Graz University of Technology

Bakkalaureatsprojekt TM (Statistik), PR 506.558

Zeitreihen PM_{10} und NO_x für Graz-Süd 2005/2006

ausgeführt am

Institut für Statistik der TU Graz

unter Anleitung von

Univ.-Prof. DI Dr. Ernst Stadlober

durch

Sarah Karasek

Matr. Nr. 0630917

Graz, 26. August 2011

Eidesstattliche Erklärung

„Ich erkläre an Eides statt, dass ich die vorliegende Bachelorarbeit selbstständig verfasst, und in der Bearbeitung und Abfassung keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt, sowie wörtliche und sinngemäße Zitate als solche gekennzeichnet habe.“

....

Datum, Ort Unterschrift

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich bei der Erstellung dieser Bachelorarbeit unterstützt haben.

Danken möchte ich Univ.-Prof. Dipl.-Ing. Dr.techn. Ernst Stadlober, der mir diese Arbeit ermöglicht hat.

Ein besondere Dank gilt außerdem Frau Brigitte Pfeiler, die mir bei kleineren und größeren Problemen immer zur Seite gestanden ist und mir eine große Hilfe war.

Nicht zuletzt möchte ich mich bei meinen Eltern bedanken, die mir dieses Studium überhaupt erst ermöglicht und vor allem die gesamte Zeit über moralisch unterstützt haben.

Abstract

This bachelor thesis deals with the analysis of the two pollution time series PM_{10} and NO_x. The underlying data was collected at the station Graz-South over a period of two years (1/1/2005 - 31/12/2006). Originally the data is available in half-hourly values, but has also been converted into 4-hour- and daily-averages.
In a brief introduction the measurement station and the air pollutant are described. Furthermore some important statistical formulas are added. The time series are looked at from different perspectives. Therefore they are split according to categorial characteristics like season (summer/winter), month and day type. The graphic representation in the first chapters is mainly represented by boxplots.
In the subsequent correlation analysis scatterplots were used to show the correlation between PM_{10} and NO_x.
The book called „Statistical Methods for Environmental Epidemiology with R“ by Peng/-Domenici served as a basis for the last chapter, in which the autocorrelation of the time series is described on different time scales.
The statistical program R (version 2.12.1) was used for all calculations and graphics.

Übersicht

Diese Bachelorarbeit befasst sich mit der Analyse der zwei Schadstoffzeitreihen PM_{10} und NO_x. Die zugrunde liegenden Werte wurden über einen Zeitraum von zwei Jahren an der Station Graz-Süd erfasst (1.1.2005 bis 31.12.2006). Sie liegen im Originalen zwar als Halbstundenwerte vor, wurden aber für die Arbeit auch in Vierstunden- und Tagesmittelwerte umgerechnet.
Nach einer kurzen Einleitung, in der die Messstation und die Schadstoffe beschrieben, sowie einige wichtige statistische Formeln eingeführt werden, werden die Zeitreihen aus unterschiedlichen Sichtweisen betrachtet. Dafür werden die Zeitreihen in Abhängigkeit von kategorialen Merkmalen wie Saison (Sommer/Winter), Monat und Tagestyp aufgespalten. Die graphische Darstellung erfolgt hier hauptsächlich durch Boxplots.
Für die anschließende Korrelationsanalyse, die Zusammenhänge zwischen PM_{10} und NO_x aufzeigen soll, wurden hingegen Scatterplots verwendet.
Für das letzte Kapitel, in dem die Autokorrelation der Zeitreihen auf verschiedenen Zeitskalen untersucht wird, wurde außerdem als Grundlage das Buch „Statistical Methods for Environmental Epidemiology with R“ von Peng/Domenici verwendet.
Für alle Berechnungen und graphischen Darstellungen wurde das Statistikprogramm R (Version 2.12.1) verwendet.

Inhaltsverzeichnis

Abbildungsverzeichnis **VI**

Tabellenverzeichnis **VII**

1 Einleitung **1**
1.1 Die Messstation 1
1.2 PM_{10} - Feinstaub 2
1.3 NO_x - Stickstoffoxide 2
1.4 Statistische Kenngrößen 3

2 Explorative Analyse der Daten **5**
2.1 Allgemeine Darstellung der Zeitreihen 5
2.2 Aufspaltung der Zeitreihen in Abhängigkeit der Saison 11
2.3 Aufspaltung der Zeitreihen in Abhängigkeit von Monaten 16
2.4 Aufspaltung der Zeitreihen in Abhängigkeit vom Tagestyp 26
2.4.1 Gemeinsame Darstellung von PM_{10} und NO_x 37
2.5 Zusammenhangsanalyse 38
2.6 Autokorrelationsstruktur 43
2.7 Kreuzkorrelationsstruktur 46

3 Zerlegung der Zeitreihen in verschiedene Zeitskalen **47**
3.1 Zerlegung der Zeitskala mit einem Moving Average 47
3.2 Zerlegung der Zeitskala mit der Funktion `tsdecomp` 49
3.3 Untersuchung auf Störgrößen 52

4 Conclusio **55**

Literatur **56**

Abbildungsverzeichnis

1.1.1 Grazer Messstationen . . . 1
2.1.1 Darstellung der PM_{10} $^1/_2$h-, 4h- und Tagesmittelwerte . . . 7
2.1.2 Darstellung der NO_x $^1/_2$h-, 4h- und Tagesmittelwerte . . . 10
2.2.1 Darstellung PM_{10} $^1/_2$h-Werte in Abhängigkeit der Saison . . . 11
2.2.2 PM_{10} $^1/_2$h-Werte Sommer . . . 12
2.2.3 PM_{10} $^1/_2$h-Werte Winter . . . 13
2.2.4 Darstellung NO_x $^1/_2$h-Werte in Abhängigkeit der Saison . . . 14
2.2.5 Darstellung NO_x $^1/_2$h-Werte Sommer . . . 14
2.2.6 Darstellung NO_x $^1/_2$h-Werte Winter . . . 15
2.3.1 PM_{10} $^1/_2$h-Werte inkl. Monatsmittelwerte . . . 16
2.3.2 PM_{10} 4h-Werte inkl. Monatsmittelwerte . . . 18
2.3.3 PM_{10} Tagesmittelwerte inkl. Monatsmittelwerte . . . 19
2.3.4 NO_x $^1/_2$h-Werte inkl. Monatsmittelwerte . . . 20
2.3.5 NO_x 4h-Werte inkl. Monatsmittelwerte . . . 22
2.3.6 NO_x Tagesmittelwerte inkl. Monatsmittelwerte . . . 23
2.3.7 PM_{10} im Monat März 2005-2006 . . . 24
2.3.8 NO_x im Monat Februar 2005-2006 . . . 25
2.4.1 PM_{10}: Tagestypdarstellung bzw. Detailansicht . . . 28
2.4.2 PM_{10}: Tagestypdarstellung bzw. Detailansicht Winter-Sommer . . . 30
2.4.3 Durchschnittlicher PM_{10} Tagesverlauf (Werktag) . . . 31
2.4.4 NO_x: Tagestypdarstellung bzw. Detailansicht . . . 33
2.4.5 NO_x: Tagestypdarstellung bzw. Detailansicht Winter-Sommer . . . 35
2.4.6 Durchschnittlicher NO_x Tagesverlauf (Werktag) . . . 36
2.4.7 Durchschnittlicher Tagesverlauf (PM_{10} und NO_x: Werktag) . . . 37
2.5.1 Scatterplot PM_{10} gegen NO_x, $^1/_2$h-, 4h-, Tagesmittelwerte . . . 38
2.5.2 Scatterplot PM_{10} gegen NO_x bzgl. Saison, Tagesmittelwerte . . . 39
2.5.3 Scatterplot PM_{10} gegen NO_x bzgl. Tagestyp, 4h- und Tagesmittelwerte, Sommer und Winter . . . 40
2.5.4 Scatterplot PM_{10} und NO_x bzgl. Tagestyp, Tagesmittelwerte . . . 41
2.5.5 Scatterplot NO_x gegen LUTE, Tagesmittelwerte . . . 42
2.6.6 ACF PM_{10}, $^1/_2$h-, 4h- und Tagesmittelwerte . . . 44
2.6.7 ACF NO_x, $^1/_2$h-, 4h- und Tagesmittelwerte . . . 45
2.7.8 Kreuzkorrelation: PM_{10} und NO_x, $^1/_2$h-Werte . . . 46
3.1.9 Zerlegung der Zeitskala für PM_{10} und NO_x, Halbstundenwerte . . . 48
3.2.10 Zerlegung der Zeitskala für PM_{10} und NO_x mit `tsdecomp`, Halbstundenwerte . . . 49
3.2.11 Zerlegung der Zeitskala (detailliert) für PM_{10} und NO_x mit `tsdecomp`, Halbstundenwerte . . . 51
3.3.12 Windgeschwindigkeit Graz-Süd, Halbstundenwerte . . . 52
3.3.13 Zerlegung von PM_{10} nach Entfernung des Einflusses der Windgeschwindigkeit, Halbstundenwerte . . . 53
3.3.14 Zerlegung von NO_x nach Entfernung des Einflusses der Windgeschwindigkeit, Halbstundenwerte . . . 53

Tabellenverzeichnis

1 Eckdaten der PM_{10} $^1/_2$h-Werte in $\mu g/m^3$ 5
2 Eckdaten der PM_{10} 4h-Werte in $\mu g/m^3$ 6
3 Eckdaten der PM_{10} Tagesmittelwerte in $\mu g/m^3$ 6
4 Eckdaten der NO_x $^1/_2$h-Werte in $\mu g/m^3$ 8
5 Eckdaten der NO_x 4h-Werte in $\mu g/m^3$ 9
6 Eckdaten der NO_x Tagesmittelwerte in $\mu g/m^3$ 9
7 PM_{10}-Mittelwerte pro Monat (Halbstundenwerte) in $\mu g/m^3$ 17
8 NO_x-Mittelwerte pro Monat (Halbstundenwerte) in $\mu g/m^3$ 21
9 Einfache lineare Regression von PM_{10} und NO_x 47
10 Lineare Regression von PM_{10} und NO_x (aufgespalten) 47
11 Lineare Regression von PM_{10} und NO_x (mit tsdecomp) 50
12 Lineare Regression von PM_{10} und NO_x (detailliert) 50
13 Lineare Regression von PM_{10} und NO_x nach Entfernung des Einflusses durch die Windgeschwindigkeit 54
14 Mittelwerte der Schadstoffe (Halbstundenwerte) bzgl. Tagestyp in $\mu g/m^3$ 55

1 Einleitung

1.1 Die Messstation

In Graz gibt es insgesamt acht Luftgütemessstellen, an denen die Belastung verschiedener Schadstoffe gemessen wird. Die Daten, die in dieser Bakkalaureatsarbeit untersucht werden, stammen dabei aus der Messstation Graz-Süd. Diese befindet sich im Tiergartenweg 10, 8055 Graz, auf etwa 350 Meter Seehöhe und liegt recht nahe zu den Autobahnen A9 und A2. Die Inbetriebnahme dieser Station erfolgte im April 2003 und hat zum Ziel die „Erfassung der Imissionsstruktur im gewerblich-industriell dominierten Bereich der Landeshauptstadt“[1]. Neben PM_{10} und NO_x werden an dieser Station zusätzlich noch Schwefeldioxid, Ozon und Kohlenmonoxid gemessen. Außerdem werden Windrichtung und -geschwindigkeit erfasst.

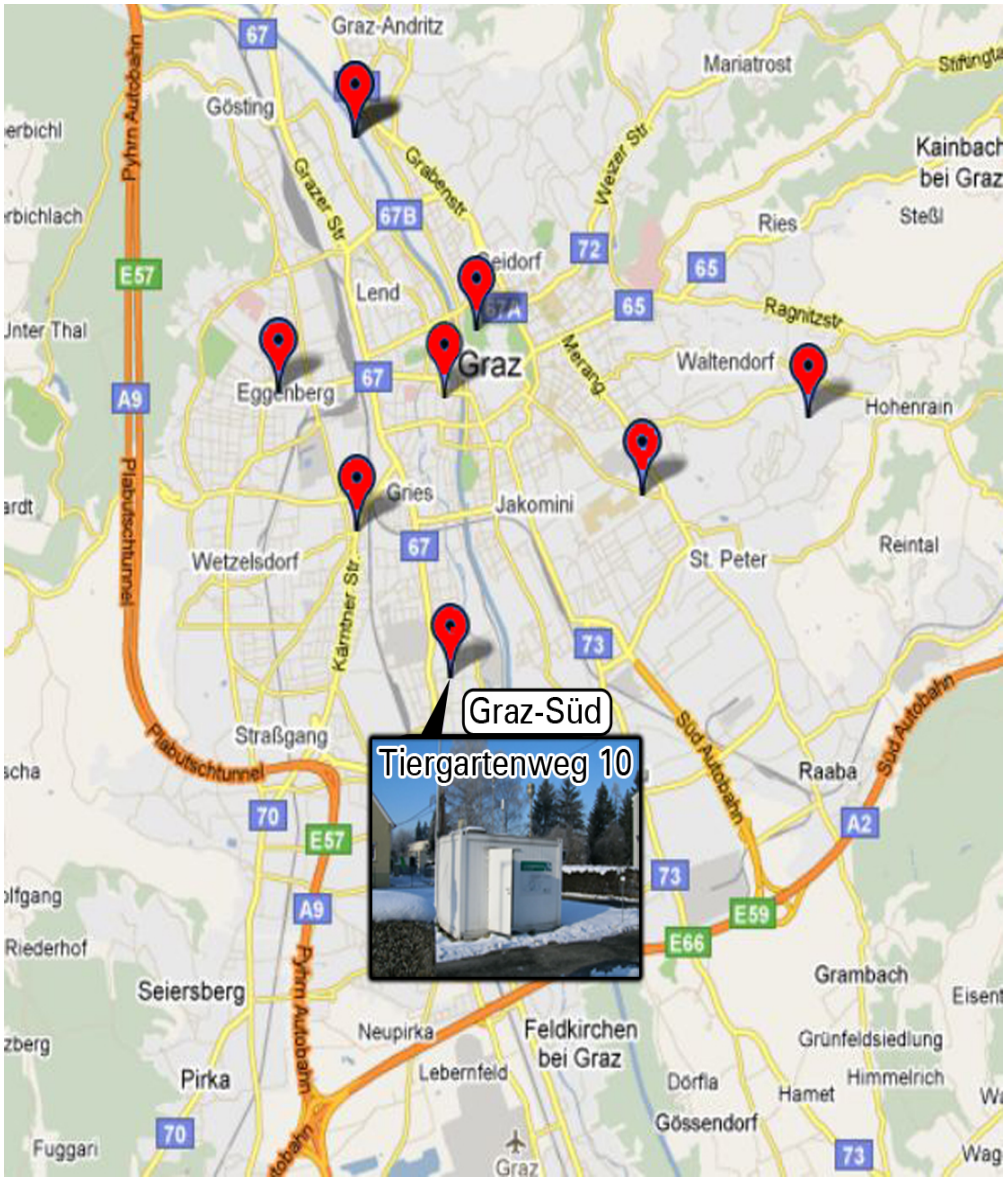


Abbildung 1.1.1: Grazer Messstationen

[1]Umweltinformation Steiermark (http://www.umwelt.steiermark.at/cms/beitrag/10021935/2061730/)

1.2 PM_{10} - Feinstaub

Der sogenannte Schwebstaub umfasst alle festen Teilchen in der Luft und teilt sich in Russ, Rauch und Staub auf. Der Staub besteht aus vielen unterschiedlich großen Partikeln, die sich in Grob- und Feinstaub unterteilen lassen. Teilchen, die dabei kleiner als 10 μm sind werden als inhalierbarer (atembarer) Feinstaub bezeichnet und tragen den Namen PM_{10}, wobei „PM“ für **P**articulate **M**atter steht. Feinstaub kann aus natürlichen, aber auch menschlichen Quellen stammen. So sind einige Hauptverursacher für Feinstaub:

- Pflanzenpollen
- Vulkanausbrüche
- Waldbrände
- Holzheizungen
- Wirtschaft (Industrie)
- Verkehr (vor allem Dieselmotoren)

Es gilt bereits als erwiesen, dass zahlreiche gesundheitliche Probleme, wie etwa Atemwegsbeschwerden, Lungenkrebs und Zunahme von asthmatischen Anfällen, mit der Feinstaubbelastung in Verbindung stehen. Auch ein Zusammenhang mit Herz-Kreislauf-Erkrankungen (z.B. Herzinfarkt) wird vermutet. Dabei spielt vor allem die Größe der Teilchen eine Rolle: Je kleiner das Teilchen, desto gefährlicher ist es für den Körper, weil es tiefer in die Lunge eindringen kann.
Aufgrund dieser Risiken wurde durch die Europäische Union ein Grenzwert festgelegt, der seit dem 1.Jänner 2005 bei 50 μg/m^3 für den Tagesmittelwert von PM_{10} liegt. Dabei sind seit 2010 nur mehr 25 Überschreitungen pro Jahr erlaubt, während es früher 30 waren.
Außerdem wurde ein Grenzwert von 40 μg/m^3 für den Jahresmittelwert von PM_{10} bestimmt. Sollten die Grenzwerte deutlich überschritten werden, so sind die zuständigen Behörden verpflichtet Luftreinhaltepläne aufzustellen. Aus diesem Grund gibt es auf der Südautobahn A2 bei Graz einen Streckenabschnitt, auf dem die Geschwindigkeitsbeschränkung von 130 km/h auf 100km/h gesenkt wird, sobald die Belastung durch PM_{10} zu groß geworden ist (Immissionsschutzgesetz Luft (IG-L)- Gebiet).

1.3 NO_x - Stickstoffoxide

Stickstoffoxide NO_x entstehen hauptsächlich bei der Verbrennung von Brenn- und Treibstoffen bei hoher Temperatur. Dabei werden in Österreich 60% der Stickoxide durch den Verkehr verursacht. Besonders Lkw und Dieselfahrzeuge tragen zu hohen Stickoxidwerten in der Luft bei.
Stickstoffoxide sind mitverantwortlich für die Bildung von Feinstaub, verstärken die Erderwärmung und die schädliche Wirkung von bodennahem Ozon. Sie führen aber auch zu starken gesundheitlichen Problemen, da sie die Atmungsorgane reizen und schädigen können (besonders NO_2 (Stickstoffdioxid)).
Die Grenze für den erlaubten Tagesmittelwert für Stickstoffdioxid (NO_2) liegt bei 80 Mikrogramm pro Kubikmeter Luft, während sich der vom Imissionsschutzgesetz vorgeschriebene maximale Jahresmittelwert auf 40 μg/m^3 beläuft. Ab dem Jahr 2012 gilt aber sogar der Jahresgrenzwert von 30 Mikrogramm pro Kubikmeter. Für Stickstoffmonoxid (NO) gibt es zurzeit noch keinen gesetzlichen Grenzwert.

1.4 Statistische Kenngrößen

In dieser Arbeit werden einige wichtige statistische Kenngrößen und Formeln verwendet, die der Vollständigkeit halber hier angeführt werden sollten.

Definition 1. (Lagemaße)

1. **Arithmetisches Mittel**
   Sei $(x_1, ..., x_n)$ eine Stichprobe aus metrischen Daten.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{heißt arithmetisches Mittel.} \tag{1}$$

2. **Median (Zentralwert)**
   Für die geordnete Stichprobe $(x_{(1)}, ..., x_{(n)})$ unterscheidet man zwei Fälle:

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

$$n\ gerade: \ \ Median\ x_{med} = \tfrac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}). \tag{3}$$

Definition 2. (Streuungsmaße)

1. **Stichprobenvarianz**
   Sei $(x_1, ..., x_n)$ metrisch skaliert.

$$s_L^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \quad \text{heißt Stichprobenvarianz} \tag{4}$$

2. **Standardabweichung**

$$s_L = \sqrt{s_L^2} \quad \text{definiert die Standardabweichung} \tag{5}$$

3. **Quantile**
   Analog zum Median, der die Daten in etwa gleich große Hälften teilt, kann man das so genannte p-Quantil q_p (0<p<1) der Stichprobe definieren. Dieses teilt die Daten so, dass mindestens p% der Daten $\leq q_p$ und mindestens $(1-p)\% \geq q_p$ sind.
   Allgemein definiert man das p-Quantil q_p als:

$$q_p = (1-g)x_{(\lfloor(n-1)p\rfloor+1)} + gx_{(\lfloor(n-1)p\rfloor+2)} \tag{6}$$

   mit $g = (n-1)p - \lfloor(n-1)p\rfloor$.

   Spezielle Quantile:
   $q_{0.25}$ ... 1. Quartil, $q_{0.5}$ ...2. Quartil (Median), $q_{0.75}$ ...3. Quartil

4. **IQR-Standardabweichung**
   iqr=$q_{0.75} - q_{0.25}$ heißt interquartiler Bereich.

$$s_q = \frac{iqr}{1.349} = \frac{q_{0.75} - q_{0.25}}{1.349} \quad \text{definiert die iqr-Standardabweichung.} \tag{7}$$

Definition 3. Schiefemaße
Diese Maße geben Abweichungen von der Symmetrie einer Verteilung an.

1. **Momentkoeffizient der Schiefe**

$$g_1^m = \frac{m_3}{s^3} \quad mit \;\; m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3 \tag{8}$$

Definition 4. Zusammenhangsmaß
Der Pearson-Korrelationskoeffizient gibt die Stärke des linearen Zusammenhangs zwischen zwei Merkmalen an.

1. **Korrelationskoeffizient von Pearson**
   Es liege eine paarige Stichprobe $(x_1, y_1), ..., (x_n, y_n)$ von metrischen Merkmalen X und Y vor.

$$r = r_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{s_{XY}}{s_X s_Y} \tag{9}$$

   heißt Korrelationskoeffizient von Pearson mit

$$s_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \quad \text{empirische Standardabweichung von X,}$$

$$s_Y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad \text{empirische Standardabweichung von Y,}$$

$$s_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \quad \text{empirische Kovarianz von (X,Y).}$$

Definition 5. Korrelationsfunktion

1. **Autokorrelationsfunktion**
   Die Autokovarianzfunktion c_k beschreibt die Korrelation einer Funktion mit sich selbst zu unterschiedlichen Zeitpunkten (lag k) und ist definiert durch:

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(x_t - \overline{x})(x_{t+k} - \overline{x}), k = 0, 1, \ldots, K \tag{10}$$

   Die Autokorrelationsfunktion r_k ergibt sich als $r_k = \frac{c_k}{c_0}$.

2. **Kreuzkorrelationsfunktion**
   Mithilfe der Kreuzkorrelationsfunktion können lineare Zusammenhänge zwischen zwei Funktionen aufgedeckt werden.

$$r_{XY}(k) = \frac{c_{XY}(k)}{N \cdot s_X \cdot s_Y} \tag{11}$$

   heißt Kreuzkorrelationsfunktion zweier Zeitreihen $(x_1, \ldots, x_N)$ und $(y_1, \ldots, y_N)$ mit

$$c_{XY}(k) = \frac{1}{N}\sum_{t=1}^{N-k}(x_t - \overline{x})(y_{t+k} - \overline{y}), k = 0, 1, \ldots \tag{12}$$

2 Explorative Analyse der Daten

2.1 Allgemeine Darstellung der Zeitreihen

Beginnen wir nun mit den ersten Analysen der Feinstaubdaten aus den Jahren 2005 und 2006 in Form von Halbstundenwerten, Vierstundenmittelwerten und Tagesmittelwerten. **PM_{10}**

In der ersten Darstellung (oberste Graphik in Abb. 2.1.1) ist der jährliche Verlauf der Zeitreihe leicht ersichtlich: Zu Beginn des Jahres sind die Werte recht hoch, genauso wie am Ende des Jahres, während die Belastung in den Sommermonaten am Geringsten zu sein scheint. Diese Verhalten ist sowohl im Jahr 2005, als auch 2006 erkennbar. In folgender Tabelle sind außerdem die wichtigsten Eckdaten der Halbstundenwerte aufgelistet:

	2005	2006	Gesamt
Minimum	0.01	0.03	0.01
Maximum	426.53	798.33	798.30
Mean	43.66	43.59	43.63
Median	31.41	32.73	32.17
Varianz	1432.73	1525.64	1479.51
Standardabweichung	37.85	39.60	38.46
1.Quartil	17.75	19.63	18.67
3.Quartil	57.98	53.75	55.87
IQR-Standardabweichung	29.83	25.29	27.58
Momentkoeff. d. Schiefe	1.90	3.22	2.59
Summe der beobachteten Werte	17520	17520	35040
Summe der fehlenden Werte	1056	797	1853

Tabelle 1: Eckdaten der PM_{10} $^1/_2$h-Werte in μg/m^3

Der maximale Wert liegt bei 798.3 μg/m^3, während der niedrigste nur 0.01 μg/m^3 erreicht. Interessant ist auch der Mittelwert der Halbstundenwerte, der sich bei 43.63 μg/m^3 befindet (Bemerkung: der Grenzwert für den Tagesmittelwert liegt bekanntlich bei 50 Mikrogramm pro Kubikmeter Luft).

Die Daten schwanken außerdem mit 38.46 μg/m^3 um den Mittelwert. Zusätzlich ist noch erwähnenswert, dass man anhand der Schiefekoeffizienten erkennen kann, dass die Daten aus dem Jahr 2005 nicht ganz so stark rechtsschief verteilt sind, wie jene aus dem Jahr 2006.

Im Bezug auf die oberste Graphik in Abb. 2.1.1 bleibt außerdem noch zu sagen, dass es insgesamt noch zwei Werte gibt, die deutlich über dem für die Graphik gewählten Bereich von 600 μg/m^3 liegen und deshalb nicht mehr dargestellt werden. Allerdings wird hier auch deutlich, dass diese erste Darstellung nur dafür geeignet ist, sich einen Überblick über die Daten zu verschaffen. Möchte man nähere Details betrachten, so sollte man die extrem hohen Werte (Ausreißer) nicht berücksichtigen und außerdem auf eine andere Darstellungsart zurückgreifen (siehe Kapitel 2.2).

Aus diesem Grund berechnen wir nun aus den vorhandenen Halbstundenmittelwerten erneut Mittelwerte aus jeweils 8 Werten, d.h. wir berechnen Vierstundenmittelwerte. Dadurch reduziert sich die Datenmenge von 35040 auf 4380 Werte und die Darstellung wird übersichtlicher, wie an der mittleren Graphik zu erkennen ist. Hier wurde außerdem die y-Skala auf 400 μg/m^3 verkürzt, um eine detaillierte Ansicht zu bekommen. Man erkennt denselben Jahresverlauf wie zuvor, allerdings sind die Daten aufgrund der Mittelung geglättet. Der

Maximalwert liegt hier z.B. nur mehr bei 550.3 μg/m^3 (Tab. 2).

	2005	2006	Gesamt
Minimum	0.74	3.14	0.74
Maximum	343.66	550.27	550.27
Mean	43.48	43.49	43.48
Median	31.90	33.19	32.64
Varianz	1287.01	1387.89	1337.48
Standardabweichung	35.87	37.25	36.57
1.Quartil	18.49	20.18	19.40
3.Quartil	57.58	53.40	55.64
IQR-Standardabweichung	28.97	24.63	26.86
Momentkoeff. d. Schiefe	1.84	3.02	2.46
Summe der beobachteten Werte	2190	2190	4380
Summe der fehlenden Werte	110	81	191

Tabelle 2: Eckdaten der PM_{10} 4h-Werte in μg/m^3

Außerdem erkennt man, dass der maximale Wert der Vierstundenmittelwerte deutlich niedriger liegt, als jener der Halbstundenwerte. Median und Mittelwert sind hingegen in etwa gleich geblieben, wie auch z.B. die IQR-Standardabweichung.
Fehlende Werte bei den 4h-Werten bedeuten, dass acht aufeinanderfolgende Datenpunkte fehlen, und somit kein Mittelwert daraus berechnet werden konnte, was hier noch 191 Mal der Fall war.
Anschließend nehmen wir noch eine weitere Glättung vor und berechnen die Tagesmittelwerte, sodass wir schlussendlich nur mehr über 730 Daten verfügen, die in der untersten Graphik in Abb. 2.1.1 dargestellt werden.
Bei der Darstellung der Tagesmittelwerte konnte nun schon eine y-Skalierung von 250 μg/m^3 gewählt werden, da der größte Wert nur mehr bei 228.4 μg/m^3 liegt.

	2005	2006	Gesamt
Minimum	6.78	5.60	5.60
Maximum	195.33	228.41	228.41
Mean	43.54	43.32	43.43
Median	33.02	34.94	33.90
Varianz	958.44	1041.55	998.76
Standardabweichung	30.96	32.27	31.60
1.Quartil	21.41	22.99	22.35
3.Quartil	56.74	52.78	54.06
IQR-Standardabweichung	26.19	22.08	23.51
Momentkoeff. d. Schiefe	1.48	2.17	1.85
Summe der beobachteten Werte	365	365	730
Summe der fehlenden Werte	14	11	25

Tabelle 3: Eckdaten der PM_{10} Tagesmittelwerte in μg/m^3

Die fehlenden Werte bedeuten hier, dass für komplette 25 Tage keine Messwerte vorhanden sind, was vor allem für den September im Jahr 2005 der Fall ist, da in diesem Zeitraum 11 von 25 Tagesmittelwerte fehlen.

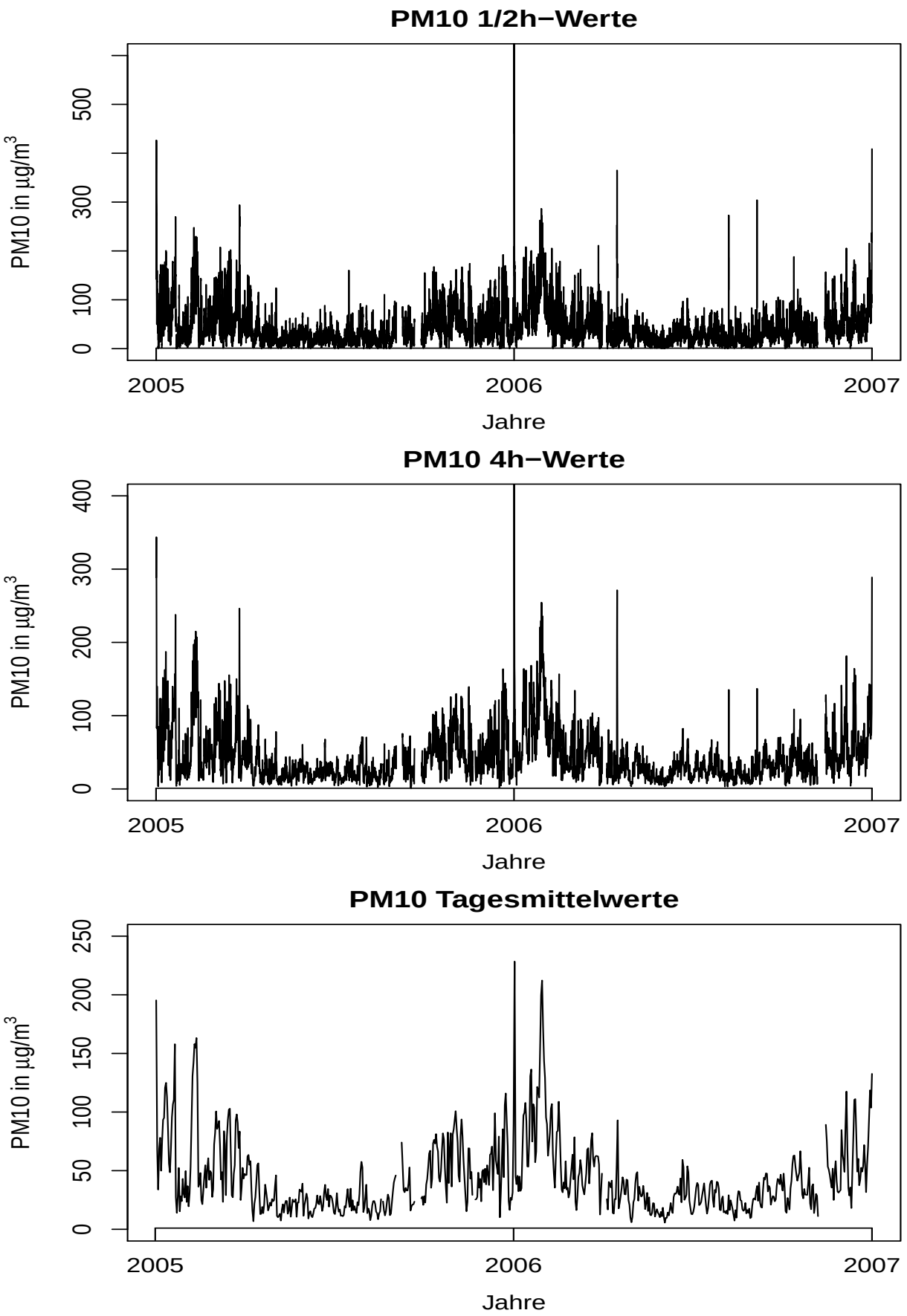


Abbildung 2.1.1: Darstellung der PM_{10} $^1/_2$h-, 4h- und Tagesmittelwerte

NO_x Für die Betrachtung der NO_x-Werte wurden abermals Graphiken der $^1/_2$h-, 4h- und Tagesmittelwerte erstellt.
Auch für die $^1/_2$h-Werte des Schadstoffes NO_x erkennen wir bei der Betrachtung der graphischen Darstellung (oberste Graphik aus Abb. 2.1.2) einen typischen jährlichen Verlauf, bei dem die Belastung in den Wintermonaten höher ist als in den Sommermonaten. Im Vergleich zu den PM_{10}-Werten sind die Stickoxidwerte höher und man erkennt eine größere Schwankung der Daten. Der höchste Wert liegt hier bei 550.8 μg/m^3 und der niedrigste bei 0.33 μg/m^3. Weitere Details sind wiederum der nachfolgenden Tabelle zu entnehmen.

	2005	2006	Gesamt
Minimum	0.53	0.33	0.33
Maximum	529.90	550.8	550.8
Mean	52.17	54.31	53.24
Median	24.99	29.35	27.01
Varianz	4402.45	4095.85	4250.05
Standardabweichung	66.35	64.00	65.19
1.Quartil	12.74	14.40	13.49
3.Quartil	63.31	69.68	66.72
IQR-Standardabweichung	37.49	40.98	39.46
Momentkoeff. d. Schiefe	2.40	2.38	2.39
Summe der beobachteten Werte	17520	17520	35040
Summe der fehlenden Werte	851	826	1677

Tabelle 4: Eckdaten der NO_x $^1/_2$h-Werte in μg/m^3

Im Vergleich zu den Feinstaubdaten weisen die NO_x Daten eine höhere Varianz auf, was wir in der Graphik an den starken Schwankungen erkennen können. Zudem sind die Stickstoffoxiddaten offensichtlich noch etwas mehr rechtsverteilt, als zuvor die Feinstaubdaten, was man anhand des Schiefekoeffizienten erkennt.

Betrachtet man die Darstellung der 4h-Werte in der mittleren Graphik, so erkennt man sofort, dass die Daten durch die Mittelung geglättet wurden. Außerdem konnten wir die y-Skala auf einen Höchstwert von 500 μg/m^3 senken, da darüber keine Werte mehr vorhanden sind. Da wir bei den Halbstundenwerten einige ziemlich hohe Werte in unmittelbarer Nähe zueinander aber keine eindeutigen Spitzen haben, hat sich auch das Maximum nicht stark verändert und liegt nun bei 475.4 μg/m^3.

	2005	2006	Gesamt
Minimum	1.59	1.69	1.59
Maximum	475.37	426.41	475.37
Mean	52.40	54.30	53.35
Median	27.22	31.03	29.20
Varianz	3900.52	3541.02	3720.60
Standardabweichung	62.45	59.51	61.00
1.Quartil	13.77	16.07	14.77
3.Quartil	66.00	71.25	68.56
IQR-Standardabweichung	38.71	40.90	39.88
Momentkoeff. d. Schiefe	2.36	2.25	2.31
Summe der beobachteten Werte	2190	2190	4380
Summe der fehlenden Werte	55	50	105

Tabelle 5: Eckdaten der NO_x 4h-Werte in $\mu g/m^3$

In der untersten Graphik in Abb. 2.1.2, die die Tagesmittelwerte des Schadstoffes NO_x zeigt, wurde eine y-Skalierung von 300 gewählt, da hier die Daten weiter geglättet wurden und das Maximum nur mehr bei 293 $\mu g/m^3$ liegt. Außerdem haben wir nun nur noch 15 fehlende Werte, von denen rund 50% aus dem Oktober im Jahr 2005 stammen. Desweiteren erkennt man eine Reduktion der Varianz und somit der Standardabweichung. Die Daten schwanken also nicht mehr so stark. Zudem wurde außerdem die Rechtsschiefe etwas ausgeglichen, was man am Schiefekoeffizient von 2.01 erkennt.

	2005	2006	Gesamt
Minimum	4.69	4.78	4.69
Maximum	293.02	283.49	293.02
Mean	52.46	54.44	53.45
Median	32.63	35.80	34.46
Varianz	2606.68	2427.76	2514.56
Standardabweichung	51.06	49.27	50.15
1.Quartil	19.75	22.75	21.08
3.Quartil	66.28	71.90	70.77
IQR-Standardabweichung	34.50	36.44	36.83
Momentkoeff. d. Schiefe	2.02	1.99	2.01
Summe der beobachteten Werte	365	365	730
Summe der fehlenden Werte	8	7	15

Tabelle 6: Eckdaten der NO_x Tagesmittelwerte in $\mu g/m^3$

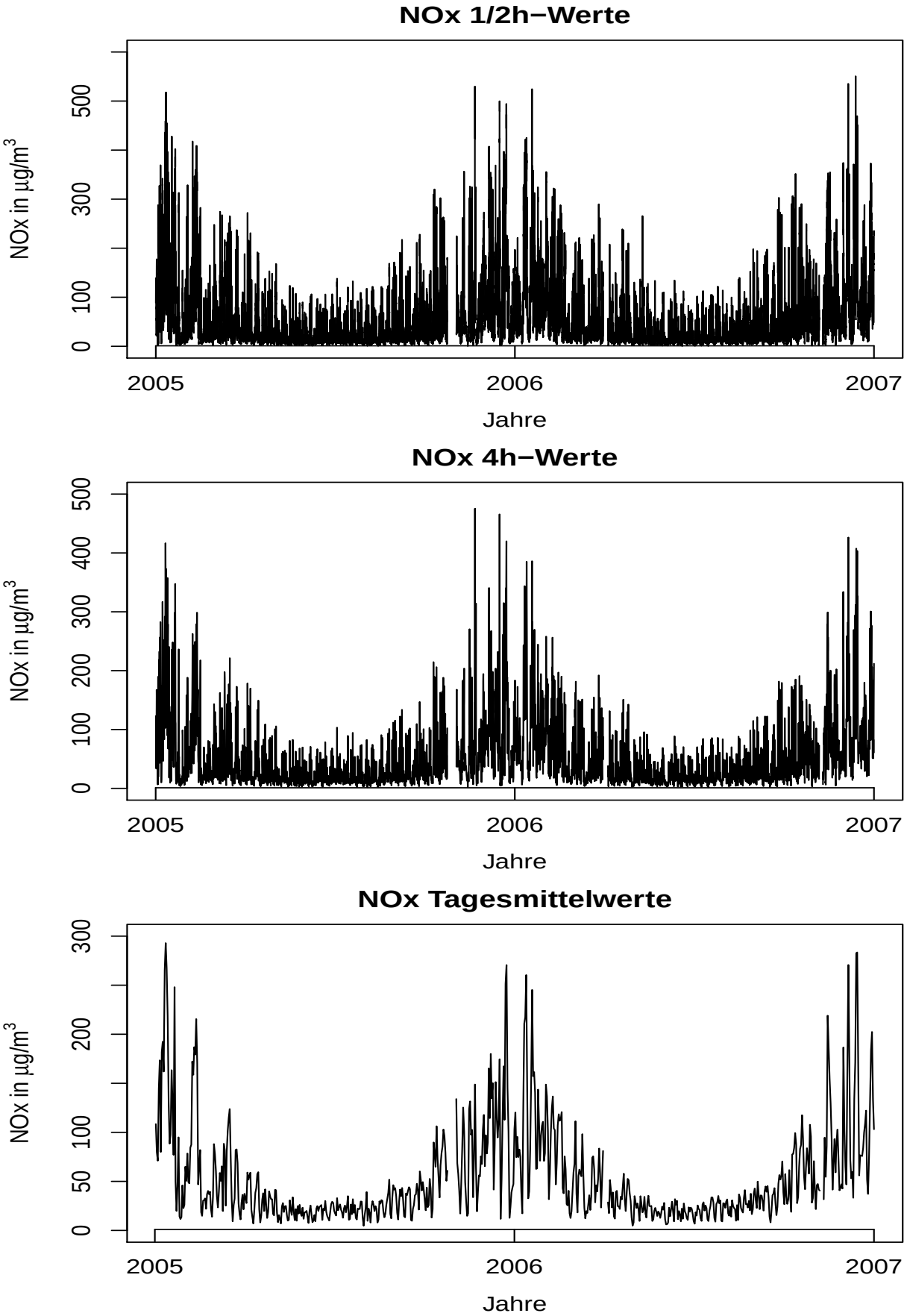


Abbildung 2.1.2: Darstellung der NO_x $^1/_2$h-, 4h- und Tagesmittelwerte

Um die Verteilung der Daten besser untersuchen zu können, entscheiden wir uns für die graphische Darstellung durch Boxplots. Diese haben den großen Vorteil, dass wir mit einem Blick erkennen können, wo sich die zentralen 50% der Daten und der Median befinden und ob bzw. wo viele Ausreißer vorhanden sind.

2.2 Aufspaltung der Zeitreihen in Abhängigkeit der Saison

Zu Beginn teilen wir die vorhandenen Daten in Sommer- und Winterdaten, also nach Saison, ein. Dabei unterscheiden wir auch, ob die Daten im Jahr 2005 oder 2006 erhoben wurden. Als „Sommer“ bezeichnen wir hierbei die Daten vom 1.April bis zum 30.September. Demnach bedeutet „Winter“ die Zeit vom 1.Oktober bis zum 31.März. Mit dieser Aufspaltung wollen wir die unterschiedliche Belastung in den Sommer- und Wintermonaten verdeutlichen.

Die ersten Boxplots stellen die $^1/_2$h-Werte von PM_{10}, aufgeteilt in Sommer und Winter, dar (Abb. 2.2.1) **PM_{10}**

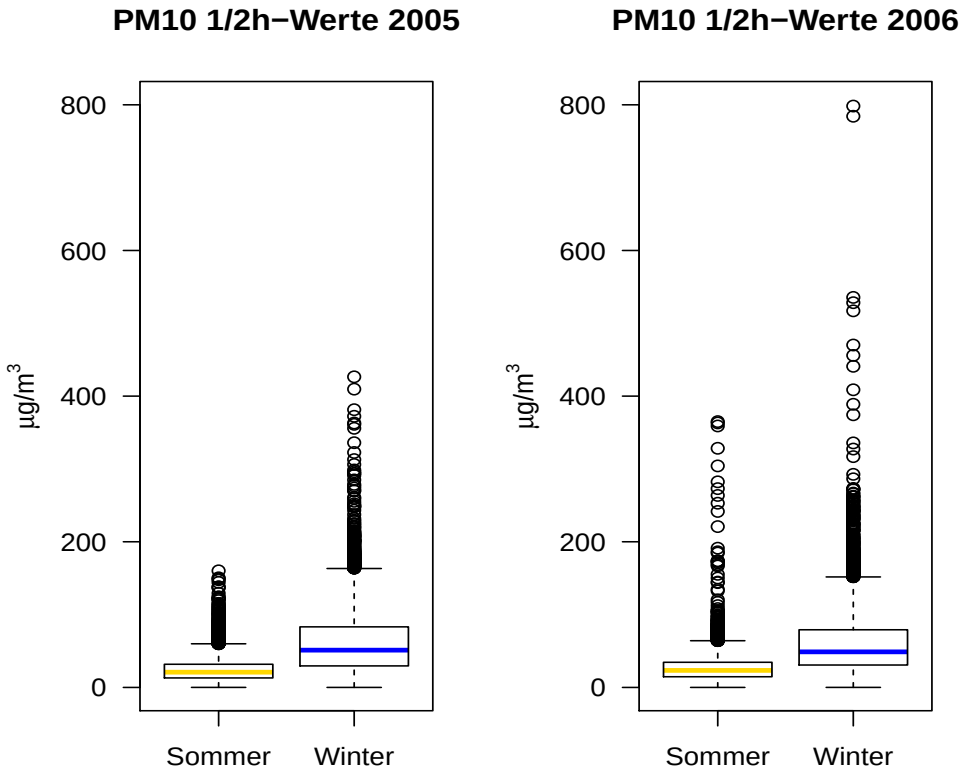


Abbildung 2.2.1: Darstellung PM_{10} $^1/_2$h-Werte in Abhängigkeit der Saison

Dabei fällt sofort auf, dass die Belastung durch PM_{10} im Sommer um einiges geringer ist, als im Winter. Außerdem streuen die Daten im Sommer weit weniger als im Winter, was wir an der Höhe der Boxen erkennen können. Dieses Verhalten konnten wir auch schon in den vorigen Graphiken beobachten. Im Jahresvergleich fällt vor allem auf, dass im Jahr 2006, sowohl im Sommer als auch im Winter, extremere Ausreißer vorkommen. Markant sind dabei vor allem zwei Winter-Ausreißer, die bei 789.3 und 784.3 $\mu g/m^3$ liegen. Diese werden wir im nächsten Kapitel noch genauer untersuchen.

Aber auch die Ausreißer im Sommer fallen im Jahr 2006 deutlich höher aus als im Jahr zuvor. Denn während die Ausreißer im Jahr 2005 unter 200 μg/m^3 liegen, erreichen sie im Jahr 2006 fast 400μg/m^3.
Da sich die Sommer- von den Winterwerten aber sehr stark voneinander unterscheiden und viele hohe Ausreißer vorhanden sind, ist es schwer sie in einer gemeinsamen Graphik zu analysieren. Deshalb scheint es sinnvoll, die Sommer- und Winterdaten jeweils noch einmal in eigenen Graphiken zu betrachten. Außerdem verkürzen wir die y-Skala auf 70 bzw. 180 μg/m^3, da wir dadurch die Boxen besser erkennen und interpretieren können, wofür wir die Ausreißer nicht benötigen.

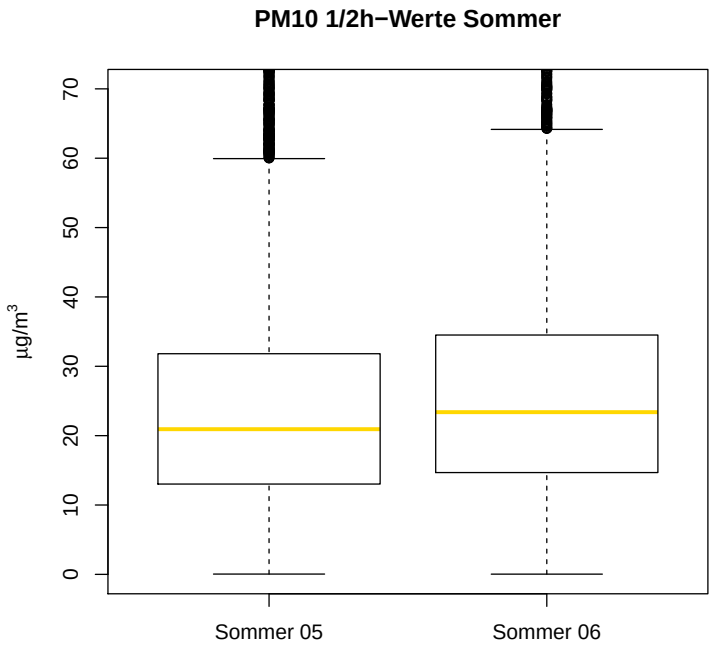


Abbildung 2.2.2: PM_{10} $^1/_2$h-Werte Sommer

In der Detailansicht (Abb. 2.2.2) erkennt man, dass die zentralen 50% der Daten im Jahr 2006 höher als im Vorjahr liegen, wobei die Höhe der Boxen ungefähr gleich zu sein scheint. Gleichzeitig hat sich auch der Median der Daten vom Jahr 2005 zum Jahr 2006 etwas nach oben verschoben. Lag dieser früher bei 20.93, so stieg er im Jahr 2006 auf 23.39 an. Aus diesen Beobachtungen lässt sich schließen, dass die Verteilungsstruktur in etwa gleichgeblieben ist, sich allerdings weiter nach oben verschoben hat. Außerdem liegt sowohl im Sommer 2005 als auch 2006 eine leicht rechtsschiefe Verteilung vor.

Auch für die genauere Betrachtung der PM_{10}-Winterdaten wurde die y-Skala verkürzt und liegt in der folgenden Graphik (Abb. 2.2.3) bei 180 $\mu g/m^3$.

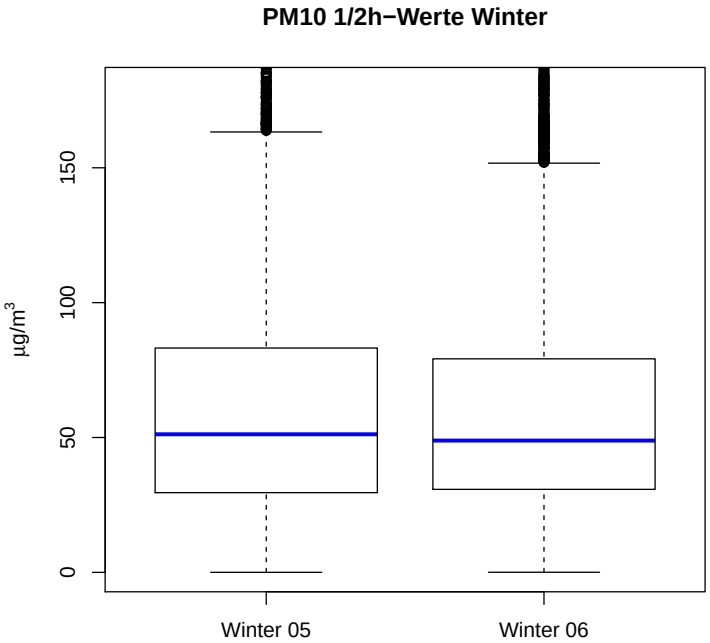


Abbildung 2.2.3: PM_{10} $^1/_2$h-Werte Winter

An der Graphik erkennt man, dass die zentralen 50% der Daten im Winter 2005 einen etwas größeren Bereich beanspruchen als im darauffolgenden Jahr. Für eine genauere Interpretation wollen wir uns aber noch die Summary der beiden Winterdatensätze genauer anschauen:

Winterdaten 2005:

```
Min.   1st Qu.  Median   Mean  3rd Qu.   Max.    NA's
0.01    29.55   51.19   61.42   83.17   426.50   235
```

Winterdaten 2006:

```
Min.   1st Qu.  Median   Mean  3rd Qu.   Max.    NA's
0.04    30.78   48.87   60.88   79.17   798.30   449
```

An den Summaries lässt sich gut ablesen, dass die Werte im Winter 2006 (ohne Ausreißer) etwas niedriger sind als im Vorjahr. Allerdings fehlen im Winter 2006 fast doppelt so viele Werte wie noch einen Winter zuvor. Da diese nicht vorhandenen Werte einiges verändern hätten können, lässt sich nicht genau sagen, ob es deutliche Unterschiede zwischen den beiden Datensätzen gibt. Zumindest anhand der gegebenen Werte lässt sich eine große Ähnlichkeit der beiden Winterdatensätze erkennen.
Auffallend ist jedoch, dass der Mittelwert der Winterdaten (Halbstundenwerte) um 60 $\mu g/m^3$ liegt, wobei 50 $\mu g/m^3$ dem Tagesmmittelgrenzwert für den Feinstaub entspricht. Auch wenn man diese Werte nicht eins zu eins vergleichen kann, lässt sich dennoch sagen, dass die Belastung im Winter zu hoch ist.

NO_x Nun werden auch die NO_x Daten in Sommer und Winter unterteilt (Abb. 2.2.4).

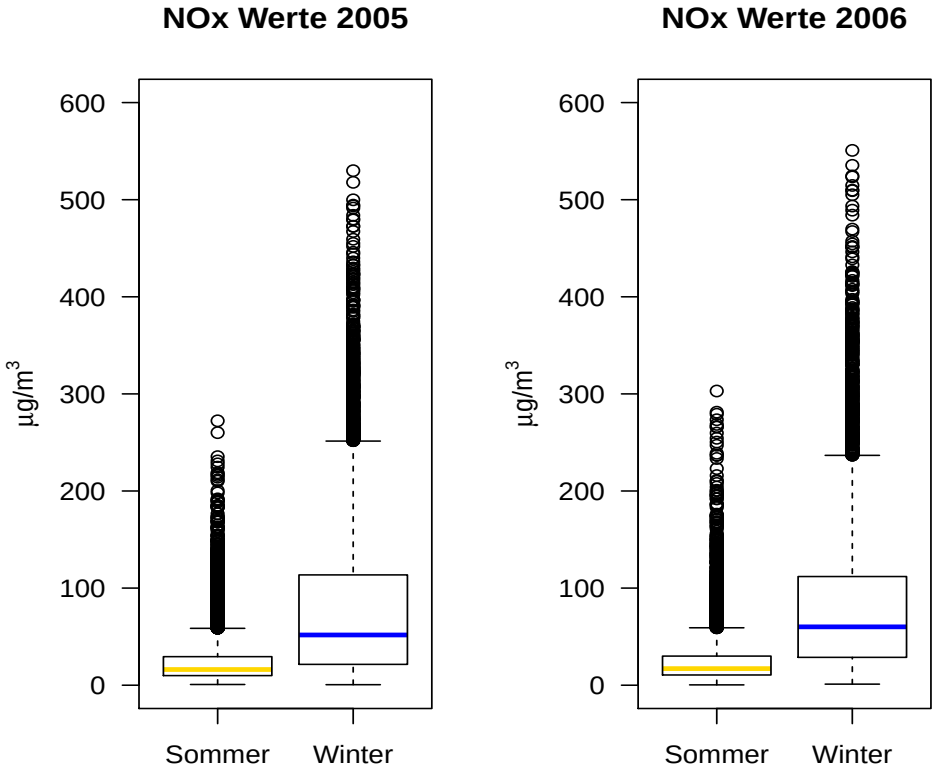


Abbildung 2.2.4: Darstellung NO_x $^1/_2$h-Werte in Abhängigkeit der Saison

Im Gegensatz zu den PM_{10} Daten sieht man hier allerdings kaum Veränderungen im Jahresvergleich, jedoch eine größere Streuung bzgl. Winter/Sommer. Die Ausreißer scheinen im Jahr 2006 nur minimal höher zu liegen, als im Jahr 2005. Um eventuelle Unterschiede anhand der Boxen feststellen zu können, sollten wir aber noch eine Detailansicht der Sommer- bzw. Wintermonate betrachten. Für die Detailansicht der Sommerdaten ergibt sich Abb. 2.2.5.

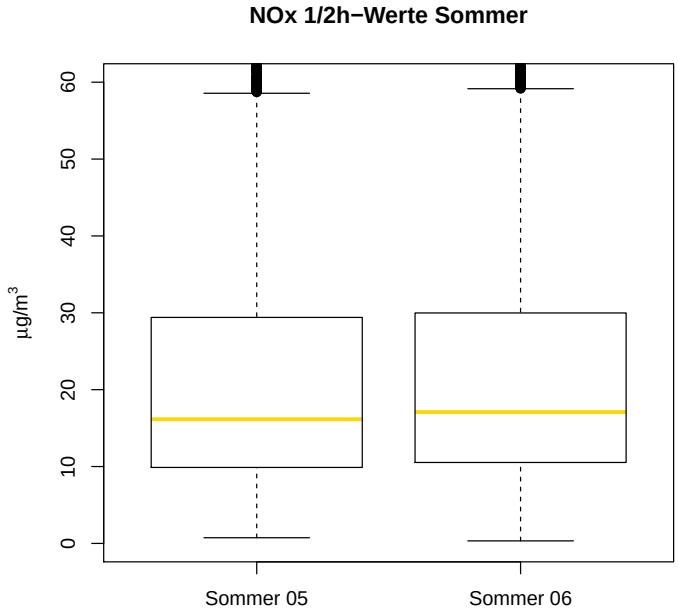


Abbildung 2.2.5: Darstellung NO_x $^1/_2$h-Werte Sommer

Die zwei Boxen der Sommerdaten scheinen tatsächlich fast identisch zu sein, was bedeutet, dass die Verteilung der NO_x Daten im Sommer 2005 und 2006 sehr ähnlich sein wird und es keine gravierenden Unterschiede gibt.

Zusätzlich betrachten wir noch die NO_x Winterdaten in einer Detailansicht (Abb. 2.2.6).

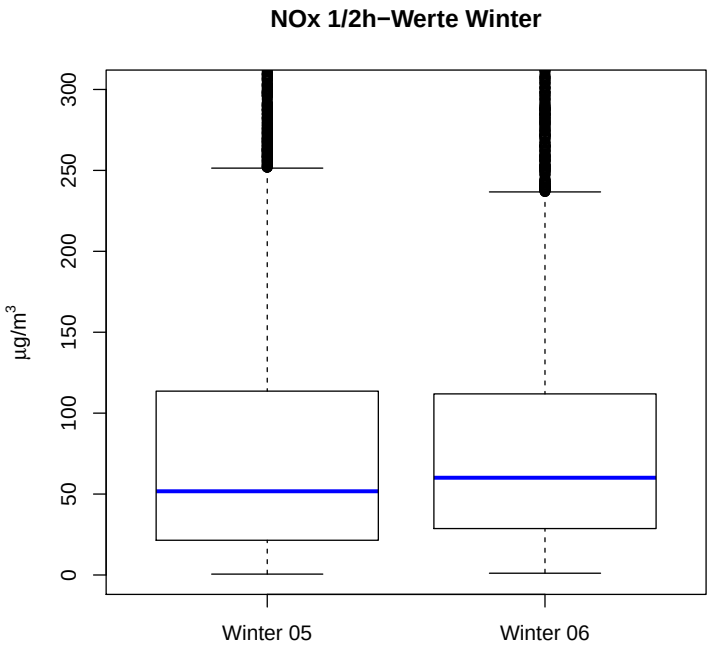


Abbildung 2.2.6: Darstellung NO_x $^1/_2$h-Werte Winter

Die Boxen sind sich hier abermals sehr ähnlich, wobei im Winter 2005 niedrigere Werte häufiger vorkommen, als im darauffolgenden Winter. Dies lässt sich am niedrigeren Median erkennen. Die Verteilung ist aber in beiden Fällen noch leicht rechtsschief. Ansonsten erkennt man nur, dass der obere Whisker im Winter 2005 etwas länger ist, als im Folgejahr, was aber auf die Größe der Box zurückzuführen ist.

Auf weitere Darstellungen mit 4h- und Tagesmittelwerten soll an dieser Stelle verzichtet werden. Stattdessen gehen wir zum nächsten Kapitel über, in dem wir die Daten nach Monaten unterteilen.

2.3 Aufspaltung der Zeitreihen in Abhängigkeit von Monaten

Nachdem wir die Daten zuerst in Sommer und Winter geteilt und anschließend analysiert haben, wenden wir uns jetzt einer anderen Unterteilung zu. Wir wissen bereits, dass die Belastung durch PM_{10} und NO_x in den Wintermonaten höher als in den Sommermonaten ist. Allerdings können wir noch nicht einschätzen, wie hoch die Belastung in den einzelnen Monaten tatsächlich ist. Dafür teilen wir die Daten in Monate ein und plotten diese Monatsdaten.

PM_{10} Beginnen wir bei PM_{10}, so erkennen wir wieder den typischen Jahresverlauf mit geringerer Belastung im Sommer und hohen Werten in den Wintermonaten (Abb. 2.3.1).

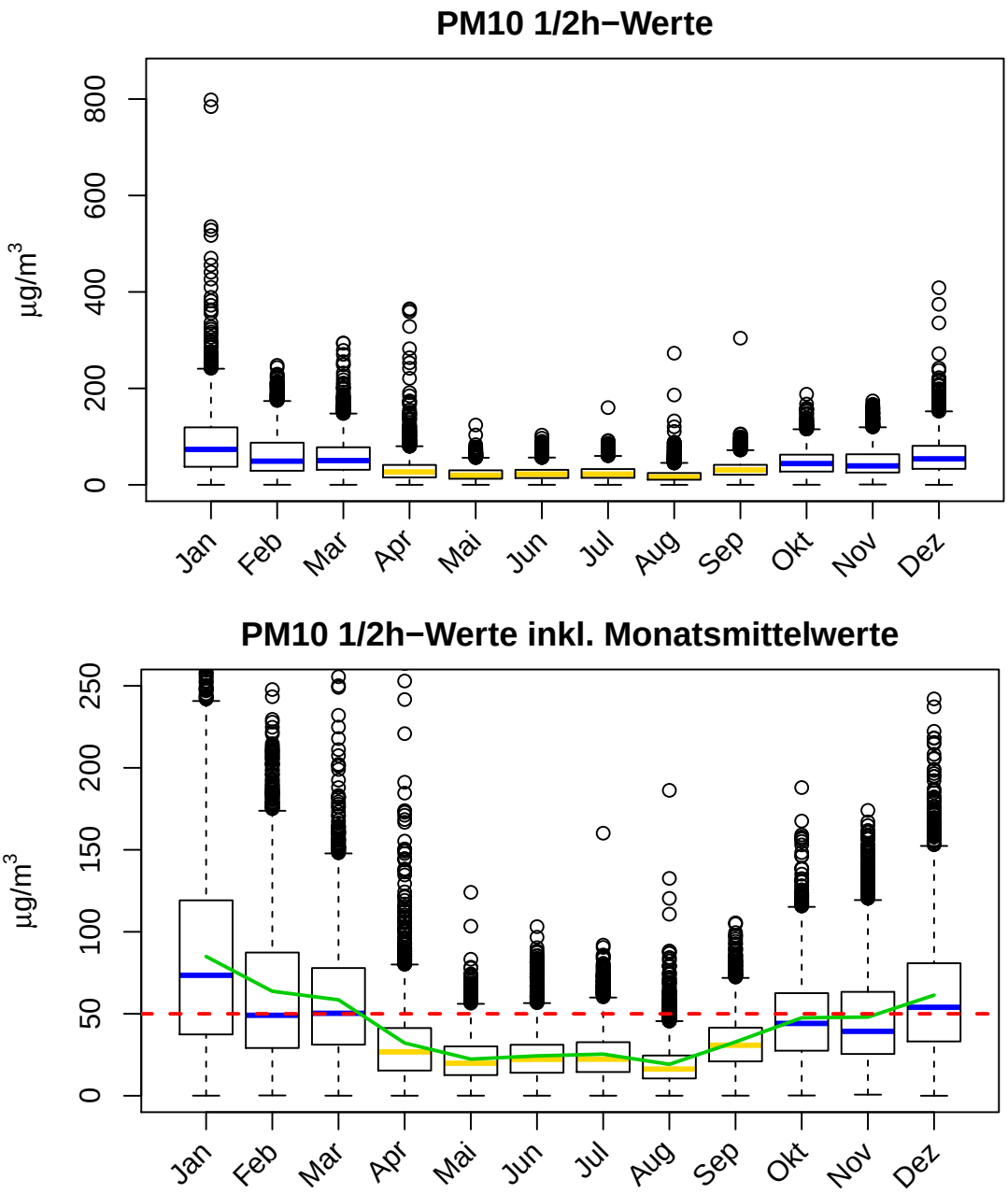


Abbildung 2.3.1: PM_{10} $^1/_2$h-Werte inkl. Monatsmittelwerte

Die obere Darstellung in Abb. 2.3.1 mit derart hoch gewählten Grenzen dient hauptsächlich der Lokalisierung von Ausreißern, über die wir auch tatsächlich einen guten Überblick bekommen. Besonders auffallend sind zwei Ausreißer die wir schon zuvor beobachten konnten und die mit den Werten 798.33 und 784.29 μg/m^3 deutlich höher sind als alle anderen Ausreißer. Anhand dieser Darstellung erkennen wir nun, dass diese Ausreißer im Jänner gemessen wurden. Erinnern wir uns an dieser Stelle kurz daran, dass die Grenze der erlaubten Schadstoffbelastung für PM_{10} bei 50 μg/m^3 liegt, so haben wir eine Vorstellung davon, wie stark die Belastung zu den Zeitpunkten der gemessenen Ausreißern wirklich war. Diese wurden am 1.1.2006 um 0:30 Uhr bzw. am selben Tag um 1 Uhr früh (Spitzenwert) gemessen. Auch die nächsthöheren Ausreißer in diesem Monat wurden an diesem Tag verzeichnet. Die höchsten Ausreißer im Dezember kommen außerdem am 31. Dezember vor. Hier liegt also die Vermutung nahe, dass die zahlreichen Feuerwerke zum Jahreswechsel die Ursache für derart hohe Werte sind.
Des weiteren fallen recht hohe Ausreißer im Monat April auf. Diese liegen um die 350 μg/m^3 und wurden um Mitternacht herum gemessen. Wirft man allerdings einen Blick in den Kalender, so wird man feststellen, dass der Tag dieser Beobachtungen (16.04.2006) der Ostersonntag war. Daher ist es sehr wahrscheinlich, dass die hohe Schadstoffbelastung an diesem Tag auf die traditionellen Osterfeuer zurückzuführen ist, die erst im Jahr 2007 in Graz verboten wurden. Die viel zu hohen Feinstaubwerte am Karsamstag bzw. Ostersonntag sind auch genau der Grund dafür, dass dieses Verbot eingeführt wurde.
Im September gibt es noch einen einzelnen auffälligen Ausreißer, der nachmittags am 5.9.2006 gemessen wurde. Dieser Tag liegt in der letzten Woche der Sommerferien, jedoch ist nicht ersichtlich, wieso hier ein so hoher Wert gemessen wurde.

Zur genaueren Betrachtung der Boxplots wird die y-Skala nun auf 250 μg/m^3 verkürzt, um die Boxen der Boxplots besser erkennen zu können (untere Grafik in Abb. 2.3.1)
Da die Größe der Box anzeigt, wie stark die Daten schwanken, sehen wir hier wieder eine recht große Schwankungsbreite in den Wintermonaten (vor allem im Jänner), und eine geringe in den Sommermonaten. Der Median befindet sich in den Sommermonaten zwischen 16.32 und 30.81 μg/m^3, während er in den Wintermonaten zwischen 39.29 und 73.46 μg/m^3 schwankt.
Des weiteren sind die unterschiedlichen Mittelwerte pro Monat noch interessant und wurden deshalb in der Graphik in Farbe (grün) eingezeichent. Die expliziten Werte werden in folgender Tabelle auch noch dargestellt:

Monat	Jänner	Februar	März	April	Mai	Juni
Mittelwert	84.94	63.74	58.52	32.17	22.35	24.32
Monat	Juli	August	September	Oktober	November	Dezember
Mittelwert	25.39	19.20	32.84	47.55	48.00	61.34

Tabelle 7: PM_{10}-Mittelwerte pro Monat (Halbstundenwerte) in μg/m^3

Man erkennt hier noch, dass die Mittelwerte für die Monate Juni und Juli etwas höher sind, als eigentlich durch die Parabelform der Mittelwerte zu erwarten wäre.
Zusätzlich zu den Mittelwerten pro Monat wurde außerdem der gesetzlich erlaubte Grenzwert von 50 μg/m^3 eingezeichnet (rot). Dieser gilt zwar für die Tagesmittelwerte, allerdings lässt sich damit trotzdem erahnen, in welchen Monaten die Feinstaubbelastung zu hoch ist. Das gilt dabei offensichtlich vor allem für die Wintermonate.

Betrachten wir nun dieselbe Darstellung für die 4h-Mittelwerte (Abb. 2.3.2).

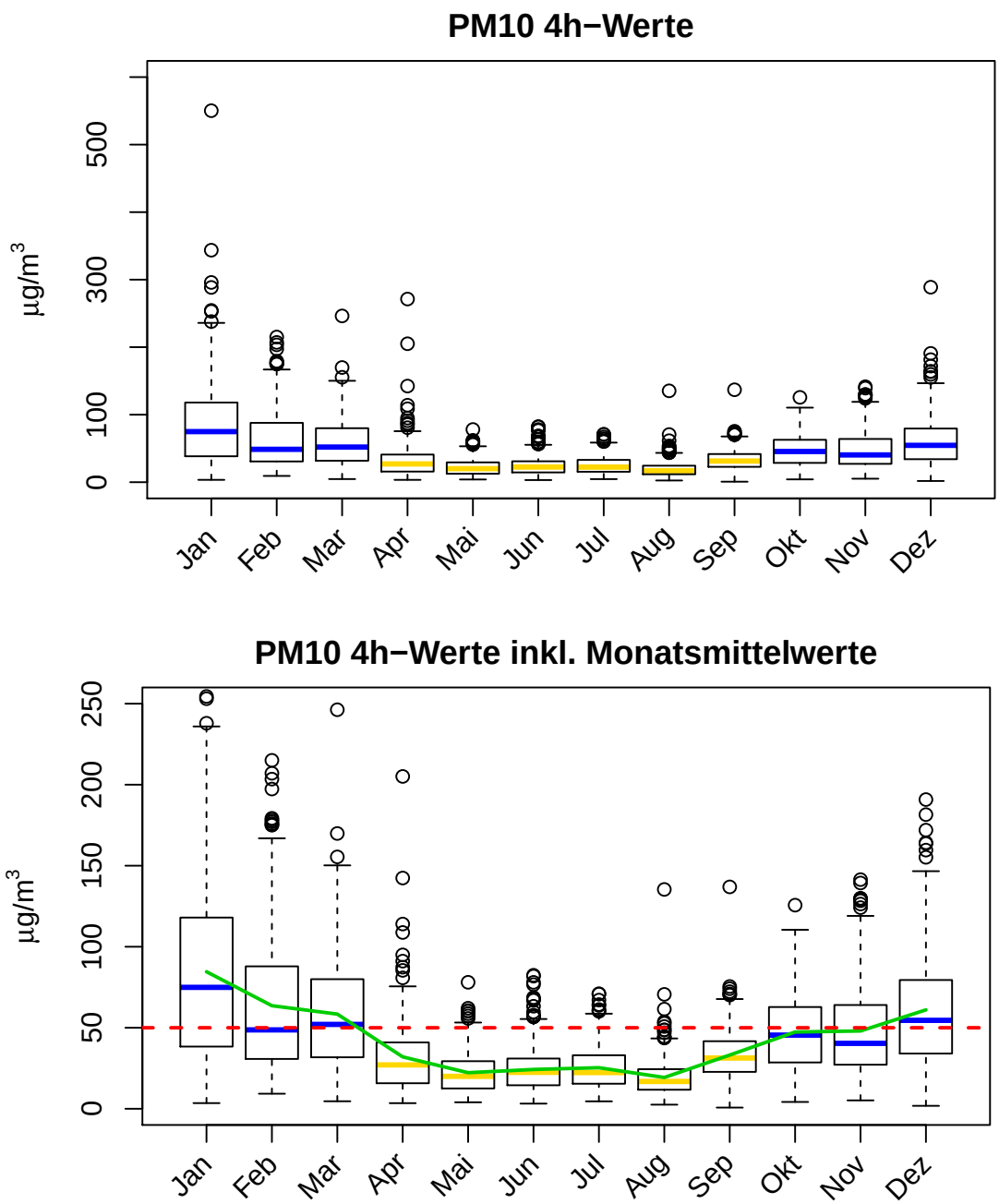


Abbildung 2.3.2: PM_{10} 4h-Werte inkl. Monatsmittelwerte

In der Darstellung der 4h-Mittelwerte sind zwar noch einige Ausreißer zu erkennen, allerdings weit weniger als bei den Halbstundenwerten.
Auch hier erstellen wir eine Detailansicht mit eingezeichneten Mittelwerten pro Monat (grün) und dem gesetzlich erlaubten Grenzwert (für Tagesmittelwerte) von 50 μg/m^3 (rot), um die Boxen besser interpretieren zu können (Abb. 2.3.2).
Im Wesentlichen kommt man bei dieser Graphik zu denselben Erkenntnissen wie zuvor: Es lässt sich eine hohe Belastung in den Wintermonaten und eine geringere Belastung im Sommer feststellen. Hier haben wir allerdings nicht mehr 35040, sondern 4380 Daten.

Zu guter Letzt werfen wir noch einen Blick auf die Tagesmittelwerte, die in Abb. 2.3.3 dargestellt werden.

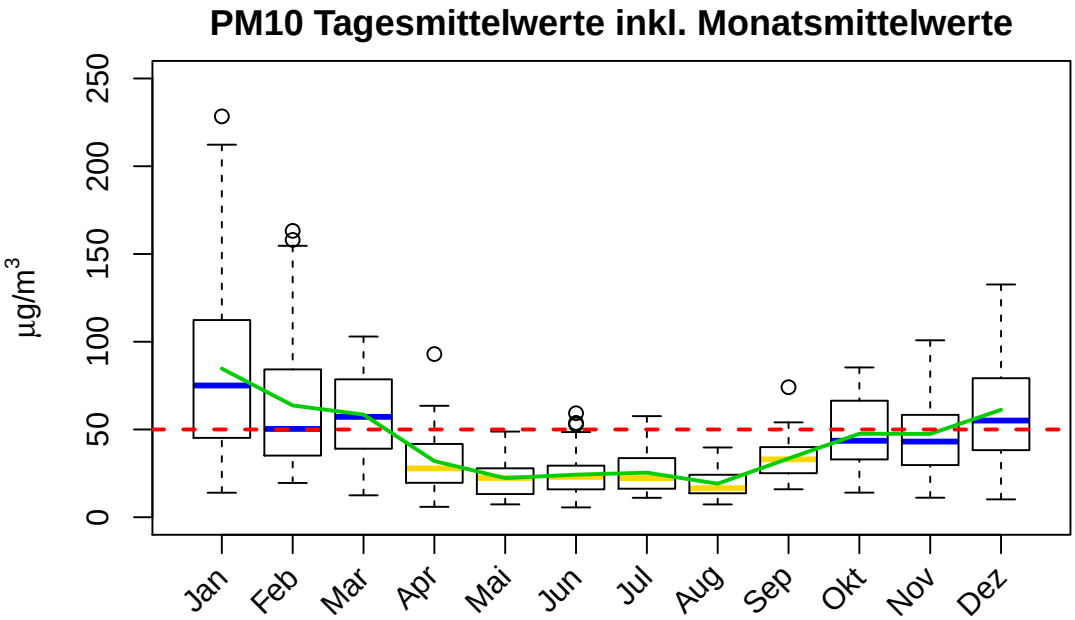


Abbildung 2.3.3: PM_{10} Tagesmittelwerte inkl. Monatsmittelwerte

Natürlich erwarten wir bei dieser Darstellung keine großartigen Überraschungen mehr. Es gibt hier nun kaum Ausreißer und man sieht einen recht schönen Jahresverlauf. Wie sich anhand der Lage des Medians feststellen lässt, liegen rund 50% der Tagesmittelwerte im Februar unter 50 $\mu g/m^3$, was auch für die anderen Wintermonate, bis auf den Jänner, annähernd zutrifft. Das sich der Mittelwert im Februar aber über der gesetzlichen Grenze befindet wird daran liegen, dass es einige recht hohe Werte in diesem Monat gibt, die den durchschnittlichen Wert deutlich anheben.
Im Gegensatz zu den Wintermonaten bleiben in den Sommermonaten so gut wie alle Tagesmittelwerte unter der 50 $\mu g/m^3$ Grenze.

NO_x Dieselben Darstellungen verwenden wir nun auch für die NO_x-Werte, beginnend mit Halbstundenwerten in Abb. 2.3.4.

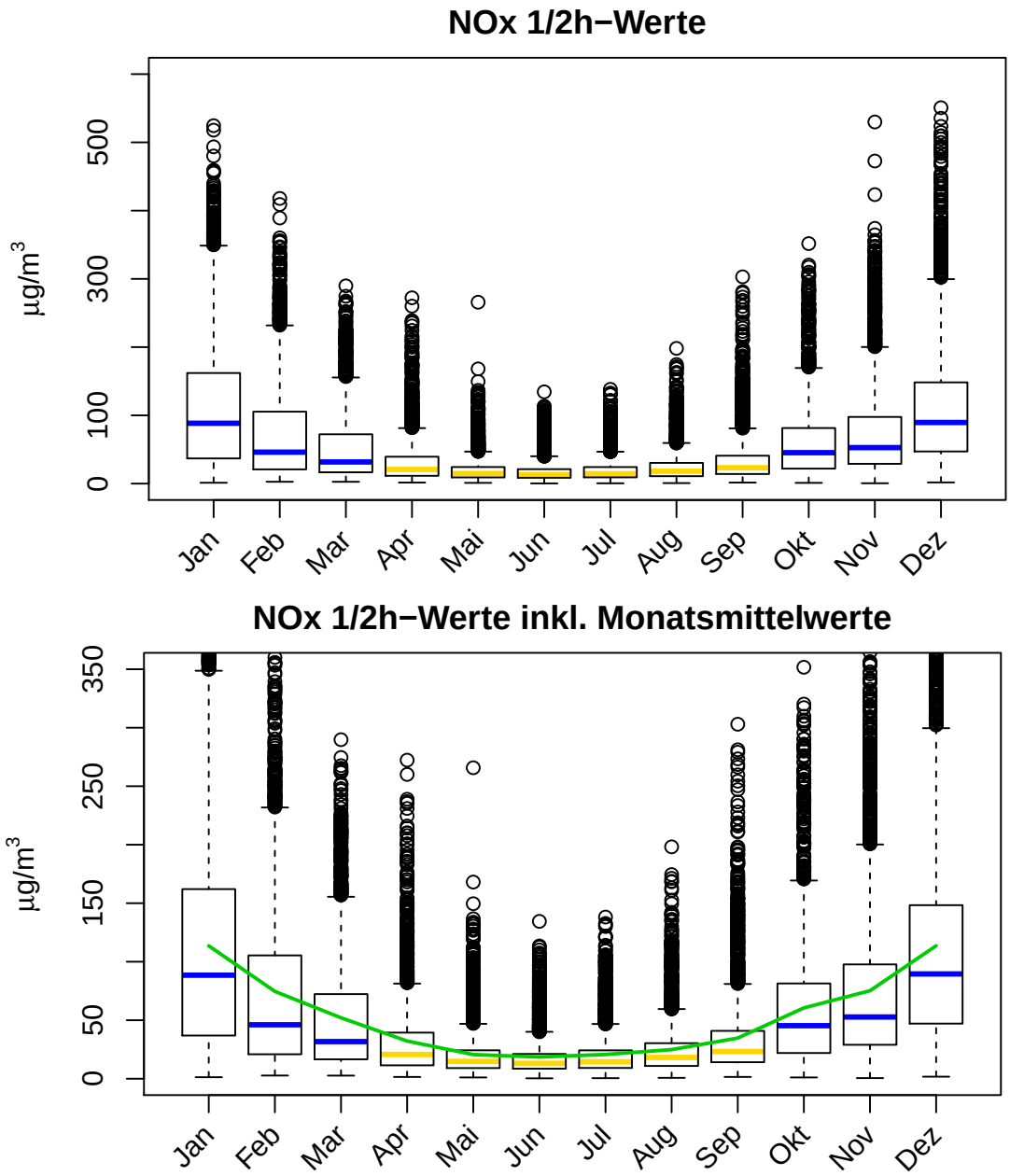


Abbildung 2.3.4: NO_x $^1/_2$h-Werte inkl. Monatsmittelwerte

Im Gegensatz zu den PM_{10} Daten kommen bei den NO_x Daten nicht nur im Jänner, sondern auch im Dezember und November hohe Ausreißer vor. Außerdem sind die Ausreißer im Allgemeinen höher und zahlreicher.
Neben den hohen Ausreißern im Jänner und Dezember, die, wie schon zuvor die PM_{10}-Ausreißer, zum Jahreswechsel beobachtet wurden, ist ein Ausreißer im Mai auffallend, der am 10.5.2006 nachmittags beobachtet wurde. Dieser Tag scheint aber auf den ersten Blick kein signifikantes Datum zu haben, und da die Werte vor und nach diesem Extremwert mit 37.49 bzw. 19.77 $\mu g/m^3$ nicht auffallen, lässt sich dieser hohe Wert nicht ohne weitere Informationen erklären.

Auch im November gibt es drei außergewöhnliche Ausreißer, die allesamt in der Früh des selben Tages, dem 21.11.2005, gemessen wurden. Ein Grund für diese hohen Werte könnte die Tatsache sein, dass dieser Tag ein Montag war, und der Verkehr montags in den Morgenstunden bekanntlich sehr stark ist.
Im Allgemeinen sieht man aber auch in dieser Graphik einen deutlichen Jahresverlauf, bei dem die Werte am Jahresbeginn sehr hoch sind, bis zur Mitte des Jahres abfallen, um dann wieder bis zum Ende des Jahres anzusteigen.
Die zweite Graphik (untere Darstellung in Abb. 2.3.4) zeigt die Detailansicht der monatlichen NO_x Halbstundenwerte, wobei die y-Skala auf 350 $\mu g/m^3$ verkürzt wurde, da abermals die höchsten Ausreißer aus der Graphik entfernt wurden, um die zentralen 50% pro Monat besser interpretieren zu können. Allerdings sind die oberen Whiskers (vor allem im Jänner) so hoch, dass die Boxen der Sommermonate dennoch sehr klein bleiben.
Zusätzlich wurden die Monatsmittelwerte eingezeichnet, die auch in der folgenden Tabelle nachzulesen sind.

Monat	Jänner	Februar	März	April	Mai	Juni
Mittelwert	113.58	74.63	52.08	32.17	20.64	18.59
Monat	Juli	August	September	Oktober	November	Dezember
Mittelwert	20.72	24.88	34.74	60.52	75.22	113.63

Tabelle 8: NO_x-Mittelwerte pro Monat (Halbstundenwerte) in $\mu g/m^3$

Die Monatsmittelwerte weisen eine deutliche Struktur auf, da sie von Jänner bis Juni monoton fallen um anschließend wieder monoton anzusteigen.
Anhand der Graphik erkennt man außerdem, dass sich die zentralen 50% der Daten vor allem im Jänner und Dezember (aber auch Februar und November) in einem recht großen Bereich befinden, d.h. die Werte der Wintermonate schwanken sehr stark, während die Werte im Sommer nicht nur um einiges niedriger sind, sondern auch eine geringere Variabilität aufweisen.

Reduzieren wir nun die Datenmenge und betrachten die Darstellung der 4h-Mittelwerte für den Schadstoff NO_x (Abb. 2.3.5).

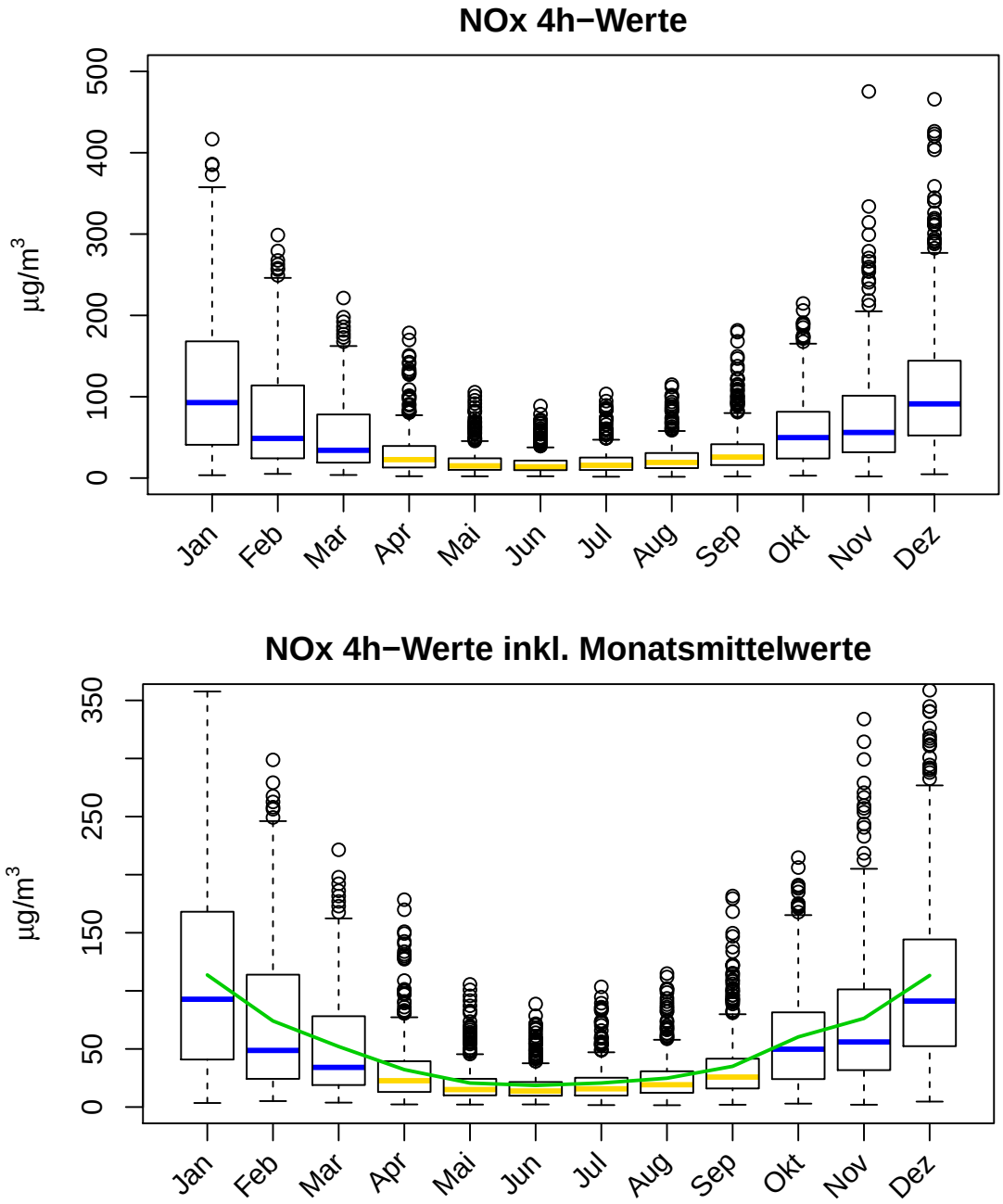


Abbildung 2.3.5: NO_x 4h-Werte inkl. Monatsmittelwerte

Auf den ersten Blick sieht man, dass wir nun deutlich weniger Ausreißer haben als zuvor. Doch für weitere Interpretationen ist es unerlässlich, eine Detailansicht zu betrachten, in der wir abermals die Monatsmittelwerte eingezeichnet haben (untere Graphik in Abb. 2.3.5). Hier lässt sich wieder derselbe jährliche Verlauf mit hohen Winter- und niedrigen Sommerwerten erkennen. Auch der parabelförmige Verlauf der monatlichen Mittelwerte der 4h-Werte ist gut zu sehen.

Zum Schluss plotten wir noch die Tagesmittelwerte des Schadstoffes NO_x pro Monat und zeichnen die Mittelwerte der einzelnen Monate ein (Abb. 2.3.6).

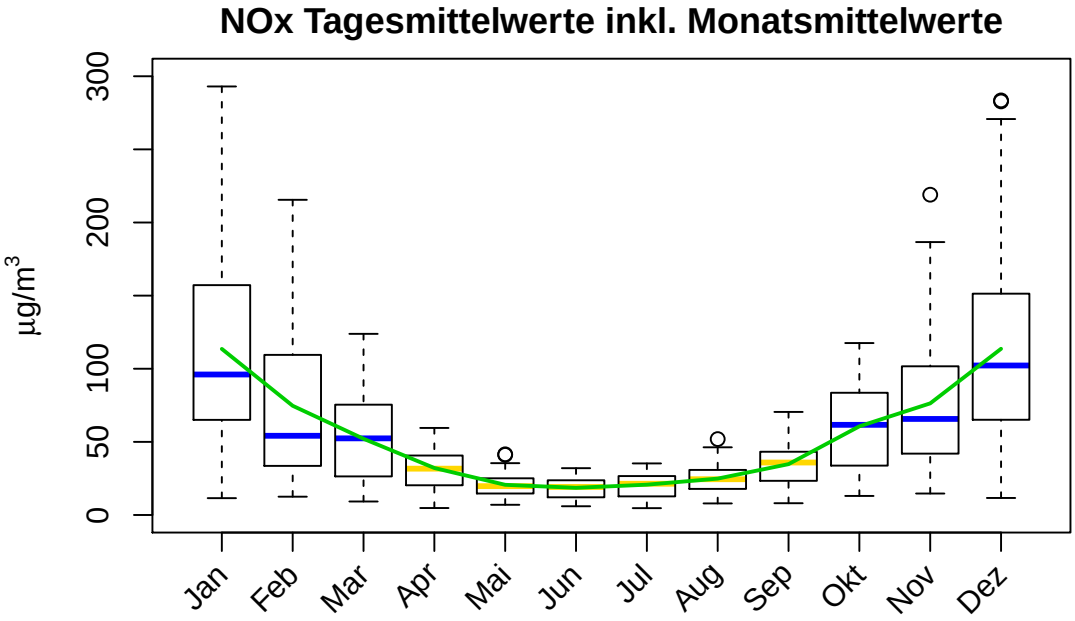
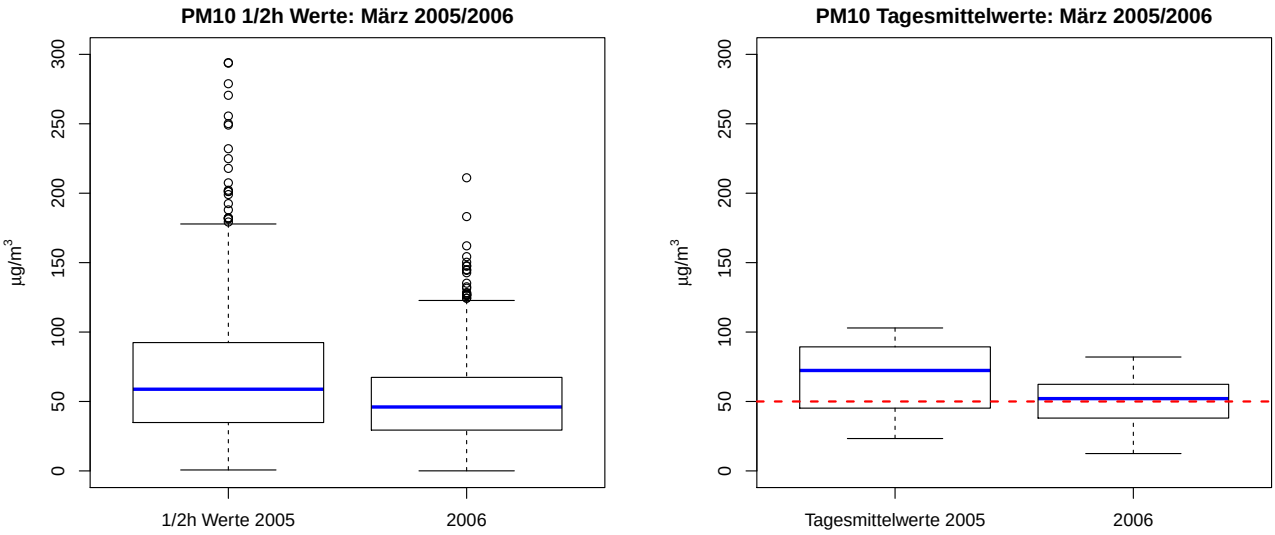


Abbildung 2.3.6: NO_x Tagesmittelwerte inkl. Monatsmittelwerte

Man erkennt hier wieder, dass die Belastung durch Stickstoffoxide von Mai bis August sehr gering ist und sogar unter 50 $\mu g/m^3$ liegt. Am Ende bzw. Anfang des Jahres werden hingegen teilweise sehr hohe Werte gemessen und auch die Mittelwerte befinden sich teilweise weit über 50 $\mu g/m^3$.
Die verbleibenden Ausreißer in dieser Graphik wurden im Mai, August, November und Dezember gemessen, wobei es sich nur mehr um sechs Werte handelt. Abgesehen davon erkennt man wieder den großen Unterschied zwischen den Sommer- und Wintermonaten aufgrund der unterschiedlichen Höhe der Boxen und deren Mediane.

Nun haben wir einen Eindruck davon gewonnen, wie die Daten innerhalb des Jahres in etwa verteilt sind, wann die höchsten und niedrigsten Schadstoffwerte gemessen wurden und wo viele Ausreißer vorhanden sind. Dabei haben wir allerdings keine Unterscheidung zwischen Daten aus dem Jahr 2005 und 2006 gemacht. Daher wollen wir noch zwei Monate genauer betrachten, bei denen es einen sichtbaren Unterschied zwischen Werten aus dem Jahr 2005 und 2006 gibt.
Zuerst untersuchen wir die Schadstoffbelastung durch PM_{10} im Monat März (Abb. 2.3.7):

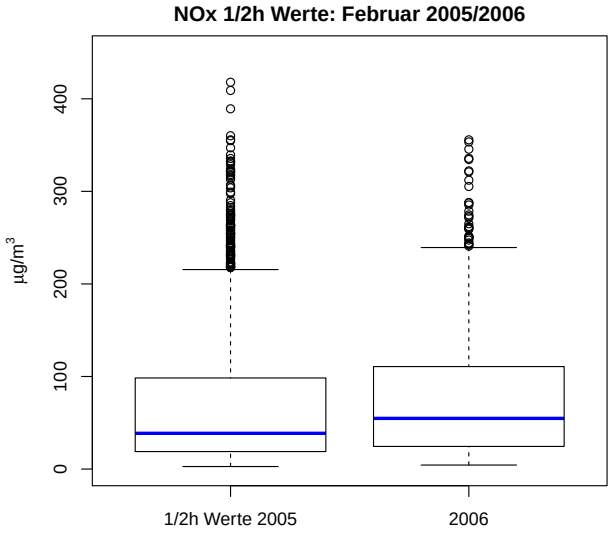


(a) PM_{10} $^1/_2$h-Werte: März 2005/2006 **(b)** PM_{10} Tagesmittelwerte: März 2005/2006

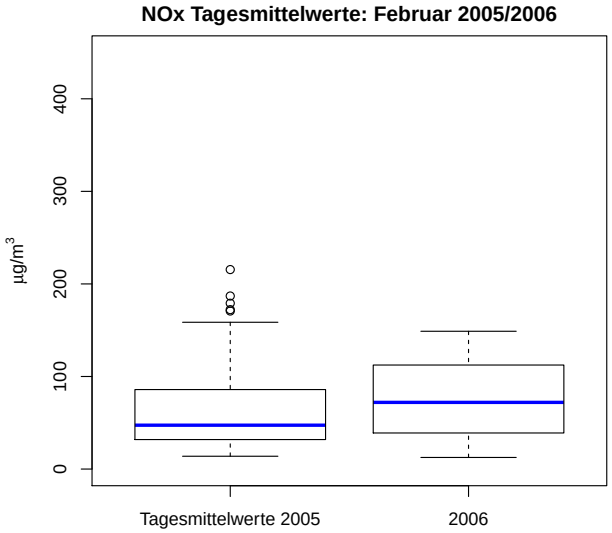
Abbildung 2.3.7: PM_{10} im Monat März 2005-2006

Anhand der Halbstundenwerte erkennen wir, dass im März 2005 extremere Ausreißer vorkommen, als ein Jahr darauf im selben Monat. Die sechs höchsten Ausreißer im Jahr 2005 wurden dabei in der Nacht vom 26. auf den 27.März gemessen und liegen zwischen 250 und 300 μg/m^3. Da es sich hierbei um die Karsamstagnacht handelt, liegt der Grund für die hohen Schadstoffwerte höchstwahrscheinlich an den zahlreichen Osterfeuern.
Auch die zwei höchsten Ausreißer im März 2006 wurden in einer Nacht, vom 27. auf den 28.3., gemessen. Da das Osterwochenende in diesem Jahr allerdings erst Mitte April stattfand, muss es einen anderen Grund als die Osterfeuer für die hohe Schadstoffbelastung geben.
Die Daten im Jahr 2005 weisen außerdem eine größere Variabilität auf, was man anhand der größeren Box erkennen kann. Dadurch ist auch der obere Whisker größer als im Jahr 2006.
Vergleicht man die beiden Mediane der Daten, so erkennt man, dass der Median im März 2005 mit 58.83 μg/m^3 doch deutlich größer ist als der Median im März 2006 mit 46.03 μg/m^3. Das bedeutet, dass im Monat März im Jahr 2005 höhere Werte gemessen wurden als im Folgejahr.
Betrachtet man zusätzlich die Tagesmittelwerte dieser Daten, so erkennt man, dass sich der Abstand der Mediane der beiden Beobachtungsmonate noch vergrößert hat und der Median im März 2005 höher liegt als bei den Halbstundenwerten. Das liegt daran, dass im März 2005 viele hohe Werte gemessen wurden, die die Tagesmittelwerte in die Höhe heben. Vergleicht man die vorhandenen Daten außerdem mit dem gesetzlich erlaubten Tagesmittelgrenzwert (rot), so erkennt man, dass im März 2005 dieser Grenzwert an drei von vier Tagen überschritten wurde, während es im März 2006 nur an jedem zweiten Tag zu einer Übertretung gekommen ist. Mögliche Gründe könnten dabei sowohl in der unterschiedlichen Temperatur, als auch in unterschiedlichem Windverhalten liegen (vgl. Kapitel 2.5).

Für die Überprüfung einer möglicherweise unterschiedlichen Belastung durch Stickstoffoxide im Jahr 2005 im Gegensatz zu 2006 wurde der Monat Februar herangezogen (Abb. 2.3.8).



(a) NO_x $^1/_2$h-Werte: Februar 2005/2006



(b) NO_x Tagesmittelwerte: Februar 2005/2006

Abbildung 2.3.8: NO_x im Monat Februar 2005-2006

Auch hier erkennt man, dass im Jahr 2005 höhere Ausreißer gemessen wurden als im nächsten Jahr in diesem Monat. Außerdem scheinen die Ausreißer im Februar 2005 zahlreicher zu sein. Dennoch liegt der Median mit 38.61 μg/m^3 sehr niedrig, was zwar bedeutet, dass 50% der Daten unter dieser Grenze liegen, da der obere Whisker allerdings sehr lang ist und es eben darüber noch viele Ausreißer gibt, liegen die oberen 50% der Daten in einem sehr großen Bereich zwischen ca. 70 und 400 μg/m^3 und sind sehr verstreut. Die Verteilung im Februar 2005 ist zudem extrem rechtsschief, während die Daten im Februar 2006 zwar auch rechtsschief verteilt sind, allerdings nicht ganz so stark.
Die höchsten Werte im Februar 2005, sowie einige hohe Werte im Februar 2006, wurden in den Morgenstunden (etwa acht Uhr) gemessen, weswegen der Frühverkehr als Ursache in Frage kommen könnte. Die höchsten Werte im Februar 2006 wurden allerdings an einem bestimmten Abend (1.Februar) gemessen. Ob es für die hohen Werte an diesem Mittwochabend einen besonderen Grund gibt bleibt hier offen. Eine mögliche Ursache wäre hier die extreme Inversionswetterlage, die es im Februar 2006 gegeben hat.
Vergleicht man nun die Graphik der Halbstundenwerten mit jener der Tagesmittelwerte, so fällt in erster Linie auf, dass für den Februar 2005 noch einige Ausreißer vorhanden sind. Das deutet darauf hin, dass bei den Halbstundenwerten einige sehr hohe Werte in unmittelbarer zeitlicher Umgebung gemessen wurden, sodass durch die Bildung der Tagesmittelwerte einige Werte entstanden sind, die noch immer weit von den restlichen Daten entfernt liegen.
Außerdem ist noch interessant, dass die Tagesmittelwerte im Februar 2006 nahezu symmetrisch verteilt zu sein scheinen. Daher wurden über den Tag verteilt sehr unterschiedliche Werte (sehr hohe und sehr niedrige) gemessen, sodass sich die Tagesmittelwerte aller Februartage im Jahr 2006 nahezu symmetrisch um 72.02 μg/m^3 befinden.

2.4 Aufspaltung der Zeitreihen in Abhängigkeit vom Tagestyp

In diesem Kapitel werden wir die vorhandenen Schadstoffwerte aus einem weiteren Blickwinkel betrachten, indem wir die Daten in Tagestypen einteilen. Dadurch erhalten wir drei Gruppen: Werktage, Samstage und Sonn- und Feiertage. Diese Darstellung interessiert uns besonders, da wir dadurch einen genauen Überblick über die unterschiedliche Belastung an Werktagen im Vergleich zum Wochenende erlangen können. Wir vermuten zwar, dass die Belastung an Werktagen am Höchsten ist, können aber noch nicht sagen, um welchen Unterschied im Vergleich zu Samstagen und Sonn- und Feiertagen es sich tatsächlich handelt.

PM_{10} Die folgende Graphik (Abb. 2.4.1, links oben) stellt die PM_{10} Halbstunden-, 4h- und Tagesmittelwerte eingeteilt nach Tagestyp dar.
In der ersten Graphik erkennen wir wieder unsere zwei schon bekannten Spitzenwerte um 800 μg/m^3, die in der Silvesternacht auf den 1.1.2006 gemessen wurden. Aber auch ansonsten findet man die höchsten Ausreißer in der Kategorie „Sonn- und Feiertag“. An dieser Stelle bleibt jedoch offen, ob diese Werte eher in den Sommer- oder den Wintermonaten vorkommen. Auf diese Frage werden wir jedoch etwas später noch eingehen.
Bis auf die Lokalisierung der Ausreißer können wir allerdings anhand dieser Graphik nichts mehr erkennen, da die Boxen relativ gleich groß bzw. klein zu sein scheinen. Für eine genauere Betrachtung verkürzen wir also die y-Skala auf 125 (Graphik rechts oben). Anhand dieser Detailsansicht wird sichtbar, dass die Boxen doch nicht gleich hoch sind, wie es zuvor den Anschein erweckt hat. Die zentralen 50% der Daten in der Kategorie „Werktag“ befinden sich ca. zwischen 20 und 60 μg/m^3, während sich die Boxen der Samstage bzw. Sonn- und Feiertage ca. zwischen 18 und 38 bzw. 15 und 35 μg/m^3 befinden. Diese Eckdaten können auch folgender Summary entnehmen:

Summary PM_{10} $^1/_2$h-Werte: Werktag

```
Min.   1st Qu.   Median    Mean   3rd Qu.    Max.     NA's
0.01    20.28    34.35   46.09    60.70   304.20    1304
```

Summary PM_{10} $^1/_2$h-Werte: Samstag

```
Min.   1st Qu.   Median    Mean   3rd Qu.    Max.     NA's
0.08    17.45    30.15   38.79    48.29   293.70    245
```

Summary PM_{10} $^1/_2$h-Werte: Sonn- und Feiertag

```
Min.   1st Qu.   Median    Mean   3rd Qu.    Max.     NA's
0.05    14.65    25.84   38.02    45.49   798.30    304
```

Auf den ersten Blick mögen die 1304 fehlenden Werte aus „Werktag“ erschreckend wirken und wir fragen uns, ob das nicht zu viele sind. Bedenkt man jedoch, dass wir für diese Kategorie 23904 Daten zur Verfügung haben, können wir unsere Zweifel verwerfen, da uns somit nur 5.5% der Werte fehlen. Im Vergleich dazu fehlen in der Kategorie „Samstag“ sowie „Sonn- und Feiertag“ jeweils 4.9%.
Mithilfe der Graphiken und der Summaries können wir erkennen, dass der Wert des Medians der ersten Kategorie („Werktag“) am Größten ist und über den Wert des Median der zweiten („Samstag“) bis zur dritten Kategorie („Sonn- und Feiertag“) abfällt. Ein ähnliches

Verhalten können wir beim Mittelwert erkennen, der in der ersten Kategorie mit 46 $\mu g/m^3$ deutlich über den anderen zwei Mittelwerten (um 38 $\mu g/m^3$) liegt.
Bei der Betrachtung der 4h- und Tagesmittelwerte können wir keinen großen Unterschied feststellen, außer, dass die Daten geglättet und somit einige Ausreißer beseitigt wurden. Die Daten weisen jedoch ein sehr ähnliches Verhalten wie jenes der Halbstundenwerte auf: Auch hier besitzen die Werktagsdaten eine größere Variabilität und somit eine größere Box, deren Median sich etwas höher befindet, als in den Boxen der Samstags- bwz. Sonn- und Feiertagsdaten.
Anhand der Summaries der 4h- und Tagesmittelwerte lässt sich erkennen, dass die Höhe des Medians und des Mittelwertes annähernd gleich geblieben sind. Anhand der minimalen und maximalen, sowie an den fehlenden Werten erkennt man jedoch die starke Glättung der Daten.

Summary PM_{10} 4h-Werte: Werktag

```
 Min.  1st Qu.  Median   Mean  3rd Qu.   Max.    NA's
 1.92   21.25   34.76   46.01   59.77   237.9    136
```

Summary PM_{10} Tagesmittelwerte: Werktag

```
 Min.  1st Qu.  Median   Mean  3rd Qu.   Max.    NA's
 9.46   23.65   37.04   45.87   58.42   172.6     16
```

Summary PM_{10} 4h-Werte: Samstag

```
 Min.  1st Qu.  Median   Mean  3rd Qu.   Max.    NA's
 0.74   18.25   30.30   38.53   46.65   254.5     25
```

Summary PM_{10} Tagesmittelwerte: Samstag

```
 Min.  1st Qu.  Median   Mean  3rd Qu.   Max.    NA's
 5.60   21.66   32.02   38.56   48.34   199.6      4
```

Summary PM_{10} 4h-Werte: Sonn- und Feiertag

```
 Min.  1st Qu.  Median   Mean  3rd Qu.   Max.    NA's
 2.54   14.92   26.39   37.76   45.58   550.3     30
```

Summary PM_{10} Tagesmittelwerte: Sonn- und Feiertag

```
 Min.  1st Qu.  Median   Mean  3rd Qu.   Max.    NA's
 5.92   16.11   28.00   37.82   45.71   228.4      5
```

Dadurch haben wir nun gezeigt, dass die Schadstoffbelastung durch PM_{10} an Werktagen höher ist, als an Samstagen bzw. Sonn- und Feiertagen.

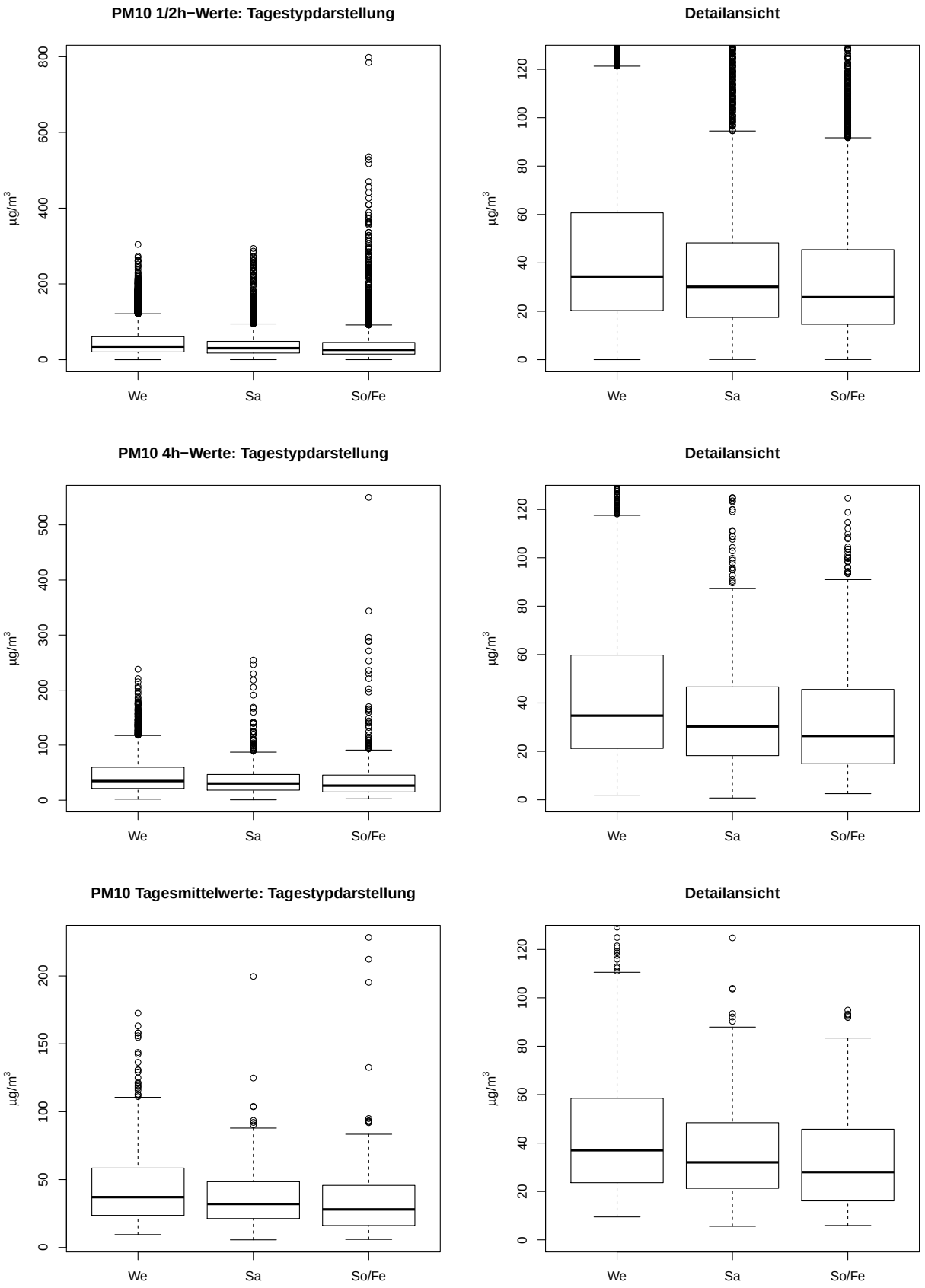


Abbildung 2.4.1: PM_{10}: Tagestypdarstellung bzw. Detailansicht

Des Weiteren interessiert uns aber, ob es einen Unterschied zwischen der Belastung an Sommer- und Wintertagen gibt, und ob dieser bedeutungsvoll ist. Aus diesem Grund haben wir in der folgenden Abbildung (Abb. 2.4.2) die Belastung durch PM_{10} erneut in die bereits bekannten Kategorien eingeteilt, zusätzlich allerdings zwischen Sommer- und Wintertagen unterschieden.

Es ist sofort erkennbar, dass es einen großen Unterschied macht, ob wir Sommer- oder Wintertage betrachten. Diese Erkenntnis ist alles andere als neu, da wir schon im Kapitel 2.2 festgestellt haben, dass sich die Sommer- von den Winterdaten sehr stark unterscheiden.
Beschränken wir uns zuerst nur auf die Interpretation der Ausreißer der Graphik der Halbstundenwerte (Abb. 2.4.2, links oben), so lässt sich sagen, dass es an den Sommertagen ein paar einzelne recht hohe Ausreißer gibt, während in den Wintermonaten die Ausreißer gleichmäßig verstreuter vorkommen. Eine Ausnahme bilden hierbei zwei uns schon bekannte Ausreißer, die in der Kategorie Sonn- und Feiertag vorkommen. Diese Werte sind natürlich die Silvester-Spitzenwerte.
Man kann zwar außerdem noch erkennen, dass die Boxen der Sommerwerte kleiner als die der Winterwerte sind, jedoch möchten wir auch hier eine Detailansicht zur besseren Begutachtung heranziehen (Abb. 2.4.2, rechts oben).
Hier erkennt man den deutlichen Unterschied zwischen den Werten an Sommer- und Wintertagen. Während sich die zentralen 50% der Sommerdaten deutlich unter dem Grenzwert von 50 $\mu g/m^3$ befinden, so sind die Boxen der Winterdaten nicht nur höher, sondern beginnen teilweise erst auf Höhe des oberen Randes der Boxen bezüglich der Sommertage. Da höhere Boxen bedeuten, dass die Daten mehr streuen, können wir sagen, dass bei den Daten der Wintertage eine größere Streuung vorliegt, als bei den Sommerdaten.
Auch die Darstellungen der 4h- und Tagesmittelwerte liefern dieselben Erkentnisse und zeigen uns, dass unabhängig vom Werktag, die Wintertage eine stärkere Belastung durch PM_{10} aufweisen. Anhand der Tagesmittelwerte lässt sich der Unterschied sogar noch deutlicher feststellen, als bei deren Halbstundenwerten.

Insgesamt lässt sich erkennen, dass die stärkste Belastung durch PM_{10} an Werktagen im Winter gemessen wurde, und die niedrigsten Schadstoffwerte an Sonn- und Feiertagen im Sommer erhoben wurden.

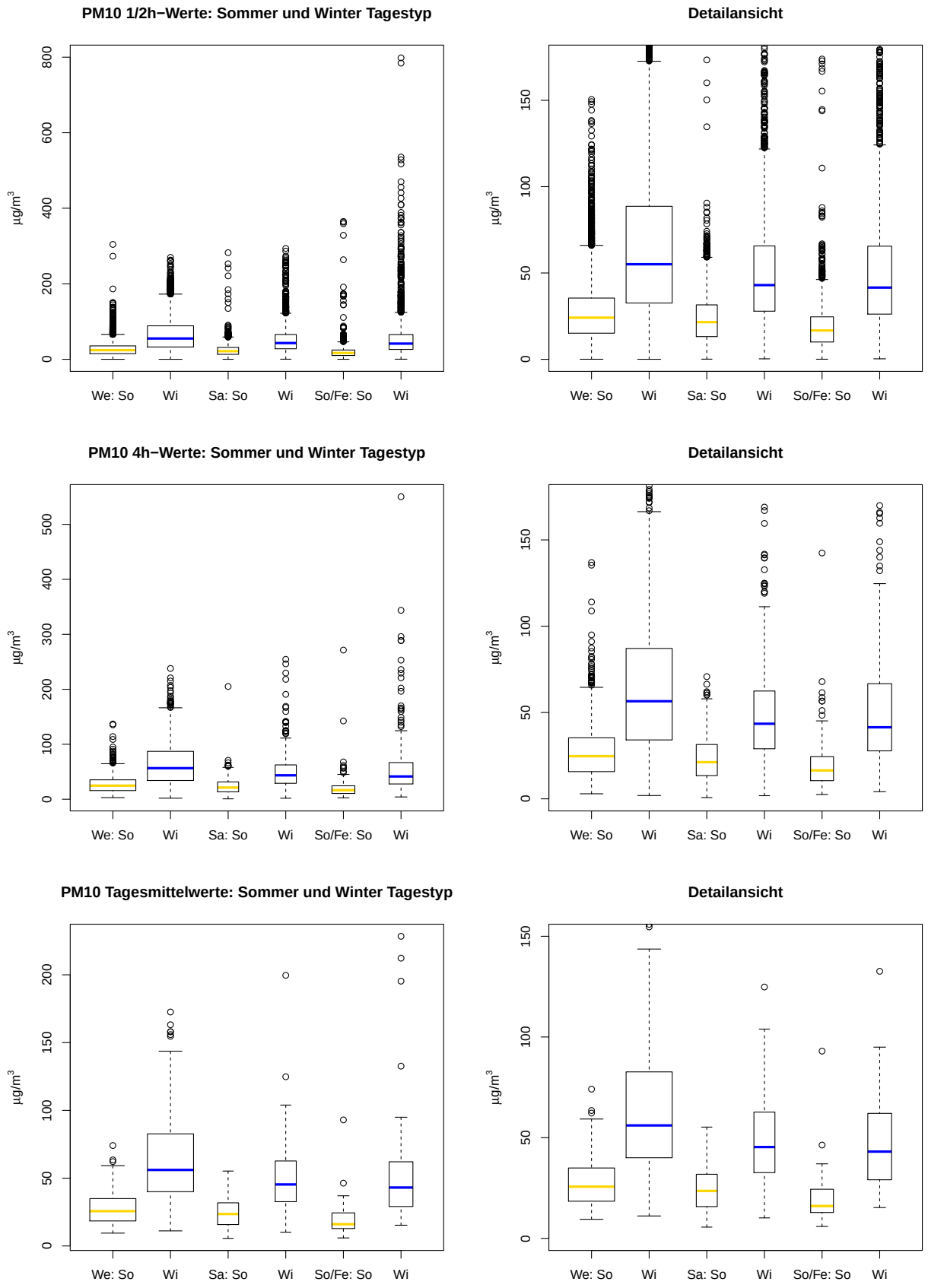


Abbildung 2.4.2: PM_{10}: Tagestypdarstellung bzw. Detailansicht Winter-Sommer

An dieser Stelle möchten wir noch mehr in die Tiefe gehen und fragen uns, wie sich die Schadstoffwerte im Laufe eines Tages verändern und wo sie ihr Maximum und Minimum erreichen. Dafür verwenden wir alle Halbstundenwerte der PM_{10} Daten und berechnen daraus für jede halbe Stunde einen Mittelwert, wodurch wir 48 gemittelte Halbstundenwerte erhalten. Dargestellt wird in Abb. 2.4.3 sowohl der durchschnittliche Verlauf eines Werktages, als auch der eines Samstages und eines Sonn- oder Feiertages. Zusätzlich wurde der Tagesgrenzwert von 50 $\mu g/m^3$ eingezeichnet.

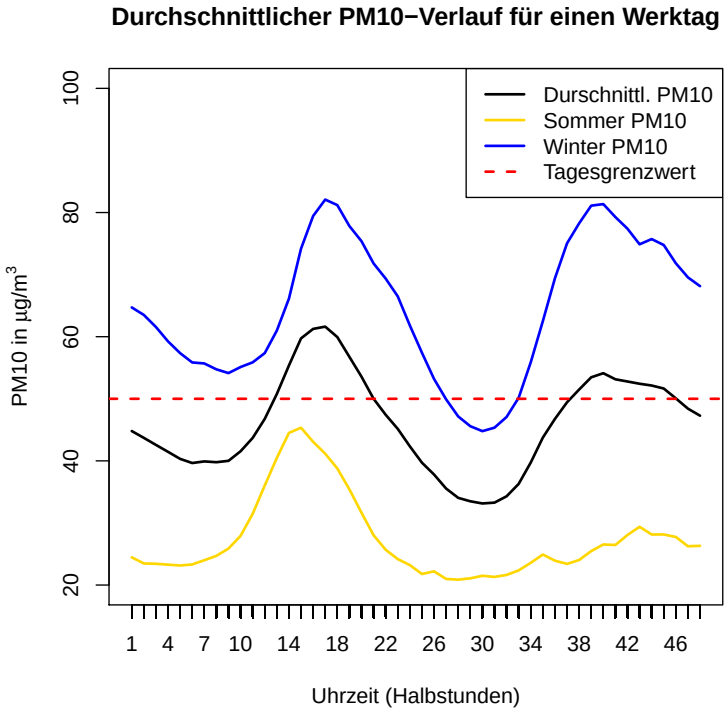


Abbildung 2.4.3: Durchschnittlicher PM_{10} Tagesverlauf (Werktag)

Interpretiert man zuerst den allgemeinen durchschnittlichen PM_{10}-Verlauf, so erkennt man, dass dieser sein erstes Hoch um etwa acht Uhr früh und eine weitere, allerdings kleinere Spitze, bei etwa acht Uhr abends hat. Hier scheint es naheliegend, dem Früh- bzw. Abendverkehr die Schuld für diese hohen Werte zu geben.
Zusätzlich wurden die durchschnittlichen PM_{10}-Werte in Sommer- und Winterdaten aufgeteilt, sodass man den deutlichen Unterschied hier noch einmal erkennen kann. Im Winter liegen die Schadstoffwerte fast durchgehend über dem Grenzwert (der allerdings für den Tagesmittelwert gilt), während an einem Sommerwerktag die 50 $\mu g/m^3$ Grenze nie erreicht oder überschritten wird.
Auffallend ist außerdem, dass das zweite Hoch im Tagesverlauf im Winter fast gleich hoch ist, wie das erste Hoch, wohingegen an einem durchschnittlichen Werktag im Sommer gar keine zweite Spitzenbelastung vorhanden ist. Ein Grund dafür könnte der Hochnebel im Winter sein, der die Schadstoffe länger im Grazer Becken festhält, als es im Sommer der Fall ist.

NO_x Auch die NO_x Daten werden nun in die drei Tagestypen eingeteilt und wir sind gespannt, wie groß die Unterschiede hier ausfallen werden.
Zu Beginn betrachten wir die Summaries der Stickstoffoxidwerte ($^1/_2$h-Werte), eingeteilt nach Tagestyp:

Summary NO_x $^1/_2$h-Werte: Werktag

```
 Min.   1st Qu.   Median    Mean   3rd Qu.    Max.     NA's
 0.33    15.83    31.87    61.33    77.84   550.80    1161
```

Summary NO_x $^1/_2$h-Werte: Samstag

```
 Min.   1st Qu.   Median    Mean   3rd Qu.    Max.     NA's
 1.76    11.02    22.88    39.83    50.82   289.80    199
```

Summary NO_x $^1/_2$h-Werte: Sonn- und Feiertag

```
 Min.   1st Qu.   Median    Mean   3rd Qu.    Max.     NA's
 0.48     7.52    16.24    32.70    40.45   269.50    317
```

Auch hier beträgt der Prozentsatz der fehlenden Daten nur 4-5% und ist mit dieser Höhe nicht weiter beunruhigend. Anhand der Summaries lässt sich außerdem herauslesen, dass der Mittelwert in der Kategorie Werktag deutlich größer ist, als in den anderen beiden. So befindet sich dieser in der ersten Kategorie bei 61.33 $\mu g/m^3$, während er bei der zweiten bzw. dritten Kategorie nur mehr 39.83 bzw. 32.7 $\mu g/m^3$ ausmacht. Des Weiteren liegt der Median in allen Fällen näher beim 1.Quantil als beim 3., was bedeutet, dass wir es hier mit rechtsschiefen Verteilungen zu tun haben.
Zusätzlich zu den Summaries ist die graphische Darstellung jedoch unerlässlich für eine gute Interpretation, weshalb hier nicht darauf verzichtet werden soll. Geplottet wurden hier allerdings nicht nur die NO_x Halbstundenwerte, sondern auch die 4h- und Tagesmittelwerte (Abb. 2.4.4). Im Unterschied zu den PM_{10} Daten, bei denen die meisten Ausreißer an Sonn- und Feiertagen vermerkt wurden (Abb. 2.4.2), findet man die meisten NO_x- Ausreißer an Werktagen. Die Ausreißer von PM_{10} und NO_x verhalten sich also im Bezug auf Werktag und Sonn- und Feiertag genau gegensätzlich.
In der Detailansicht (Abb. 2.4.4, rechts oben) erkennen wir ein ähnliches Verhältnis zwischen den Boxen, wie auch schon bei den PM_{10} Daten. So liegen die zentralen 50% der Werktagsdaten im größten Bereich, während die Boxen für Samstag und Sonn- und Feiertag kleiner ausfallen.
Die 4h- und Tagesmittelwerte verhalten sich hier gleich, sodass wir bis auf die Anzahl der Ausreißer keine großen Veränderungen feststellen können.

Auch die NO_x Belastung ist also an Werktagen am stärksten, wobei der Unterschied zu den Samstagen bzw. Sonn-und Feiertagen größer ist, als zuvor bei den PM_{10} Daten.

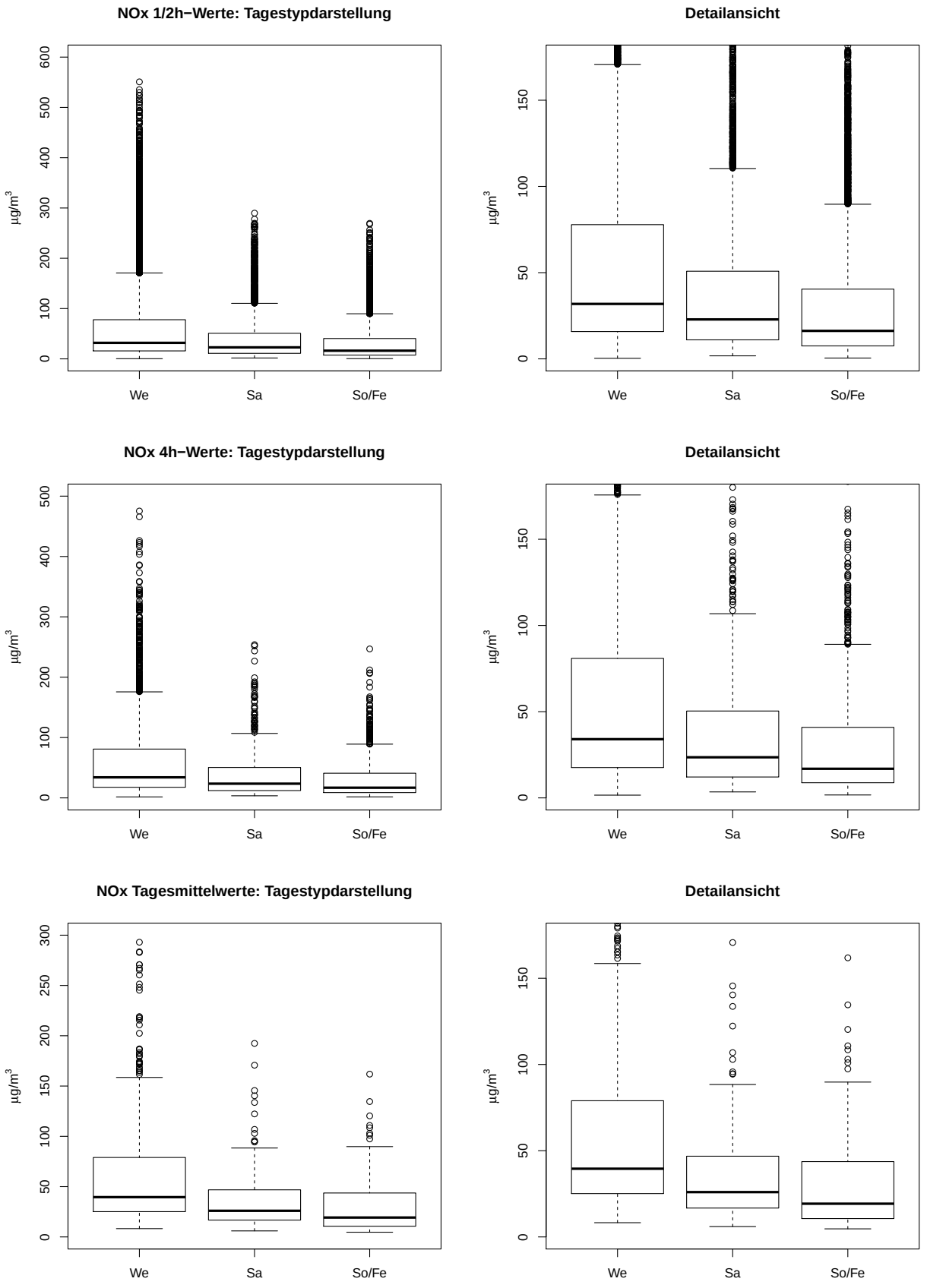


Abbildung 2.4.4: NO_x: Tagestypdarstellung bzw. Detailansicht

Wollen wir nun noch den Unterschied zwischen Sommer- und Wintertagen aufzeigen, so wird die folgende Graphik sehr nützlich sein (Abb. 2.4.5).

Wie wir zuvor schon bei den PM_{10} Daten, sowie auch in Kapitel 2.2 entdeckt haben, sind die Werte im Sommer deutlich niedriger und schwanken weniger als im Winter. Außerdem erkennt man Darstellung der $^1/_2$h-Werte (Abb. 2.4.5, links oben) die hohen Ausreißer in der Kategorie „Werktag“, während diese in den anderen beiden Kategorien nicht so hoch ausfallen. Die größten Ausreißer wurden eindeutig an Winter-Werktagen gemessen, während zum Beispiel die Ausreißer an einem Sommer-Sonntag im Vergleich dazu nahezu winzig erscheinen. Auch die Box und der obere Whisker sind an Winter-Werktagen am dominantesten.
Für weitere Interpretationen betrachten wir allerdings wieder eine Detailansicht (Abb. 2.4.5, rechts oben). Die Boxen der Sommertage wirken hier sehr klein im Vergleich zu denen der Wintertage. Tatsächlich machen die Boxen der Sommertage nur 22.34%, 26.63% und 18.06% der Größe der Winterboxen aus. Außerdem beginnen die Boxen der Wintertage dort, wo die Boxen der Sommertage bereits enden. Es gibt also einen sehr großen Unterschied in der Lage der zentralen 50% der Winter- bzw. Sommertage.
Interessant ist außerdem, dass die Winterdaten durch ihre hohen Ausreißer immer weiter nach oben gezogen werden, je mehr gemittelt wird. So sieht man zum Beispiel bei den Tagesmittelwerten, dass die Box der Winterwerktage erst beginnt, wenn die zentralen 50% der Sommerwerktage schon längst vorbei sind. Das lässt sich auch gut aus den Summaries herauslesen:

Summary NO_x Tagesmittelwerte: Werktag - Sommer

```
 Min.  1st Qu.  Median   Mean  3rd Qu.   Max.    NA's
 8.25   21.58   27.17   29.61   35.85   70.48     2
```

Summary NO_x Tagesmittelwerte: Werktag - Winter

```
 Min.  1st Qu.  Median   Mean  3rd Qu.   Max.    NA's
 11.52  50.34   81.57   94.69   121.50  293.00    7
```

Das 3.Quantil der Sommerdaten liegt hier bei 35.85 μg/m^3, während sich das 1.Quantil der Winterdaten bei 50.34 μg/m^3 befindet. Dadurch wird und abermals der große Unterschied zwischen Sommer- und Wintertagen im Bezug auf die Schadstoffbelastung bewusst.

Wie auch schon bei den PM_{10} Daten können wir hier erkennen, dass die geringste Belastung durch NO_x an Sonn-und Feiertagen im Sommer gemessen wurde, während die höchsten Schadstoffwerte an Werktagen im Winter vorkommen.

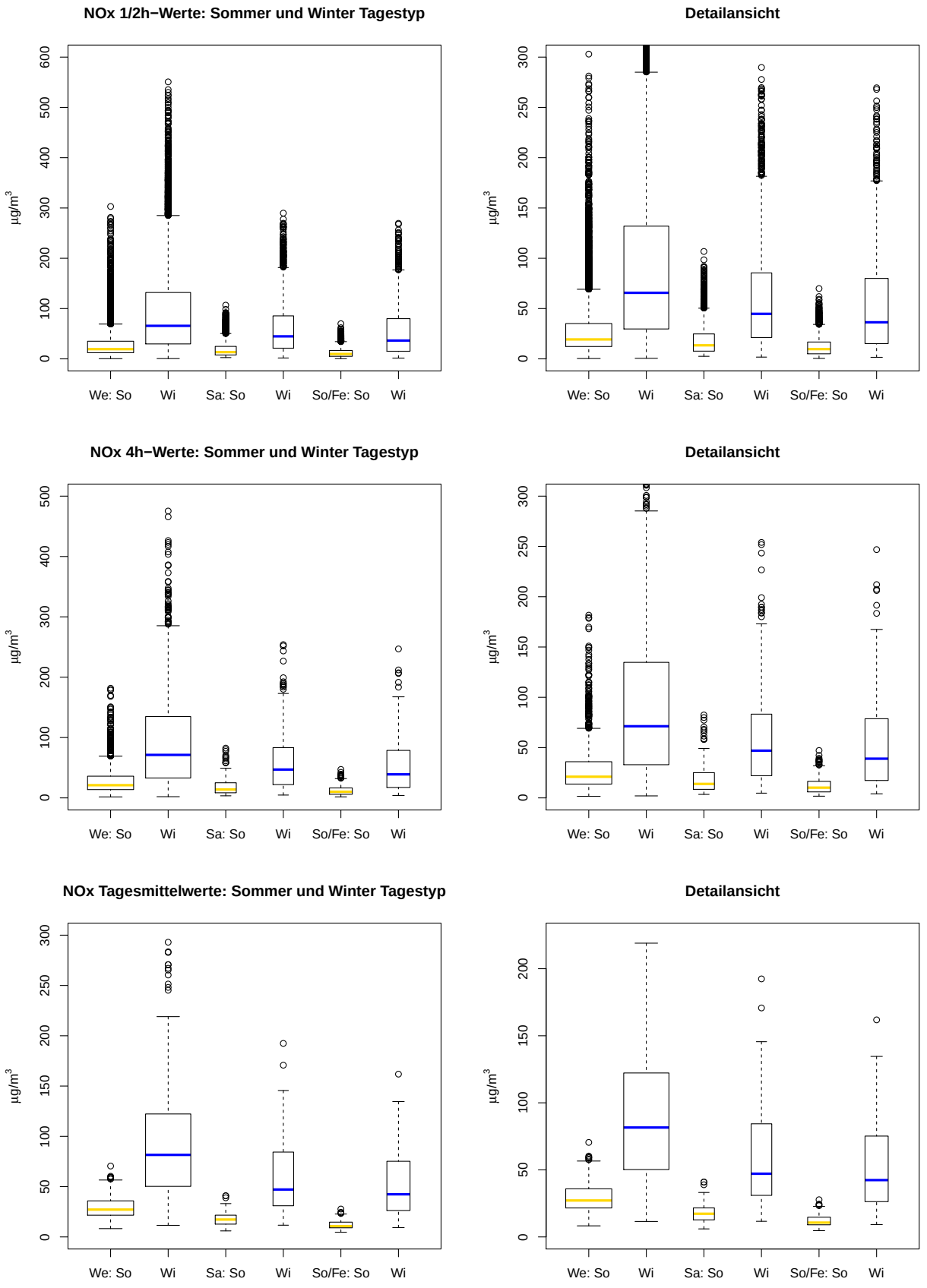


Abbildung 2.4.5: NO_x: Tagestypdarstellung bzw. Detailansicht Winter-Sommer

Auch hier sind wir daran interessiert, wie sich die Belastung im Verlauf eines Tages verhält. Aus diesem Grund berechnen wir aus allen Stickstoffoxiddaten die Mittelwerte für jede halbe Stunde, sodass wir die Höhen und Tiefen eines durchschnittlichen Tages im Bezug auf die Schadstoffbelastung erkennen können. Dabei beschränken wir uns auf die Kategorie Werktag (Abb. 2.4.6).

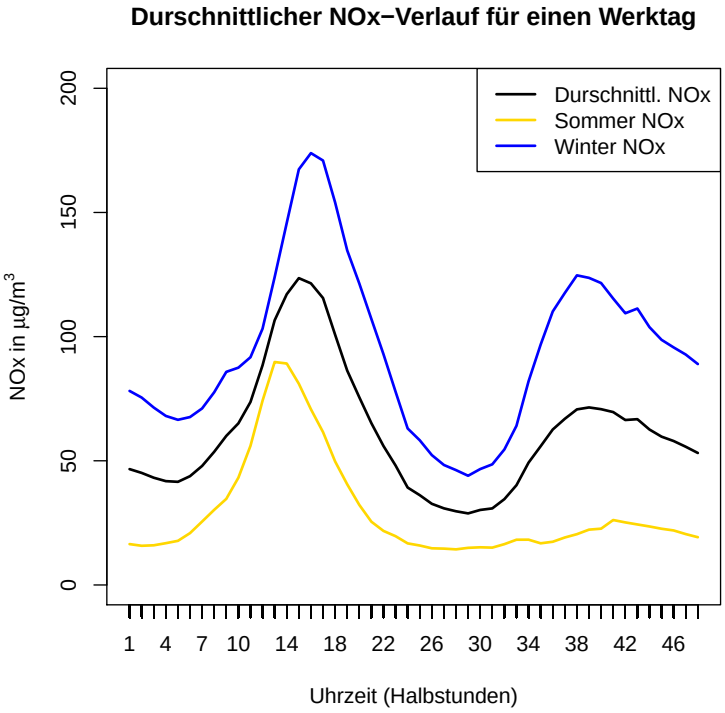


Abbildung 2.4.6: Durchschnittlicher NO_x Tagesverlauf (Werktag)

Man erkennt sofort, dass die höchsten Werte in der Früh um etwa 8 Uhr gemessen wurden. Hier steigen die Werte sogar bis zu 175 $\mu g/m^3$ an (Winter). Das ist ein sehr großer Unterschied zu den Werten im Sommer zur gleichen Zeit, da hier nur Werte von unter 100 $\mu g/m^3$ erhoben wurden. Anschließend fallen die Stickstoffoxidwerte bis zu einem Tiefpunkt bei ungefähr 3 Uhr nachmittags um dann wieder bis kurz vor 8 Uhr abends zu steigen. Dies gilt zumindest für die Winterwerte. Die Werte an Sommerwerktagen steigen abends nämlich nicht mehr an sondern bleiben recht konstant um die 20 $\mu g/m^3$.

2.4.1 Gemeinsame Darstellung von PM_{10} und NO_x

Bisher haben wir zwar den typischen Tagesverlauf der PM_{10}-Schadstoffwerte und jenen der NO_x-Werte betrachtet, jedoch noch nicht beide in direktem Zusammenhang. Plotten wir nämlich beide Verläufe in einer Graphik, so können wir erkennen, ob es Zeitpunkte gibt, in denen sich die Zeitreihen ähnlich, oder vielleicht sogar gegenläufig verhalten. Wir wählen hier also eine Darstellung der durchschnittlichen Halbstundenwerte aller Daten (Abb. 2.4.7).

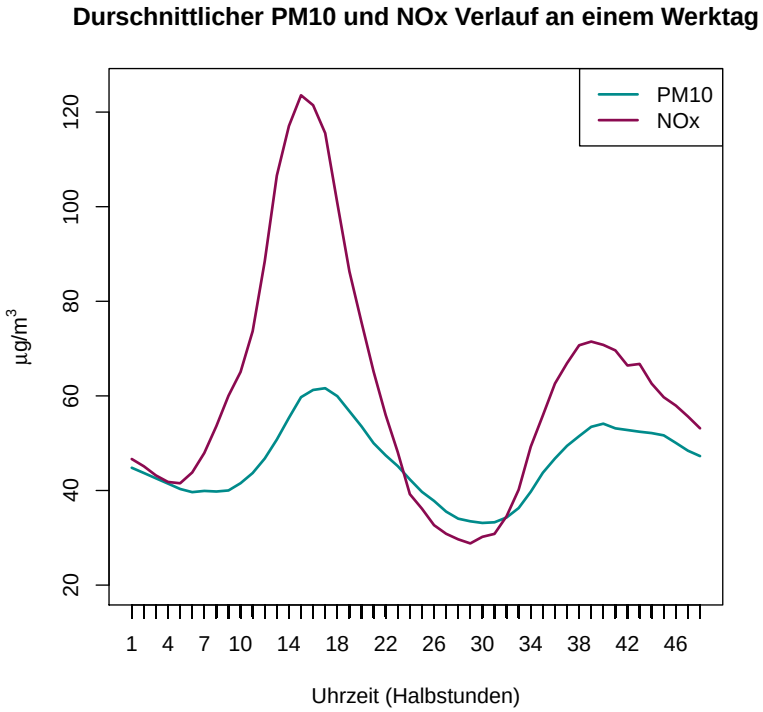


Abbildung 2.4.7: Durchschnittlicher Tagesverlauf (PM_{10} und NO_x: Werktag)

Auf den ersten Blick erkennen wir, dass sich die zwei Zeitreihen sehr ähnlich sind, wobei die NO_x-Werte fast ausschließlich über den PM_{10}-Werten liegen. Um Mitternacht eines typischen Werktages liegen beide Werte zunächst bei etwa 40 $\mu g/m^3$ und sinken leicht. Um ca. halb drei in der Nacht beginnen die NO_x-Werte bis zu einem Spitzenwert von über 120 $\mu g/m^3$, der um halb acht herum gemessen wurde, stark zu steigen. Währenddessen fallen die PM_{10}-Werte zunächst noch ein wenig ab, um dann um etwa fünf Uhr auch anzusteigen, allerdings bei weitem nicht so stark. Sie erreichen um etwa acht Uhr ihr Maximum bei 60 $\mu g/m^3$. Nach dem Erreichen der Spitzenwerte sinken beide Schadstoffwerte wieder ab, um bei ungefähr halb drei Uhr nachmittags ihren Tiefpunkt zu erreichen, wobei hier die NO_x-Werte das einzige Mal unter die PM_{10}-Werte sinken. Die tiefsten Werte liegen dabei um die 30 $\mu g/m^3$. Anschließend steigen beide Schadstoffwerte wieder an, wobei NO_x wieder schneller wächst als PM_{10}. Nach dem zweiten, allerdings kleineren Hoch zwischen halb acht und acht Uhr am Abend fallen beide Werte wieder auf knapp über 40 $\mu g/m^3$ ab.
Die beiden Schadstoffwerte PM_{10} und NO_x weisen also im Großen und Ganzen ein sehr ähnliches Verhalten auf und ihre Höhen und Tiefen wurden etwa zur gleichen Zeit gemessen.

2.5 Zusammenhangsanalyse

Bis hierher wurden die beiden Zeitreihen PM_{10} und NO_x hauptsächlich getrennt voneinander betrachtet. Doch nun wollen wir mögliche Zusammenhänge bzw. Abhängigkeiten zwischen diesen beiden untersuchen. Dafür stehen uns unter anderem einfache Scatterplots und Scatterplotmatrizen zur Verfügung.
Zeichnet man zunächst einen einfachen Scatterplot für die Halbstundenwerte, so erkennt man schnell, dass es zu einem Overplotting kommt, d.h. es werden so viele Datenpunkte geplottet, dass man nur noch eine kompakte schwarze Masse sieht (Abb. 2.5.1, links oben). Auch der Scatterplot für die 4h-Werte ist alles andere als übersichtlich. Aus diesem Grund werden wir uns in diesem Kapitel hauptsächlich auf die Untersuchung der Tagesmittelwerte beschränken.
Wichtig hierbei ist noch, dass in den ersten beiden Graphiken der folgenden Abbildung Ausreißer durch die Wahl der Achsenlänge abgeschnitten wurden. Da es sich hierbei aber nur um sehr wenige handelt, beeinflusst das unsere weitere Analyse nicht.

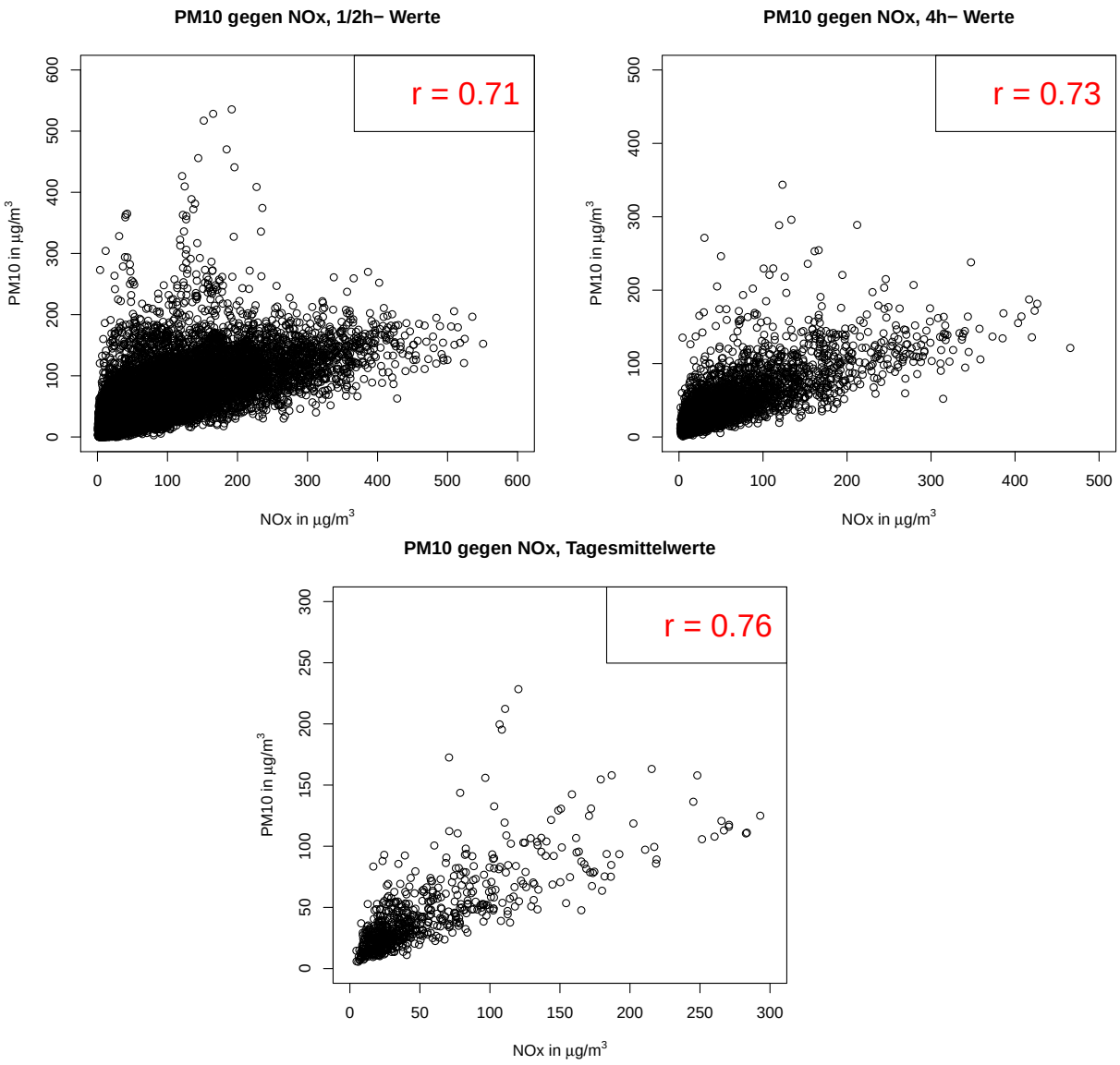


Abbildung 2.5.1: Scatterplot PM_{10} gegen NO_x, $^1/_2$h-, 4h-, Tagesmittelwerte

Man erkennt hier eine positive Korrelation zwischen PM_{10} und NO_x. Um aber genaueres über die Stärke des linearen Zusammenhanges sagen zu können, benötigt man eine wichtige Kenngröße, den Pearson Korrelationskoeffizienten. Dieser liegt zwischen -1 und +1 und gibt Auskunft darüber, wie stark zwei Variablen linear voneinander abhängen. Er wurde mit der in Kapitel 1.4 angeführten Formel berechnet und in die jeweilige Graphik geschrieben.
Für den Zusammenhang zwischen PM_{10} und NO_x ergibt sich ein Pearson Korrelationskoeffizient von 0.71, was auf einen recht starken positiven linearen Zusammenhang hinweist. Steigen also die PM_{10}-Werte, so werden auch die NO_x-Werte höher und umgekehrt. Auch die Korrelationskoeffizienten der 4h- und Tagesmittelwerte sind hoch, bzw. sogar noch höher als bei den Halbstundenwerten. Man kann also sagen: Je mehr die Daten gemittelt werden, umso stärker wird der lineare Zusammenhang zwischen PM_{10} und NO_x.

Nun sind wir aber auch daran interessiert, wie die Scatterplots bzgl. der Saison ausschauen, da wir doch zuvor schon einen großen Unterschied zwischen Sommer und Winter feststellen konnten. Aus diesem Grund zeichnen wir die Scatterplots bezüglich Sommer bzw. Winter (Abb. 2.5.2).

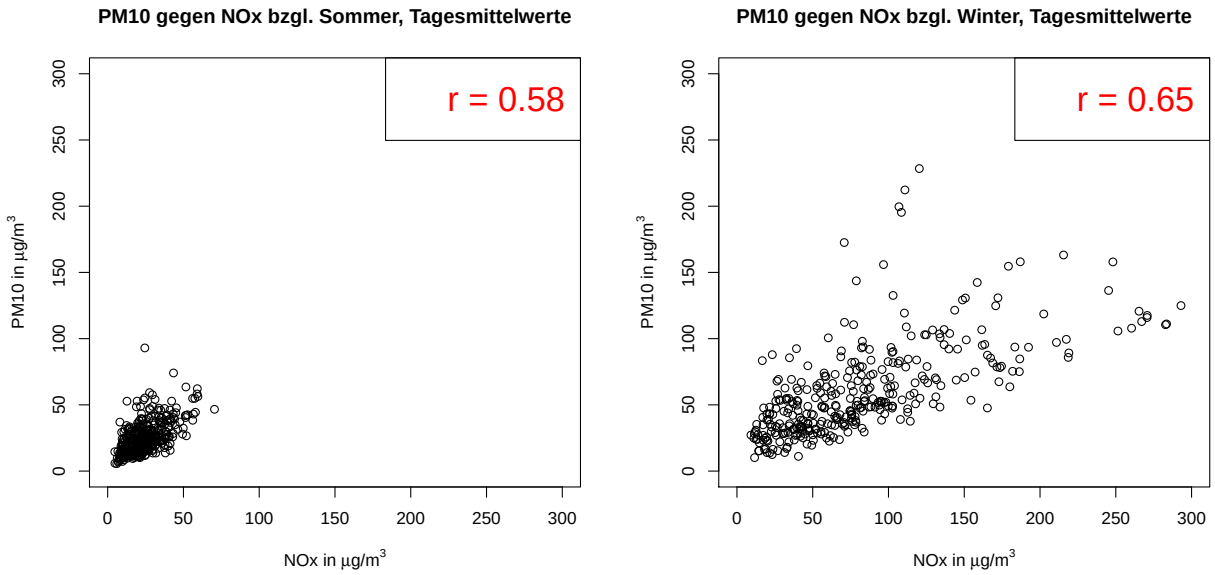


Abbildung 2.5.2: Scatterplot PM_{10} gegen NO_x bzgl. Saison, Tagesmittelwerte

Anhand der Graphik der Sommerdaten (links) erkennt man, dass sowohl die PM_{10}, als auch die NO_x-Werte nicht über 100 $\mu g/m^3$ steigen. Man kann zwar einen leichten positiven linearen Zusammenhang erkennen, der auch durch den Pearson Korrelationskoeffizienten von 0.58 bestätigt wird, jedoch weist die Trichterform der Daten darauf hin, dass die Daten auch noch von anderen Merkmalen abhängen.
In der Graphik der Winterdaten (rechts) sieht man, dass die Werte deutlich höher sind, als im Sommer. Außerdem erkennt man, dass die NO_x-Werte stärker steigen, als die PM_{10}-Werte. Dennoch gibt es im Winter einen stärkeren linearen Zusammenhang zwischen den beiden Schadstoffen als im Sommer. Allerdings sieht man auch hier wieder die Trichterform, die auf zusätzliche Abhängigkeiten hinweist.

Interessant wäre auch zu wissen, wie die Scatterplots eingeteilt nach Saison bezüglich dem Tagestyp aussieht, da wir doch in einem vorigen Kapitel (Kapitel 2.4) herausgefunden haben, dass die Belastung sowohl von PM_{10}, als auch NO_x, an Werktagen besonders hoch ist. Für diese Darstellung wurden für die jeweiligen Tagestypen nicht nur verschiedene Zeichen, sondern auch unterschiedliche Farben gewählt, um sich leichter einen Überblick verschaffen zu können (Abb. 2.5.3).

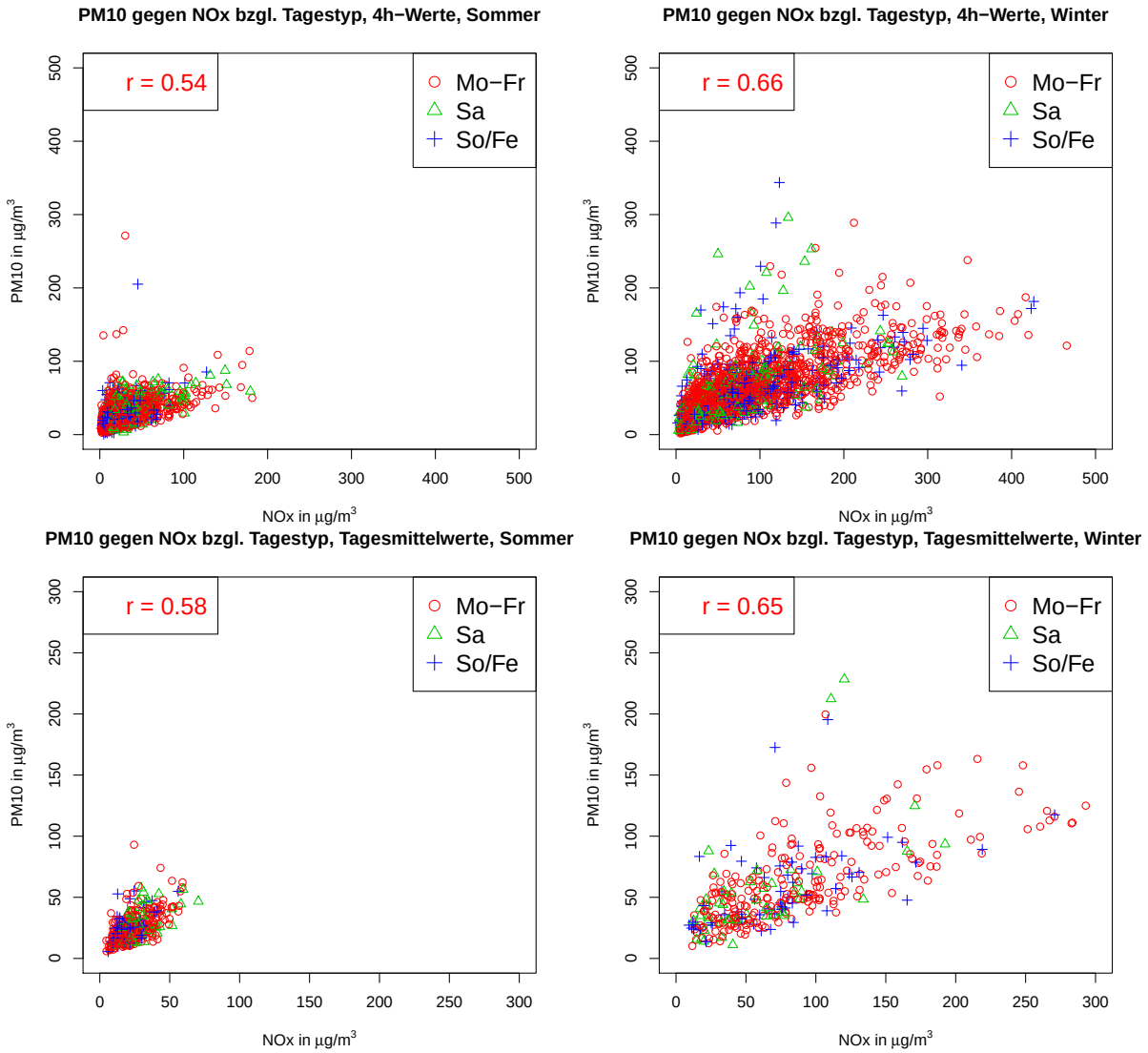


Abbildung 2.5.3: Scatterplot PM_{10} gegen NO_x bzgl. Tagestyp, 4h- und Tagesmittelwerte, Sommer und Winter

Im Scatterplot der 4h-Werte im Sommer (links oben) kann man erkennen, dass die höchsten NO_x-Werte hauptsächlich unter der Woche bzw. an Samstagen beobachtet wurden, während im Winter die höchsten Werte eindeutig an Werktagen beobachtet wurden (rechts oben). Der lineare Zusammenhang ist in beiden Fällen durch einen recht hohen Pearson Korrelationskoeffizienten gegeben, wobei dieser bei den Winterdaten signifikanter ist als im Sommer. In der Darstellung der Tagesmittelwerte im Sommer (links unten) scheinen die höchsten Werte verteilt auf die Tagestypen vorzukommen. In der Graphik der Winterdaten zeigt sich allerdings wieder, dass die höchsten Werte, bis auf ein paar Ausnahmen, an Werktagen gemessen wurden. Auch bei den Tagesmittelwerten haben wir annähernd gleich hohe Korrelationskoeffizienten wie zuvor bei den 4h-Werten.

Aufgrund unserer starken Vermutung, dass die PM_{10}- bzw. NO_x- Werte nicht nur vom jeweils anderen Wert beeinflusst werden, sondern, dass es auch andere Einflussgrößen gibt, wollen wir uns einer Scatterplotmatrix zuwenden, in der wir neben PM_{10} und NO_x auch noch die Windgeschwindigkeit, Niederschlagsmenge und Lufttemperatur mit einbinden (Abb. 2.5.4).

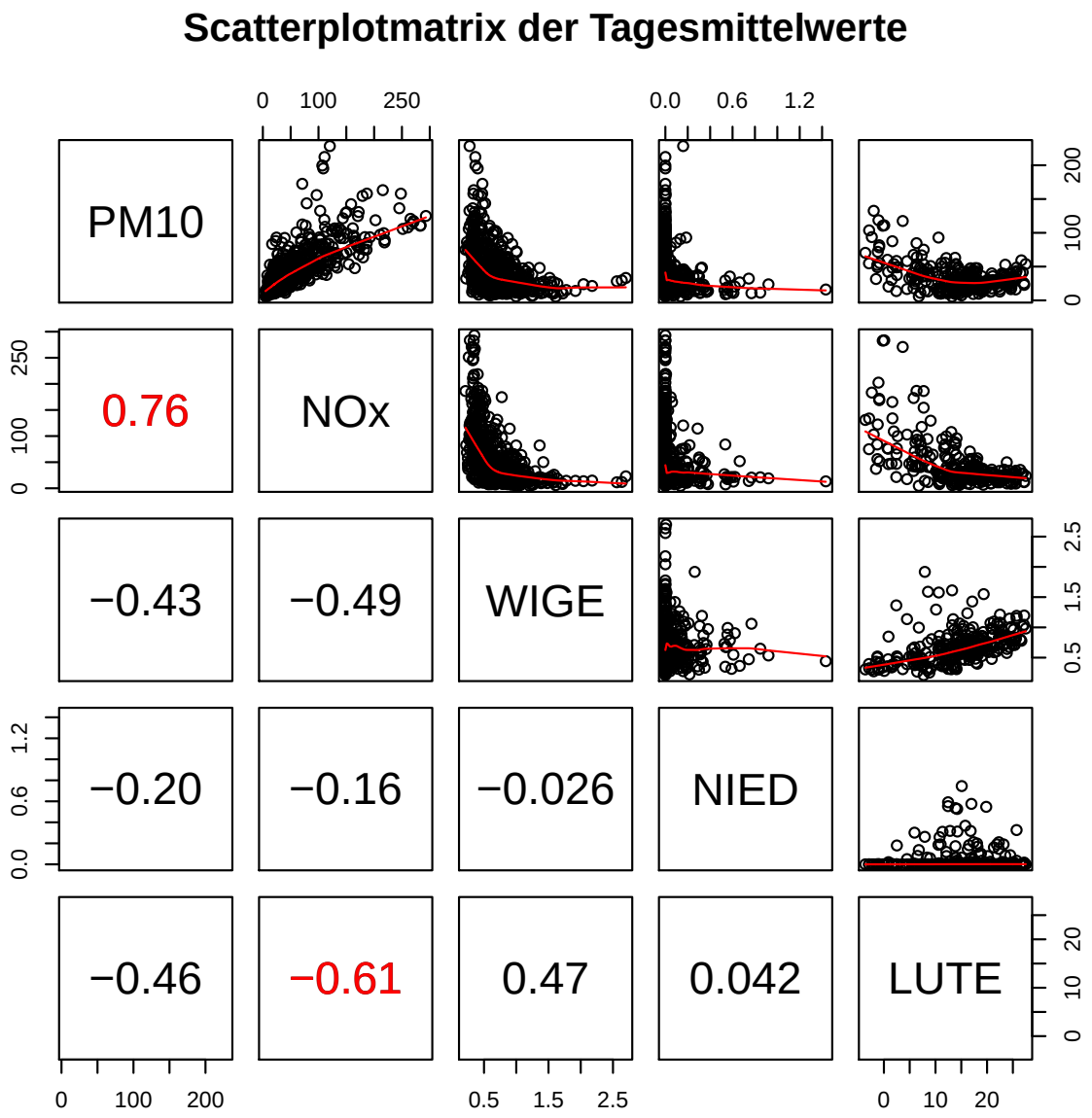


Abbildung 2.5.4: Scatterplot PM_{10} und NO_x bzgl. Tagestyp, Tagesmittelwerte

In der Diagonale stehen hierbei die beobachteten Größen, in den oberen Feldern die jeweiligen Scatterplots und in den unteren Quadraten steht der dazugehörige Pearson-Korrelationskoeffizient. Dieser wurde ab einem Wert von ± 0.5 rot markiert, um signifikante Werte sofort erkennen zu können. Aus diesem Grund sehen wir gleich, dass es den größten linearen Zusammenhang zwischen PM_{10} und NO_x gibt, welchen wir zuvor schon untersucht haben. Auffallend ist jedoch auch, dass sowohl die Windgeschwindigkeit, als auch die Lufttemperatur eine negative Korrelation im Bezug auf PM_{10} sowie NO_x aufweisen. Das bedeutet, dass mit steigender Windgeschwindigkeit bzw. Lufttemperatur die PM_{10}- und NO_x- Werte sinken. Untereinander sind Windgeschwindigkeit und Lufttemperatur allerdings auch positiv

korreliert, d.h. mit steigender Lufttemperatur wird auch die Windgeschwindigkeit stärker und umgekehrt. Aus dieser Scatterplotmatrix lässt sich außerdem herauslesen, dass die Niederschlagsmenge keinen signifikanten linearen Einfluss auf PM_{10} oder NO_x hat. Man sieht zwar, dass die Schadstoffwerte sinken, wenn die Niederschlagsmenge steigt, allerdings liegt hier bei weitem kein linearer Zusammenhang vor. Da die einzelnen Scatterplots in der Matrix allerdings recht klein sind, holen wir uns nun den Scatterplot heraus, der neben PM_{10}-NO_x am signifikantesten ist. Dabei fällt die Wahl klarerweise auf das Merkmal Lufttemperatur gegen NO_x (Abb. 2.5.5).

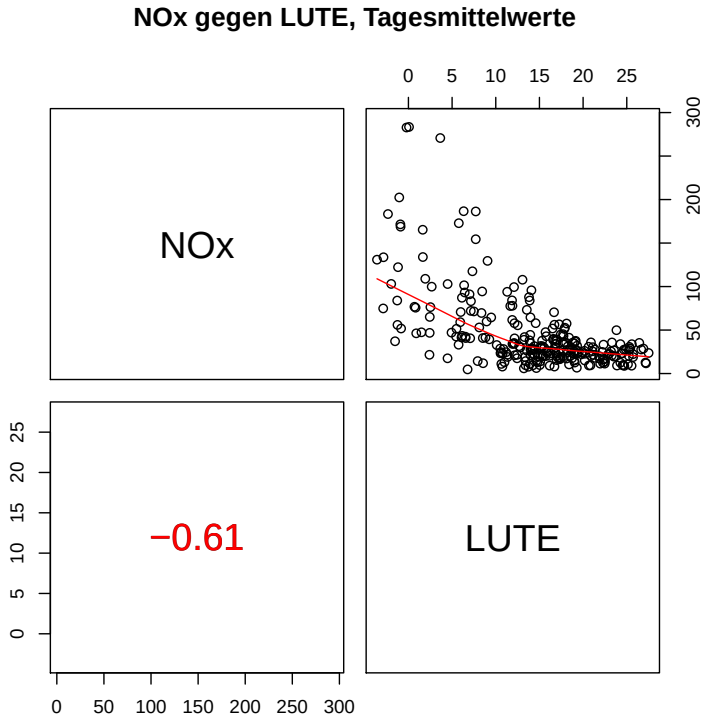


Abbildung 2.5.5: Scatterplot NO_x gegen LUTE, Tagesmittelwerte

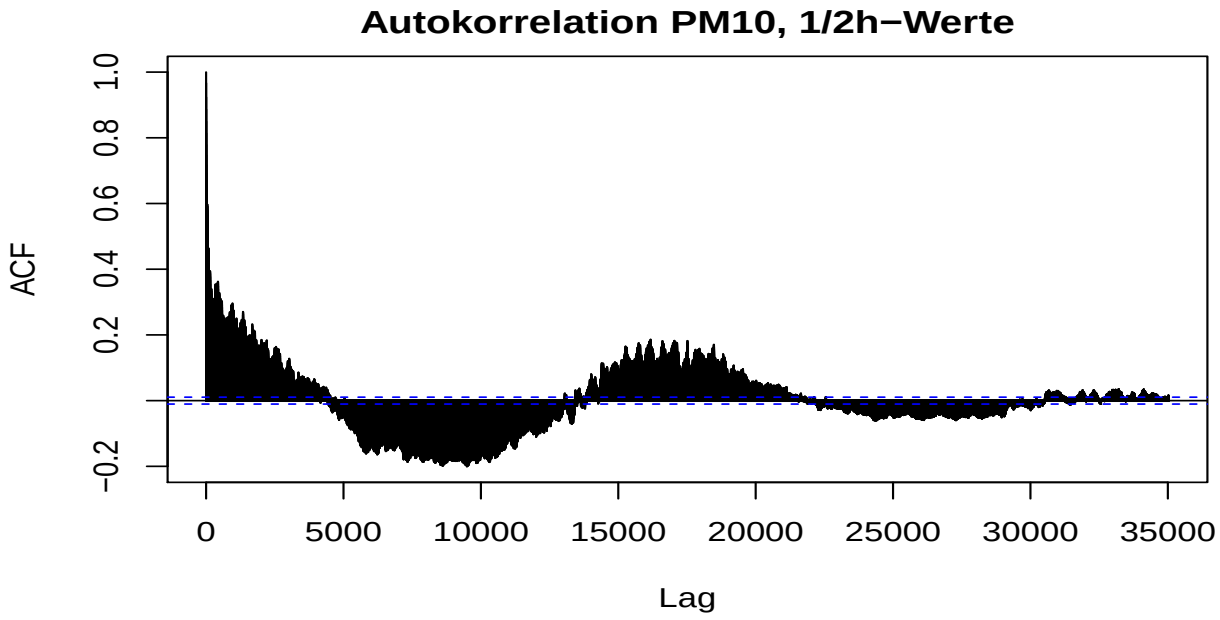
Man sieht hier noch einmal deutlich, wie diese beiden Merkmale zusammenhängen: Steigt die Lufttemperatur, so fallen die Stickstoffoxidwerte. Der Pearson- Korrelationskoeffizient von -0.61 zeigt noch einmal die recht starke, negative Korrelation.

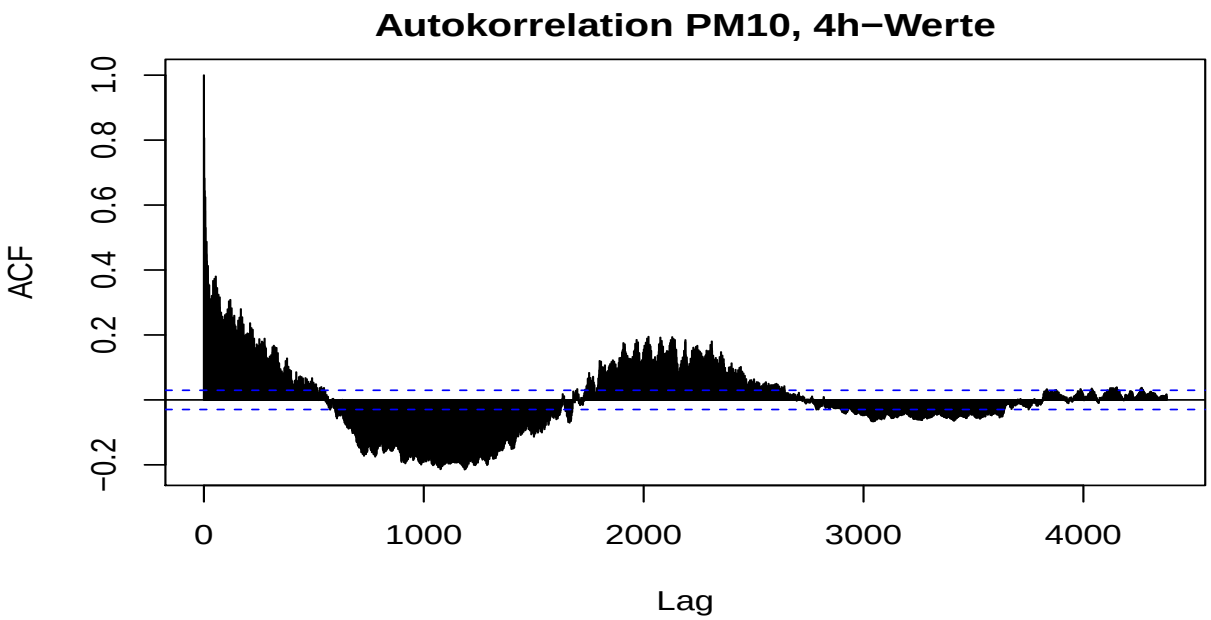
2.6 Autokorrelationsstruktur

In diesem Kapitel soll nun die Korrelation der Zeitreihen mit sich selber untersucht werden; die Autokorrelation. Die dazugehörige Formel wurde bereits in Kapitel 1.4 vorgestellt, sodass wir hier direkt mit den ersten Graphiken beginnen können.

Für die Autokorrelation der PM_{10}-Daten ergibt sich Abb. 2.6.6. Man erkennt, dass die Autokorrelation eine regelmäßige Schwingung aufweist, die mit der Zeit immer schwächer wird. Die ACF der Halbstundenwerte ist positiv bis zu einem Lag von etwa 4800, was 100 Tagen entspricht. Danach erkennen wir eine negative Korrelation bis zu einem Lag von unter 15000, die danach wieder positiv wird. Also wechselt in einem Abstand von ungefähr 100 Tagen bzw. drei Monaten die Korrelation, d.h. liegen die aktuelle Werte hoch, so werden die Werte in den darauffolgenden drei Monaten auch recht hoch sein, die Werte in den anschließenden drei Monaten (also sechs Monate nach den aktuellen Werten) werden niedrig ausfallen. Dieser Zyklus wiederholt sich in immer schwächer werdender Form. **PM_{10}**
Die ACF zeigt uns hier also den saisonalen Trend: liegen die Werte im Sommer hoch, sind sie sechs Monate darauf, im Winter, tief.

Dasselbe Verhalten findet man auch in der ACF der 4h- bzw. Tagesmittelwerte. Bei den 4h-Werten finden wir bis $\approx$ Lag 600 (= 100 Tage) eine positive Korrelation die danach bis zu einem Tiefpunkt bei Lag 1200 (= 200 Tage), um dann wieder zu steigen. Bei den Tagesmittelwerten sehen wir dasselbe Verhalten bei $\approx$ Lag 100 bzw. 200. Somit wird auch hier der saisonale Trend sichtbar.





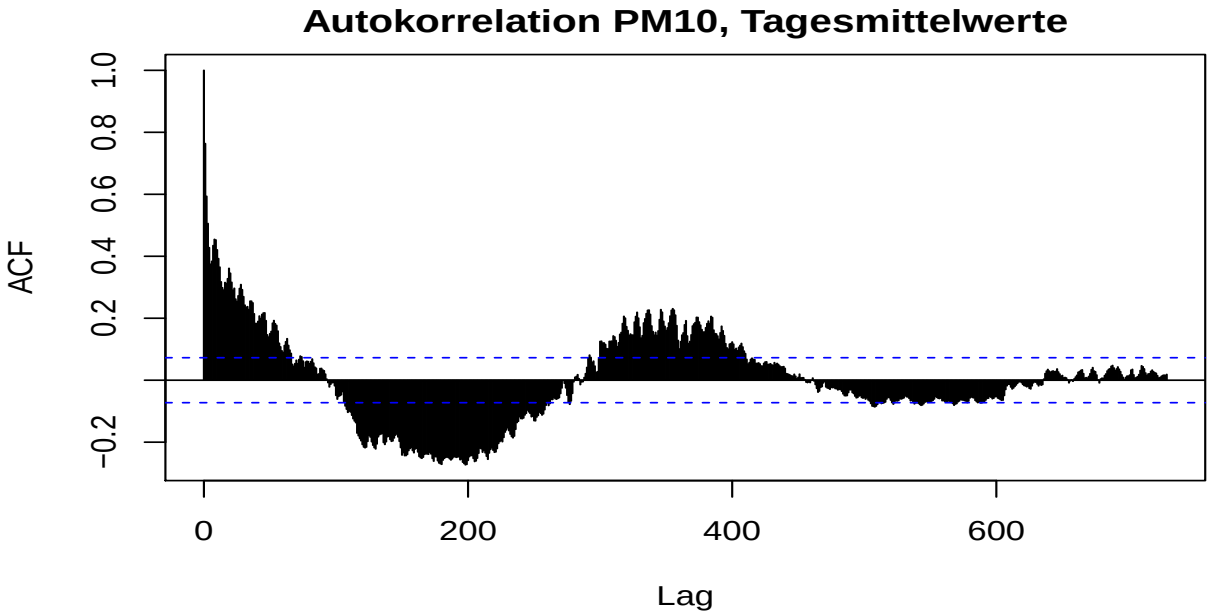


Abbildung 2.6.6: ACF PM_{10}, $^1/_2$h-, 4h- und Tagesmittelwerte

NO_x Für die NO_x-Daten ergibt sich eine sehr ähnliche Darstellung (Abb. 2.6.7). Auch hier wird das saisonale Verhalten mit wechselnden Korrelationen zu etwa den gleichen Zeitpunkten dargestellt. Hier sieht man noch zuätzlich deutlich, dass die Autokorrelation der Tagesmittelwerte am stärksten ist.

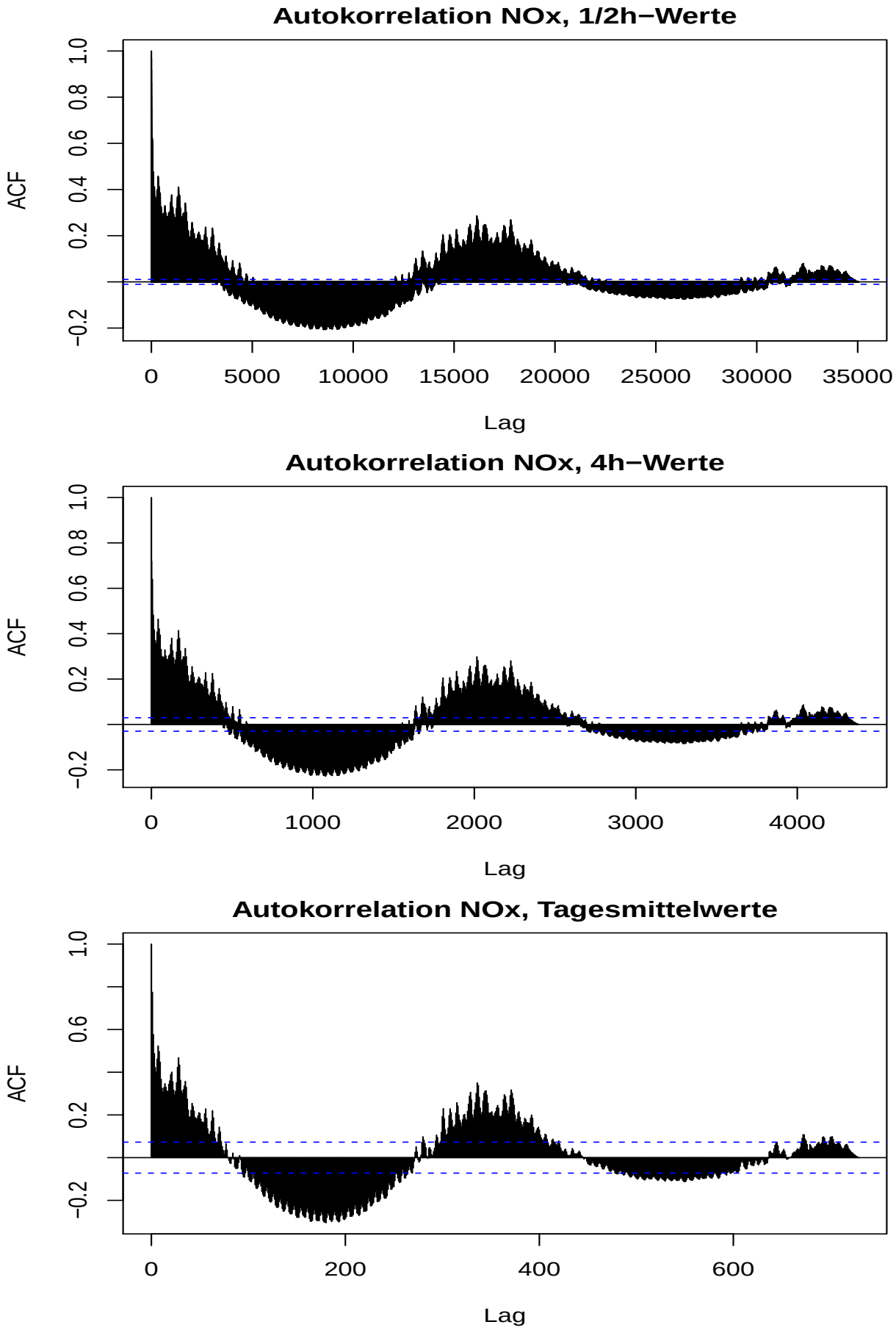


Abbildung 2.6.7: ACF NO_x, $^1/_2$h-, 4h- und Tagesmittelwerte

2.7 Kreuzkorrelationsstruktur

Nun wollen wir noch die Kreuzkorrelation zwischen PM_{10} und NO_x betrachten um zu erkennen, wie sich die zwei Zeitreihen ähneln und um wieviel sie zeitlich verschoben sind (Abb. 2.7.8).

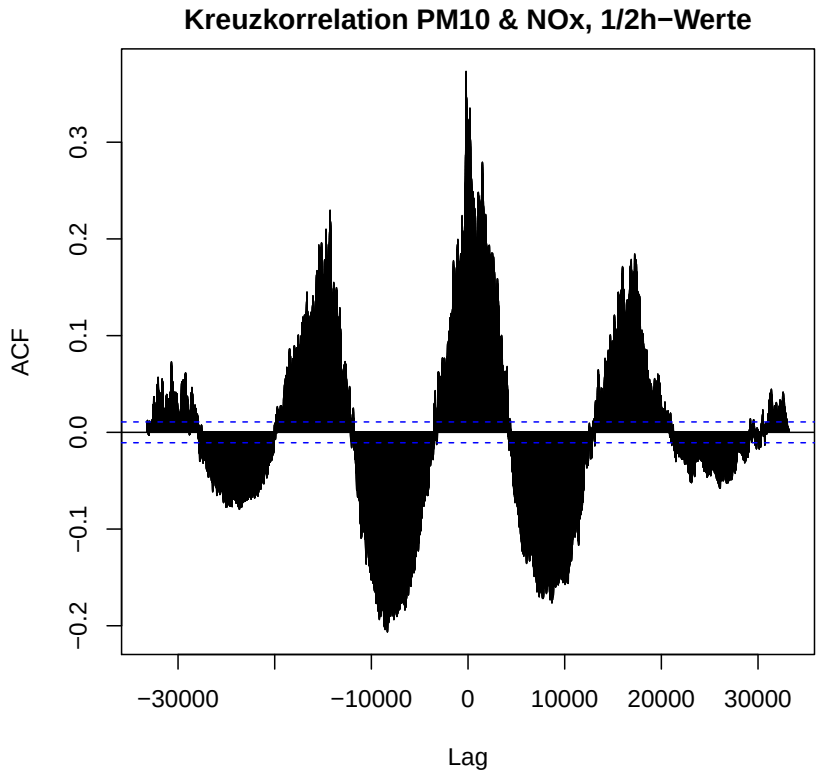


Abbildung 2.7.8: Kreuzkorrelation: PM_{10} und NO_x, $^1/_2$h-Werte

Anhand dieser Graphik lässt sich erkennen, dass die höchste Korrelation beim Lag h=0 liegt. Dies entspricht genau dem Korrelationskoeffizienten von Pearson, den wir bereits zuvor berechnet haben. Des weiteren erkennt man einen sich wiederholenden Verlauf mit einem Zyklus von etwa 17000 Datenpunkte ($\approx$ 12 Monate). Das bedeutet: sind die PM_{10}-Werte in einer Saison hoch (bzw. niedrig), so sind die NO_x-Werte in der darauffolgenden Saison niedrig (bzw. hoch) und so weiter.

3 Zerlegung der Zeitreihen in verschiedene Zeitskalen

Bisher haben wir herausgefunden, dass sowohl die PM_{10}-, als auch die NO_x-Werte eine starke saisonale Abhängigkeit und ähnliche Trends aufweisen. Nun wollen wir noch genauer auf die Beziehung zwischen den beiden Zeitreihen eingehen und auch eventuell vorhandene Störterme ausfindig machen. Die Grundlage für dieses Kapitel bildet dabei ein Abschnitt aus dem Buch von Peng/Domenicini.

3.1 Zerlegung der Zeitskala mit einem Moving Average

Zu Beginn zerlegen wir die Zeitreihe in ihre einzelnen Komponenten, um diese besser untersuchen und mit jenen der anderen Zeitreihe vergleichen zu können. Dazu möchten wir einen Moving Average verwenden. Da unsere Methode allerdings nicht für Zeitreihen funktioniert, die fehlende Werte beeinhaltet, füllen wir diese Stellen mit dem Mittelwert der Zeitreihe.
Eine lineare Regression von PM_{10} und NO_x liefert uns folgendes Ergebnis (Tab. 9):

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	22.03	0.19	116.6	<2e-16
NO_x	0.41	0.00	178.1	<2e-16

Tabelle 9: Einfache lineare Regression von PM_{10} und NO_x

Die Regressionsgerade lautet somit:

$$P\hat{M}_{10} = 22.03 + 0.41 \times NO_x$$

Das deutet auf einen starken positiven Zusammenhang zwischen PM_{10} und NO_x hin. Um ein noch genaueres Bild vom Zusammenhang der zwei Zeitreihen zu erhalten, zerlegen wir nun die Zeitreihen mithilfe der Funktion `filter`. Diese Funktion verwendet die ersten Beobachtungen zur Berechnung, weswegen in der Graphik die ersten Daten fehlen. Plottet man die Zerlegung der Zeitreihen PM_{10} und NO_x nebeneinander, so kann man deren Unterschiede und Gemeinsamkeiten besser erkennen (Abb. 3.1.9).
Der jährliche Trend der beiden Zeitreihen scheint sich nicht sehr ähnlich zu sein, während man sowohl beim saisonalen, als auch beim wöchentlichen Trend große Gemeinsamkeiten erkennen kann. Die Interpretation der Residuen wird allerdings schwierig, sodass wir auf jeden Fall die Ergebnisse der linearen Regression unter Zerlegung der Zeitskalen zuhilfe nehmen sollten (Tab. 10).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-3.634e+02	3.781e+01	-9.614	<2e-16
NO_x.yearly	7.667e+00	7.203e-01	10.644	<2e-16
NO_x.seasonal	3.963e-01	1.435e-02	27.625	<2e-16
NO_x.weekly	2.030e-01	1.299e-02	15.629	<2e-16
NO_x.random	3.098e-01	3.316e-03	93.425	<2e-16

Tabelle 10: Lineare Regression von PM_{10} und NO_x (aufgespalten)

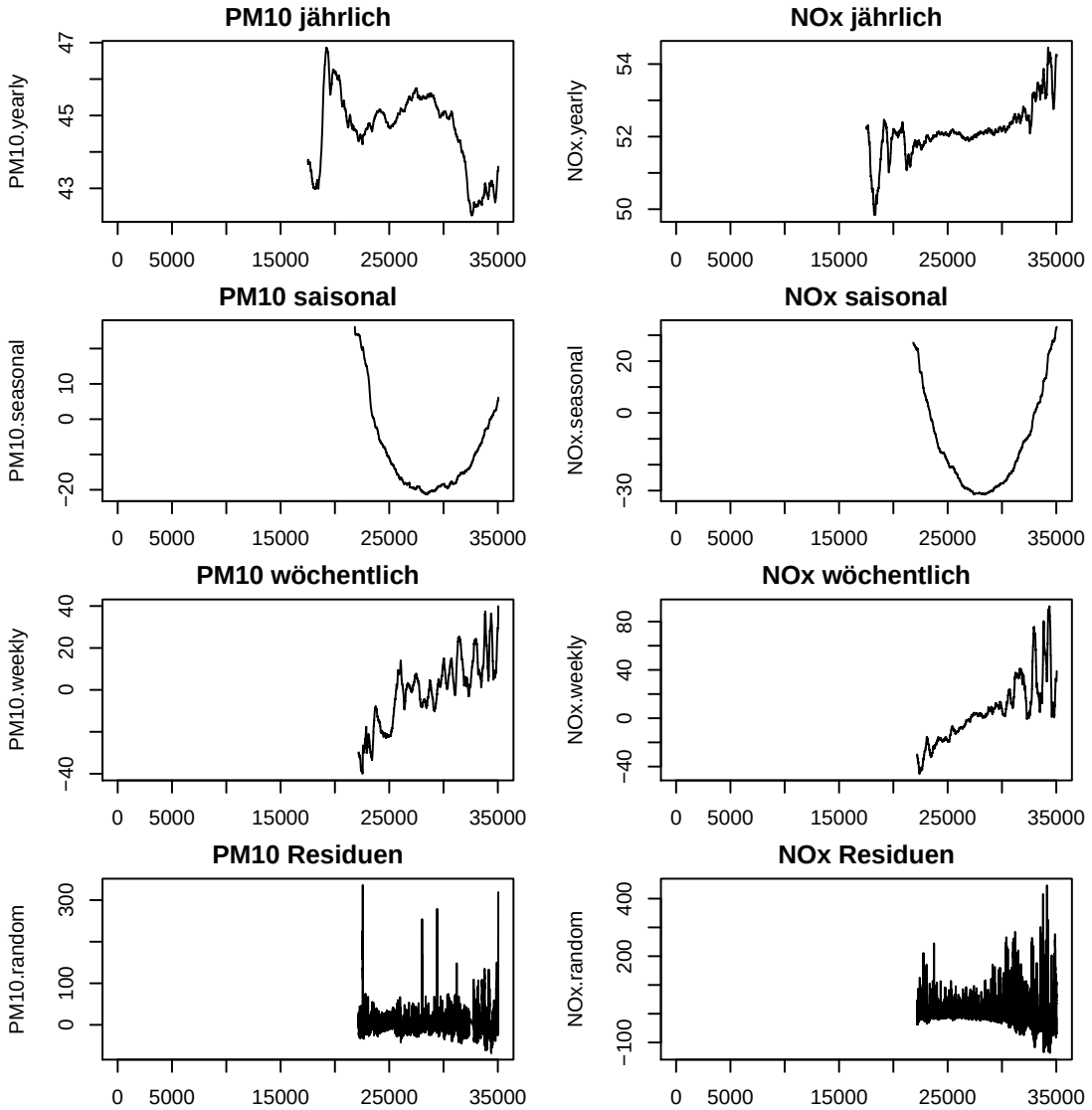


Abbildung 3.1.9: Zerlegung der Zeitskala für PM_{10} und NO_x, Halbstundenwerte

Bis auf den Intercept sind alle Werte positiv, sodass eine positive Korrelation zwischen PM_{10} und den verschiedenen Aufspaltungen der NO_x-Werte existiert. Damit lautet die Regressionsgleichung:

$$\hat{PM}_{10} = -363 + 7.67 \times NO_x.yearly + 0.40 \times NO_x.seasonal + 0.20 \times NO_x.weekly + 0.31 \times NO_x.random$$

Der negative Intercept wirkt zwar sehr groß, allerdings wird er durch die hohen NO_x.yearly-Werte sofort ausgeglichen.

3.2 Zerlegung der Zeitskala mit der Funktion `tsdecomp`

Für die Zerlegung einer Zeitreihe durch eine Fourriertransformation, anstatt mit einem Moving Average, wurde bereits eine Funktion geschrieben, die wir nun verwenden wollen. Mithilfe dieser Funktion (`tsdecomp`) zerlegen wir die Zeitreihen (Halbstundenwerte) in die folgenden drei Zeitskalen:

- ein großer Zyklus über den gesamten Beobachtungszeitraum für den langfristigen Trend
- 2-24 Zyklen für den saisonalen Trend
- über 25 Zyklen für den kurzzeitigen Trend.

Die drei Zeitskalen für die Zeitreihen PM_{10} und NO_x werden in (Abb. 3.2.10) dargestellt.

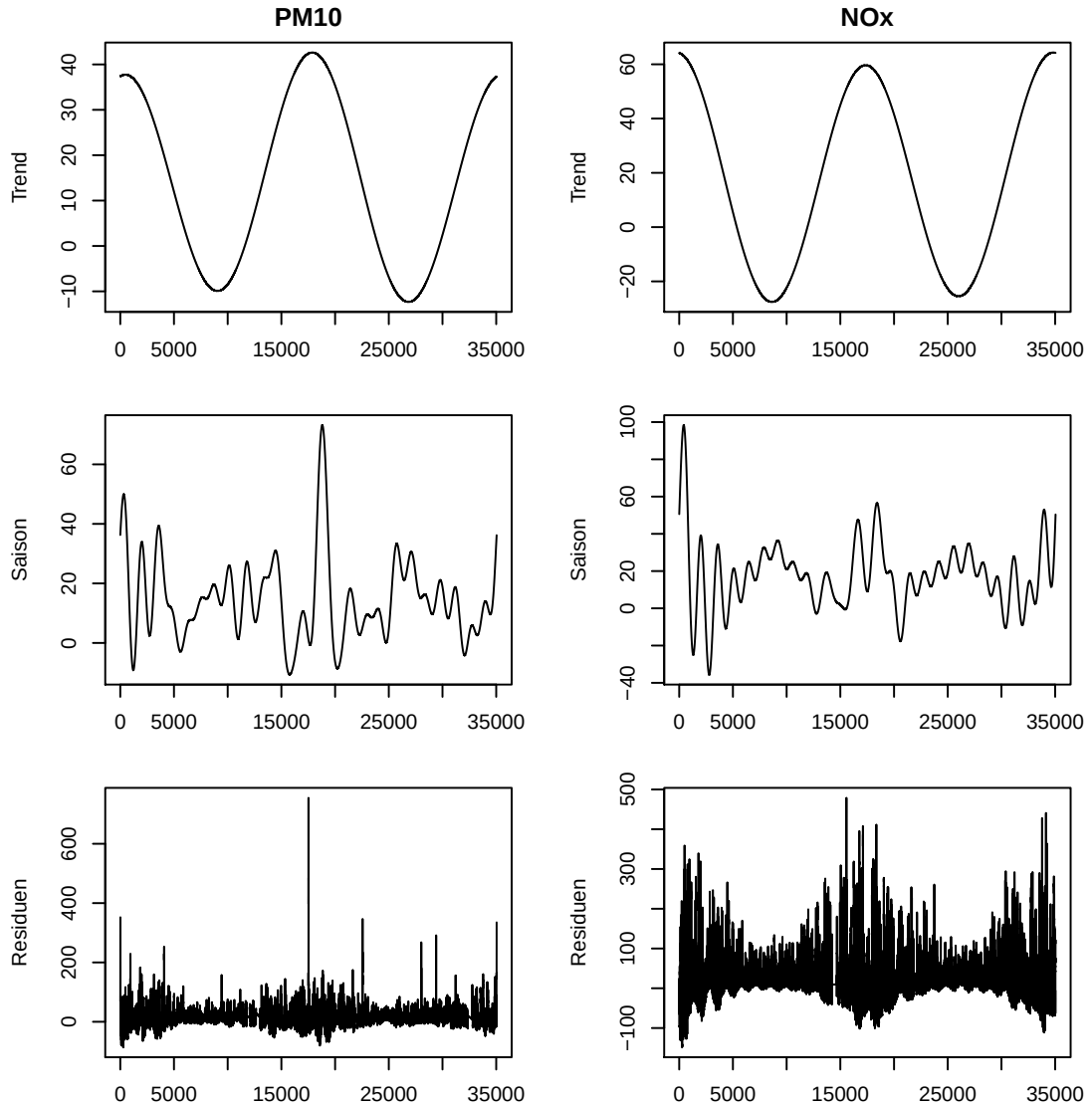


Abbildung 3.2.10: Zerlegung der Zeitskala für PM_{10} und NO_x mit `tsdecomp`, Halbstundenwerte

Auch hier betrachten wir neben der Graphik auch noch eine lineare Regression deren Resultat in Tab. 11 dargestellt ist.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	20.24	0.22	92.48	<2e-16
Trend	0.56	0.00	123.70	<2e-16
Saison	0.41	0.01	52.48	<2e-16
kurzer Trend	0.35	0.00	129.08	<2e-16

Tabelle 11: Lineare Regression von PM_{10} und NO_x (mit tsdecomp)

Die Ergebnisse unter der Berechnung mit der Funktion `tsdecomp` sind denen unter Zuhilfenahme eines Moving Average sehr ähnlich. Auch hier erkennen wir eine positive Korrelation zwischen PM_{10} und dem Trend, Saison bzw. dem kurzzeitigen Trend von NO_x.

Die Funktion `tsdecomp` bietet aber auch die Möglichkeit andere Zeitskalen zu betrachten. Aus diesem Grund werden wir die Zeitreihen in folgende Zeitskalen einteilen:

- langzeitiger Trend
- saisonaler Trend
- 2-4 Wochen
- 1-2 Wochen
- 1-7 Tage
- < 1 Tag

In diese Zeitskalen werden nun sowohl PM_{10}, als auch NO_x eingeteilt und geplottet (Abb. 3.2.11). Man sieht, dass sich die Graphiken der beiden Zeitreihen sehr ähnlich sind und positiv korrelieren. Außerdem erkennt man deutlich, dass die Werte zum Jahreswechsel hin deutlich höher sind, als unter dem Jahr.
Auch hier möchten wir wieder eine lineare Regression von PM_{10} in Abhängigkeit der aufgespaltetenen NO_x-Zeitreihe betrachten (Tab. 12).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	22.13	0.20	111.15	<2e-16 ***
langzeitiger Trend (NO_x)	0.56	0.00	124.94	<2e-16 ***
saisonaler Trend (NO_x)	0.41	0.01	53.01	<2e-16 ***
2-4 Wochen (NO_x)	0.33	0.01	33.95	<2e-16 ***
1-2 Wochen (NO_x)	0.45	0.01	69.00	<2e-16 ***
1-7 Tage (NO_x)	0.40	0.00	85.12	<2e-16 ***
< 1 Tag (NO_x)	0.28	0.00	67.39	<2e-16 ***

Tabelle 12: Lineare Regression von PM_{10} und NO_x (detailliert)

Auch hier erhalten wir eine positive Korrelation von PM_{10} mit den einzelnen Zerlegungen der NO_x-Zeitreihe. Außerdem sind alle Komponenten höchst signifikant, da die p-Werte, sowie die Standardfehler sehr klein sind.

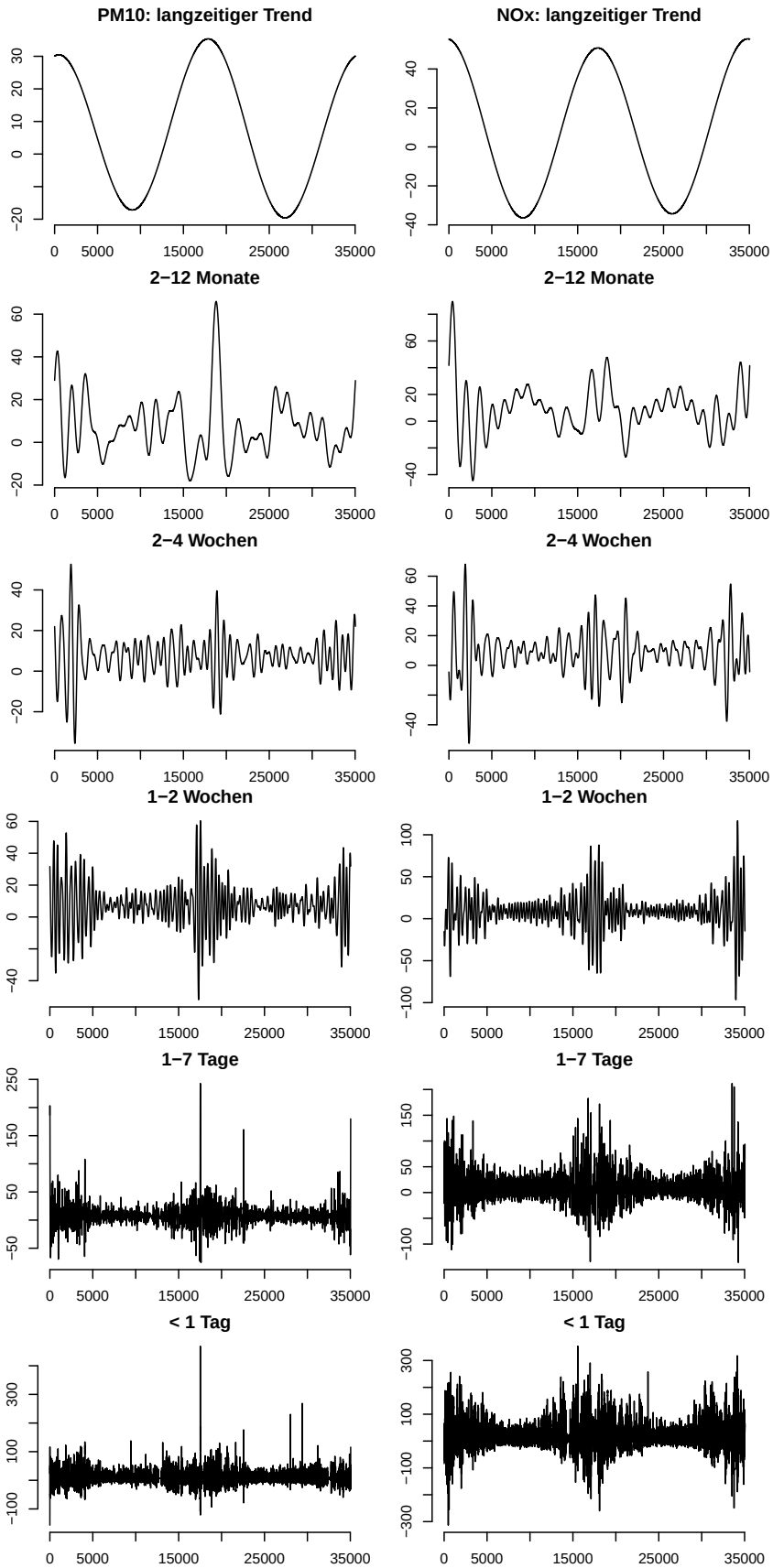


Abbildung 3.2.11: Zerlegung der Zeitskala (detailliert) für PM_{10} und NO_x mit `tsdecomp`, Halbstundenwerte

3.3 Untersuchung auf Störgrößen

Wir haben für die Zerlegung von NO_x in verschiedene Zeitskalen jedes Mal eine positive Korrelation mit PM_{10} feststellen können. Dennoch könnte es sein, dass die Beziehung zwischen PM_{10} und NO_x auch noch durch eine dritte Variable beeinflusst wird. Dies wäre eine sogenannte Störgröße. Da die Lufttemperatur an der Messstation Graz-Süd erst ab dem Jahr 2006 gemessen wurde und wir daher nur Daten für ein Jahr haben, entscheiden wir uns für die Untersuchung der Windgeschwindigkeit.
Zuerst plotten wir die Daten über die Windgeschwindigkeit, um deren Struktur zu erkennen (Abb. 3.3.12).

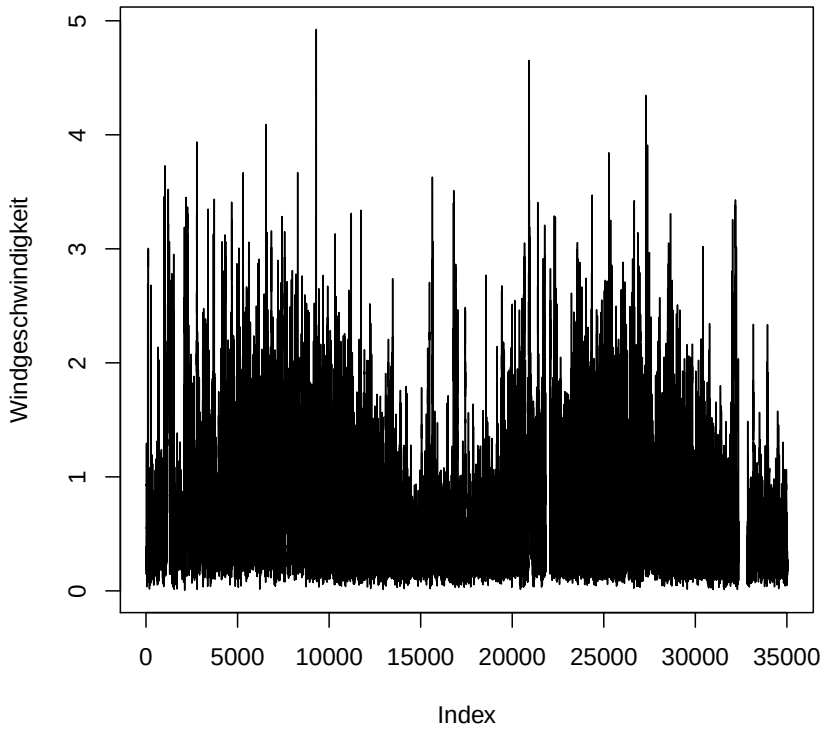


Abbildung 3.3.12: Windgeschwindigkeit Graz-Süd, Halbstundenwerte

Die Zeitreihe der Windgeschwindigkeit weist etwas niedrigere Werte zum Jahreswechsel, und höhere Werte unter dem Jahr auf. Sie ist also in den Sommermonaten etwas stärker, als im Winter.

Nun können wir den Effekt der Windgeschwindigkeit mithilfe einer linearen Regression entfernen. Die daraus resultierenden Zeitreihen werden in Abb. 3.3.13 und Abb. 3.3.14 dargestellt.

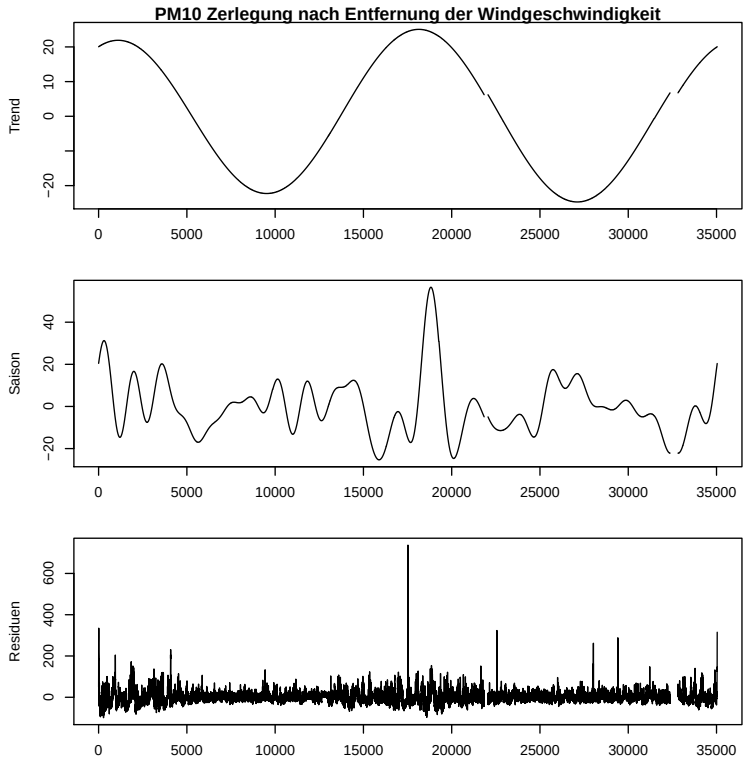


Abbildung 3.3.13: Zerlegung von PM_{10} nach Entfernung des Einflusses der Windgeschwindigkeit, Halbstundenwerte

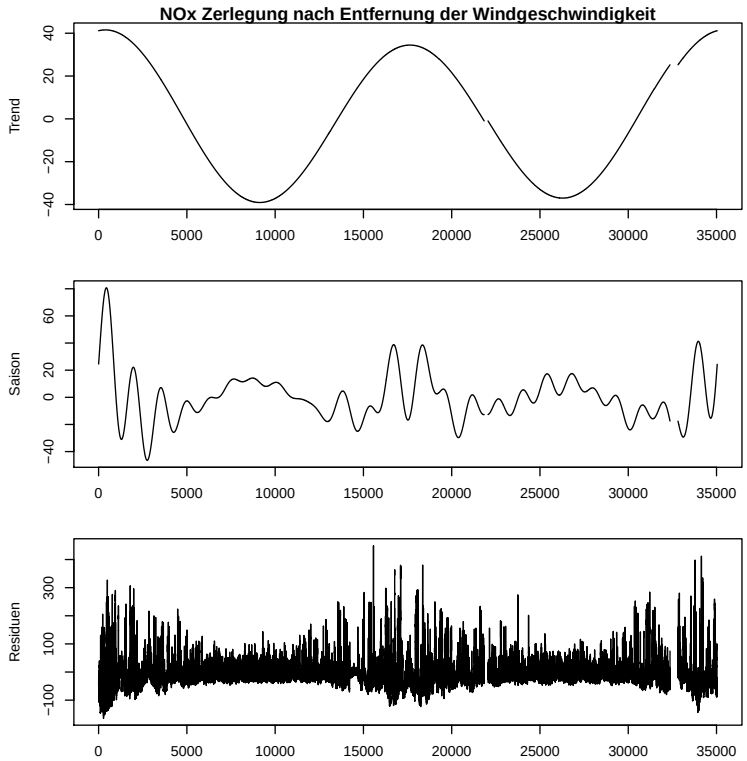


Abbildung 3.3.14: Zerlegung von NO_x nach Entfernung des Einflusses der Windgeschwindigkeit, Halbstundenwerte

Die Trends bleiben recht gut erhalten, allerdings merkt man, dass die Werte nicht mehr so stark schwanken wie zuvor.
Nun kann man noch eine lineare Regression von PM_{10} in Abhängigkeit der auf verschiedene Zeitskalen zerlegten NO_x-Werte mit entfernter Windgeschwindigkeit betrachten. Die Ergebnisse sind in Tab. 13 dargestellt.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	43.55	0.15	285.69	<2e-16 ***
Trend (NO_x)	0.66	0.01	117.52	<2e-16 ***
Saison (NO_x)	0.37	0.01	43.52	<2e-16 ***
kurzzeitiger Trend (NO_x)	0.32	0.00	106.20	<2e-16 ***

Tabelle 13: Lineare Regression von PM_{10} und NO_x nach Entfernung des Einflusses durch die Windgeschwindigkeit

Durch die Entfernung des Einflusses der Windgeschwindigkeit hat sich der Intercept vergrößert. Außerdem wird der Einfluss des Trends von NO_x etwas stärker gewichtet als zuvor. Im Wesentlichen hat sich aber nicht viel verändert und wir sehen die positive Korrelation zwischen PM_{10} und NO_x.

4 Conclusio

Aufgrund der ausführlichen Analyse und der unterschiedlichen Betrachtungsweisen konnte herausgefunden werden, dass die beiden Schadstoffe im Winter höhere Werte aufweisen als im Sommer. Da dies sowohl für das Jahr 2005, als auch für 2006 der Fall ist, wird das ein allgemeines Phänomen sein. Interessant ist hier vor allem, dass die PM_{10}-Werte im Sommer so gut wie nie den Grenzwert (Tagesmittelwerte) von 50 $\mu g/m^3$ überschreiten, während dies im Winter nahezu täglich der Fall ist, wobei der Monat Jänner eindeutig der Spitzenreiter zu sein scheint. So gab es im Winter 2005 insgesamt 96 Überschreitungen, während es im Winter 2006 auch noch 90 waren. Im Sommer wurden hingegen nur an 11 bzw. 5 Tagen zu hohe Tagesmittelwerte für das Jahr 2005 bzw. 2006 gemessen. Da die erlaubte Anzahl an Überschreitungen pro Jahr in diesen Jahren bei 30 lag, konnte diese sowohl im Jahr 2005 (107 Überschreitungen), als auch im darauffolgenden Jahr (95 Überschreitungen) nicht eingehalten werden.
Die Jahresmittelwerte liegen mit 41.84 (2005) bzw. 42.02 $\mu g/m^3$ (2006) knapp über der erlaubten Grenze von 40 $\mu g/m^3$.
Im Süden von Graz fallen die Werte für PM_{10} also doch recht deutlich höher aus, als gesetzlich erlaubt. Aus diesem Grund wurde in bestimmten Streckenabschnitten auf der Südautobahn eine Geschwindigkeitsbegrenzung festgelegt, wodurch die Schadstoffwerte etwas sinken sollten.
Auch die NO_x-Werte fallen im Winter deutlich höher aus als im Sommer. Da aber nur für die NO_2-Werte, nicht jedoch für die NO_x-Werte, ein Grenzwert vorhanden ist, kann man hier nicht von einer bestimmten Anzahl an Überschreitungen sprechen. Dennoch kann man zumindest sagen, dass die Werte im Winter, vor allem im Jänner und Dezember, deutlich höher sind, als in den Sommermonaten.
Gründe für diese saisonalen Unterschiede könnten unter anderem die ungünstige Beckenlage der Stadt Graz und die hohe Anzahl an Fahrzeugen mit Dieselmotoren ohne Katalysatoren sein, da die Benützung von PKWs im Winter häufiger ist als im Sommer.
Zusätzlich zu den Unterschieden im Bezug auf die Saison fällt auf, dass die Schadstoffbelastung an Werktagen höher ausfällt, als an Samstagen oder Sonn- und Feiertagen. Die relevanten Mittelwerte (Halbstundenwerte) wurden in Tab. 14 zusammengefasst.

	PM_{10}			NO_x		
	Werktag	Samstag	So/Fe	**Werktag**	Samstag	So/Fe
Sommer	27.48	24.37	19.60	29.63	18.83	12.42
Winter	64.55	53.07	54.82	94.74	61.71	52.36
Gesamt	**46.09**	38.79	38.02	**61.33**	39.83	32.70

Tabelle 14: Mittelwerte der Schadstoffe (Halbstundenwerte) bzgl. Tagestyp in $\mu g/m^3$

Das ist jedoch nicht überraschend wenn man bedenkt, dass die Abgase der Autos einen großen Beitrag zur Luftverschmutzung leisten. Das könnte auch ein Grund dafür sein, dass die Schadstoffwerte ihre zwei Tagesspitzenwerte in den Morgen- und den Abendstunden haben. Also zu Zeiten, in denen viel Verkehr auf den Straßen herrscht.
Bei der Betrachtung der Korrelation zwischen den zwei Schadstoffzeitreihen konnte ein recht starker positiver linearer Zusammenhang festgestellt werden, wobei dieser im Winter etwas stärker als im Sommer war. Außerdem scheint sich die Lufttemperatur negativ auf die Schadstoffwerte auszuwirken, d.h. bei steigenden Temperaturen sinken die Schadstoffwerte.

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