.2005 bis 31.12.2006 ist ein leichter Anstieg der Messwerte von  $PM_{10}$  und  $NO_x$  festzustellen.

## Quellenverzeichnis

- [1] Peng R. D. and Dominici F., 2008, 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, TU Graz, 2. Auflage, September 2008, 212S.
- [3] Shumway R. H. and Stoffer D.S. (2006): Time Series Analysis and Its Applications (With R Examples), 2nd Edition, Springer, New York.
- [4] <http://www.umweltbundesamt.at>,  
Abrufdatum 20.09.2010
- [5] <http://www.lgl.bayern.de/gesundheit/umweltmedizin/feinstaub.htm>,  
Abrufdatum 20.09.2010