

Abschlussbericht zum Forschungs- und Entwicklungsvorhaben 201 43 250
auf dem Gebiet des Umweltschutzes „Anwendung modellgestützter
Beurteilungssysteme für die bundeseinheitliche Umsetzung der EU-
Rahmenrichtlinie Luftqualität und ihrer Tochterrichtlinien“

Konzepte zur Formulierung von Kriterien zum Nachweis der von den EU-Richtlinien verlangten Genauigkeit von Modellrechnungen Beispiele mit dem chemischen Transportmodell REM-CALGRID

Johannes Flemming

Rainer Stern



**Freie Universität Berlin
Institut für Meteorologie
Troposphärische Umweltforschung**

Mai 2004

Berichts-Kennblatt

BerichtsnummerUBA-FB	2.	3.
4. Titel des Berichts Konzepte zur Formulierung von Kriterien zum Nachweis der von den EU-Richtlinien verlangten Genauigkeit von Modellrechnungen--Beispiele mit dem chemischen Transportmodell REM/CALGRID		
5. Autor(en), Name(n), Vorname(n) Flemming, Johannes Stern, Rainer	8. Abschlußdatum Mai 2004	
	9. Veröffentlichungsdatum Mai 2004	
6. Durchführende Institution (Name, Anschrift) Freie Universität Berlin, Institut für Meteorologie Carl-Heinrich-Becker-Weg 6-10 12165 Berlin	10. UFOPLAN-Nr. <div style="text-align: center;">201 43 250</div>	
	11. Seitenzahl 83	
	12. Literaturangaben 16	
7. Fördernde Institution (Name, Anschrift) Umweltbundesamt, Postfach 33 00 22, D-14191 Berlin	13. Tabellen und Diagramme 7	
	14. Abbildungen 52	
15. Zusätzliche Angaben		
16. Kurzfassung In dieser Arbeit wurde versucht, Kriterien zu formulieren, die zum Nachweis der von den EU-Richtlinien verlangten Genauigkeit von Modellrechnungen benutzt werden können. Es wurden verschiedene Maße untersucht und anhand einer europaweiten Anwendung des chemischen Transportmodells REM-CALGRID (RCG) auf das gesamte Jahr 2002 und eines Vergleichs der Modellergebnisse mit Messungen aus den deutschen Messnetzen diskutiert. Es wird vorgeschlagen, die Modellgüte nur am Mittelwert und an dem Perzentil zu bestimmen, das der Anzahl von erlaubten Überschreitungen der Kurzfristziel- oder Grenzwerte entspricht. Das Hauptproblem einer Modellevaluierung anhand von Messungen liegt aber in der Auswahl der zur Evaluierung zu benutzenden Stationen, da deren oft unbekannte oder nur unzureichend definierbare Repräsentativität große Auswirkungen auf den Modellfehler haben kann. Die Untersuchungen haben gezeigt, dass es auf Basis der gegenwärtigen EU-Richtlinienbestimmungen nicht möglich ist, eindeutige und objektivierbare Kriterien für eine Modellevaluierung anhand von Messungen zu entwickeln. Hier besteht also noch weiterer Handlungsbedarf.		
17. Schlagwörter EU-Richtlinien zur Luftqualität, Ausbreitungsrechnung, Modellfehler, Repräsentativität von Messungen, REM-CALGRID-Modell		
18. Preis	19.	20.

Report Cover Sheet

Report No.UBA-FB ----	2.	3.
4. Report Title Formulation of criteria to be used for the determination of the accuracy of model calculations according to the requirements of the EU directives for air quality – Examples using the chemical transport model REM-CALGRID		
5. Autor(s), Family Name(s), First Name(s) Flemming, Johannes Stern, Rainer	8. Report Date May 2004	
	9. Publication Date May 2004	
	10. UFOPLAN-Ref. No. <div style="text-align: center;">201 43 250</div>	
	11. No. of Pages 83	
6. Performing Organisation (Name, Address) Freie Universität Berlin, Institut für Meteorologie Carl-Heinrich-Becker-Weg 6-10 12165 Berlin	7. Sponsoring Agency (Name, Address) Umweltbundesamt, Postfach 33 00 22, D-14191 Berlin	
	12. No. of Reference 16	
	13. No. of Tables, Diagrams 7	
	14. No. of Figures 52	
15. Supplementary Notes		
16. Abstract Several measures to define model accuracy were examined based on the results of European wide model calculations for the whole year 2002 employing the chemical transport model REM-CALGRID. Using the observations of the German measurement network, the model accuracy is discussed in context with the requirements given in the EU directives for air quality. It is proposed to calculate model accuracy for the annual mean and that percentile which corresponds to the allowed number of exceedances of the short-term limit values. However, the major problem in the determination of the model uncertainty is the selection of appropriate measurement stations, because the mostly unknown area representativity of sites has a great influence on the model accuracy. Based on the currently available guidance it seems not to be possible to develop clear and objective rules for the determination of model uncertainty.		
17. Keywords EU directives for air quality, dispersion modelling, model accuracy, representativity of stations, REM-CALGRID-model		
18. Price	19.	20.

Forschungs- und Entwicklungsvorhaben 201 43 250

auf dem Gebiet des Umweltschutzes

„Anwendung modellgestützter Beurteilungssysteme für die bundeseinheitliche Umsetzung der EU-Rahmenrichtlinie Luftqualität und ihrer Tochterrichtlinien“

**Konzepte zur Formulierung von Kriterien zum Nachweis der
von den EU-Richtlinien verlangten Genauigkeit von Modell-
rechnungen**

Beispiele mit dem chemischen Transportmodell REM-CALGRID

für:

Umweltbundesamt

II 6.1

Postfach 33 00 22

14191 Berlin

Johannes Flemming

Rainer Stern

Institut für Meteorologie

Freie Universität Berlin

Mai 2004

Inhaltsverzeichnis

1	Einführung	1
2	Die Anforderungen der EU-Richtlinien an die Modellierung.....	3
2.1	<i>Die Grenzwertbetrachtung</i>	3
2.2	<i>Aufgabe der Modellierung.....</i>	4
2.3	<i>Festlegung von allgemeinen Mindestanforderungen an die Modellrechnung</i>	6
2.4	<i>Genauigkeitsanforderungen der Tochterrichtlinien.....</i>	7
3	Die Definition anwendungsorientierter Modellgüteanforderungen	10
3.1	<i>Bestimmung und Auswahl eines Modellgütemaß</i>	10
3.2	<i>Die Definition stationsbezogener Gütemaße.....</i>	11
3.3	<i>Festlegung einer Mindestanforderung für das Modellgütemaß.....</i>	14
3.4	<i>Beobachtungsfehler als Begrenzung für die erreichbare Modellgüte.....</i>	14
3.5	<i>Genauigkeitsanforderungen an die Messung</i>	15
3.6	<i>Auswahl der Stationen für die Modellgütemaßbestimmung.....</i>	16
4	Anwendung der Modellgütemaße.....	19
4.1	<i>Vergleich von Messung und Rechnung am Beispiel einer ausgewählten Station ..</i>	20
4.2	<i>Jahresmittelwerte</i>	26
4.3	<i>Tages und Stundenwerte.....</i>	30
4.3.1	<i>Überblick.....</i>	30
4.3.2	<i>Auswirkungen der Aufgabe des Zeitbezuges</i>	31
4.3.3	<i>Konzentrationswerte beim maximalen Fehler.....</i>	33
4.3.4	<i>Konzentrationswerte beim Perzentil, das der erlaubten Überschreitungsanzahl entspricht.....</i>	34
4.3.5	<i>Vergleich der Fehlermaße max err_p und per err_p_{GW} hinsichtlich der betrachteten Konzentrationsniveaus</i>	36
4.3.6	<i>Relative Fehlermaße.....</i>	39
4.3.7	<i>Überschreitung von Grenzwert- und Beurteilungsschwellen.....</i>	44
4.4	<i>Vergleich von kurz- und langfristigen Fehlermaßen.....</i>	45
4.5	<i>Räumliche Glättung der Messung entsprechend der Modellauflösung</i>	50
5	Beobachtungsfehler und der Auflösungsfehler des „perfekten“ Modells.....	55
5.1	<i>Beobachtungsfehler - Rauschen</i>	56

5.1.1	Varianz des Beobachtungsfehlers	56
5.1.2	Die stationsbezogenen Beobachtungsfehlervarianzen für deutsche Messstationen	56
5.1.3	Vergleich der Beobachtungsfehlervarianzen mit der Forderung von 15% Messgenauigkeit in den EU-Richtlinien.....	57
5.1.4	Einfluss des erlaubten 10% Datenausfalls	59
5.2	<i>Modellauflösung</i>	64
6	Zusammenfassung	73
7	Literatur	76
8	Abbildungsverzeichnis	78
9	Tabellenverzeichnis	83

1 Einführung

Die bisherigen Anwendungen von chemischen Transportmodellen haben gezeigt, dass die Modelle inzwischen weit genug entwickelt sind, um ihren Einsatz für die Kontrolle der Luftqualität und für eine Maßnahmenplanung zu rechtfertigen. Die EU-Rahmenrichtlinie 96/62/EG über die Beurteilung und die Kontrolle der Luftqualität (EU, Amtsblatt der Europäischen Gemeinschaften, 1996) berücksichtigt diese Tatsache und gestattet ausdrücklich zur Erstellung einer Ausgangsbeurteilung zur Luftqualität, zu deren späteren regelmäßigen Beurteilung sowie implizit zur Erstellung von Maßnahmenplänen den Einsatz von Modellen.

Danach können Modelle eingesetzt werden

- Zur räumlichen Darstellung der Schadstoffverteilung in allen Skalenbereichen (National, Ballungsgebiet, Straße), insbesondere in Gebieten, wo die Schadstoffkonzentrationen unter der jeweiligen Oberen Beurteilungsschwelle liegen.
- Zur Prognose der Wirksamkeit von Maßnahmen.

Nach den Verpflichtungen der EU-Richtlinien müssen die zuständigen Behörden für Gebiete und Ballungsräume, in denen die Werte eines oder mehrerer Schadstoffe die Immissionsgrenzwerte überschreiten, Luftreinhaltepläne und Aktionspläne entwickeln, die letztendlich zur Einhaltung der Grenzwerte führen sollen. Insbesondere in solch einer Maßnahmenplanung ist der Einsatz von Modellen unerlässlich, da die Auswirkungen von Maßnahmen auf die Luftqualität nicht über Messungen bestimmt werden können.

Damit ein Modell für die Belange der EU-Richtlinien zur Luftqualität eingesetzt werden kann, muss es den in den Richtlinien definierten Datenqualitätsanforderungen genügen. In den entsprechenden Anhängen der Richtlinien sind die Datenqualitätsziele aber nur sehr allgemein und nicht eindeutig formuliert.

In den Richtlinien wird im Gegensatz zur Vorgehensweise bei der Messung weder eine Referenzmethode für die Modellierung genannt noch gibt es eine nähere Spezifikation der anwendbaren Modelltypen. Es werden jedoch analog zur Messung Anforderungen an die Mindestgenauigkeit der Modellergebnisse gestellt.

Die Bestimmung der Modellgüte soll dabei in folgender Weise erfolgen: „Die Genauigkeit von Modellberechnungen und objektiven Schätzungen ist definiert als die größte Abweichung der gemessenen und berechneten Konzentrationswerte über den betreffenden Zeitraum in Bezug auf den Grenzwert, ohne dass die zeitliche Abfolge der Ereignisse berücksichtigt wird“. Die Gütekriterien für die Genauigkeit beruhen auf stoffabhängigen und mittlungszeitraumbezogenen prozentualen Toleranzgrenzen. Insgesamt ist diese Anleitung aber nicht eindeutig interpretierbar und bedarf für eine praxisgerechte Umsetzung weiterer Klärung.

In dieser Arbeit wird daher versucht, die Vorgaben der EU-Richtlinien zu konkretisieren und Kriterien zu formulieren, die zum Nachweis der von den EU-Richtlinien verlangten Genauigkeit von Modellrechnungen benutzt werden können. Es werden verschiedene Maße vorgeschlagen und anhand einer europaweiten Anwendung des chemischen Transportmodells REM-CALGRID (RCG) auf das gesamte Jahr 2002 und eines Vergleichs der Modellergebnisse mit Messungen aus den deutschen Messnetzen diskutiert. Der Schwerpunkt der Unter-

suchung liegt in der Umsetzung der Genauigkeitsmaße für die Schadstoffe O₃, PM₁₀, SO₂ und NO₂.

Ziel der in diesem Bericht vorgestellten Arbeiten ist jedoch nicht die allgemeine Evaluierung des REM-CALGRID-Modells. Das Modell wird hier lediglich zur Demonstration der Bedeutung verschiedener Fehlermaße für Güte von Modellergebnissen benutzt. Die vorgeschlagenen Maße sind so formuliert, dass sie auf jedes Modell angewendet werden können, das in der Lage ist, die von den EU-Richtlinien geforderte Auflösung von einer Stunde bis zu mindestens einem Jahr zu liefern. Grundsätzlich sind die Gütemaße für Modellrechnungen in allen Skalenbereichen anwendbar, auch wenn sie hier nur für eine großräumige Anwendung diskutiert werden.

In Kapitel 2 werden zuerst die in den EU-Richtlinien definierten Anforderungen an die Modellrechnung diskutiert und interpretiert. Kapitel 3 beinhaltet die verschiedenen Ansätze und Überlegungen für eine praxisorientierte Modellevaluierung. In Kapitel 4 werden dann die vorgestellten Gütemaße für die Rechnungen mit dem REM-CALGRID-Modell in der großräumigen Skala für das Jahr 2002 an den deutschen Messstationen bestimmt und ausgewertet.

Im Kapitel 5 wird diskutiert, inwieweit Unzulänglichkeiten der Beobachtung die Ermittlung der Modellgüte beeinflussen. Die Richtlinie geht bei der Modellgütebestimmung nicht auf diese Unzulänglichkeiten ein, obwohl durch sie auch für ein hypothetisch perfektes Modell im Vergleich mit Messungen immer ein Modellfehler ermittelt werden würde. Dadurch wird eine untere Schranke für diesen Modellfehler und damit ein Schwellenwert für die erreichbare Modellgüte festgelegt. Sie kann als Maß verwendet werden, zu der die ermittelte Güte des untersuchten Modells in Beziehung gesetzt wird. Die wichtigste Unzulänglichkeit der Messung ist deren beschränkte bzw. unbekannte räumliche Repräsentativität. Weiterhin wird untersucht, wie weit der erlaubte Messdatenausfall von bis zu 10% den Modellfehler beeinflussen kann. Der Bericht schließt mit einer Zusammenfassung der gewonnenen Ergebnisse.

2 Die Anforderungen der EU-Richtlinien an die Modellierung

2.1 Die Grenzwertbetrachtung

Die Zielvorstellungen bezüglich der Luftgüte werden in den EU-Richtlinien mit Hilfe von Ziel- oder Grenzwerten ausgedrückt, deren Einhaltung vom Gesetzgeber innerhalb eines Zeitrahmens, d.h. bis 2005 oder 2010 gefordert wird. Die Grenzwerte beziehen sich auf verschiedene Mittelungszeiträume, die von einer Stunde, 8 Stunden, 24 Stunden bis zu einem Jahr reichen. Damit werden die kurz- und langfristigen Wirkungen der Immission auf die menschliche Gesundheit und die Vegetation erfasst.

Die Grenzwerte setzen sich aus einer Konzentrationsschwelle und einer zulässigen Anzahl von Überschreitungen¹ innerhalb eines Jahres zusammen. Dies bedeutet, dass ein Konzentrationsmaß und ein Häufigkeitsmaß (Perzentilwert) den Grenzwert auszeichnen. Neben den Grenzwerten, die einzuhalten sind, führt die Rahmenrichtlinie 1999/30/EC für NO₂, SO₂, PM₁₀ und Blei (EU, 1996), ein System von oberen und unteren Beurteilungsschwellen ein. Mit der Überschreitung dieser Schwellen sind gewisse Vorgaben verbunden, die die zeitliche und räumliche Messdichte sowie die Anwendung von weiteren Methoden der Luftgütebewertung wie Modellierung und objektive Schätzungen betreffen. Die Beurteilungsschwellen entsprechen einer Verringerung der Konzentrationsschwelle des Grenzwertes; die Anzahl der zulässigen Überschreitungen ist jedoch die gleiche wie bei den einzuhaltenden Grenzwerten.

Die Grenzwerte für die betrachteten Stoffe sind in den Tochterrichtlinien der EU-Rahmenrichtlinie zur Luftqualität festgelegt und in Tabelle 1 zu finden.

KOMPONENTE	ART DES SCHWELLENWERTES	MITTEL ÜBER	KONZENTRATIONSSCHWELLE	ZULÄSSIGE ANZAHL VON ÜBERSCHREITUNGEN	GRENZWERT EINZUHALTEN BIS
SO ₂	Grenzwert zum Gesundheitsschutz	1 h	350 µg/m ³	24 x / Jahr (99.73- Perzentil)	1.1.2005
	Grenzwert zum Gesundheitsschutz	24 h	125 µg/m ³	3 x / Jahr (99.18- Perzentil)	1.1.2005
	Grenzwert zum Ökosystemschutz	1 Jahr Winterhj.	20 µg/m ³	--	sofort
PM 10	Grenzwert zum Gesundheitsschutz	24 h	50 µg/m ³	35 x / Jahr (90.41- Perzentil)	1.1.2005
		1 Jahr	40 µg/m ³	--	1.1.2005
	Richtgrenzwert zum Gesundheitsschutz	24 h	50 µg/m ³	7 x / Jahr (98.08- Perzentil)	1.1.2010
		1 Jahr	20 µg/m ³	--	1.1.2010

¹ Für die jahresmittelwertbezogenen Grenzwerte ist die Anzahl der erlaubten Überschreitungen null.

KOMPONENTE	ART DES SCHWELLENWERTES	MITTEL ÜBER	KONZENTRATIONSSCHWELLE	ZULÄSSIGE ANZAHL VON ÜBERSCHREITUNGEN	GRENZWERT EINZUHALTEN BIS
NO ₂	Grenzwert zum Gesundheitsschutz	1 h	200 µg/m ³	18 x / Jahr (99.79- Perzentil)	1.1.2010
		1 Jahr	40 µg/m ³	--	1.1.2010
Ozon	Zielwert zum Gesundheitsschutz	8 h	120 µg/m ³	an 25 Tagen im 3 Jahresmittel (93.15- Perzentil)	1.1.2010
	Langfristziel zum Gesundheitsschutz	8 h	120 µg/m ³	--	--
	Zielwert zum Vegetationsschutz	AOT40 Mai - Juni	18.000 µg/m ³ h		1.1.2010
	Langfristziel zum Vegetationsschutz	AOT40 Mai - Juni	6.000 µg/m ³ h		--
	Informationsschwelle	1 h	180 µg/m ³		--
	Alarmschwelle	1 h	240 µg/m ³		--
	Referenzwert zum Schutz der Wälder	AOT40 Apr.-Sept.	20.000 µg/m ³ h		--

Tabelle 1 Ziel- bzw. Grenzwerte für Ozon, PM10, NO₂ und SO₂ nach den EU-Richtlinien

2.2 Aufgabe der Modellierung

In der EU-Rahmenrichtlinie über die Beurteilung und die Kontrolle der Luftqualität (EU, 1996) werden Modellrechnungen und objektive Schätzungen als Methoden zur Beurteilung in Artikel 6, Absatz 2 und 3 genannt: „Die vorgesehenen Messungen können durch Modellrechnungen ergänzt werden, um angemessene Informationen über die Luftqualität zu liefern.“ Den Modellrechnungen und objektiven Schätzungen wird dabei besonders der Bereich unter dem Grenzwert zugewiesen (Absatz 3). Nicht direkt genannt wird die Modellrechnung als Hilfsmittel für die Erstellung von Maßnahmenplänen, falls die Luftgüte nicht den Vorgaben entspricht. Insbesondere in der Maßnahmenplanung erscheint der Einsatz von Modellen aber unerlässlich, da die Auswirkungen von Maßnahmen auf die Luftqualität nicht über Messungen bestimmt werden können.

In der Tochtrichtlinie 1999/30/EG über die Grenzwerte für SO₂, NO₂, PM10 und Blei (EU, 1999) wird der Modellrechnung und der objektiven Schätzung vorrangig das Belastungsniveau unterhalb der oberen und unteren Beurteilungsschwelle zugewiesen. In Artikel 2, Punkt 13 wird die *obere Beurteilungsschwelle*“ definiert als „einen Wert unterhalb dessen ... eine Kombination von Messungen und Modellrechnungen zur Beurteilung der Luftqualität angewandt werden kann“. Weiterhin definiert Punkt 14 die *untere Beurteilungsschwelle* als „einen Wert ... unterhalb dessen für die Beurteilung der Luftqualität nur Modellrechnungen oder Techniken der objektiven Schätzung angewandt zu werden brauchen“. Dabei ist laut Tochtrichtlinie darauf zu achten, dass eine angemessene räumliche Auflösung gewählt wird (Artikel 7, Absatz 3). Die Werte der oberen und unteren Beurteilungsschwellen werden

in Anhang 1, Abschnitt 5, stoff- und grenzwertspezifisch festgelegt und betragen 40-50% und 60-80% des Grenzwertes für die untere bzw. die obere Beurteilungsschwellen.

Die Methodik der oberen und unteren Beurteilungsschwellen wird in der Tochtrichtlinie 2002/3/EG über den Ozongehalt der Luft nicht eingeführt (EU, 2002). Die Ozon-Modellierung wird hier ergänzend zu Messungen erlaubt, falls Messungen nicht für den geforderten Beurteilungszeitraum von 5 Jahren vorliegen. Weiterhin können Modellrechnungen auch Messungen ersetzen bzw. „die in Anhang V Abschnitt I festgelegte Gesamtzahl der Probenahmestellen“ (Artikel 9, Abschnitt 1) verringern, falls - zusammen mit anderen Forderungen - die „zusätzlichen Methoden (Modellierung oder orientierende Messung) ein angemessenes Informationsniveau für die Beurteilung der Luftqualität in Bezug auf die Zielwerte sowie die Informations- und Alarmschwellen zur Verfügung stellen,, (Artikel 9, Abschnitt 1, Punkt a) und weiterhin „die räumliche Auflösung anderer Techniken (Modellierung oder orientierende Messung) ausreicht, um die Ozonkonzentration im Einklang mit den in Anhang VII Abschnitt I festgelegten Datenqualitätszielen zu ermitteln“ (Artikel 9, Abschnitt 1, Punkt b).

Diese Ausführungen bedeuten, dass bei Ozon die Modellierung auch im und über dem Niveau der Grenzwertüberschreitung zur Luftgütbewertung anwendbar ist. Der höhere Stellenwert der Modellierung in der Ozon-Richtlinie in Vergleich zu 1999/30/EG ist auch an dem Umstand zu erkennen, dass 1. der Vergleich der Modellrechnung mit Beobachtungen explizit in Artikel 11 zur Überprüfung und Berichterstattung verlangt wird und, dass 2. ein Ziel der geforderten Messung von Ozonvorläuferstoffen das verbesserte Prozessverständnis und die Evaluierung photochemischer Modelle ist (Anhang VI).

Neben der Beurteilung der Luftgüte stellt die in der EU-Rahmenrichtlinie 96/62/EG geforderte Erstellung von Maßnahmenplänen (Artikel 7) bei Überschreitung von Grenzwerten ein weiteres wichtiges Anwendungsgebiet für Modellrechnungen dar. Mit den Maßnahmen soll sichergestellt werden, dass Grenzwertüberschreitung bis zu dem festgelegten Termin im Jahr 2005 oder 2010 nicht mehr auftreten. Modelle sind für die Maßnahmenplanung unverzichtbar, da mit ihnen die Wirkung immissionsmindernder Maßnahmen abgeschätzt werden kann. Die Modellierung findet aber in diesem Zusammenhang keine explizite Erwähnung in der Rahmenrichtlinie und den Tochtrichtlinien. Die Maßnahmen müssen jedoch nach Artikel 7 (96/62/EC) „einem integrierten Ansatz zum Schutz von Luft, Wasser und Boden Rechnung tragen“ und „dürfen keine größeren Beeinträchtigungen der Umwelt in den anderen Mitgliedstaaten verursachen.“ Weiterhin muss der Maßnahmenplan laut Anhang IV eine „Schätzung der zu erwartenden Verbesserung der Luftqualität und der für die Verwirklichung dieser Ziele vorgesehenen Frist“ enthalten.

Es scheint deshalb folgerichtig, die Planung von kosteneffizienten Maßnahmen zur Erreichung der langfristigen und kurzfristigen Ziele mit Modellrechnungen zu unterstützen. Eine Eignung des Modells zu Berechnung von Emissionsszenarien erscheint demzufolge eine sinnvolle Forderung an die Modellgüte.

2.3 Festlegung von allgemeinen Mindestanforderungen an die Modellrechnung

Die Rahmenrichtlinie erwähnt keine Referenzmethode für die Modellierung und sie gibt genauso wenig eine nähere Spezifikation der anwendbaren Modelltypen (Eulersche CTM, Gauß-Modelle, Lagrangesche Modelle, Statistische Modelle) an. Aus der gestellten Aufgabe der Modellierung von Überschreitungen von Grenzwerten und den Beurteilungsschwellen sowie der implizit formulierten Aufgabe der Maßnahmenplanung im Falle der Überschreitungen können jedoch allgemeine Modellkriterien abgeleitet werden. Diese sind die Voraussetzung für die Anwendbarkeit der Modelle im Sinne der EU-Richtlinien unabhängig von den Genauigkeitsanforderungen. Die Modelle müssen in der Lage sein:

1. die geforderte zeitliche Auflösung der verschiedenen Kenngrößen von Stundenmittelwerten (SO_2 , NO_2 , O_3), 8h-Stundenmitteln (O_3), Tagesmitteln (PM_{10}) bis zu Jahresmitteln zu bestimmen,
2. Simulationen für den gesamten Bezugszeitraum von meist einem Jahr aber auch für 3 und 5 Jahre bei Ozon durchzuführen (Rechenzeitaufwand, Emissionsdaten, meteorologische Daten),
3. die Immissionsdaten in der Referenzhöhe der Messung (1,5 – 4 m) zu liefern (Grenzwertdefinition, Evaluierung),
4. die Immission in der geforderten räumlichen Auflösung von ca. 15 m (Verkehr nach 99/30/EC), einigen Kilometern (Städtischer Hintergrund) bis zu mehreren 10 km zu berechnen,
5. Gebiete, die vom verkehrsnahen Regime über den Ballungsraum bis zur kontinentalen Skala (10.000 km nach 2002/3/EC) reichen, zu erfassen,
6. zum Zweck der Maßnahmenplanung Emissionsszenarien plausibel in Immissionsänderungen zu überführen.

Die Gesamtheit der Modellanforderungen 1 – 6 ist beim heutigen Stand der Modellentwicklung nur sehr schwer von einem einzigen Modell erfüllbar. Die schwierigsten Punkte sind die hohen Anforderungen an die räumliche Modellauflösung und die Gebietsgröße. Das Hauptproblem ist dabei die Bereitstellung der zum Betreiben der Modelle benötigten Emissionsdaten und meteorologischen Felder in der notwendigen Genauigkeit und Auflösung. Der numerische Aufwand, der auch noch von der Länge des Bezugszeitraums abhängt, wird in den nächsten Jahren durch schnellere Rechner sicherlich besser verkraftbar sein.

Der derzeitige Ansatz zur skalenübergreifenden Modellierung der Immission ist das Koppeln („Nesting“) von Modellen mit verschiedener Auflösung, wobei die Ergebnisse für die größere Skala als Randbedingungen für die feinskaligere Modellierung benutzt werden. Im Rahmen des FE-Vorhabens 29841252 „Modellierung und Prüfung von Strategien zur Verminderung der Belastung durch Ozon“ und des FE-Vorhabens 29943246 „Entwicklung eines Modellsystems für das Zusammenspiel von Messung und Rechnung für die bundeseinheitliche Umsetzung der EU-Rahmenrichtlinie Luftqualität“ wurde solch eine genestete Modellkette auf der Basis des chemischen Transportmodells REM-CALGRID (RCG) entwickelt und angewandt (Stern, 2003, Stern et al., 2003).

Die zeitliche Minimalauflösung von einer Stunde entspricht der von Eulerschen CT-Modellen, die die zeitlichen Änderungen der Immission auf der Basis der relevanten physika-

lisch/chemischen Prozesse simulieren. Modelle mit längeren Mittelungsintervallen oder Stationaritätsannahmen (z.B. immissionsklimatologische Gauß-Modelle) müssen auf statistische Beziehungen zurückgreifen, um die Überschreitung von Grenzwerten der Stunden- oder Tagesmittelwerte zu ermitteln (Hogrefe, 2001). Dieses Vorgehen widerspricht aber dem Ansatz der Richtlinie, für einen Stoff sowohl Lang- und Kurzzeitmittel als Grenzwerte für den Gesundheitsschutz zu definieren. Würde eine feste Beziehung zwischen Lang- und Kurzzeitmittel hinsichtlich der Werte und der Schadenswirkung bestehen, so würde die Festlegung eines Langzeit-Grenzwertes genügen.

Emissionsszenarien zur Maßnahmenplanung können nur mit Modellen durchgeführt werden, bei denen die Emission explizit berücksichtigt wird. Ein wichtiger und in der Rahmenrichtlinie fast vollkommen vernachlässigter Punkt ist jedoch der Einfluss der Jahreswitterung auf die Immission. Er findet nur bei Ozon (2003/3/EC) Berücksichtigung, da hier der Bezugszeitraum auf 3 bzw. 5 Jahre ausgedehnt wird, um zwischenjährliche Schwankungen der Witterung auszugleichen. In Anwendungen des RCG-Modells auf einen langjährigen Zeitraum mit getrennten Emissions- und Meteorologie-Szenarien konnte jedoch der große Einfluss der meteorologischen Variabilität auf die Mittelwerte und Maxima gezeigt werden (Flemming und Stern, 2004). Es erscheint für die Maßnahmenplanung unumgänglich zu sein, auch die Veränderung der Grenzwertüberschreitungen durch die Wettervariabilität zu berücksichtigen. CT-Modelle, die die Wettervariabilität berücksichtigen können, sind hierfür das geeignete Instrumentarium.

2.4 Genauigkeitsanforderungen der Tochterrichtlinien

Die hier betrachteten Tochterrichtlinien (für Ozon und für SO₂, NO₂, PM₁₀ und Blei) definieren im Gegensatz zur Messung keine Referenzmethode für die Modellierung. Es werden jedoch analog zur Messung Anforderungen für die Datenqualitätsziele der Modellierung von Ozon nach 2003/3/EG Anhang VII und von NO₂, PM₁₀ und SO₂ nach 1999/30/EG, Anhang VIII festgelegt. Dabei werden in Abhängigkeit vom Mittelungsintervall der Grenzwerte (1h, 8h, 24h, 1 Jahr) prozentuale Mindestgenauigkeiten gefordert, die in Tabelle 2 aufgeführt sind.

	Ozon	NO ₂	PM ₁₀	SO ₂
Stundenwerte	50%	50-60%		50-60%
Tagesmittel/ 8h Tagesmaximum (O ₃)	50%	-	*	-
Jahresmittel	-	30%	50%	30%

Tabelle 2 Geforderte Genauigkeit der Modellrechnung (Datenqualitätsziel) für Ozon nach 2003/3/EG Anhang VII und für NO₂, PM₁₀ und SO₂ nach EC 1999/30/EG Anhang VIII. * noch nicht festgelegt.

Die Bestimmung der Modellgüte soll dabei in folgender Weise erfolgen: „Die Genauigkeit von Modellberechnungen und objektiven Schätzungen ist definiert als die größte Abweichung der gemessenen und berechneten Konzentrationswerte über den betreffenden Zeitraum in Bezug auf den Grenzwert, ohne dass die zeitliche Abfolge der Ereignisse berücksichtigt wird“. In der englischen Fassung steht: „The uncertainty for modelling and objective estimation is defined as the maximum deviation of the measured and calculated concentra-

tion levels, over the period for calculating the appropriate threshold, without taking into account the timing of the events.”

Für die nicht jahresmittelwertbezogenen Grenzwerte ist diese Anleitung in beiden Fassungen nicht eindeutig interpretierbar und bedarf weiterer Klärung. Als unklar erweist sich der Zusatz „ohne dass die zeitliche Abfolge der Ereignisse berücksichtigt wird“. Das Aufgeben der zeitlichen Zuordnung von Messung und Rechnung ergibt in Verbindung mit der Maximalforderung ein nicht sinnvolles Qualitätsmaß, da die größte Differenz zwischen Messung und Rechnung ohne Zeitbezug der Differenz aus dem Maximum der Messung und dem Minimum der Modellrechnung (oder umgekehrt) entsprechen müsste.

Die Aufgabe der zeitlichen Zuordnung lässt vermuten, dass damit die Übereinstimmung in den Häufigkeitsverteilungen untersucht werden soll. Dies ist plausibel, da nach der EU-Rahmenrichtlinie nur die Tatsache und nicht der Zeitpunkt der Grenzwertüberschreitung innerhalb des Bezugszeitraumes (z.B. 1 Jahr, 3 Jahre oder Winterhalbjahr) von Bedeutung ist. Für die Mindestgenauigkeit der Stunden- oder Tageswerte müssen alle Perzentile bzw. der gesamte Wertebereich der Konzentrationen untersucht werden. Dabei sollten die hohen Perzentile, die den erlaubten Überschreitungshäufigkeiten entsprechen, von besonderem Interesse sein. Die jahresmittelwertbezogenen Grenzwerte liefern eine Fehlermaß, das bereits die Übereinstimmung hinsichtlich des Schwerpunkts der Häufigkeitsverteilung, d.h. der Mittelwerte, beschreibt.

Die geforderte Mindestgenauigkeit ist relativ formuliert, es erfolgt jedoch keine Angabe zu welchem Wert die „größte Abweichung“ ins Verhältnis gesetzt werden muss, um eine prozentuale Angabe für die Mindestanforderung zu erhalten. Es ist jedoch anzunehmen, dass es sich dabei um den entsprechenden Messwert handelt. Die einzuhaltende Reihenfolge bei der Bestimmung der „größten Abweichung“ ist ebenfalls nicht klar formuliert. Es bleibt offen, ob die „größte Abweichung“ der relative Maximalfehler oder der maximale relative Fehler ist.

Wird der größte relative Fehler gesucht, so ist die erlaubte Abweichung bei den niedrigen Werten besonders klein. Vom Modell wird demzufolge eine besonders hohe Genauigkeit bei den niedrigen Immissionswerten gefordert. Dies wäre aus folgenden Gründen problematisch

- niedrige Konzentrationen stellen kaum ein Gefahrenpotential dar,
- die relative Genauigkeit der Messung ist bei niedrigen Werten ebenfalls gering wegen des mit abnehmenden Immissionswerten zunehmenden Einflusses des Messfehlers bzw. der unterschiedlichen Definitionen der unteren Messbereichsgrenze²

Eine von Stern (2004) angewendete Interpretation des „without taking into account the timing of the events“ ist ein Vergleich, der auf den Abweichungen der kumulativen Häufigkeitsverteilung bzw. den Perzentilen basiert. Dazu werden die Daten von Messung und Modell der Größe nach geordnet und die Abweichung zwischen den jeweiligen Perzentilen bestimmt. Bei Stedman et al. (2003) wird die Einhaltung der Modellgenauigkeit ausschließlich an dem Perzentilwert bestimmt, der der Anzahl der erlaubten Überschreitungen entspricht (z.B. 99,79–Perzentil der Stundenwerte NO₂ oder 90,41–Perzentil der Tagesmittelwerte PM₁₀). Im Ansatz von Stern wird die auf den Messwert bezogene relative Abweichung für alle Werte gebildet und deren Maximum ermittelt. Da bei den Messreihen nur eine Min-

² Z.B. ist für SO₂ in Brandenburg der kleinste mögliche Wert 3 µg/m³, während in den benachbarten Bundesländern der häufig gemessene Minimalwert 1 µg/m³ ist

destvollständigkeit³ von 90%⁴ gefordert ist, die Modellzeitreihen aber vollständig sind, stellt sich die Frage nach einer adäquaten Zuordnung der Perzentile⁵. Neben diesem mehr formalen Problem ist unklar, inwieweit ein Ausfall von 10% der Daten die Häufigkeitsverteilung beeinflussen kann und wie das bei der Modellgüteuntersuchung berücksichtigt werden muss (siehe Kapitel 5.1.4).

Die EU-Richtlinien machen weiterhin keine Aussagen über die Auswahl der zur Modellgütebestimmung heranzuziehenden Stationen und an welchen Stationen das Modell die Gütekriterien einhalten muss. Es ist anzunehmen, dass die Modellrechnungen mit den Messreihen aller Stationen verglichen werden müssen, die die in den Richtlinien definierten Datenqualitätskriterien erfüllen und einen Repräsentativitätsbereich besitzen, der der Modellauflösung entspricht.

³ Zusätzlicher Datenausfall ist erlaubt, wenn er durch Kalibrierung und Wartung entsteht.

⁴ Im Winter ist die geforderte Mindestdatenerfassung für Ozon nur 75%.

⁵ Ist die Anzahl der Daten zweier Reihen unterschiedlich, beschreibt ein bestimmter Perzentilwert auch eine unterschiedliche Anzahl von Überschreitungen.

3 Die Definition anwendungsorientierter Modellgüteanforderungen

In diesem Kapitel werden verschiedene Ansätze und Überlegungen für eine praxisorientierte Modellevaluierung diskutiert. Die Modellgüteanforderung der Tochterrichtlinien (siehe Kapitel 2.4) sind dabei der Ausgangspunkt. Aus dem Blickwinkel der Modellierung werden verschiedene Ansätze zur möglichen Interpretation der EU-Richtlinien vorgeschlagen. In Kapitel 4 werden dann die vorgestellten Gütemaße für die Rechnungen mit dem RCG-Modell in der großräumigen Skala für das Jahr 2002 bestimmt und ausgewertet. Da Modellgüteforderungen immer in Relation zur Genauigkeit und Repräsentativität der Beobachtung gesehen werden müssen, wird in Kapitel 5 versucht, dieser Fragestellung nachzugehen.

Die Probleme einer Modellgüteanforderung können in folgende miteinander verbundene Teilaspekte eingeteilt werden:

1. Bestimmung und Auswahl eines Modellgütemaß anhand einer Messzeitreihe
2. Festlegung einer Mindestanforderung für das Modellgütemaß
3. Auswahl der Stationen, die zur Modellgütemaßbestimmung herangezogen werden

Der erste Punkt und zweite Punkt wird in den Rahmenrichtlinien aufgeführt, der dritte wird nicht erwähnt und es ist nur implizit anzunehmen, dass die Modellgütemaße an allen vorliegenden Stationen überprüft und eingehalten werden müssen.

3.1 Bestimmung und Auswahl eines Modellgütemaß

Wie in Kapitel 2.4 bereits erwähnt, ist die Definition des Modellgütemaßes in den Tochterrichtlinien nicht eindeutig hinsichtlich der Aufgabe der zeitlichen Zuordnung von Modell- und Messwert. Man kann jedoch interpretieren, dass ein Vergleich der Häufigkeitsverteilungen unter Aufgabe der zeitlichen Zuordnung angestrebt wird. Ein weiteres Problem stellt die prozentuale Formulierung der erlaubten Fehler dar, da die Mindestanforderung im Bereich der geringen Immission sehr streng ist und kaum erfüllt werden kann.

Die Übereinstimmung der gemessenen und berechneten Häufigkeitsverteilungen als Maßstab für die Modellgüte zu definieren, erscheint ein angemessener Ansatz, da die EU-Richtlinien die Zielvorstellungen zur Luftqualität über Jahresmittelwerte oder die Anzahl von Überschreitungen eines Grenzwerten mit kleineren Mittelungsintervallen (1h, 8 h oder 24 h) innerhalb eines Jahres formulieren. Mit diesem Ansatz können nicht nur die Erwartungswerte (Mittelwerte) und Streuungen, sondern auch Extremwerte verglichen werden. Für Grenzwerte, die als Jahresmittelwerte formuliert sind (bei NO_2 , PM_{10} und SO_2) stellt sich das Problem der zeitlichen Zuordnung nicht, da hier bereits die Erwartungswerte, ausgedrückt durch die Mittelwerte, miteinander verglichen werden.

Der Vergleich von Häufigkeitsverteilungen ist verteilungsunabhängig und ist dem reinen Vergleich von Mittelwerten oder Streuungen vorzuziehen, da diese Parameter bei den häufig nicht-normalverteilten Häufigkeitsverteilungen der Immissionswerte wenig bedeutungsvoll sind. Die Rahmenrichtlinie sieht die Aufgabe der Modellierung bei NO_2 , PM_{10} und SO_2 vorrangig im Bereich der oberen und unteren Beurteilungsschwelle, für Ozon aber auch im Be-

reich der Grenzwerte. Daher ist es sinnvoll, die Modellgüte besonders im Bereich der Beurteilungsschwellen und der Grenzwerte zu überprüfen.

Es sei angemerkt, dass die Aufgabe des zeitlichen Bezuges nicht ganz berechtigt ist, wenn wie in der Tochtrichtlinie für Ozon (2002/2/EG, Anhang II) oder für SO₂ und NO₂ (1999/30/EG, Anhang I und II) die Informations- oder Alarmpflicht zusammen mit der Anforderung zur Prognose formuliert wird. Bei der Immissionsprognose spielt der zeitliche Bezug eine Rolle, so dass hierfür zusätzlich RMSE⁶ und zeitliche Korrelation betrachtet werden sollten. Die Aufgabe des Zeitbezuges entspricht auch nicht dem wissenschaftlichen Anspruch der Eulerschen CTM-Modelle, da diese ja gerade die zeitliche Entwicklung der Immission wiedergeben wollen.

Der RMSE hat sich für die Überprüfung der Modellgüte bewährt, da er sowohl die mittlere und damit zeitbezugslose Abweichung (Bias) als auch die Streuung der zeitbezogenen Fehler bewertet (Flemming, 2003a, 2003c). Der Streuung der Abweichungen zwischen Modell und Messung kommt zudem eine besondere Bedeutung zu, da sie auch von den Beobachtungsfehlern (siehe Kapitel 5) beeinflusst wird. Die mangelnde räumliche Repräsentativität der Messungen wird häufig als ein „Rauschen“ im Messsignal gedeutet. Um diese Größe zu schätzen, geht man meist von der Vorstellung aus, dass der Repräsentativitätsfehler im Mittel verschwindet. Der Beobachtungsfehler trägt demzufolge zum RMSE, nicht aber zum Bias des ermittelten Modellfehlers bei. Will man den Fehler der Beobachtung bei der Modellgütebewertung beachten, so ist dazu der RMSE sehr gut geeignet, da die Standardabweichung des Modellfehlers eine untere Schranke für den RMSE darstellt.

Eine weitere, in den EU-Richtlinien nur implizit erwähnte Aufgabe der Modellierung ist die Maßnahmenplanung, bei der die Auswirkungen von Emissionsminderungen auf die Grenzwertüberschreitungen simuliert werden sollen. Eine Modellgüteforderung sollte demzufolge auch die Eignung eines Modells für Szenarioberechnungen testen. Für diese Fragestellung gibt es bis jetzt keine einheitliche Vorgehensweise. Modellvergleiche zeigen jedoch deutlich, dass komplexe Eulersche CT-Modelle auch unterschiedlich auf Emissionsszenarien reagieren (Römer et al., 2003; siehe auch <http://rea.ei.jrc.it/netshare/thunis/citydelta> und <http://rea.ei.jrc.it/netshare/thunis/eurodelta>). Eine Untersuchung der Fähigkeit des RCG-Modells, die zwischenjährliche Variabilität aufgrund von Emissionstrends und Wettereinfluss zu simulieren, ist in Fleming und Stern (2004) zu finden.

3.2 Die Definition stationsbezogener Gütemaße

Die folgenden Gütemaße werden für Zeitreihen der Beobachtung und für am Stationsort simulierte Modellzeitreihen gewonnen. Das Gütemaß soll im Sinne der Anwendung wichtige Charakteristiken des Unterschiedes zwischen beiden Zeitreihen in einem Wert zusammenfassen. Dafür wird meist ein statistischer Parameter der Zeitreihe der Abweichungen verwendet. Bei der Untersuchung der Modellgüte für die Simulation von Jahresmittelwerten ist dieses Zusammenfassen bereits durch den Jahresmittelwert selbst gegeben. Das Gütemaß für den Jahresmittelwert kann gewissermaßen als die Differenz eines Perzentils in der Nähe des Medians interpretiert werden. Die folgenden Gütemaße orientieren sich demnach vorrangig

⁶ Root mean square error, mittlerer quadratischer Fehler

an dem Vergleich von Perzentilen der Stunden- oder Tagesmittel. Es werden folgende Definitionen verwendet:

Perzentilbezug

Zeitbezug

$p = 1, n_p$ Perzentile

$n_p = n_t$

$t = 1, n_t$ Zeitpunkte

$$(p_1 < p_2 \longleftrightarrow o_{p_1} \leq o_{p_2} \text{ und } r_{p_1} \leq r_{p_2})$$

$o_p \dots$ Beobachtung, der Größe nach geordnet $o_t \dots$ Beobachtung zum Zeitpunkt t

$r_p \dots$ Modellwert, der Größe nach geordnet $r_t \dots$ Modellwert zum Zeitpunkt t

Werden die Abweichungen zwischen Modell und Beobachtung aus Daten des gleichen Zeitpunktes gewonnen, so werden sie mit **Zeit-Bezug** ermittelt. Ordnet man stattdessen die Werte der Größe nach und bildet die Differenz immer zwischen den Werten gleicher Rangfolge, so spricht man von einem Gütemaß mit **Perzentil-Bezug**. Bei gewissen Gütemaßen mit linearer Mittelung wie dem Bias ergibt sich kein Unterschied zwischen Zeit- und Perzentil-Bezug. Die Differenz bei gleicher Zeit oder gleichem Perzentil entspricht immer einem Konzentrationsunterschied. Im Rahmen des Perzentil-Bezuges ist es jedoch unter Umständen auch sinnvoll, bei einem gewissen Konzentrationswert, z.B. einer Grenzwertschwelle, die Unterschiede in den Perzentilen, d.h. die Häufigkeit, wie oft dieser Wert überschritten wird, zu vergleichen.

Die Aufgabe des Zeitbezuges nach der Rahmenrichtlinie legt die Betrachtung im Perzentil-Bezug nahe. Das Fehlermaß nach den Tochterrichtlinien wäre dann der größte absolute Fehler im Perzentil-Bezug (**max err_p**), der mit dem entsprechenden Messwert normiert werden muss und damit den Fehler **rel max err_p** ergibt. Alternativ könnte der größte relative Fehler mit Perzentil-Bezug (**max rel err_p**) bestimmt werden. Maximalbestimmung und Normierung sind zwischen beiden Varianten vertauscht:

Perzentilbezug

Zeitbezug

$p = 1, n_p$ Perzentile

$n_p = n_t$

$t = 1, n_t$ Zeitpunkte

$$\max rel err_p = \max \left(\left| \frac{o_p - r_p}{o_p} \right| \right)$$

$$\max rel err_t = \max \left(\left| \frac{o_t - r_t}{o_t} \right| \right)$$

$$rel max err_p = \frac{\max(|o_p - r_p|)}{o_{p_{\max(o_p - r_p)}}}$$

$$rel max err_t = \frac{\max(|o_t - r_t|)}{o_{t_{\max(o_t - r_t)}}}$$

Ein weiterer sinnvoller Ansatz ist nicht die Betrachtung des Maximums der Differenzen aus allen Perzentilen, sondern die Betrachtung der Übereinstimmung der Häufigkeitsverteilungen nur an bestimmten Perzentilen. Hierfür eignet sich die Konzentrationsdifferenz bei dem Perzentil, das der Anzahl von erlaubten Grenzwertüberschreitungen entspricht, und die Konzentrationsdifferenz beim Median. Mit ersterer kann die Übereinstimmung im Bereich der höheren Konzentrationen bewertet werden. Der Vergleich im mittleren Konzentrationsniveau erfolgt durch die Differenz der jeweiligen Mediane. Dieser Unterschied im Bereich

der mittleren Werte wird jedoch in gewisser Weise schon durch den Jahresmittelwertvergleich erfasst, wie er in 1999/30/EC für SO₂, NO₂ und PM₁₀ gefordert ist. Eine Modellüberprüfung im Bereich der kleinen Perzentilwerte scheint nicht sinnvoll, da die niedrigen Konzentrationswerte häufig nahe oder gleich Null sind, relative Maße demzufolge ungeeignet sind und die Definition der unteren Messbereichsgrenze den Vergleich zwischen Modell und Beobachtung dominieren kann. Darüber hinaus besitzen die geringen Konzentrationen meist keine umweltpolitische Relevanz. Es ergibt sich als Gütemaß:

Perzentilbezug

$p = 1, n_p$ Perzentile

$$rel\ per\ err - p = \frac{|o_p - r_p|}{o_p}, p = p_{GW}$$

p_{GW} wäre hier z.B. das 99,79-Perzentil der Stundenwerte NO₂ oder das 90,41-Perzentil der Tagesmittelwerte PM₁₀.

Die bisherigen Fehlermaße mit und ohne Zeit-Bezug basieren immer auf einer Differenz der Konzentrationen zwischen Modell und Beobachtung. Im Rahmen der Betrachtung im Perzentil-Bezug macht es auch Sinn, nicht die Konzentrationsunterschiede bei gleichem Perzentil, sondern die Unterschiede in der Häufigkeit der Überschreitung von festgelegten Konzentrationswerten zu bewerten. Das Fehlermaß ist dann eine Perzentil-Differenz Δp_{GW} bei einem gewissen Konzentrationswert, normalerweise dem des Grenzwerts GW oder dem Wert für die obere und untere Beurteilungsschwelle. Der Vergleich der Perzentile bei festgehaltenem Konzentrationswert ist jedoch nur sinnvoll für den Konzentrationsbereich, der von Modellrechnung und Beobachtung mit Werten belegt ist. Damit ergibt sich als weiteres Gütemaß:

Perzentilbezug

$p = 1, n_p$ Perzentile

Perzentildifferenz Δp bei Konzentration $r, o = GW$:

$$\Delta p_{GW} = |p_{GW}^o - p_{GW}^r| \quad p_{GW}^o = p \text{ falls } o_p \geq GW \text{ und } p_{GW}^r = p \text{ falls } r_p \geq GW$$

Sehr häufig wird bei der Überprüfung der Modellgüte auf die EPA Richtlinien hingewiesen, die auch im EEA-Report (Kielland et al, 1998) zur vorläufigen Einschätzung der Luftqualität nach EU Rahmenrichtlinie benutzt werden. Es wird dabei davon ausgegangen, dass der Modellfehler immer aus Mess- und Modellwert zum gleichen Zeitpunkt gebildet wird. Die EPA-Richtlinien umfassen Mittelwert und Streuung von Modell und Messungen, sowie den absoluten und relativen Bias, RMSE, die Streuungen und die zeitliche Korrelation. Im Folgenden werden die Formel für Bias und RMSE mit und ohne Zeitbezug vorgestellt. Der Bias unterscheidet sich nicht zwischen Zeit- und Perzentil-Bezug:

Perzentilbezug

 $p = 1, n_p$ Perzentile

Zeitbezug

 $n_p = n_t$ $t = 1, n_t$ Zeitpunkte

$$bias_p = \frac{1}{n} \sum_{p=1}^{n_p} o_p - \frac{1}{n} \sum_{p=1}^{n_p} r_p = \frac{1}{n_p} \sum_{p=1}^{n_p} (o_p - r_p) = bias_t$$

$$rmse_p = \sqrt{\frac{1}{n_p} \sum_{p=1}^{n_p} (o_p - r_p)^2}$$

$$rmse_t = \sqrt{\frac{1}{n_t} \sum_{t=1}^{n_t} (o_t - r_t)^2}$$

3.3 Festlegung einer Mindestanforderung für das Modellgütemaß

Nachdem der Wahl eines geeigneten Modellgütemaßes stellt sich das Problem der anzustrebenden Mindestanforderung an die Modellgenauigkeit. Die Mindestanforderungen werden in den EU-Richtlinien (siehe Tabelle 2) ohne Begründung der gewählten Schwellenwerte festgeschrieben, müssten aber prinzipiell von folgenden Randbedingungen abhängig gemacht werden können:

- Wahl des Modellgütemaßes und der Aufgabenstellung
- Güte und Repräsentativität der Messungen.
- Stand der Forschung auf dem Gebiet der Modellierung.

Die Mindestanforderung an das Modellgütemaß muss auf die Aufgabe der Modellierung im Rahmen der Richtlinie abgestimmt sein, d.h. auf die Modellierung im Bereich der Beurteilungsschwellen und Schwellenwerte. Ein weiterer wichtiger Aspekt bei der Formulierung der Mindestanforderungen für die Modellgüte ist abhängig vom politischen Willen, die Modellierung als kostengünstige Ergänzung zur Messung zu fördern. Werden die Modellgüteanforderungen so formuliert, dass sie auf absehbare Zeit nicht bzw. prinzipiell nur schwer erfüllt werden, so würde eine wesentliche Motivation für die wissenschaftliche Weiterentwicklung der Modelle ausgeschaltet werden. In gleicher Weise bieten aber zu große bzw. zu leicht erfüllbare Toleranzgrenzen keinen Anreiz für zur Weiterentwicklung. In Analogie zu den Toleranzmargen der Grenzwerte von der Gegenwart bis 2010 könnten progressiv verschärfende Modellgüteanforderungen gestellt werden.

3.4 Beobachtungsfehler als Begrenzung für die erreichbare Modellgüte

Ein vernünftiger Zugang für das Problem der Quantifizierung der Mindestanforderungen liegt in dem Versuch, die theoretisch mögliche Modellgüte zu bestimmen. Damit würde man gewissermaßen eine nachvollziehbare „Maßeinheit“ für die Modellgüte gewinnen. Die Modellgüte eines „perfekten“ Modells, d.h. eines fehlerfreien Modells, würde in Bezug auf die Messungen nur durch den immer vorhandenen Beobachtungsfehler begrenzt sein. Der Beo-

bachtungsfehler ist dabei die Summe aus dem reinen Messgerätefehler und dem sog. Repräsentanzfehler, der die Abweichung der Messung von einem für ein Gebiet als repräsentativ angesehenen Wert angibt. Ein perfektes Modell müsste theoretisch eine unendliche große Auflösung besitzen, da die Natur oberhalb der molekularen Skala keine Diskretisierung kennt. Selbst unter der Annahme einer perfekten Simulation der physikalisch/chemischen Prozesse hat auch ein perfektes Modell immer Auflösungsfehler, da Modelldaten (z.B. die Emissionen) nur diskret vorliegen und die numerische Lösung der Gleichungssysteme bei Gittermodellen nur in einem diskretisierten Raum praktisch durchführbar ist. Beobachtungsfehler und Auflösungsfehler eines hypothetisch „perfekten“ Modells werden in Kapitel 5 im Detail diskutiert.

3.5 Genauigkeitsanforderungen an die Messung

Die Rahmenrichtlinie gibt wie für die Modellrechnung auch für die Genauigkeit der Messungen ein prozentuales Gütemaß an (siehe Tabelle 3 und Annex VIII in 99/30/EC und Annex VII in 2003/3/EC). Die Prozentangaben gelten für Einzelmessungen, gemittelt über den betreffenden Zeitraum, wobei die geforderte Genauigkeit von kontinuierlichen Messungen im Bereich des jeweiligen Grenzwertes gilt. Bei dem in den Richtlinien angegebenen Vertrauensbereich von 95% entspricht diese Spanne dem zweifachen der Standardabweichung des als biasfrei angenommenen Messfehlers. Der Messfehler des Messgeräts wird mit Hilfe einer Referenzmessmethode im Labor bestimmt.

	Ozon	NO ₂	PM ₁₀	SO ₂
Einzelmessung	15%	15%	25%	15%

Tabelle 3 Geforderte Genauigkeit der kontinuierlichen Messungen bezogen auf die Einzelmessung gemittelt über den betreffenden Zeitraum und in der Nähe des jeweiligen Grenz- oder Schwellenwertes für Ozon nach 2003/3/EG Anhang VII und für NO₂, PM₁₀ und SO₂ nach EC 1999/30/EG Anhang VIII.

Die Richtlinien enthalten auch Hinweise zur räumlichen Repräsentativität der Messungen, ohne jedoch Angaben zu machen, wie diese Kriterien überprüfbar sind. Tabelle 4 listet die Größe der nach den Richtlinien zu betrachtenden Gebiete auf und gibt die für eine Modellierung solch eines Gebiets notwendige Modellauflösung an. So müsste bei der geforderten Repräsentativität von Messungen im Verkehrsregime von 200 m² ein Eulersches Modell mit einer Mindestauflösung von 14 m arbeiten. Städtische Hintergrundberechnungen müssten mit einer Gitterauflösung von 2-3 Kilometern durchgeführt werden.

Regime nach 1999/30/EC oder 2003/3/EC	Ozon	NO ₂ , PM ₁₀ , SO ₂	entsprechende Modellauflösung
Verkehr	-	200 m ²	14 m
Städtischer Hintergrund	Einige km ²	Einige km ²	2-3 km
Vorstädtisch	Einige Dutzend km ²	-	5-10 km
Ländlich	Ein paar 100 km ²	-	10-30 km
Ländlicher Hintergrund	1000-10000 km ²	-	30-100 km

Tabelle 4 Geforderte räumliche Repräsentativität der kontinuierlichen Messungen und die entsprechende Modellauflösung eines Eulerschen CTM für Ozon nach 2003/3/EG Anhang VII und für NO₂, PM₁₀ und SO₂ nach EC 1999/30/EG Anhang VIII.

Es stellt sich die Frage, wie diese Mindestanforderungen für die Genauigkeit und Repräsentativität der Messungen hinsichtlich der Modellgüteanforderung verwendet werden können. Zunächst ist es problematisch, dass die Güteanforderungen für Modell und Messung nicht vergleichbar festgelegt sind. Bei den Messungen bezieht sich die prozentual angegebene Fehlerbreite auf die mit Zeitbezug ermittelte Differenz zur Referenzmessung im Bereich der Grenzwerte, während bei der Modellgüte der Zeitbezug aufgehoben wird und es unklar ist, mit welchem Wert die Normbildung erfolgen soll (siehe Kapitel 2.4). Für die in Kapitel 4 durchgeführten praktischen Tests mit verschiedenen Genauigkeitsmaßen empfiehlt es sich demzufolge, den Modellfehler auch entsprechend der Angaben für die Messfehlerbestimmung, d.h. mit Zeitbezug und bezogen auf Stundenwerte, zu ermitteln.

Die Ungenauigkeiten, die durch die räumlich eingeschränkte Repräsentativität der Messungen entstehen, sind bei den normalerweise stark strukturierten Immissionsfeldern von wesentlich größerer Bedeutung als der Messgerätefehler. Die Überprüfung der geforderten Repräsentativität ist zudem wesentlich aufwendiger als die Überprüfung der Messgenauigkeit. Dazu wären zusätzliche Messungen im angenommenen Repräsentativitätsbereich der betreffenden Station notwendig. Grundsätzlich muss man sogar feststellen, dass jede Messstation, insbesondere die städtischen Hintergrundstationen, zeitweise soweit von subskaligen Phänomenen beeinflusst werden, dass der Repräsentativitätsbereich stark eingeschränkt wird. Die Repräsentativität einer Station kann demzufolge auch nur in einem mittleren Sinne bzw. ohne Zeitbezug definiert werden.

Trotz diesen Einschränkungen ist eine der Modellauflösung angemessene Auswahl von Stationen von entscheidender Bedeutung für die Einschätzung der Modellgüte.

3.6 Auswahl der Stationen für die Modellgütemaßbestimmung

Wie bereits ausgeführt machen die Richtlinien keine Aussagen über eine mögliche Auswahl der zur Modellgütebestimmung heranzuziehenden Stationen und an welchen Stationen das Modell die Gütekriterien einhalten muss. Nimmt man alle verfügbaren Stationen, hat man Probleme mit der Repräsentativität. Auch die Festlegung einer absoluten oder relativen Mindestanzahl der Messungen, die mit hinreichender Genauigkeit simuliert werden, ist problematisch, da die Anzahl der verfügbaren Messungen in den einzelnen Regimes bzw. Gebieten von der Messnetzstrategie der Messnetzbetreiber abhängt. So ist z.B. bekannt, dass in Bayern PM₁₀-Messungen vorrangig in mehr städtisch geprägten Gebieten gemacht werden,

während im Nachbarland Baden-Württemberg die Mehrzahl der Stationen in eher ländlichen Gebieten zu finden ist. Die Tochterrichtlinien geben eine Mindestanzahl von kontinuierlich zu betreibenden Messstationen in Ballungsgebieten in Abhängigkeit von Bevölkerungszahl an (2003/3/EC Anhang V und 1999/30/EG Anhang VII). Die Anzahl hängt davon ab, ob Grenzwerte oder Beurteilungsschwellen überschritten werden. Eine Stadt von der Größe Berlins (ca. 3,5 Mio. Einwohner) hätte danach, falls die obere Beurteilungsschwelle überschritten wird, mindestens 7 ansonsten mindestens drei NO₂-Messstationen zu betreiben. Die Anzahl der messenden Stationen ist jedoch deutlich höher und dieser Umstand gilt höchstwahrscheinlich für die gesamte Bundesrepublik. Eine höhere Messdichte würde wahrscheinlich auch eine höhere Anzahl von Stationen bedeuten, an denen das Modellgütekriterium nicht erfüllt ist. Eine niedrige Messnetzdicke ist jedoch nicht im Sinne der Modellierung, da mit dichten Netzen viel besser die möglichen Repräsentativitätsfehler der Beobachtungen ermittelt werden können.

Die Richtlinie fordert die Messungen in verschiedenen Immissionsregimes und eine genaue Dokumentation der Messumgebung. Diese Dokumentation kann in Verbindung mit Tabelle 4 benutzt werden, um die der verwendeten Modellskala entsprechenden Stationen auszuwählen.

Darüber hinaus existieren die von Flemming (2003b) auf Basis der gemessenen Zeitreihen entwickelten objektiven Klassifikationsansätze in Immissionsregimes. Diese Klassifikation erfolgt separat für jede Spezies mit einem hierarchischen Clusterverfahren für ausgewählte klimatologische Parameter der Jahreszeitreihen, die durch Tests der linearen Unabhängigkeit und der Stabilität des Clusterergebnisses für die Clusterung ausgewählt und skaliert wurden. Die Klassifikation beruht für alle Spezies auf einem Maß für die mittlere Belastung und einem für die Ausprägung des Tagesganges (normierte Tagesschwankung). Für NO, NO₂, SO₂ und PM₁₀ spielt der Tagesgang eine untergeordnete Rolle, für Ozon ist er jedoch für eine stabile Clusterung von Bedeutung.

Für Ozon unterscheidet die mit der Cluster-Methode abgeleitete Klassifikation für die Messstationen die Regime:

- B = Berg
- R = Land
- U1 = Vorstadt
- U2 = Stadt
- U3 = belastete Stadt
- S = Strasse

Für NO und NO₂ werden mit Ausnahme der Kategorie „Berg“ die gleichen Regimetypen wie beim Ozon ermittelt. Zusätzlich ergibt die Clusterung bei den Stickoxiden noch das Regime „Straße extrem“. Die SO₂- und PM₁₀-Immissionen werden in fünf Regime mit zunehmender Belastung eingeteilt. Eine Zusammenstellung der mit Hilfe der Clusterung ermittelten Immissionsregime zeigt die Tabelle 5.

Die Modellgüteprüfung kann bei Verwendung der Klassifikation als Unterscheidungskriterium für Messstationen auch als die Zuweisung von Aufgabengebieten bzw. typischen Skalen verstanden werden. Es ist z.B. absehbar, dass ein Modell von 30 km Auflösung die Skala der

verkehrsnahen Messung (Regime Straße oder # 5, siehe Tabelle 5) nicht wiedergeben kann, d.h. der Modellfehler wird für solche Stationen sehr groß sein. Trotzdem kann das Modell zur Ermittlung des städtischen Hintergrund und der ländlichen Regimes eingesetzt werden, wenn das festzulegende Gütekriterium für Stationen in diesem Regime eingehalten wird.

Die Klassifikation gibt also Aufschluss darüber, welche Modellskala benötigt wird, um eine Messreihe angemessen zu simulieren. Eine genaue Zuordnung ist jedoch im Einzelfall problematisch, da eine Messung des städtischen Regimes im Ruhrgebiet flächenrepräsentativ ist und von einer Modellrechnung mit einer Auflösung von wenigen Dekakilometern simuliert werden kann, während dies im eher ländlichen Mecklenburg-Vorpommern mit kleinen städtischen Inseln nicht der Fall ist.

Trotz diesen Einschränkungen wird vorgeschlagen, die von Flemming (2003b) entwickelte Klassifizierung von Messstationen als Grundlage für die Beurteilung der Güte einer bestimmten Modellanwendung heranzuziehen.

		Name und Abkürzung des Regimes						
Spezies	Kriterium:							
Ozon	Log (P50ME) P50TA	Berg B	Land R	Vorstadt U1	Stadt U2	Belastete Stadt U3	Straße S	
NO ₂	P50ME	Land R		Vorstadt U1	Stadt U2	Belastete Stadt U3	Straße S	Straße ex- trem S2
PM10	P50ME	#1		#2	#3	#4	#5	
SO ₂	Log (P50ME)	#1		#2	#3	#4	#5	

Tabelle 5 Immissionsregimes nach Flemming (2003). Clusterparameter sind: P50ME = Median des Tagesmittelwertes und P50TA = Median der normierten Tagesschwankung, Log = Logarithmische Transformation

Bei der Interpretation der Immissionsdaten werden verschiedene zeitliche Mittelungsintervalle angewendet, die von einer Stunde bis zu einem Jahr reichen. Entsprechend der zeitlichen Mittelung wäre es sinnvoll, auch mit räumlichen Mittelungen zu arbeiten. Solch ein Mittelungsgedanke entspricht zum einen dem Ansatz der Eulerschen Modelle, die immer mit einer endlichen Auflösung, d.h. Mittelung über eine Gitterzelle, arbeiten. Die räumliche Zusammenfassung entspricht auch dem Grundgedanken der Rahmenrichtlinie, die Einteilungen in „Gebiete“ und „Ballungsräume“ mit einer Mindestanzahl von Messungen kennt.

Durch eine räumliche Mittelung wird der Einfluss einzelner Stationen, d.h. der lokalen unrepräsentativen Störungen vermindert, und somit die „statistische“ Qualität, auch im Sinne der Datenverfügbarkeit, erhöht.

Es erscheint demzufolge untersuchenswert, die Modellgütebewertung auch mit fiktiven Stationen durchzuführen, deren Zeitreihe aus den stündlichen Mittelwerten besteht, die aus den Einzelreihen allen Stationen innerhalb einer Gitterzelle des Modells gebildet werden.

4 Anwendung der Modellgütemaße

In diesem Kapitel werden die in Kapitel 3.2 vorgestellten Gütemaße in der praktischen Anwendung getestet. Für diesen Zweck wurde das RCG-Modell mit einer horizontalen Auflösung von $0.5^\circ \times 0.25^\circ$ (ca. 25 bis 30 km) auf das ganze Jahr 2002 angewandt. Das Untersuchungsgebiet umfasst ganz Zentraleuropa. Beobachtung und Modellrechnung werden dann stationsweise unter Anwendung der Gütemaße verglichen. Die Fehlermaße werden hinsichtlich der Zielgrößen der EU-Rahmenrichtlinie, d.h. Jahresmittelwerte und Zeitreihen mit stündlichen bzw. tagesbezogenen Konzentrationswerten für Ozon, NO₂, SO₂ und PM₁₀, gebildet. Die Gütemaße werden für alle vorliegenden Stationen, die der Mindestanforderung der Datenerfassung genügen, ermittelt und in ihrer Gesamtheit dargestellt. Nach der Diskussion eines Beispiels an einer Station und der Fehlermaße hinsichtlich der Jahresmittelwerte wird der Schwerpunkt auf die Tages- und Stundenwerte gelegt. Dabei wird die Aufgabe des Zeitbezuges erörtert und die Eignung verschiedener relativer Gütemaße zur Formulierung von Mindestanforderung der Modellgüte gezeigt. Abschließend erfolgt ein Vergleich der Modellgüte bei den Jahresmittelwerten und der Zeitreihengrößen.

Die für die Güteprüfung verwendeten Messungen stammen aus den Messnetzen der Bundesländer und des Umweltbundesamts und wurden vom Umweltbundesamt zur Verfügung gestellt. Diese Datenreihen sind oft nicht vollkommen vollständig (siehe Abbildung 45). Bei fehlenden Messungen wurden die entsprechenden Modellwerte aus der berechneten Zeitreihe entfernt. Durch die damit gleiche Anzahl der Modell- und Messwerte sind Messungen und Berechnungen sowohl im Zeitbezug als auch in der Perzentil-Darstellung immer eindeutig vergleichbar. Bei der Modellgütebestimmung im Perzentil-Bezug muss jedoch bei dieser Vorgehensweise bedacht werden, dass gerade die Berücksichtigung der Modelldaten aus dem Zeitraum, für den keine Messungen zur Verfügung stehen, zu einer besseren Übereinstimmung der Häufigkeitsverteilung im Bereich der hohen Werte führen könnte. Dieser Umstand wird in Kap. 5.1.4 näher untersucht.

Der Vergleich von Modell und Beobachtung erfolgt immer stationsbezogen, d.h. es wird immer die Beobachtungszeitreihe mit der zugehörigen Modellreihe verglichen. Die Modellreihe wird aus den Werten der Gitterbox, in der sich die Station befindet, gewonnen. Dargestellt werden dann die Gütemaße an den Stationen. Aufgrund der Tatsache, dass in vielen Fällen mehrere Stationen in einer Gitterbox liegen, wird auch die Modellgüte dieser Gitterbox mehrmals bewertet. Die Stationsdichte hat demzufolge einen Einfluss auf die zusammengefasste Modellgütebewertung, da eine hohe Stationsdichte in einer Gitterbox mit schlechterem Simulationsergebnis dieses entsprechend häufiger bewertet als wenn dort nur eine oder sogar gar keine Station liegen würde. Dieses Problem könnte vermieden werden, wenn, wie in Flemming et al. (1999) für die Ozonprognose durchgeführt, zunächst eine Immissionsfeldanalyse, d.h. eine gitterpunktbezogene Felddarstellung der Beobachtung, erzeugt würde, die dann gitterpunktbezogen mit dem Modellfeld verglichen wird. Da aber durch die räumliche Glättung der Beobachtung bei der Analyse die Feldwerte nicht mehr hundertprozentig mit den tatsächlichen Beobachtungswerten übereinstimmen, wird hier auf dieses Vorgehen verzichtet. Es ist anzunehmen, dass eine Veränderung der Messung als Maß für die „wahre“ Immissionssituation nicht mit den Grundsätzen der Rahmenrichtlinie vereinbar ist.

4.1 Vergleich von Messung und Rechnung am Beispiel einer ausgewählten Station

In diesem Kapitel wird die gemessene Zeitreihe der PM₁₀ Tagesmittelwerte an der Station Burg in Brandenburg (BB001) im Jahr 2002 mit der entsprechenden Zeitreihe aus einer Modellsimulation mit RCG in der horizontalen Auflösung von $0.5^\circ * 0.25^\circ$ (ca. 30 km) verglichen, um die Bedeutung der in Kapitel 3.2 vorgestellten Gütemaße zu demonstrieren. Zu diesem Zweck werden die folgenden Größen in Abbildungen dargestellt:

1. Die Zeitreihen von Modellrechnung und Beobachtung mit Zeit- und Perzentil-Bezug,
2. die absoluten Fehler und die relativen Fehler mit Zeit- und Perzentil-Bezug ,
3. die Perzentilabweichung bei gleichem Konzentrationswert ,
4. die Häufigkeitsverteilung der Werte von Modellrechnung und Beobachtung,
5. die Häufigkeitsverteilung der Fehler mit Zeit- und Perzentil-Bezug ,
6. die Häufigkeitsverteilung der relativen Fehler mit Zeit- und Perzentil-Bezug .

Abbildung 1 zeigt die Zeitreihe der Beobachtung und der Modellierung für die Tage des Jahres 2002. Dabei zeigt sich zunächst, dass das Modell die Beobachtung leicht unterschätzt, den grundsätzlichen Verlauf der Schwankungen, z.B. die erhöhten Werte nach Tag 219, aber gut wiedergibt. Die offensichtlichste Abweichung tritt im Dezember 2002 (Tag 347) auf, wo ein extrem hohes und in dieser Höhe vorher nicht aufgetretenes Maximum vom Modell nicht erfasst wird. Dies wird besonders deutlich aus Abbildung 2, die die Differenz zwischen den Zeitreihen zeigt. Während in der überwiegenden Zeit des Jahres der tägliche Fehler zwischen -25 und $10 \mu\text{g}/\text{m}^3$ schwankt, ergibt sich bei dem hohen gemessenen Wert am Tag 347 ein Modellfehler von ca. $-80 \mu\text{g}/\text{m}^3$. Normiert man den Fehler mit dem entsprechenden Messwert, so erhält man die Zeitreihe des relativen Fehlers, die in Abbildung 3 dargestellt ist. Die meisten täglichen relativen Fehler liegen unterhalb der in der Tochterrichtlinie 1999/30/EC vorgegebenen Mindestgenauigkeitsanforderung von 50%. Der relative Fehler des größten absoluten Fehlers liegt bei ca. 75%, d.h. größter relativer und größter absoluter Fehler fallen zeitlich nicht zusammen. Eine beliebig große Unterschätzung des Modells kann bei der Normierung mit dem Messwert keine relativen Fehler von mehr als 100% zu Folge haben. Die maximalen relativen Fehler von bis zu 170% am Beginn des Jahres sind demzufolge immer mit Überschätzungen des Modells bei kleinen Messwerten verbunden. Die zugehörigen absoluten Fehler bei den größten relativen Fehlern reichen bis zu $15 \mu\text{g}/\text{m}^3$.

Gibt man den Zeit-Bezug auf und ordnet die Daten der Größe nach, so erhält man die Darstellung der Zeitreihe mit Perzentil-Bezug in Abbildung 4. Sie zeigt an, dass bei einem bestimmten Perzentil-Wert x $(100-x) \%$ der Tage eine Konzentration haben, die größer gleich dem auf der y-Achse abzulesenden Konzentrationswert ist. Multipliziert man den Perzentilwert mit 3.65, ergibt sich die Anzahl der Tage, an denen die Tagesmittelwerte kleiner gleich dem dazugehörigen Konzentrationswert sind. Mit vertauschten Achsen entspricht Abbildung 4 also genau den akkumulierten Häufigkeitsverteilungen der Mess- und Modellwerte. Abbildung 5 zeigt die Differenz der beiden akkumulierten Häufigkeiten hinsichtlich der Konzentrationswerte und stellt somit die Fehler mit Perzentil-Bezug dar. Der Fehler mit Perzentilbezug ist immer negativ, d.h. der Modellbias ist negativ. Der überwiegende Teil des Fehlers (95%) mit Perzentil-Bezug liegt zwischen 0 und $-15 \mu\text{g}/\text{m}^3$. Der größte absolute Wert

tritt beim Perzentilwert 100 auf und stellt die Differenz der Maxima aus der Beobachtung (Tag 347 mit $118 \mu\text{g}/\text{m}^3$) und der Modellierung (Tag 76 mit $48 \mu\text{g}/\text{m}^3$) dar.

Normiert man den Fehler mit Perzentil-Bezug mit der entsprechenden Beobachtung, so ergeben sich relative Fehler, die mit einer Ausnahme deutlich unter 50% liegen (Abbildung 6). Höhere relative Fehler treten bei den sehr kleinen Perzentilen auf, sehr geringe relative Fehler ergeben sich bis zum 20-Perzentil. In Richtung höherer Perzentile wächst der relative Fehler dann mehr oder weniger kontinuierlich an, um mit einem sprunghaften Anwachsen bei den sehr hohen Perzentilen zu enden.

Abbildung 7 zeigt die alternative Betrachtung zur Darstellung der akkumulierten Häufigkeitsverteilungen, bei der der Unterschied im Perzentil bei einem festgelegten Konzentrationswert untersucht wird. Die Perzentil-Differenz (y-Achse) kann durch die Multiplikation mit 3,65 in die Anzahl der Tage umgerechnet werden, an denen eine gewisse Konzentrationschwelle in der Beobachtung (positive Perzentildifferenz) häufiger überschritten wird als in der Berechnung. Ein PM10-Tagesmittelwert von $50 \mu\text{g}/\text{m}^3$ wird demnach an der ausgewählten Station an 11 Tagen mehr überschritten als in der Rechnung. Für die Darstellung wurden die Konzentrationswerte in Klassen von $1 \mu\text{g}/\text{m}^3$ eingeteilt und der jeweils kleinste Perzentilwert wurde der entsprechenden Klasse zugeordnet. Um die Perzentildifferenz auch bei den Konzentrationswerten bestimmen zu können, die von der Modellrechnung nicht erreicht werden, wurde in diesen Fällen der Modellrechnung das Perzentil $(100 + 100/365)$ zugewiesen. Dadurch wird erreicht, dass die Überschreitung dieser Werte im Diagramm sichtbar ist.

Begleitend zu den diskutierten Abbildungen wird in Abbildung 8 die einfache Häufigkeitsverteilung von Beobachtung und Modell gezeigt. Dabei wird deutlich, dass Modellierung und Beobachtung hauptsächlich den Konzentrationsbereich bis $40 \mu\text{g}/\text{m}^3$ überdecken, in der Beobachtung jedoch vereinzelt deutlich höhere Werte auftreten können.

Abbildung 9 stellt die Häufigkeitsverteilungen der Fehler mit und ohne Zeit-Bezug dar. Während die Mehrheit der Fehler mit Zeitbezug zwischen -30 und $10 \mu\text{g}/\text{m}^3$ streuen, wird durch den Perzentil-Bezug die Bandbreite deutlich von -20 bis $0 \mu\text{g}/\text{m}^3$ verringert. Dieser Umstand ist auch bei der Bandbreite der relativen Fehler zu erkennen, die in Abbildung 10 dargestellt sind. Insgesamt zeigt die Auswertung der Gütemaße, dass die Modellfehler mit Perzentil-Bezug deutlich kleiner sind als diejenigen mit Zeitbezug.

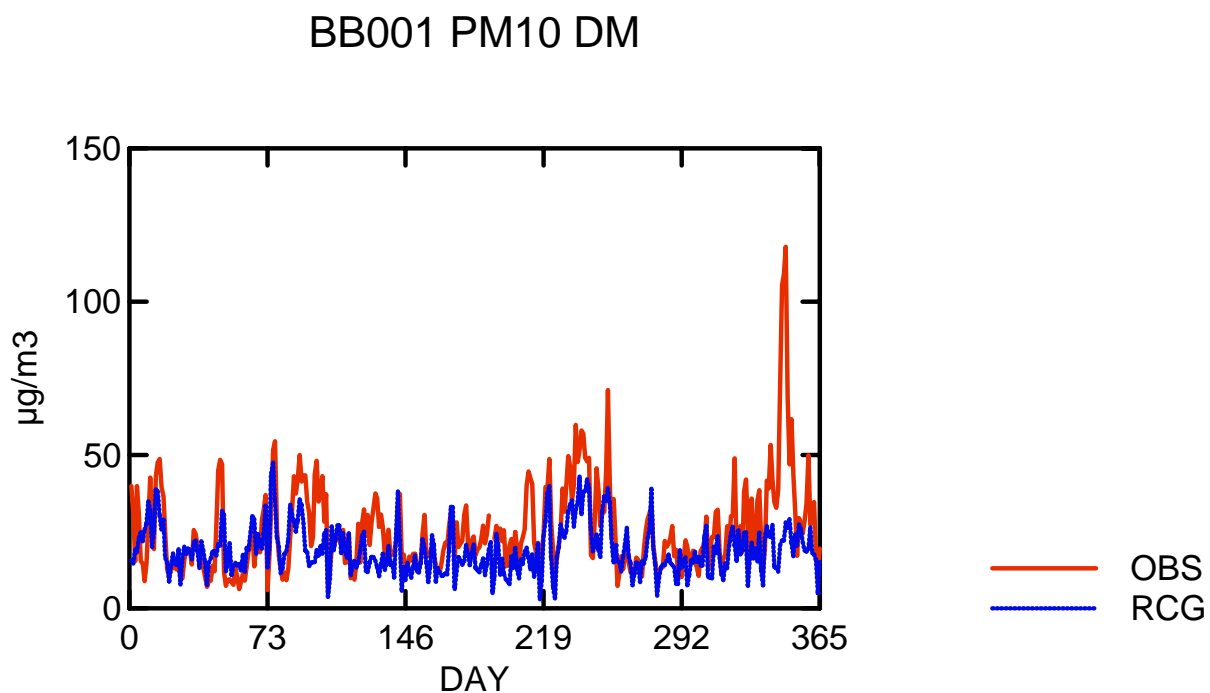


Abbildung 1 Zeitreihen des PM10 Tagesmittelwertes an der Station BB001 aus der Beobachtung OBS und der Modellrechnung RCG

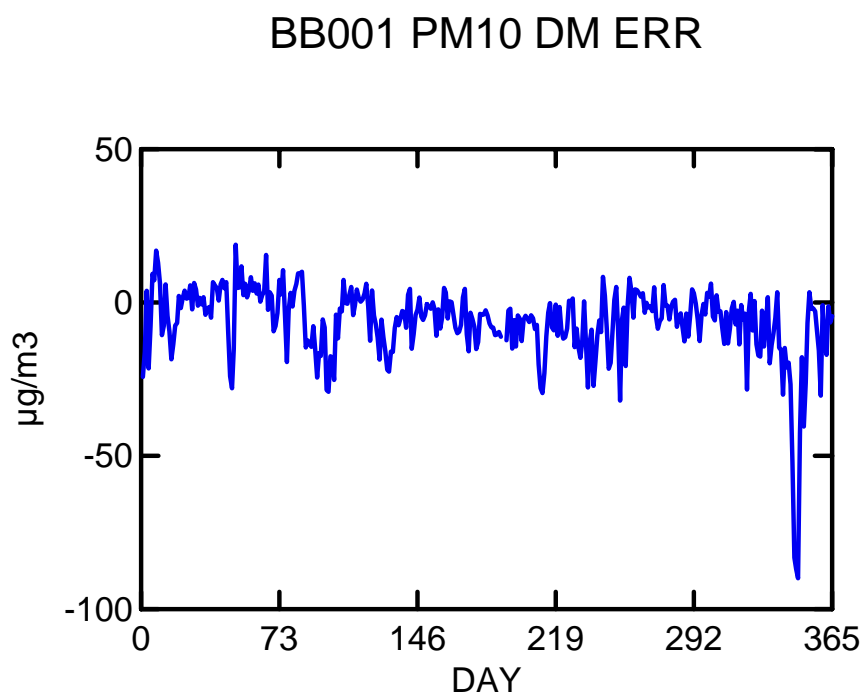


Abbildung 2 Zeitreihen der Differenz zwischen Beobachtung und Modellrechnung an der Station BB001, PM10 Tagesmittelwerte.

BB001 PM10 DM REL. ERR

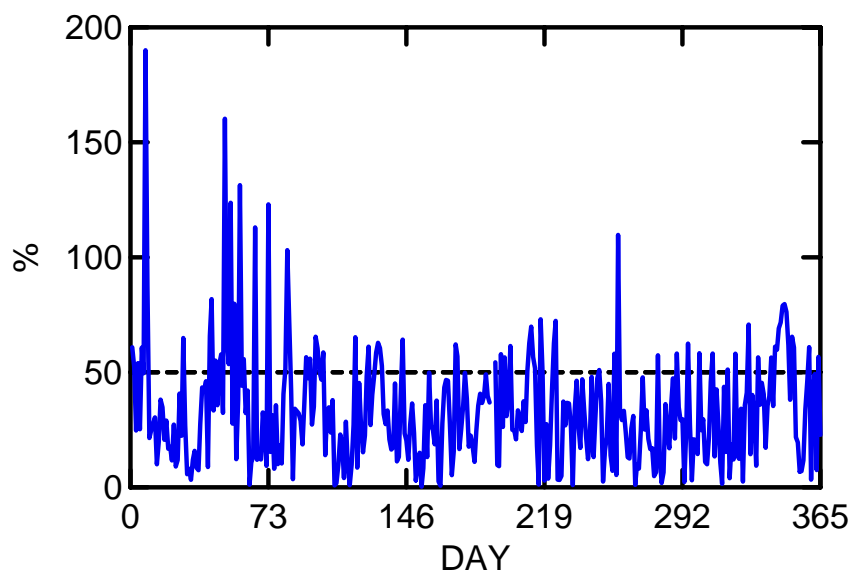


Abbildung 3 Zeitreihen der relativen Differenz zwischen Beobachtung und Modellrechnung an der Station BB001, PM10 Tagesmittelwerte.

BB001 PM10 DM

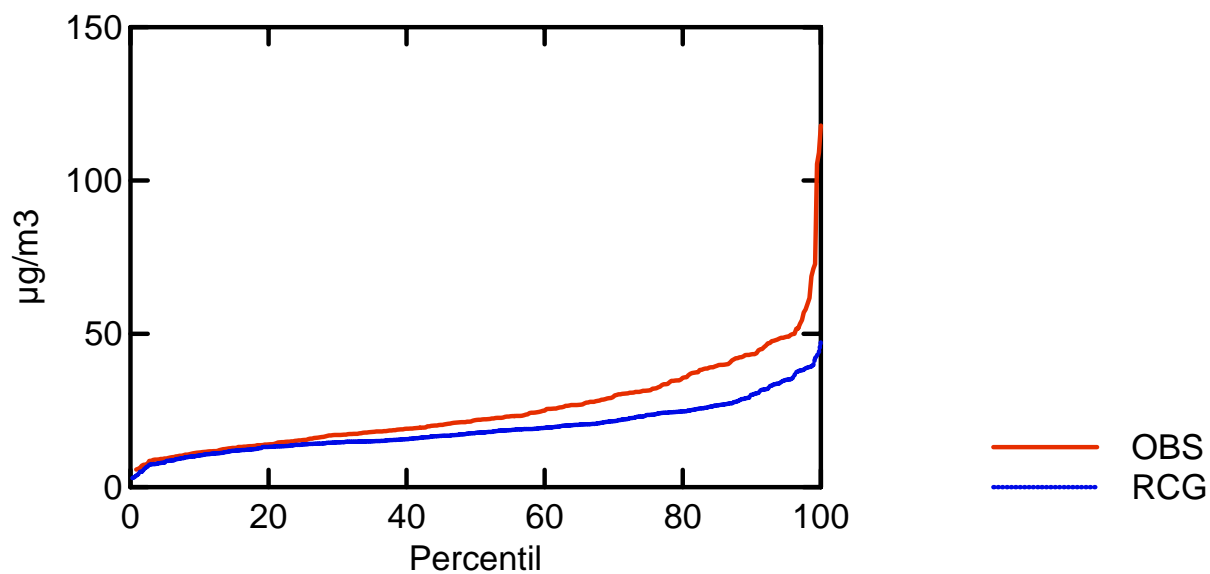


Abbildung 4 Akkumulierte Häufigkeitsverteilung der PM10 Tagesmittelwerte an der Station BB001 für die Modellrechnung RCG, die Beobachtung OBS und eine zusätzliche Schwankungsbreite von +/- 50%.

BB001 PM10 DM ERR

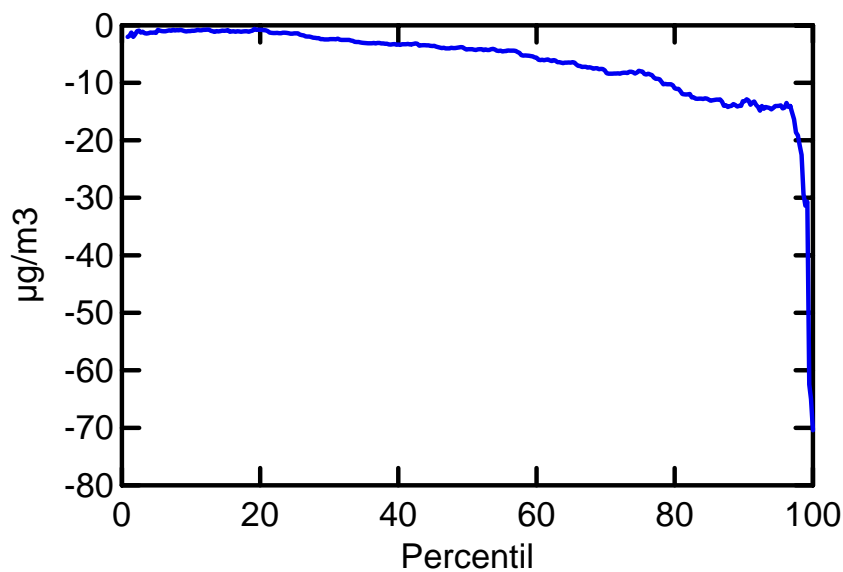


Abbildung 5 Absolute Fehler mit Perzentil-Bezug der PM10 Tagesmittelwerte an der Station BB001

BB001 PM10 DM REL. ERR

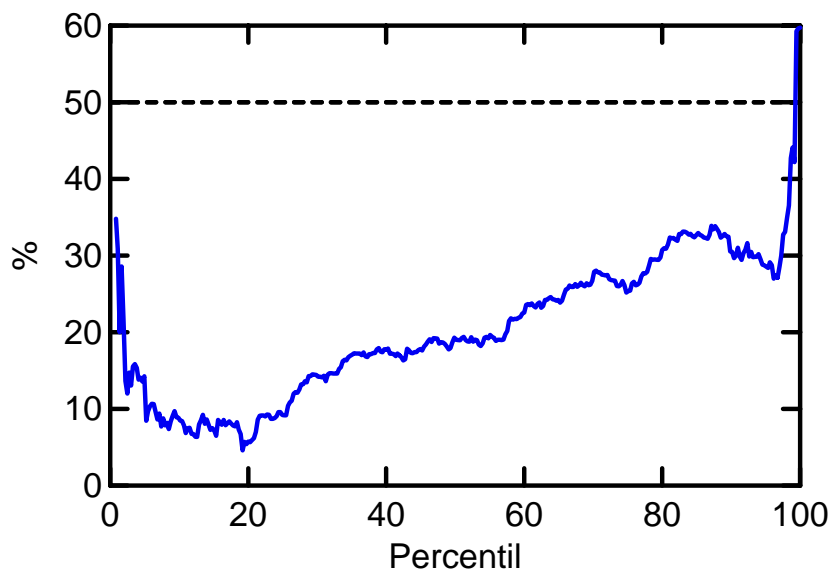


Abbildung 6 Relativer Fehler mit Perzentil-Bezug der PM10 Tagesmittelwerte an der Station BB001

BB PM10 DM Per RCG - Per OBS

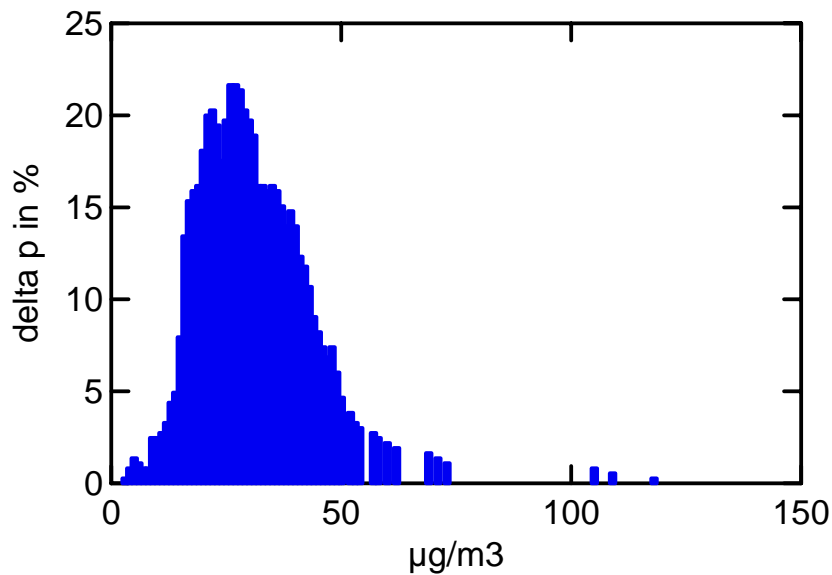


Abbildung 7 Differenz zwischen dem Perzentilwert der Modellierung und dem der Beobachtung bei einem festen Konzentrationswert. PM10-Tagesmittelwerte an der Station BB001.

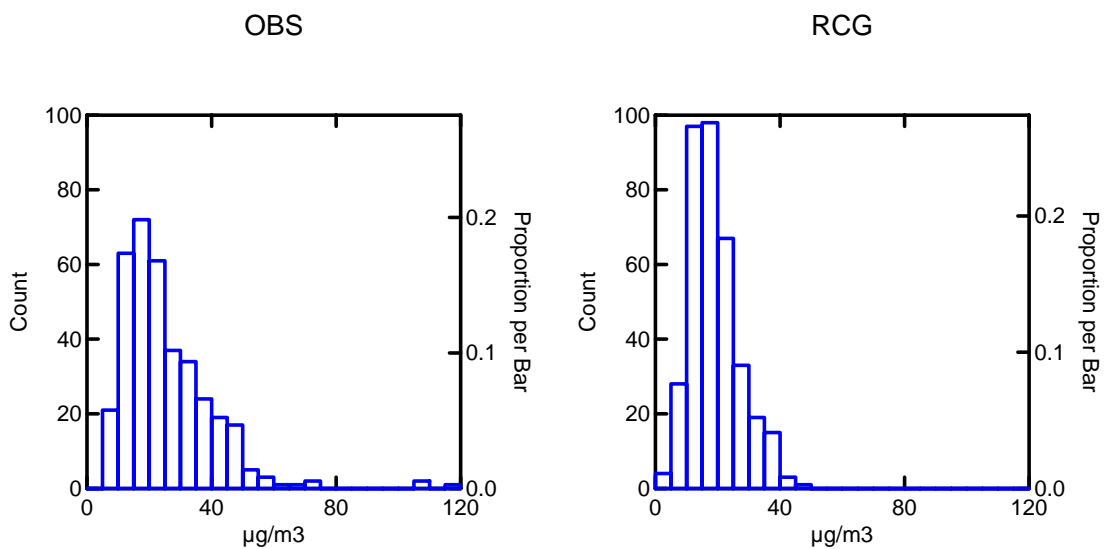


Abbildung 8 Häufigkeitsverteilung der PM10 Tagesmittelwerte an der Station BB001 aus der Beobachtung (OBS) und der Modellrechnung (RCG) .

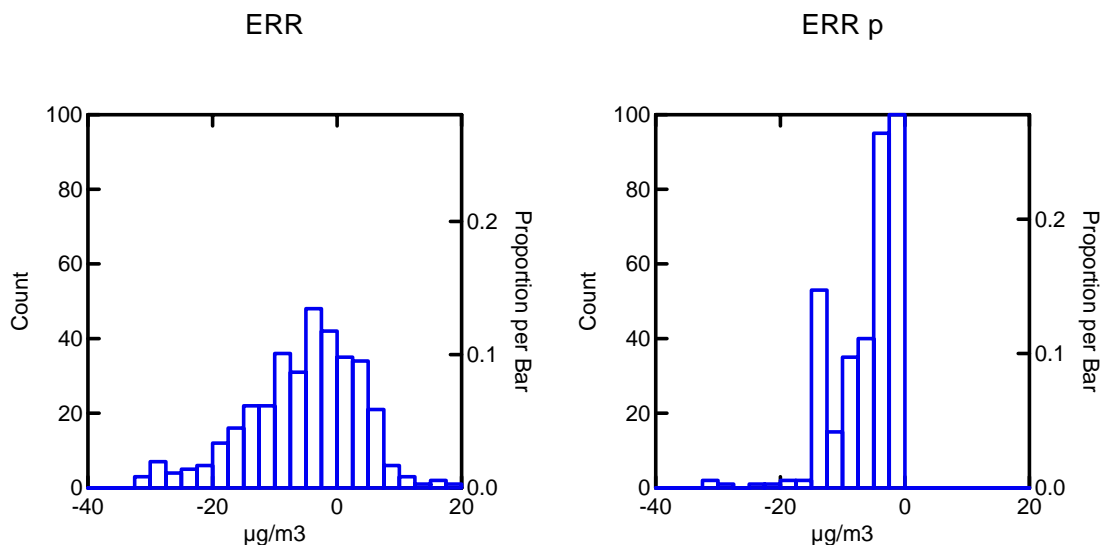


Abbildung 9 Häufigkeitsverteilung der Modellfehler (RCG-OBS) mit Zeitbezug (ERR) und ohne Zeitbezug (Differenz der entsprechenden Perzentile, ERR_p).

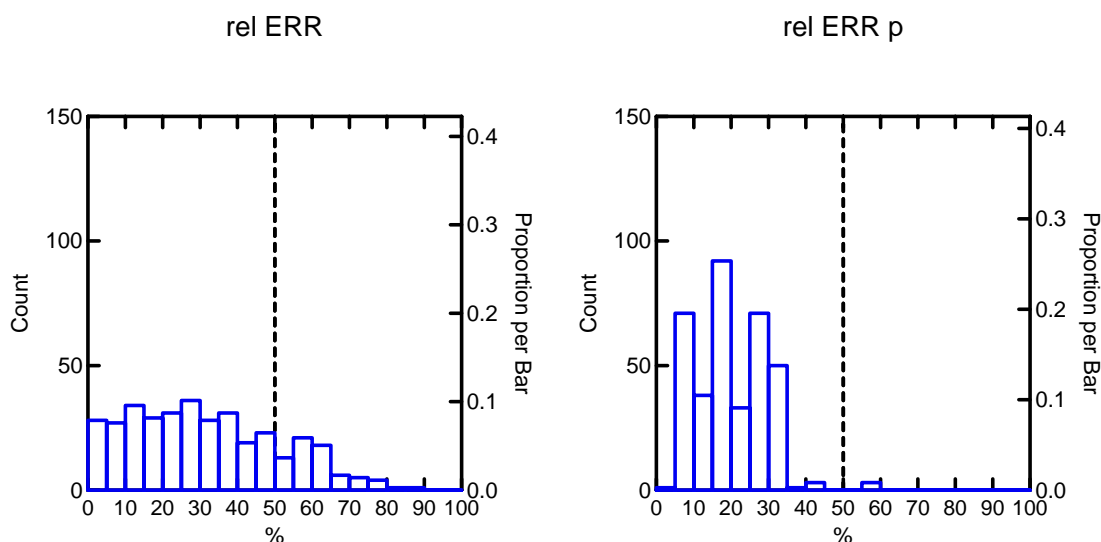


Abbildung 10 Häufigkeitsverteilung der relativen Modellfehler ($(\text{RCG-OBS})/\text{OBS}$) mit Zeitbezug (rel ERR) und ohne Zeitbezug (Differenz der entsprechenden Perzentile, rel ERR_p). Die EU-Richtlinien erlauben einen relativen Maximalfehler von 50%.

4.2 Jahresmittelwerte

Hinsichtlich der Jahresmittelwerte ist die Modellgüteanforderung der EU-Richtlinien eindeutig. Die Trennung von Zeit- und Perzentil-Bezug ist nicht notwendig und auch die Reihenfolge von Maxima-Suche und Relativwertbildung spielt keine Rolle. Grenzwerte der Jahresmittelwerte werden für NO_2 , SO_2 und PM_{10} in 1999/30/EC eingeführt. Im Sinne der Modellevaluierung im Perzentil-Bezug kann die Überprüfung der Jahresmittelwerte als der Ver-

gleich in der Nähe des 50-Perzentil gewertet werden. Abbildung 11 bis Abbildung 13 stellen im Streudiagramm den Zusammenhang zwischen den Jahresmittelwerten der Beobachtungen (x-Achse) und denen der Modellrechnung, deren Abweichung von der Beobachtung sowie den relativen Fehler (y-Achse) für PM₁₀, NO₂ und SO₂ dar. In den Streudiagrammen sind an Achsen mit Konzentrationswerten die Werte der unteren Beurteilungsschwelle und des Grenzwertes markiert. Für die relativen Fehler sind die Grenzen für die Mindestgenauigkeit von 30% bzw. 50% (PM₁₀) angegeben. In allen Streudiagrammen wird das Immissionsregime der Beobachtung mit unterschiedlich eingefärbten Punkten wiedergegeben. Abbildung 15 zeigt für die einzelnen Immissionsregimes nach Flemming (2003b) die Anzahl der Stationen, die die Genauigkeitsanforderung erfüllen bzw. nicht erfüllen.

Die Jahresmittelwerte von PM₁₀ (Abbildung 11) liegen bei fast allen Stationen übereinstimmend für Messung und Rechnung im Bereich zwischen unterer Beurteilungsschwelle und Grenzwert. Die Streuung ist jedoch groß und die hohen Werte mit Grenzwertüberschreitung stimmen in Beobachtung und Modellrechnung nicht überein. Die direkten und relativen Fehler wachsen mit zunehmender mittlerer Belastung der Beobachtung. Der relative Fehler wird jedoch nur in wenigen Fällen größer als die geforderte Genauigkeit von 50%. Relative Fehler größer als 50% treten jedoch in den Immissionsregimes mit geringster (#1) und hoher Belastung (#4/5) auf.

Die gemessenen Jahresmittelwerte von NO₂ (Abbildung 12) überdecken einen großen Bereich, der unter der unteren Beurteilungsschwelle beginnt und deutlich über den Grenzwert reicht. Die Modellrechnung weist geringere Werte auf, obwohl auch hier Grenzwertüberschreitungen auftreten, die jedoch dann die zugehörige Messung überschätzen. Der direkte Fehler nimmt mit wachsender Immission der Beobachtung zu, beim relativen Fehler kann jedoch kein eindeutiger Zusammenhang zur Höhe des Messwertes festgestellt werden. Bei hohen und geringen Jahresmittelwerten der Beobachtung liegt der relative Fehler über der Marke von 30 %. Bei den Land-Stationen werden knapp die Hälfte, bei den Straßen-Stationen ca. 10% mit hinreichender Genauigkeit vorhergesagt. Die hohen relativen Fehler ergeben sich zum einen durch die niedrigen Beobachtungswerte (meistens Immissionsregime R) und zum anderen durch die extrem hohen Belastungen (meistens Immissionsregime S/S2), die von der Modellskala nicht erfasst werden können.

Während bei NO₂ und PM₁₀ das Modell die Gesamtheit der vorliegenden Stationen eher unterschätzt, treten bei SO₂ häufiger negative Modellfehler auf, also eine Überschätzung der Messwerte. Das Modell liefert mehr Jahresmittelwerte über der unteren Beurteilungsschwelle, vereinzelt wird sogar der Grenzwert erreicht. Die Beobachtung ist jedoch fast ausschließlich unter der unteren Beurteilungsschwelle angesiedelt. Aufgrund der niedrigen Messwerte ist die erreichte Genauigkeit bei ca. 2/3 der Stationen geringer als die der Mindestanforderungen. Dies gilt für alle vertretenen Regimeklassen. Grund der Überschätzung der Modellrechnung dürfte die Konzentration der emittierten SO₂-Masse auf relativ wenige Punktquellen sein, die in der hier verwendeten großräumigen Modellauflösung nicht angemessen simuliert werden können.

Die Fehler der Ozon-Jahresmittelwerte, d.h. der Modellbias, sind angenähert proportional zu den Werten der Beobachtung. Die niedrigen Immissionswerte der Verkehrsregime werden überschätzt und die hohen der Landstationen werden leicht unterschätzt (Abbildung 14). Die größte Unterschätzung wird bei den Bergstationen festgestellt, die aber aufgrund ihrer Repräsentativität für die obere Grenzschicht besser mit den Modellwerten der 2. Modellschicht verglichen werden sollten. Aus Gründen der Vereinfachung wurde hierauf verzichtet, denn

die Rahmenrichtlinie EC 2002/2/EC benennt keine Grenzwerte für die Jahresmittelwerte von Ozon.

Beim NO_2 und SO_2 sinkt der relative Anteil der Stationen in den einzelnen Immissionsklassen, an denen das Gütekriterium der Richtlinie vom Modell eingehalten wird, mit zunehmender Klassennummerierung, d.h. mit zunehmender Belastung (Abbildung 15). Dies ist ein Hinweis, dass eine Modellauflösung von 25 bis 30 km nicht geeignet ist, um die Mehrzahl der Messungen in den höher belasteten Regimes nachzuvollziehen. Beim PM_{10} ist dieser Zusammenhang nicht so eindeutig, da die Konzentrationsunterschiede zwischen den anthropogenen Emissionsgebieten und den eher emissionsarmen Gebieten bei diesem Stoff deutlich geringer sind als bei NO_2 und SO_2 . Ursache dafür ist der hohe Anteil sekundärer Aerosole am PM_{10} . Diese Aerosole werden während des großräumigen Transports gebildet. Wie bei den anderen Stoffen wird auch beim PM_{10} deutlich, dass insbesondere Stationen der hoch belasteten Klasse #5 nicht zur Bestimmung der Güte einer Modellanwendung mit einer Auflösung im Dekakilometerbereich verwendet werden sollten.

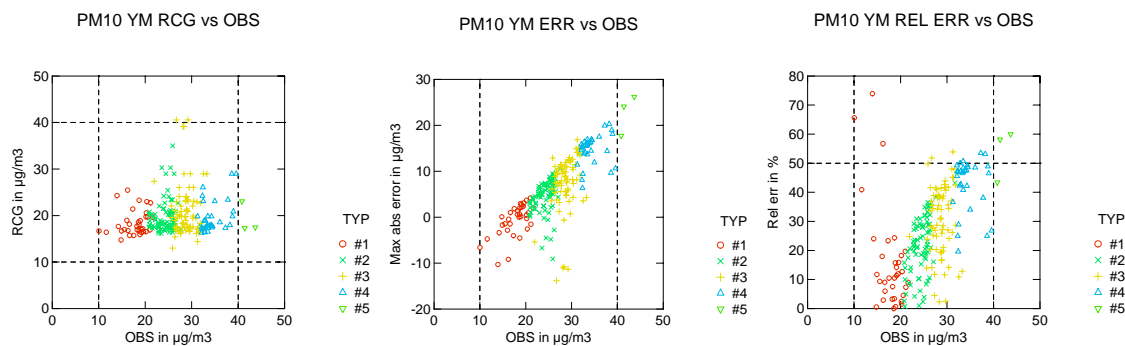


Abbildung 11 Stationsbezogene Jahresmittelwerte von PM_{10} aus Beobachtung und Modellrechnung mit RCG. Links: Beobachtung vs. Modellrechnung, Mitte: Beobachtung vs. absoluten Fehler, Rechts: Beobachtung vs. relativer Fehler. Die gestrichelten Linien bei Konzentrationsangaben entsprechen der unteren Beurteilungsschwelle bzw. dem Grenzwert. Die 50% Line (rechts) markiert die Genauigkeitsanforderung nach 1999/30/EC.

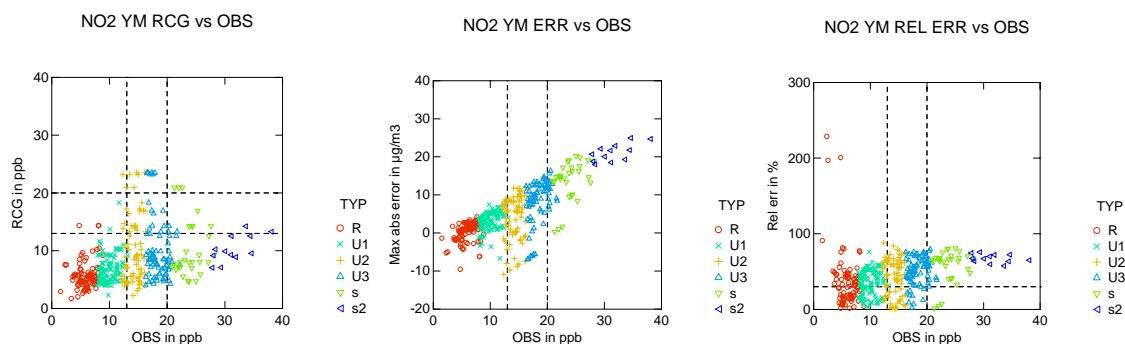


Abbildung 12 Stationsbezogene Jahresmittelwerte von NO_2 aus Beobachtung und Modellrechnung mit RCG. Links: Beobachtung vs. Modellrechnung, Mitte: Beobachtung vs. Fehler, Rechts: Beobachtung vs. relativer Fehler. Die gestrichelten Linien bei Konzentrationsangaben entsprechen der unteren Beurteilungsschwelle bzw. dem Grenzwert. Die 30% Line (rechts) markiert die Genauigkeitsanforderung nach 1999/30/EC.

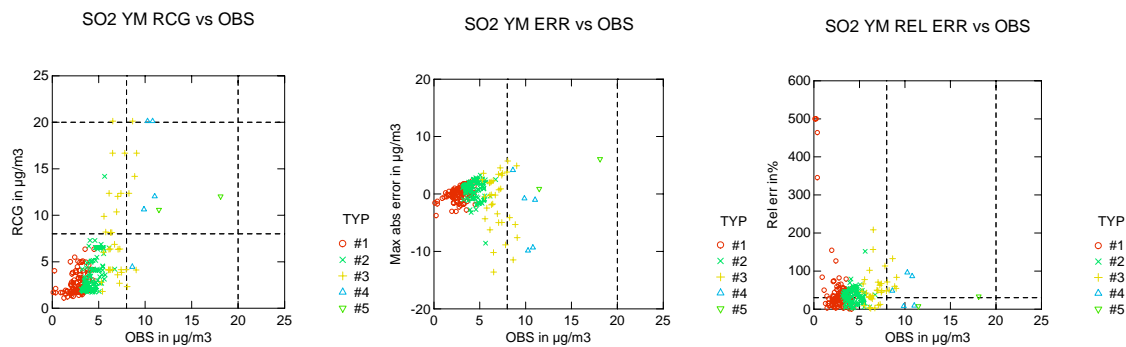


Abbildung 13 Stationsbezogene Jahresmittelwerte von SO₂ aus Beobachtung und Modellrechnung mit RCG. Links: Beobachtung vs. Modellrechnung, Mitte: Beobachtung vs. Fehler, Rechts: Beobachtung vs. relativer Fehler. Die gestrichelten Linien bei Konzentrationsangaben entsprechen der unteren Beurteilungsschwelle bzw. dem Grenzwert. Die 30% Line (rechts) markiert die Genauigkeitsanforderung nach 1999/30/EC.

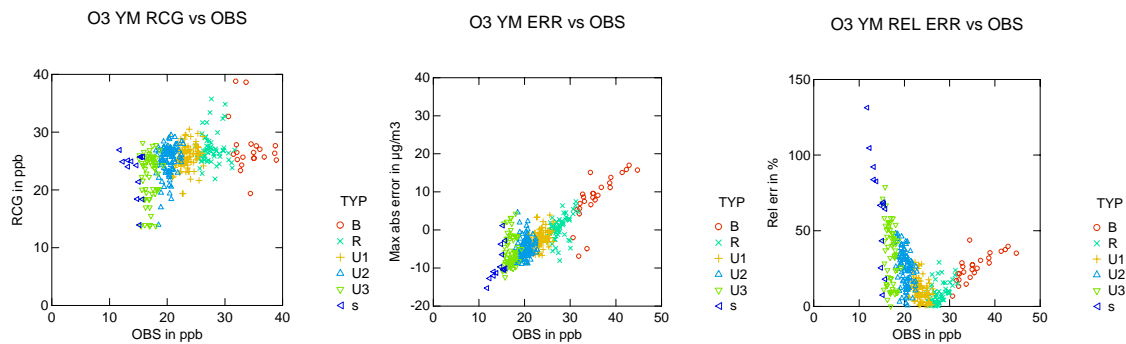


Abbildung 14 Stationsbezogene Jahresmittelwerte von Ozon aus Beobachtung und Modellrechnung mit RCG. Links: Beobachtung vs. Modellrechnung, Mitte: Beobachtung vs. absolutem Fehler, Rechts: Beobachtung vs. relativer Fehler.

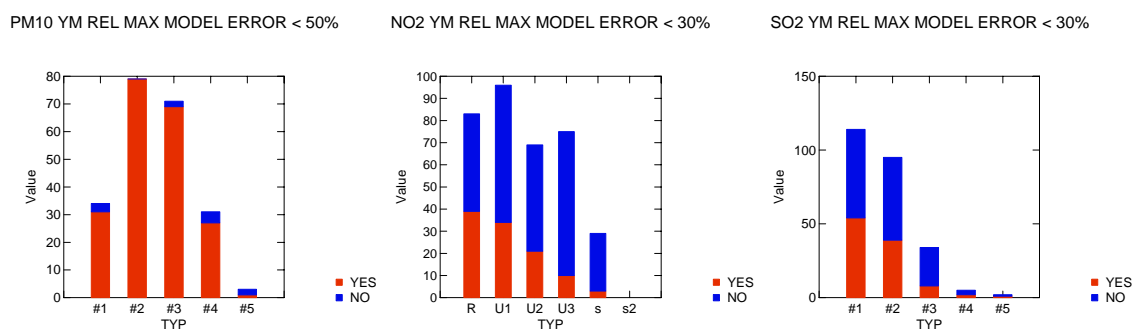


Abbildung 15 Anzahl der Stationen, die die Genauigkeitsanforderung hinsichtlich der Modellierung der Jahresmittelwerte nach 1999/30/EC erfüllen (YES), geordnet nach Regimes. Links: PM₁₀, Mitte: NO₂, rechts: SO₂ (Regime R = Land, U1 = Vorstadt, U2 = Stadt, U3 = belastetet Stadt, S = Verkehr, S2 = Verkehr extrem bzw. Belastungsstufen #1 - #5 für PM₁₀ und SO₂)

4.3 Tages und Stundenwerte

4.3.1 Überblick

Bei der Modellevaluierung hinsichtlich der Stunden- und Tageswerte kommen die Unklarheiten bei der Formulierung der Mindestgüteanforderung der Rahmenrichtlinie zur Geltung. Aus diesem Grund werden in diesem Kapitel die in Kapitel 3.2 definierten Gütemaße vergleichend diskutiert.

Jede berechnete Ozon-, PM10- und NO₂-Zeitreihe wird untersucht, wie sie der Definition der Grenzwerte zum Gesundheitsschutz (siehe Tabelle 1) entspricht, d.h.

- Tägliches Maximum des 8h-Mittelwertes für Ozon
- Stundenwerte für NO₂
- Tagesmittelwerte für PM10
- Stundenwerte für SO₂

Auf die Darstellung der SO₂-Tagesmittelwerte wird aus Gründen der Vereinfachung verzichtet, die Schlussfolgerungen aus der Betrachtung der Stundenwerte sind jedoch genauso für die Tagesmittelwerte gültig.

Für jeden der Stoffe und jede Station in Deutschland werden die Gütemaße ermittelt und in ihrer Gesamtheit dargestellt. Dabei wird zunächst die Modellgüte mit und ohne Zeitbezug mit Hilfe des RMSE untersucht. Ein besonderer Schwerpunkt liegt dann auf dem Vergleich der drei Gütemaße, die in Kapitel 3.2 als Interpretation der Mindestanforderung der Rahmenrichtlinie vorgeschlagen wurden. Es handelt sich dabei um:

1. Vergleich der Konzentrationswerte von Modellrechnung und Beobachtung beim größten absoluten Fehler $\max err_p$, normiert mit der Beobachtung (Fehlermaß nach den Tochterrichtlinien, $rel \max err_p$)
2. Vergleich der Konzentrationswerte von Modellrechnung und Beobachtung bei dem Perzentil der zulässigen Überschreitungshäufigkeit des Grenzwerts GW: $GW_{per\ err_p} = p_{GW}$
3. Vergleich der Perzentile, die dem Grenzwert GW oder den Beurteilungsschwellen UB und OB entsprechen $\Delta p, o = r = GW, UB, OB$

Der Vergleich von 1. und 2. steht im Mittelpunkt der folgenden Unterkapitel. Dabei wird besonders untersucht, bei welchem Konzentrationsniveau die Übereinstimmung von Modell und Beobachtung bewertet wurde. Der dritte Ansatz hat sich als wenig praxistauglich erwiesen, da die Konzentrationswerte häufig das Niveau der Beurteilungsschwellen gar nicht überschreiten.

Die Mindestanforderung für die Modellgüte wird in den EU-Richtlinien als ein relatives Maß formuliert. Aus diesem Grund werden die normierten Formen von 1 und 2, d.h. $rel \max err_p$ und $rel per err_p$ zusammen mit weiteren relativen Maßen mit und ohne Zeitbezug diskutiert. Der Vergleich soll verdeutlichen, wie die groß die Bandbreite der relativen Modellgüte aufgrund unterschiedlicher Gütemaße werden kann.

4.3.2 Auswirkungen der Aufgabe des Zeitbezuges

Zur Darstellung der Unterschiede der Abweichungen zwischen Modell und Beobachtung beim Übergang vom Zeit- zum Perzentil-Bezug wurde der RMSE gewählt, da er die Gesamtheit der Abweichungen zu allen Zeitpunkten bzw. an allen möglichen Perzentilwerten⁷ zusammenfasst.

Abbildung 16 zeigt die Histogramme aller stationsbezogenen RMSE-Werte mit und ohne Zeit-Bezug für O₃, NO₂, SO₂ und PM10 in Form eines Box-Whiskers-Plots⁸. In allen Fällen ist eine deutliche Verringerung des RMSE durch die Aufgabe des Zeitbezuges zu erkennen, die sich durch eine konstante Verschiebung des Histogramms zu den niedrigeren Werten bemerkbar macht. Der Median der Abweichungen verringert sich dabei bei O₃, NO₂ und SO₂ etwa auf die Hälfte und bei PM10 auf 2/3. Die größte Verringerung ist bei den Tagesmaxima der 8h-Mittelwerte von Ozon zu erkennen. Der RMSE wird durch den Bias und durch die Kovarianz der Anomalien⁹ gekennzeichnet. Durch die Aufgabe des zeitlichen Bezuges wird jedoch nur der Beitrag der Kovarianz der Anomalien zum RMSE zum Teil sehr deutlich verringert, während der Bias erhalten bleibt. Die stärkere Abnahme bei Ozon ist dadurch zu erklären, dass die Modellsimulation bereits biasfrei ist und demzufolge der zeitbezogene RMSE hauptsächlich durch die zeitliche Nichtübereinstimmung bzw. die unterschiedliche Ausprägung der Anomalien gekennzeichnet ist. Die relativ geringe Änderung beim PM10 lässt sich mit dem bestehenden Bias und der dazu offensichtlich relativ kleinen bzw. richtig prognostizierten Variabilität erklären. Genauere Untersuchungen mit Hilfe der zeitlichen Korrelationskoeffizienten könnten klären, ob die bereits gute zeitliche Übereinstimmung oder die geringe Variabilität dabei den größeren Einfluss haben.

⁷ Die Anzahl der Datenpaare ist beim Perzentil-Bezug die gleiche wie beim Zeit-Bezug. Bei den Stundenwerten beträgt der Schrittweite der Perzentilwerte 8760/100 und bei den Tageswerten 365/100.

⁸ Der Box-Whiskers Plot besteht aus einem Kasten, dessen Mittellinie den Median und die äußeren Begrenzungen das 1. und 3. Quartil sind. Damit befinden sich in der Box 50% der betrachteten Werte. Zu beiden Seiten der Box sind zwei weitere Grenzen definiert (inner and outer fences), die jeweils die 1.5- und 3-fache Länge des Quartilabstandes (Differenz zwischen 1. und 3. Quartil) von dem 1. bzw. 3. Quartil entfernt sind. Mögliche Daten werden innerhalb der ersten Grenze von einer durchgezogenen Linie überdeckt. Werte außerhalb der 1. Grenze werden als Ausreißer betrachtet und individuell dargestellt.

⁹ Die Abweichungen vom stationsbezogenen Mittelwert

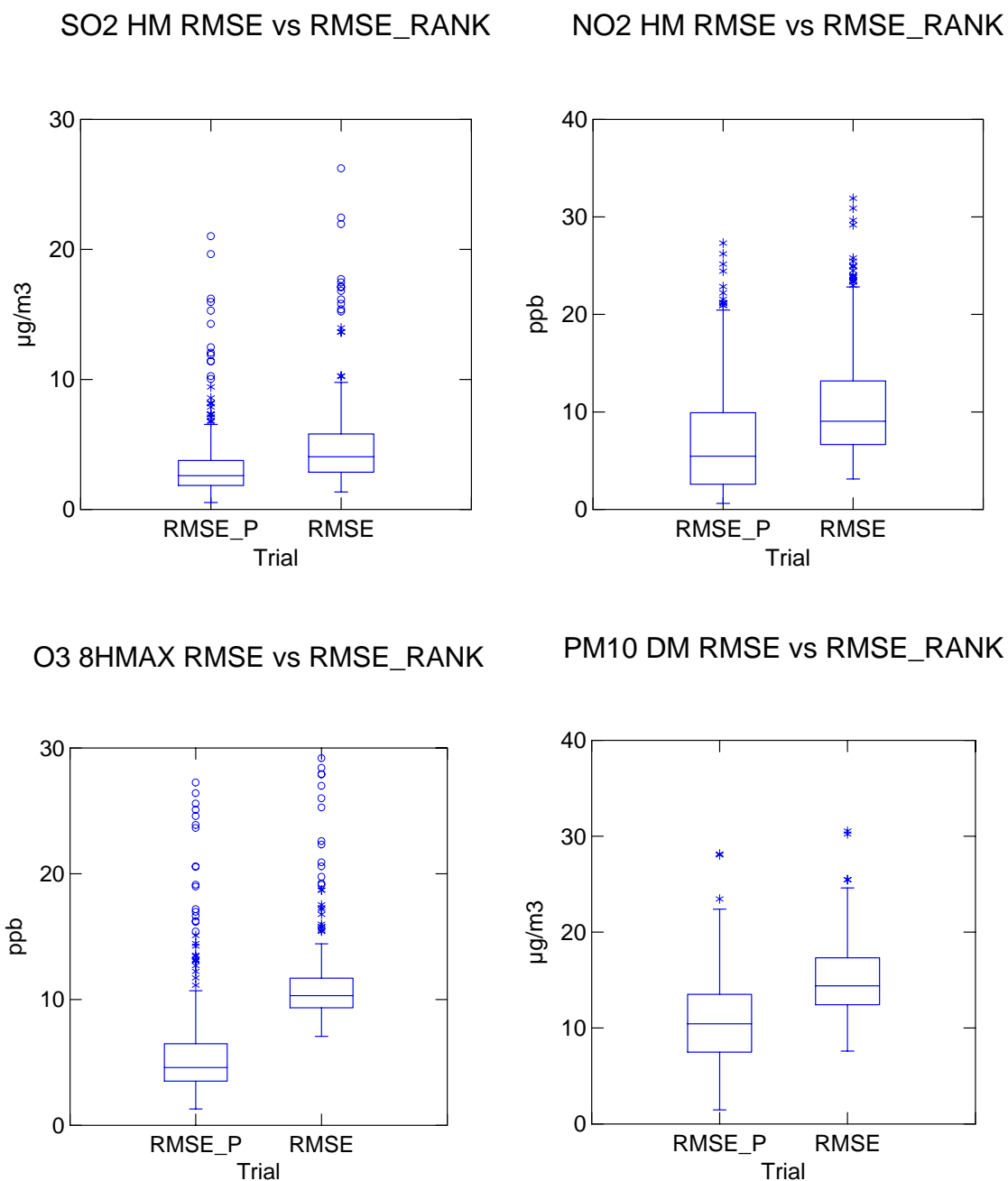


Abbildung 16 Häufigkeitsverteilung (Box-Whiskers-Plot) des stationsbezogenen RMSE mit (RMSE) und ohne Zeitbezug (RMSE_P) für die Stundenwerte von NO_2 (oben rechts), die Stundenwerte von SO_2 (oben links), die täglichen 8h-Mittelwerte von Ozon (unten links) und die Tagesmittelwerte von PM_{10} (unten rechts).

4.3.3 Konzentrationswerte beim maximalen Fehler

Abbildung 17 zeigt die Streudiagramme von Beobachtungs- und Modellwerten beim größten absoluten Fehler. Der größte absolute Fehler wird mit Perzentil-Bezug ermittelt und entspricht damit der hier vorgenommenen Interpretation der Modellfehlerdefinition nach den Richtlinien 1999/39/EC und 2002/3/EC. Die Zeitreihe von Messung und Rechnung wird der Größe nach geordnet, danach wird die Differenz zwischen den Werten gleicher Rangfolge gebildet. Die größte Differenz ergibt den größten absoluten Fehler.

Der größte absolute Fehler entsteht für alle Stoffe in den meisten Fällen durch eine deutliche Unterschätzung des Modells bei sehr hohen Werten der Beobachtung. Bei Ozon gibt es auch etliche Stationen, bei denen der größte Fehler bei relativ niedrigen Messwerten eine Folge einer Überschätzung des Modells ist. Eine Überschätzung des Modells beim Perzentilwert des größten Fehlers ist auch bei NO₂ an einigen Stationen feststellbar. In diesen Fällen stimmen auch beim maximalen Fehler Modellwert- und Beobachtungswert aber recht gut überein, da die entsprechenden Punkte nahe der 1:1-Geraden liegen. Dies bedeutet, dass der maximale Fehler gering ist. Aus der Darstellung der Konzentrationswerte für den maximalen Fehler ist nicht sofort erkennbar, ob der größte absolute Fehler im höheren oder niedrigeren Perzentilbereich auftritt. Der Fehler für einen festgelegten höheren Perzentilbereich wird bestimmt bei der Ermittlung des Fehlers beim Perzentil, das die Anzahl von erlaubten Überschreitungen des Grenzwertes angibt (siehe Kapitel 4.3.5)

In Abbildung 17 ist auch die Regimezugehörigkeit der Stationen dargestellt. Durch die Regimezugehörigkeit können starke Abweichungen zwischen Modell- und Beobachtung aufgrund nicht übereinstimmenden Skalen (Repräsentativität der Messstationen) erklärt werden. Die Stationen des NO₂-Regimes „Verkehr belastet“ werden aufgrund der groben Auflösung des Modells und dem geringen Repräsentativitätsbereich der Stationen dieser Klasse nicht im richtigen Konzentrationsniveau wiedergegeben. Beim Ozon sind die Stationen mit einem großen maximalen Fehler bei den 8h-Mittelwerten über 60 ppb meistens Bergstationen, für die die hier herangezogenen Konzentrationswerte der Modellbodenschicht nicht zum Vergleich geeignet sind. Wichtig für die Einschätzung des Gütemaßes der EU-Richtlinien ist der Umstand, dass eine große Streuung der Regimes offenbar wird. So treten große Maximalfehler auch bei Landstationen bei einem Konzentrationsniveau auf, das für diese Stationsklasse untypisch ist. Dies deutet daraufhin, dass der maximale Fehler eine Folge von sehr hohen Konzentrationsmessungen mit „Ausreißer“-Charakter sein kann. Es ist dabei unklar, ob es sich um einen Messfehler oder um eine spezielle Immissionssituation handelt.

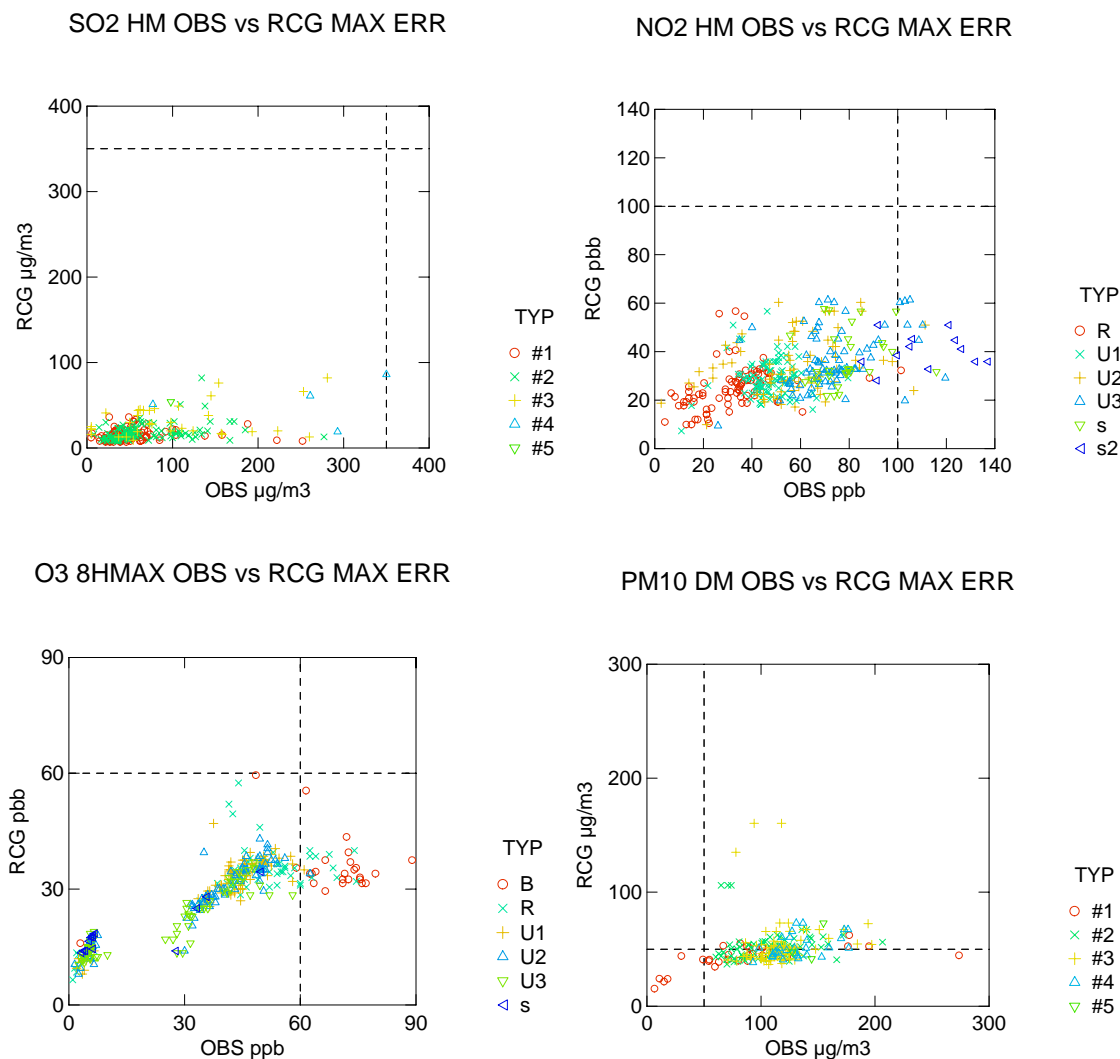


Abbildung 17 Konzentrationswerte aus Beobachtung (OBS) und Modellrechnung (RCG) beim maximalen absoluten Fehler für die Stundenwerte von NO_2 (oben rechts), die Stundenwerte von SO_2 (oben links), die täglichen 8h-Mittelwerte von Ozon (unten links) und die Tagesmittelwerte von PM_{10} (unten rechts). Markiert ist das Konzentrationsniveau des Grenzwertes.

4.3.4 Konzentrationswerte beim Perzentil, das der erlaubten Überschreitungsanzahl entspricht

Analog zu der Darstellung in Kapitel 4.3.3 werden in Abbildung 18 die Konzentrationswerte von Modell- und Beobachtung – und damit der Fehler – bei demjenigen Perzentil dargestellt, das der Anzahl der erlaubten Überschreitungen des Grenzwerts entspricht (per_err_p). Im Gegensatz zum maximalen Fehler mit Perzentil-Bezug (*max_err_p*) ist hier das Perzentil also bekannt, bei dem die Abweichung ermittelt wird. Es handelt sich dabei bei allen Stoffen um einen Wert aus den höheren Bereichen, da dieser Perzentilwert immer über 90 liegt (siehe auch Tabelle 1). Im Vergleich zu Abbildung 17 fällt in Abbildung 18 auf, dass die Werte im Streudiagramm besser um die 1:1-Gerade verteilt sind. Dies bedeutet, dass der Fehler für das Überschreitungsperzentil geringer ist als der maximale Fehler ermittelt aus allen Perzentilen.

Dies gilt im stärksten Maße für PM₁₀. Dieses Maß, das von Stedman et al. (2003) für die Bewertung der Modellgüte im Sinne der EU-Richtlinien verwendet wird, führt also zu einer höheren Güteeinschätzung der Modellrechnung als die Verwendung des maximalen Fehlers aus dem gesamten Perzentilbereich.

Da das Perzentil der erlaubten Häufigkeit der Überschreitungen des Grenzwertes entspricht, stellen Datenpunkte jenseits der markierten Konzentrationsschwelle eine echte Grenzwertüberschreitung dar. Aus Abbildung 18 wird deutlich, dass die Stationen mit einer hohen Anzahl von beobachteten Überschreitungen bei den NO₂-Stundenwerten auch in der Modellrechnung die höchsten Werte haben. Der Immissionsgrenzwert von 200 µg/m³ wird in der Rechnung aber kaum überschritten. Die Grenzwertüberschreitungen treten aber durchgängig an Stationen in den höher belasteten urbanen Regimen auf. Im Sinne einer skalengerechten Auswahl von Stationen zur Bestimmung der Modellgüte sollten solche Stationen auch nicht zur Güteeinschätzung eines Modells mit einer Auflösung im Dekakilometerbereich herangezogen werden. Im Gegensatz dazu decken sich die Stationen mit Grenzwertüberschreitungen beim PM₁₀ nicht in allen Fällen mit den Stationen, an denen das Modell die höchsten Werte berechnet. Das Modell berechnet die höchsten PM₁₀-Werte im Ruhrgebiet, einem Ballungsgebiet, das von der Modellskala erfasst werden kann. Die Modellwerte sind dort jedoch teilweise etwas höher als die zugehörigen Messungen.

Im Vergleich zu Abbildung 17 mit dem maximalen Fehler mit Perzentil-Bezug wird eine bessere „Gruppierung“ der Regime deutlich. Weniger belastete Regime haben auch geringere Maxima. Wie in Flemming (2003b) erwähnt, sind die Mittelwerte, die im Wesentlichen die Basis für die Klassifikation der Messstationen in Immissionsregime sind, und die 98-Perzentile der stündlichen Immissionswerte aller Stoffe recht gut korreliert. Erst bei deutlich höheren Perzentilen tritt die im vorangegangenen Kapitel dargestellte starke Streuung auf.

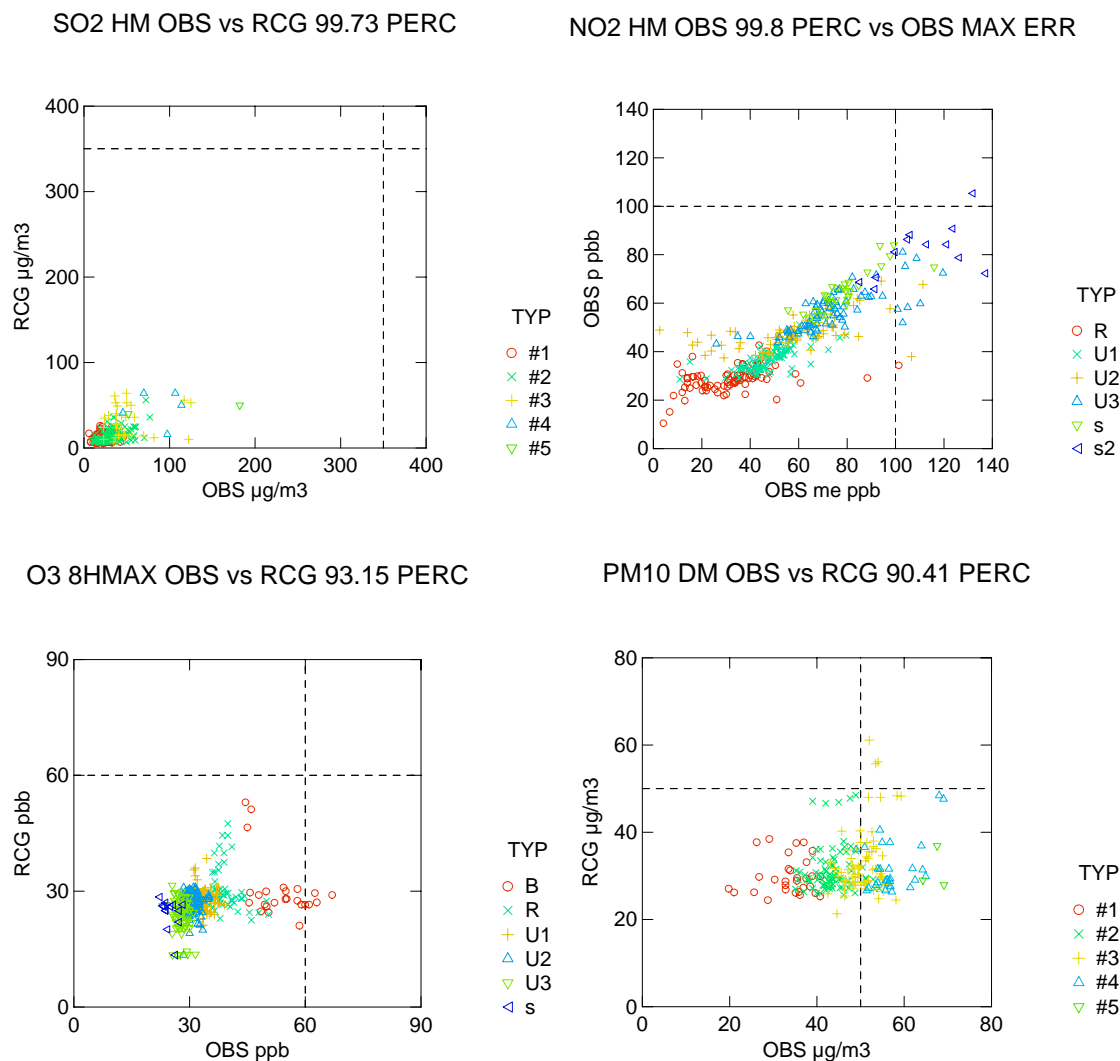


Abbildung 18 Konzentrationswerte aus Beobachtung (OBS) und Modellrechnung (RCG) beim Perzentil der erlaubten Überschreitungsanzahl des Grenzwertes für die Stundenwerte von NO₂ (oben rechts), die Stundenwerte von SO₂ (oben links), die täglichen 8h-Mittelwerte Ozon (unten links) und die Tagesmittelwerte von PM₁₀ (unten rechts). Markiert ist das Konzentrationsniveau des Grenzwertes.

4.3.5 Vergleich der Fehlermaße max err_p und per err_p_{GW} hinsichtlich der betrachteten Konzentrationsniveaus

In diesem Kapitel wird untersucht, bei welchem Konzentrationsniveau das Fehlermaß per err_p_{GW} (Fehler am Überschreitungsperzentil p_{GW}) bzw. der maximale Perzentil-Fehler max err_p bestimmt wird. Dieser Vergleich ist notwendig für die Einschätzung, welcher „Prüfung“ das Modell unterzogen wird, um seine Eignung zur Luftgütebewertung nach den EU-Richtlinien zu erlangen.

Abbildung 19 zeigt eine Gegenüberstellung der jeweiligen Messwerte, bei denen die beiden verschiedenen Fehlermaße ermittelt werden. Der Konzentrationswert an dem der erlaubten Anzahl von Grenzwertüberschreitungen entsprechenden Perzentil ist in vielen Fällen niedriger als der Konzentrationswert für den der maximale Fehler ermittelt wird. Zudem streuen die zu diesem Perzentil gehörigen Konzentrationswerte weniger als diejenigen für den maximalen Fehler. Bei SO₂ und PM10 liegt das Konzentrationsniveau beim maximalen Fehler sehr deutlich über dem bei der erlaubten Häufigkeit der Grenzwertüberschreitung. Dies deutet darauf hin, dass beim SO₂ der maximale Fehler meistens für die höchsten beobachteten Werte berechnet wird. Bei den PM10-Tagesmittelwerten sind 35 Überschreitungen (bis 2005) des Grenzwerts von 50 µg/m³ erlaubt, was einem 90,41-Perzentil entspricht. Damit handelt es sich bei den Konzentrationswerten für den maximalen Fehler nicht unbedingt um untypisch hohe Werte.

Bei allen Werten unterhalb der 1:1 Geraden ist das Perzentil, für das der maximale Fehler bestimmt wird, höher als das Perzentil, das der erlaubten Anzahl von Grenzwertüberschreitungen entspricht (Abbildung 19). Einen genaueren Vergleich erlaubt Abbildung 20, die die Häufigkeitsverteilung der Perzentile zeigt, für die der maximale Fehler $\max\ err_p$ berechnet wird. Es ist offensichtlich, dass der maximale Fehler an vielen Stationen beim 100-Perzentil, d.h. dem jeweils höchsten gemessenen Stunden- oder Tageswert auftritt. Für diesen Wert besteht eine gewisse Wahrscheinlichkeit, dass es sich um einen fehlerhaften Ausreißer handelt. Bei den stundenbezogenen Konzentrationswerten und der relativ geringen erlaubten Überschreitungsanzahl sind die Perzentile beim maximalen Fehler von NO₂ und SO₂ sonst überwiegend niedriger als das Perzentil, das der erlaubten Anzahl von Grenzwertüberschreitungen entspricht. Bei den Tageswerten von Ozon und PM10 sind die Perzentile für den maximalen Fehler dagegen fast immer größer als das Überschreitungsperzentil, da dieses bei diesen beiden Stoffen relativ niedrig (d.h. relativ viele Überschreitungen sind zugelassen) angesetzt ist. Es muss trotzdem nochmals betont werden, dass der maximale Fehler bei allen Stoffen an der Mehrzahl der Stationen für den höchsten gemessenen Wert berechnet wird.

Zusammenfassend kann festgestellt werden, dass das Fehlermaß $\text{per}\ err_p_{\text{GW}}$, das am Überschreitungsperzentil gewonnen wird, sinnvoller ist als das aus den Vorschriften der EU-Richtlinien abgeleitete Fehlermaß $\max\ err_p$, das den maximalen Fehler, gewonnen aus allen Perzentilbereichen, angibt. Die Hauptgründe dafür sind:

1. geringere Ausreißer-Anfälligkeit des Fehlermaßes $\text{per}\ err_p_{\text{GW}}$, da der maximale Fehler an vielen Stationen für den höchsten gemessenen Wert auftritt,
2. Überprüfung der Modellierung im Konzentrationsbereich, der durch die Grenzwerte und Beurteilungsschwellen festgelegt ist, was einen direkten Bezug zu den EU-Richtlinien herstellt,
3. das 50% Kriterium ist für einen größeren Anteil an Stationen erfüllt (siehe Kapitel 4.3.6).

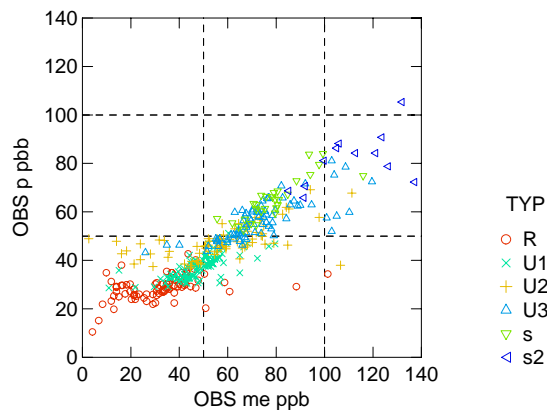
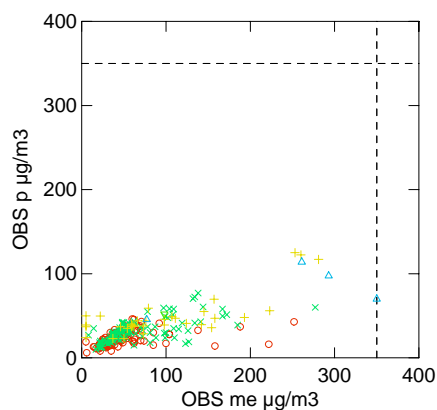
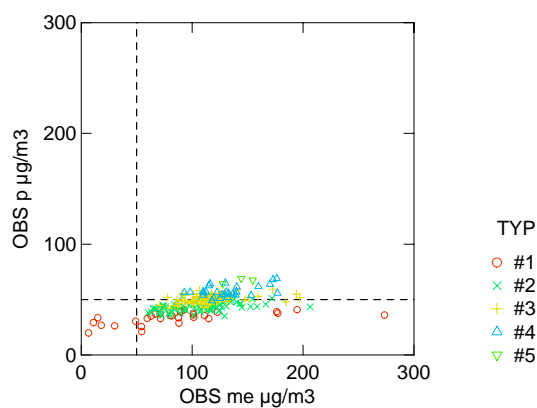
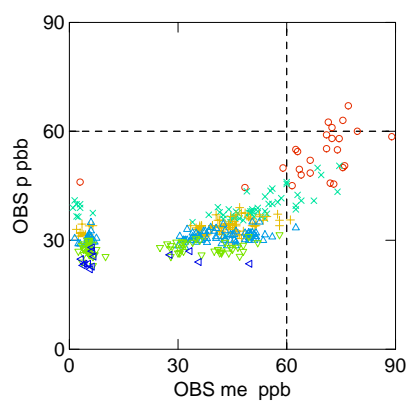
SO₂ HM OBS 99.73 PERC vs OBS MAX ERRNO₂ HM OBS 99.8 PERC vs OBS MAX ERRO₃ 8HMAX OBS 93.15 PERC vs OBS MAX ERRPM₁₀ DM OBS 90.41 PERC vs OBS MAX ERR

Abbildung 19 Beobachtete Konzentrationswerte beim maximalen absoluten Fehler (OBS me) und bei dem Perzentilwert der erlaubten Anzahl von Überschreitungen des Grenzwertes (OBS p) für die Stundenwerte von NO₂ (oben rechts), die Stundenwerte von SO₂ (oben links), die täglichen 8h-Mittelwerte von Ozon (unten links) und die Tagesmittelwerte von PM₁₀ (unten rechts).

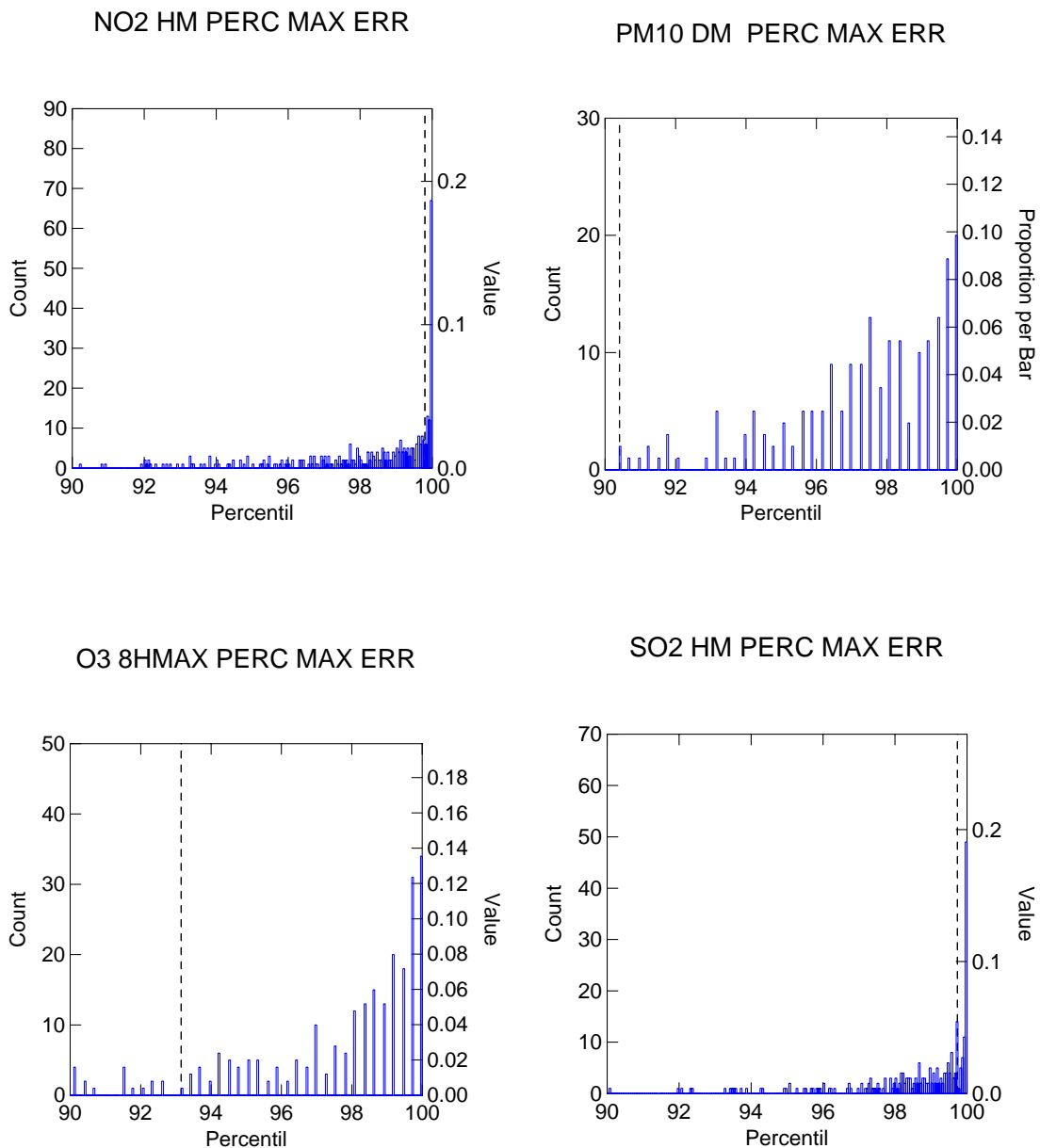


Abbildung 20 Histogramm der Perzentile, für die der größte absolute Fehler bestimmt wird (max err_p). Markiert ist das Perzentil, das der erlaubten Häufigkeit von Grenzwertüberschreitungen entspricht. Aus Gründen der Darstellung ist nur der Bereich der Perzentile 90-100 dargestellt, in der fast alle Werte zu finden sind.

4.3.6 Relative Fehlermaße

Nach der Diskussion der verschiedenen absoluten Fehlermaße werden diese mit dem entsprechenden Messwert normiert, um zu einer relativen Fehlerangabe zu gelangen. Dies ist notwendig, da in den EU-Richtlinien 1999/30/EC und 2003/3/EC eine relative Formulierung der Mindestgenauigkeit gewählt wurde (siehe Tabelle 2).

Dazu werden in Abbildung 21 die stationsbezogenen relativen Fehlermaße für die Tages- und Stundenwerte dargestellt. Die relativen Fehlermaße beziehen sich auf die Stundenwerte

von NO₂ und SO₂, die Tagesmittelwerte von PM₁₀ und die Tagesmaxima des 8-Stundenmittelwertes von Ozon. Die Histogramme aller relativen Fehlermaße für jede Station in Deutschland werden in Form von Box-Whiskers-Plots (siehe Fußnote 8, S. 31) gezeigt. Es handelt sich dabei um folgende Größen (siehe Kap. 3.2):

1. maximaler relativer Fehler mit Zeitbezug (max rel err_t)
2. maximaler relativer Fehler mit Perzentilbezug (max rel err_p)
3. Relativer maximaler Fehler mit Zeitbezug (rel max err_t)
4. Relativer maximaler Fehler mit Perzentilbezug (rel max err_p)
5. Relativer Fehler bei dem Perzentil, das der erlaubten Überschreitungshäufigkeit des Grenzwertes entspricht (rel per err_p)

Die Fehler nach Punkt 4. und 5. sind in den vorangegangenen Kapiteln ohne Relativierung behandelt worden. Der Punkt 2. entspricht Punkt 4., dem Fehlermaß nach der EU-Richtlinie, nur dass Maximalsuche und Relativierung vertauscht wurden und somit der maximale relative Fehler ermittelt wird. Um die Wirkung des Übergangs vom Zeit- zum Perzentil-Bezug zu demonstrieren, werden mit den Punkten 1. und 3. die Variante im Zeitbezug zu den Fehlermaßen 2. und 4. dargestellt.

Der maximale relative Fehler (Punkt 1.) ist bei allen Stoffen und an allen Stationen zum Teil sehr deutlich über 100%. Durch Aufgabe des Zeitbezuges verringern sich die maximalen relativen Fehler sehr stark, doch auch hier sind die meisten Werte mit Ausnahme von PM₁₀ über 100%, bei PM₁₀ liegen sie für die Stationen in Deutschland im Bereich von 50-100%. Eine weitere Verringerung und eine Verkleinerung der Streuung sind beim relativen Maximalfehler festzustellen. Wird auch hier nur der Perzentil-bezogene Fehler rel max err_p untersucht, so sind die Werte meist kleiner als 100% und bei Ozon sind 75% der Stationen mit der geforderten Mindestgenauigkeit von 50% simuliert, bei NO₂ beträgt dieser Anteil ca. die Hälfte. Bei PM₁₀ können die Mindestgenauigkeitsforderungen der EU-Richtlinien nur an ca. ¼ der Stationen und bei SO₂ nur an einigen Stationen erfüllt werden. Bei dem neu vorgeschlagenem Fehlermaß nach Punkt 5., d.h. der Abweichung beim Perzentil der Häufigkeit der erlaubten Grenzwertüberschreitung, liegt der relative Fehlerwert für fast alle Stationen bei Ozon und PM₁₀, für ca. ¾ der Stationen bei NO₂ und für ca. 1/3 der Stationen bei SO₂ unter 50%.

Die Abbildung 22 bis Abbildung 25 zeigen genauer die Regimezugehörigkeit und Anzahl der Stationen, die nach den relativen Fehlermaßen nach Punkt 4 und 5 einen Fehler von weniger als 50% haben. Beim relativen Maximalfehler sind das die Stationen, bei denen nach EU-Richtlinien-Interpretation die Mindestgenauigkeit erfüllt ist. Beim Vergleich der Fehlermaße nach Punkt 4. und 5. wird zunächst der geringere relative Fehler beim Perzentil der Häufigkeit der erlaubten Grenzwertüberschreitungen im Vergleich zum maximalen Fehler deutlich. Bei PM₁₀, NO₂ und SO₂ wird erwartungsgemäß die Abnahme der Modellgüte der hier verwendeten großräumigen RCG-Modellversion (Auflösung nur circa 25 bis 30 km) bei zunehmender Belastung deutlich, bzw. für Ozon beim Übergang vom Landregime, über die städtischen Regimes bis hin zu den Verkehrsregimes. Die Bergstationen werden aufgrund des Umstands, dass hier nur Modellwerte aus der Bodenschicht verwendet werden, häufig nicht mit der erforderlichen Genauigkeit simuliert. Es ist jedoch bei allen Stoffen keinesfalls festzustellen, dass die Modellgüte an allen ländlichen oder suburbanen und damit der Modellskala entsprechenden Stationen immer ausreichend ist. Umgekehrt gilt auch, dass einige Stationen der Verkehrsregime und die deswegen durch die Modellskala eigentlich nur unzu-

reichend wiedergegeben werden sollten, mit ausreichender Genauigkeit simuliert werden. Die Gründe hierfür sind vielfältig. Zum einen muss berücksichtigt werden, dass die Regime-einteilung im Wesentlichen auf der Bewertung von Jahresmitteln beruht und bei der Bewertung der Klassen implizit davon ausgegangen wird, dass hoch belastete Stationen dem urbanen oder Straßenregime zuzuordnen sind. Insbesondere in ausgedehnten Ballungsgebieten führt dies großflächig zu einem Übergewicht der Stationen in den hoch belasteten Klassen. Solche Regionen, wie z.B. das Ruhrgebiet, können bereits mit einer Modellauflösung von 25 bis 30 km ausreichend genau simuliert werden, falls die der Simulation zugrunde liegenden Emissionsdaten der realen Situation entsprechen. Damit kann ein von der Auflösung her großräumiges Modell in solchen Regionen auch Messungen von Stationen in den hoch belasteten Klassen mit ausreichender Genauigkeit simulieren. Es muss auch berücksichtigt werden, dass die Normierung mit dem Messwert bei hohen Belastungen, wie sie an den urbanen Stationen im Allgemeinen auftreten, größere absolute Fehler gestattet.

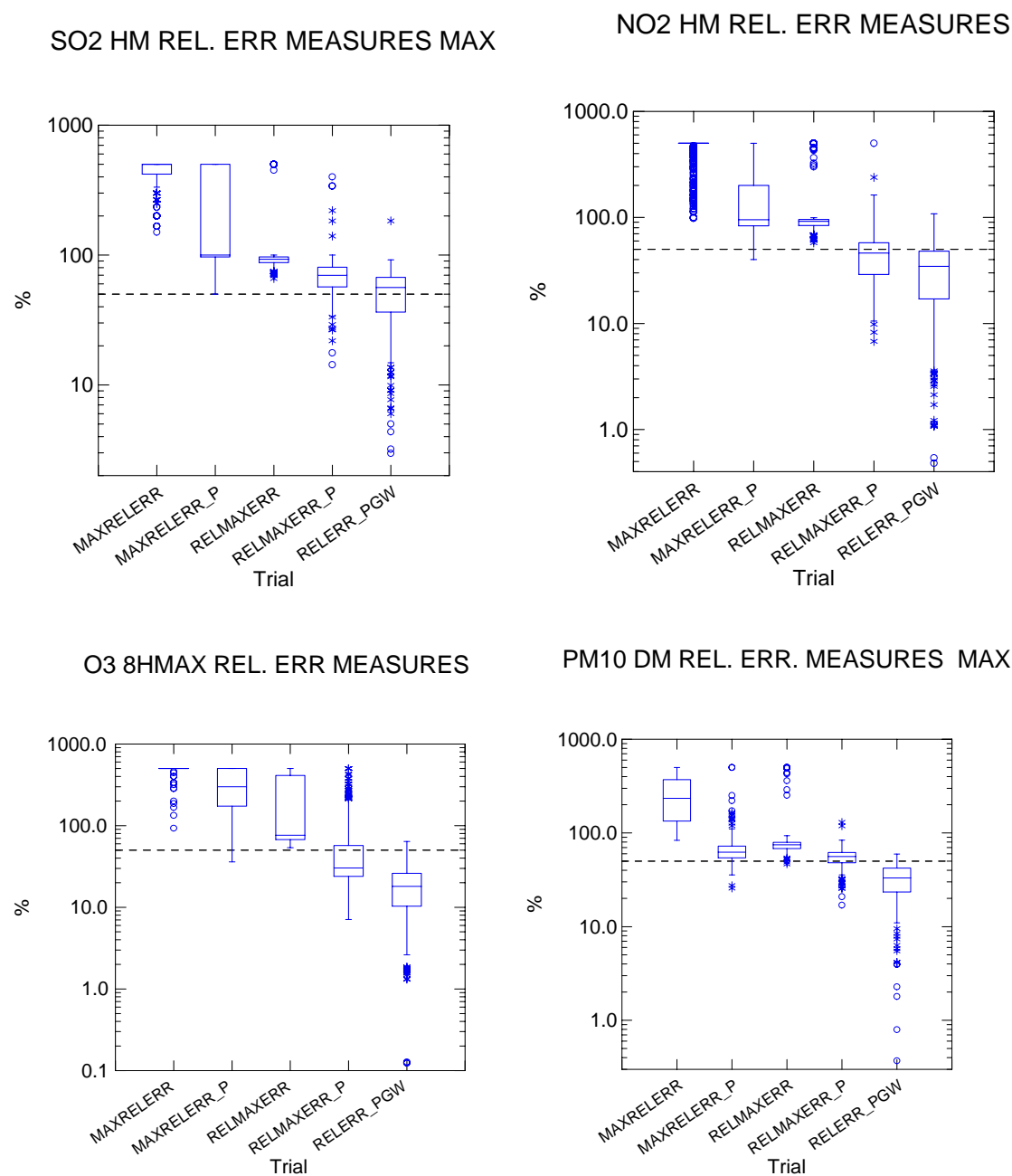


Abbildung 21 Häufigkeitsverteilung (Box-Whiskers-Plot) von stationsbezogenen relativen Fehlermaßen mit und ohne Zeitbezug (RMSE_P) für die Stundenwerte von NO₂ (oben rechts), die Stundenwerte von SO₂ (oben links), die täglichen 8h-Mittelwerte von Ozon (unten links) und die Tagesmittelwerte von PM10 (unten rechts). Maximaler relativer Fehler mit Zeitbezug (max rel err), maximaler relativer Fehler mit Perzentilbezug (max rel err_p), relativer maximaler Fehler mit Zeitbezug (rel max err), relativer maximaler Fehler mit Perzentilbezug (rel max err_p), relativer Fehler bei dem Perzentil, das der erlaubten Überschreitungshäufigkeit des Grenzwertes entspricht (rel err pGW). Alle Stationen in Deutschland.

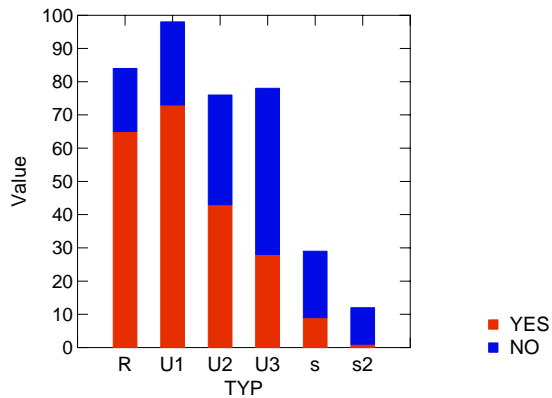
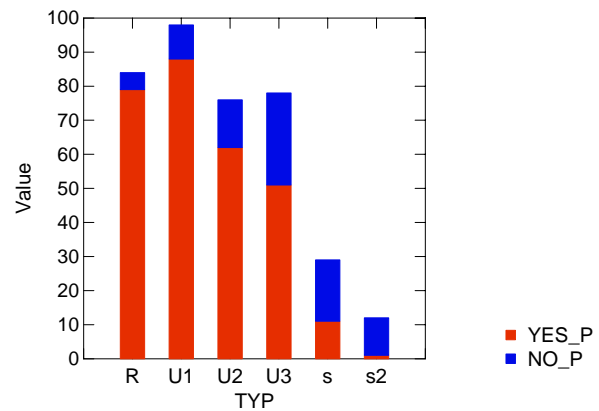
NO₂ HM REL MAX MODEL ERROR < 50%NO₂ HM REL MODEL ERROR P_GW< 50%

Abbildung 22 Anzahl der Stationen, die die Genauigkeitsanforderung von 50% hinsichtlich der NO₂ Stundenmittelwerte erfüllen (YES) für den relativen maximalen Fehler (links) und den Fehler am Perzentil der Häufigkeit der erlaubten Grenzwertüberschreitungen (rechts).

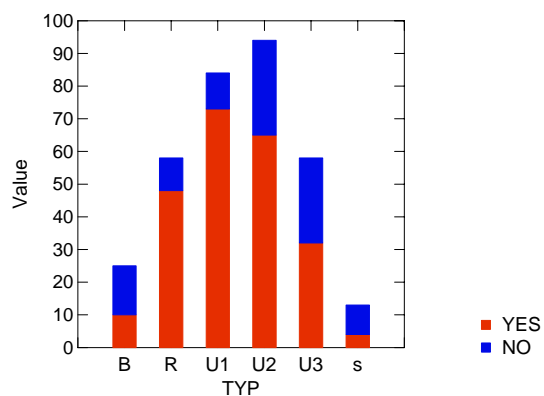
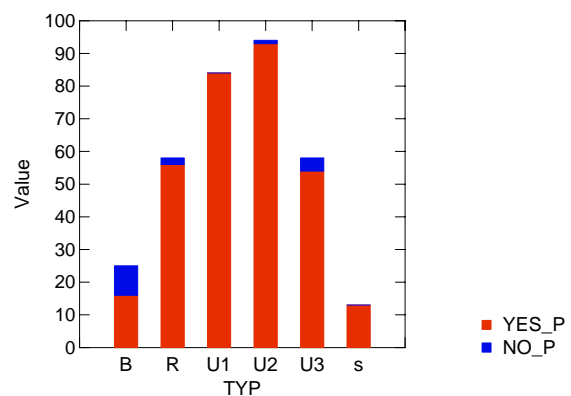
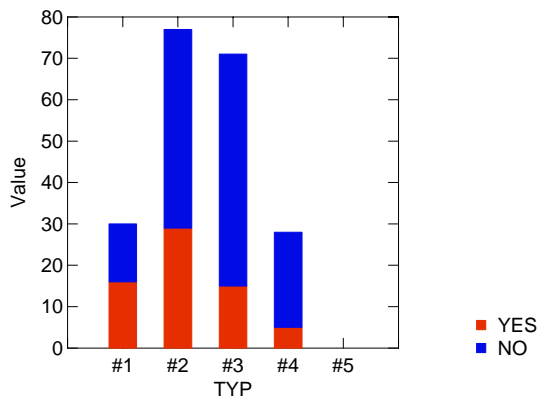
O₃ 8HMAX REL MAX MODEL ERROR < 50%O₃ 8HMAX REL MODEL ERROR P_GW< 50%

Abbildung 23 Anzahl der Stationen, die die Genauigkeitsanforderung von 50% hinsichtlich der täglichen Maxima der 8h-Mittel von Ozon erfüllen (YES) für den relativen maximalen Fehler (links) und den Fehler am Perzentil der Häufigkeit der erlaubten Grenzwertüberschreitungen (rechts).

PM10 DM REL MAX MODEL ERROR < 50%



PM10 DM REL MODEL ERROR P_GW< 50%

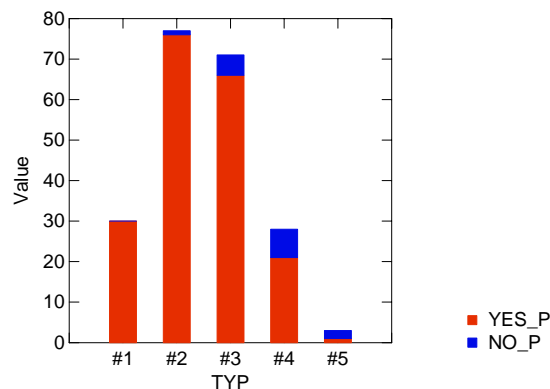
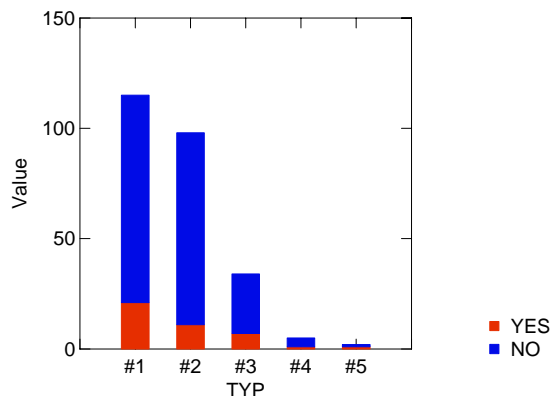


Abbildung 24 Anzahl der Stationen, die die Genauigkeitsanforderung von 50% hinsichtlich der Tagesmittelwerte von PM10 erfüllen (YES) für den relativen maximalen Fehler (links) und den Fehler am Perzentil der Häufigkeit der erlaubten Grenzwertüberschreitungen (rechts).

SO2 HM REL MAX MODEL ERROR < 50%



SO2 HM REL MODEL ERROR P_GW< 50%

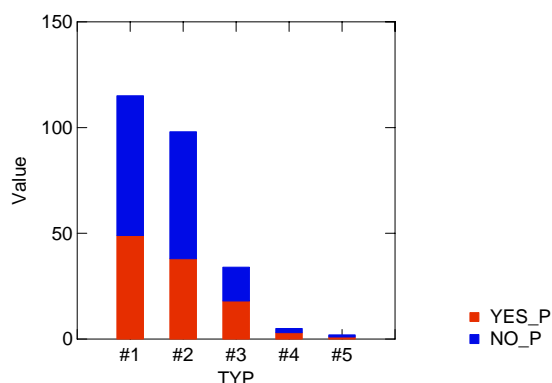


Abbildung 25 Anzahl der Stationen, die die Genauigkeitsanforderung von 50% hinsichtlich der Stundenmittelwerte von SO₂ erfüllen (YES) für den relativen maximalen Fehler (links) und den Fehler am Perzentil der Häufigkeit der erlaubten Grenzwertüberschreitungen Grenzwertüberschreitung (rechts).

4.3.7 Überschreitung von Grenzwert- und Beurteilungsschwellen

Komplementär zu dem Fehler der Konzentrationswerte beim Perzentil der erlaubten Grenzwertüberschreitungen wurde in Kapitel 3.2 vorgeschlagen, die Unterschiede im Perzentil bei einem gewissen Konzentrationswert, nämlich dem Grenzwert- und den beiden Beurteilungsschwellen zu betrachten. Bei der Untersuchung dieser Differenz ergab sich jedoch an vielen Stationen das Problem, dass diese Konzentrationswerte gar nicht erreicht werden und somit mit ihnen eine Modellüberprüfung nicht möglich ist. Nur bei PM10 ergibt sich eine große

Anzahl von Überschreitungen aufgrund der niedrigen Konzentrationsschwellen der unteren Beurteilungsschwelle, die an fast allen Stationen in Deutschland überschritten wird. Die zugehörigen Perzentile aus Modellrechnung und Beobachtung sind in Abbildung 26 dargestellt für Grenzwert und Beurteilungsschwellen.

Liegen die Punkte in Abbildung 26 links (Messung) bzw. unter (Rechnung) der gestrichelten Linie, die dem durch die Anzahl der erlaubten Überschreitungen definierten Perzentil entspricht, wird an der betreffenden Station der entsprechende Schwellenwert an mehr Tagen überschritten als erlaubt ist. Bei einer vollständigen Übereinstimmung zwischen Messung und Rechnung in der Zahl der Stationen, an denen der Schwellenwert öfters überschritten wird als in der Richtlinie zugelassen, müssten alle Punkte entweder im rechten oberen (Keine unzulässigen Überschreitungen) oder im linken unteren Quadranten (Unzulässige Überschreitungen) liegen. Je höher die Perzentile, desto seltener tritt der entsprechende Konzentrationswert auf. Liegen die Punkte auf der 100 % -Linie, so wurde der Wert nicht erreicht. Beim Grenzwert (Abbildung 26 links) wird nur an 2 Stationen in Messung und Rechnung die zulässige Anzahl von Überschreitungen nicht eingehalten. Bei der unteren Beurteilungsschwelle ergeben sich zwischen Messung und Rechnung große Unterschiede in den Perzentilen, in der Aussage, ob eine Station mehr Überschreitungen aufweist als zugelassen, stimmen Messung und Rechnung aber gut überein.

Aufgrund der oben genannten Beschränkung und weil es für Ozon keine Beurteilungsschwellen gibt, wird dieser Ansatz jedoch nicht weiter verfolgt.

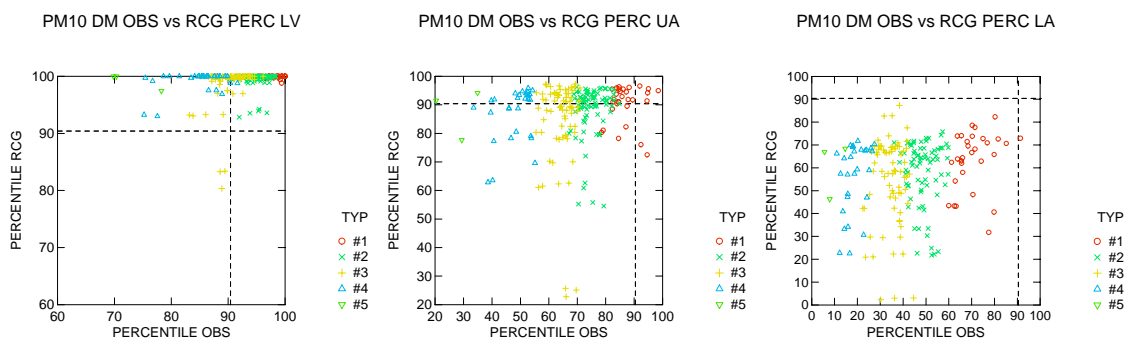


Abbildung 26 Perzentile der Tagesmittelwerte von PM10 beim Überschreiten der Grenzwertschwelle von $40 \mu\text{g}/\text{m}^3$ (links), der oberen Beurteilungsschwelle von $30 \mu\text{g}/\text{m}^3$ (Mitte) und der unteren Beurteilungsschwelle von $20 \mu\text{g}/\text{m}^3$ (rechts) in Beobachtung (x-Achse) und Modellrechnung (y-Achse). Die gestrichelte Linie entspricht dem Perzentil der erlaubten Überschreitungsanzahl. Kleinere Perzentilwerte bedeuten eine häufigere Überschreitung.

4.4 Vergleich von kurz- und langfristigen Fehlermaßen

Die EU-Richtlinien benennen Grenzwerte und Modellgüteanforderungen für kurzfristige Tages- und Stundenwerte und die langfristigen Jahresmittelwerte. Bei Ozon gibt es keinen Grenzwert für den Jahresmittelwert. Die langfristigen Wirkungen der Ozonbelastung werden mit den AOT-bezogenen Zielwerten erfasst. Zwischen den beiden Typen der Grenz- bzw. Zielwerte wird in der Rahmenrichtlinie kein Zusammenhang hergestellt, da für beide Werte die Einhaltung erforderlich ist, unabhängig davon, ob dies auch beim jeweils anderen Typ der Fall ist. Aus Sicht der Modellgütemessung erscheint jedoch ein Vergleich der Gütema-

ße hinsichtlich der kurz- und langfristigen Konzentrationswerte aus den folgenden Gründen sinnvoll:

1. Die Modellgüte ist bei kurz- und langfristigen Konzentrationswerten unterschiedlich
2. Die Modellgüteanforderungen der EU-Richtlinien sind bei kurz- und langfristigen Konzentrationswerten unterschiedlich.
3. Die kurzfristigen Grenzwerte entsprechen hohen Perzentilen (> 90) und die Jahresmittelwerte mittleren Perzentilen (≈ 50) der Häufigkeitsverteilung der Tages- oder Stundwerte.

Die Simulation von Jahresmittel- und Maximalwerten sind aus Sicht der CT-Modellierung zwei verschiedene Aufgaben, die mit unterschiedlicher Qualität gelöst werden. Der Fehler bei der Modellierung von Jahresmittelwerten entspricht dem Modellbias, der in hohem Maße von den jahresbezogenen Emissionswerten und den Randbedingungen abhängt. Beide Parameter sind oft nur mit einer großen Unsicherheit bekannt. Im Gegensatz dazu hängt die Modellierung der Maximalwerte weitgehend von der Modelldynamik ab, die die Konzentrationsänderungen in Abhängigkeit von den meteorologischen Feldern und der zeitbezogenen Emissionsmodellierung beschreibt. Der Einfluss des Modellbias ist dabei häufig gering.

Die Modellgüteanforderungen für die Jahresmittelwerte sind bei NO_2 und SO_2 mit 30% schärfer formuliert als bei den Tages- und Stundenwerten, bei denen nur 50% erreicht werden müssen. Da, wie gezeigt wurde, der Maximalfehler immer in den hohen Perzentilbereichen auftritt und die Jahresmittelwerte in etwa dem 50-Perzentil entsprechen, ist somit die erlaubte absolute Abweichung bei den Jahresmittelwerten deutlich geringer. Aufgrund des Modellbias und der asymmetrischen Verteilung der Konzentrationswerte gilt nicht unbedingt, dass die Fehler sich bei einer Verlängerung des Mittelungszeitraums immer besser kompensieren.

Das neu vorgeschlagene Fehlermaß $\text{rel per err}_{\text{pGW}}$ der relativen Abweichung bei demjenigen Perzentil, das der Anzahl zugelassener Grenzwertüberschreitungen entspricht, ist immer ein Maß, das den hohen Konzentrationsbereich untersucht, das Fehlermaß der EU-Richtlinien (so wie es hier interpretiert wird) des relativen maximalen Fehlers mit Perzentil-Bezug (rel max_err_p) entspricht gleichfalls meistens einem hohen Perzentil-Wert, der in vielen Fällen über dem der erlaubten Grenzwertüberschreitung liegt und oft sogar beim 100-Perzentil, d.h. dem höchsten beobachteten Wert auftritt (siehe Kapitel 4.3.5). Beide Fehlermaße testen demzufolge die Übereinstimmung der Häufigkeitsverteilungen der Tages- und Stundenwerte von Modell und Beobachtung im Bereich der Maximalwerte. Der Vergleich der Jahresmittelwerte würde einem Vergleich im Bereich des Medians entsprechen und damit eine weitere „Stützstelle“ bei der Untersuchung der Abweichung der Histogramme darstellen. Eine kombinierte Modellgüteanforderung wäre demzufolge eine Möglichkeit, die Eignung des Modells zur Simulation der Gesamtheit der Immissionssituation, d.h. der gesamten Häufigkeitsverteilung, zu testen.

Abbildung 27 bis Abbildung 29 zeigen eine Gegenüberstellung der relativen Fehler der Jahresmittelwerte und der relativen maximalen Fehler mit Perzentil-Bezug bzw. der Fehler beim Perzentil der Stundenwerte für SO_2 und NO_2 und der Tagesmittel für PM_{10} , das der erlaubten Anzahl von Überschreitungen entspricht. Ergänzend werden für Ozon in Abbildung 30 die relativen Fehler der Jahresmittelwerte mit denen des Tagesmaximums der 8h-Mittelwerte zusammengeführt. Dabei muss beachtet werden, dass hierbei der Mittelwert dieser Tageswerte nicht dem Jahresmittelwert entspricht.

Beim NO₂ (Abbildung 27) besteht ein ausgeprägter Zusammenhang zwischen den relativen Fehlermaßen der Stundenwerte und dem für die Jahresmittelwerte. Beide sind zudem nahezu gleich hoch. Aufgrund der niedrigen Schwelle für die Modellgüte bei den Jahresmittelwerten (30% gegenüber 50% bei den Stundenwerten) ist der Anteil der ausreichend gut simulierten Stationen hierbei demzufolge deutlich geringer. Der relative Maximalfehler (rel max err_p) ist im Mittel etwas, für viele Land-Stationen jedoch doch deutlich höher als der Fehler für das Perzentil der erlaubten Grenzwertüberschreitungen, rel per err_{p_{GW}}. Wie bereits ausgeführt, ist der Fehler für das Perzentil der erlaubten Grenzwertüberschreitungen stabiler als der relative Maximalfehler, da der letztere an vielen Stationen mit dem absolut höchsten gemessenen Wert verknüpft ist.

Beim PM₁₀ stimmen die relativen Fehler der Jahresmittelwerte und die beim Perzentil der erlaubten Grenzwertüberschreitungen berechneten Fehler sehr gut überein (Abbildung 28 rechts). Das Modell genügt an fast allen Stationen den Güteanforderungen für die Jahresmittelwerte. Die gute Übereinstimmung deutet darauf hin, dass der Modellfehler zwischen dem Median und dem 90,41-Perzentil durch einen in etwa konstanten relativen Wert (Bias) beschrieben werden kann. Der relative maximale Fehler ist jedoch deutlich höher als der der Jahresmittelwerte und streut viel stärker, so dass nur an wenigen Stationen das Modellgütekriterium erfüllt wird (Abbildung 28 links).

Der Zusammenhang der verschiedenen relativen Fehlermaße ist beim SO₂ nur schwach ausgeprägt (Abbildung 29). Es herrscht eine große Streuung, die Genauigkeitsanforderung an die Jahresmittelwerte wird jedoch an mehr Stationen eingehalten als diejenige für die Stundenwerte, insbesondere wenn man für die Stundenwerte das Maß des maximalen relativen Fehlers nimmt. Die relativen Fehler beim Perzentil der erlaubten Grenzwertüberschreitungen entsprechen im Niveau ungefähr denen der Jahresmittelwerte.

Die Ozon-Tochterrichtlinie 2002/3/EC kennt keinen jahresmittelwertbezogenen Grenzwert und demnach auch keine Modellgüteanforderung für diese Größe. Mit Ausnahme weniger Straßen-Stationen liegt der relative Fehler der Jahresmittelwerte unter 50%. Dies gilt auch für ca. 80% der Stationen beim relativen maximalen Fehler der Tageswerte (Abbildung 30 links). Die relativen Fehler beim 93,15-Perzentil, das der erlaubten Anzahl von Überschreitungen für den täglichen maximalen 8h-Mittelwert entspricht, sind deutlich kleiner als diejenigen des maximalen relativen Fehlers (vergleiche Abbildung 30, links und rechts), was wiederum aufzeigt, dass der maximale relative Fehler der Modellrechnung meistens bei den höchsten gemessenen Konzentrationen auftritt.

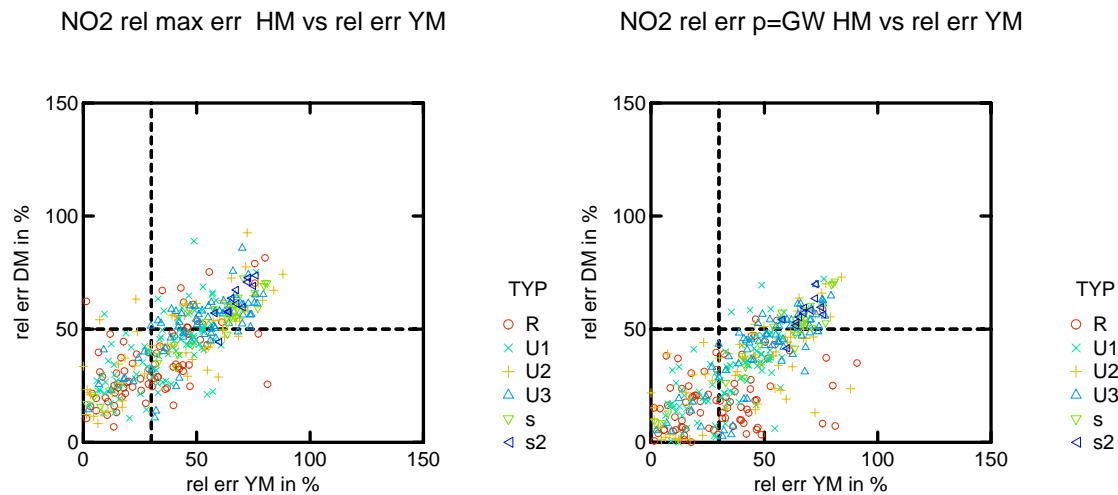


Abbildung 27 Streudiagramm der relativen Modellfehler der Jahresmittelwerte (x-Achse) und der relativen maximalen Fehler mit Perzentil-Bezug (y-Achse links) bzw. der Fehler beim Perzentil der Überschreitungshäufigkeit des Grenzwertes (y-Achse rechts) der Stundenwerte von NO₂. Markiert sind die Mindestgenauigkeitsanforderungen für die Modellrechnung nach 1999/30/EC. Die Darstellung ist beschränkt auf Fehler ≤ 150%.

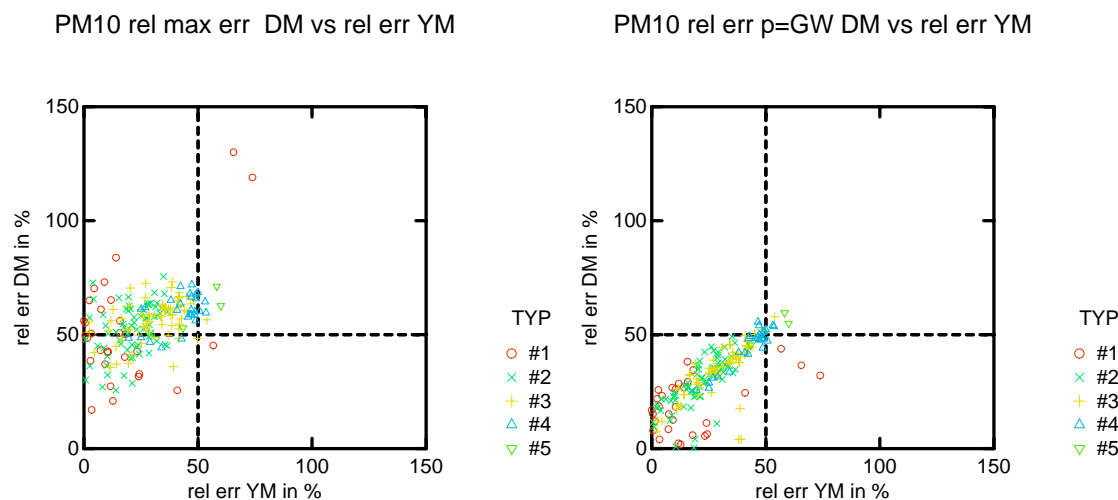


Abbildung 28 Streudiagramm der relativen Modellfehler der Jahresmittelwerte (x-Achse) und der relativen maximalen Fehler mit Perzentil-Bezug (y-Achse links) bzw. der Fehler beim Perzentil der Überschreitungshäufigkeit des Grenzwertes (y-Achse rechts) der Tagesmittelwerte von PM₁₀. Markiert sind die Mindestgenauigkeitsanforderungen für die Modellrechnung nach 1999/30/EC. Die Darstellung ist beschränkt auf Fehler ≤ 150%.

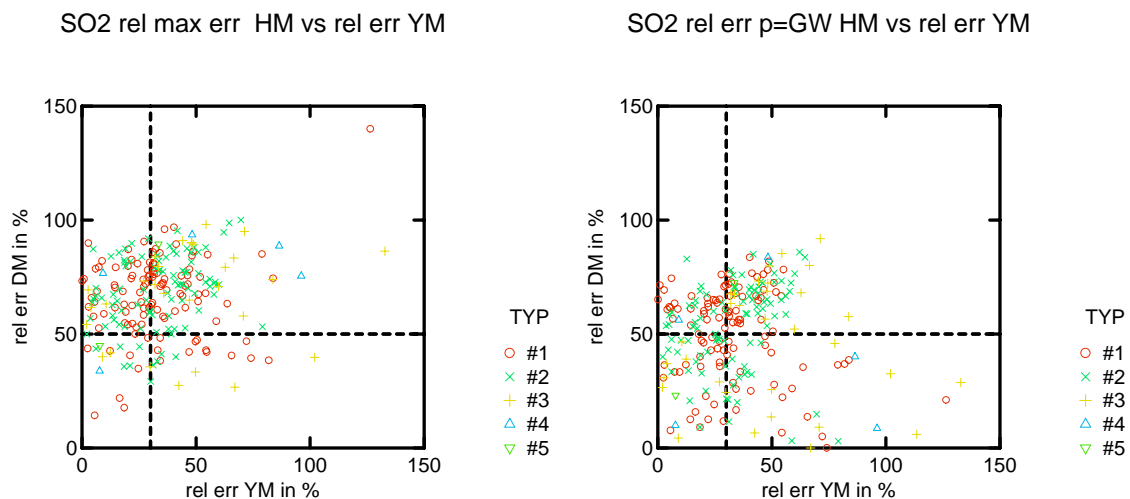


Abbildung 29 Streudiagramm der relativen Modellfehler der Jahresmittelwerte (x-Achse) und der relativen maximalen Fehler mit Perzentil-Bezug (y-Achse links) bzw. der Fehler beim Perzentil der Überschreitungshäufigkeit des Grenzwertes (y-Achse rechts) der Stundenwerte von SO₂. Markiert sind die Mindestgenauigkeitsanforderungen für die Modellrechnung nach 1999/30/EC. Die Darstellung ist beschränkt auf Fehler ≤ 150%.

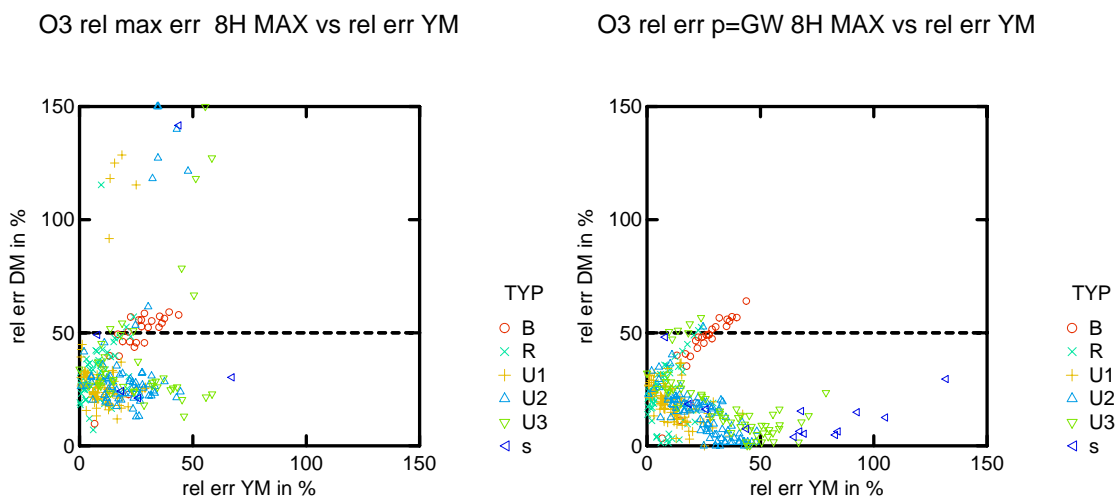


Abbildung 30 Streudiagramm der relativen Modellfehler der Jahresmittelwerte (x-Achse) und der relativen maximalen Fehler mit Perzentil-Bezug (y-Achse links) bzw. der Fehler beim Perzentil der Überschreitungshäufigkeit des Zielwertes (y-Achse rechts) der Tagesmaxima des 8h-Mittels von Ozon. Markiert sind die Mindestgenauigkeitsanforderung an die Tageswerte nach 2002/3/EC. Die Darstellung ist beschränkt auf Fehler ≤ 150%.

4.5 Räumliche Glättung der Messung entsprechend der Modellauflösung

Es wurde bereits in Kap. 3.6 angedeutet, dass unter Umständen zur Modellevaluierung eine räumliche Glättung der Messungen vorgenommen werden sollte, um den Einfluss einzelner Stationen im Glättungsgebiet zu vermindern. Mit solch einer Glättung soll also der Einfluss lokaler, und damit für das Glättungsgebiet nicht notwendigerweise repräsentativer Störungen, vermindert werden.

Es erscheint demzufolge untersuchenswert, die Modellgütebewertung auch mit fiktiven Stationen durchzuführen, deren Zeitreihe aus den stündlichen Mittelwerten besteht, die aus den Einzelreihen allen Stationen innerhalb eines Gebietes um die Station besteht, dessen Durchmesser der Auflösung des Modells entspricht. Dieser Ansatz wurde für die deutschen Stationen umgesetzt. Als Einflussradius zur Bestimmung der zur Mittelung heranzuziehenden Stationen wurde die halbe Maschenweite der Modellrechnung, d.h. 15 km, benutzt. Die Modellevaluierung wird dann nicht mit den Einzelmessungen, sondern mit Mittelwerten jedes Glättungsgebiets durchgeführt. Jede Station ist damit Mittelpunkt eines Glättungsgebiets. Liegt keine andere Station in diesem Gebiet, so stimmen die Mittelwerte des Gebiets mit den Einzelwerten der Station im Mittelpunkt überein. Die Mittelung wird dabei regimeübergreifend durchgeführt, d. h. es werden alle Stationen in einem Glättungsgebiet zur Mittelung benutzt, unabhängig von dem der Station zugeordneten Immissionsregime. Eine Auswahl der Stationen, die zur räumlichen Mittelbildung verwendet werden ist problematisch, da dafür meist nicht genügend Stationen zu Verfügung stehen.

Abbildung 31 bis Abbildung 38 zeigen die beiden relativen Fehlermaße $\text{rel per err}_{\text{pGW}}$ (relative Abweichung bei demjenigen Perzentil, das der Anzahl zugelassener Grenzwertüberschreitungen entspricht) und rel max_err_p (relativer maximaler Fehler mit Perzentil-Bezug, Maß der EU-Richtlinien) sowie den RMSE mit Perzentil-Bezug für die Tageswerte von Ozon und PM10 und die Stundenwerte von SO₂ und NO₂. Die stationsbezogenen Fehlermaße werden zum einem im Streudiagramm dargestellt und zum anderen wird die Veränderung des direkten Fehlermaßes (BASE) durch die Glättung (SMOOTH) als Mittelwert pro Immissionsregime dargestellt. Punkte unterhalb der 1:1 Linie und negative mittlere Veränderungen bedeuten eine Verbesserung des Fehlermaßes durch die Glättung. Liegen die Diagrammpunkte auf der 1:1-Linie, so enthält das Glättungsgebiet der Station im Allgemeinen keine andere Station. Dies zeigt sich z.B. bei den meisten Ozon-Stationen des Typs „Berg“, da diese Stationen oft weit entfernt von anderen Stationen liegen (Abbildung 31).

Für Ozon und PM10 verringert sich durch die modellskalengemäße Glättung der RMSE mit Perzentil-Bezug in allen Regimes. Bei NO₂ gilt dies nur für die urbanen und Verkehrsregime und bei SO₂ tritt eine Verbesserung nur bei den Stationen des Regimes #3 auf. Die Veränderungen der relativen Fehlermaße rel max_err_p und $\text{rel per err}_{\text{pGW}}$ haben qualitativ einen ähnlichen Charakter wie beim RMSE. Die Änderungen sind jedoch bei den Regimes mit geringer Belastung verhältnismäßig stärker ausgeprägt, da hier durch die Normierung mit dem geringen Messniveau gleiche absolute Abweichungen stärker betont werden. Das trifft besonders auf die mittlere Verschlechterung der relativen Fehlermaße der NO₂ Land- und Vorstadt-Regimes zu. Die Vergrößerung der mittleren relativen Fehler um 15 bzw. 10 % ist im Wesentlichen eine Folge von der starken Verschlechterung an einigen Stationen. In einigen Regimes, wie bei den urbanen Regimes U2 und U3 von Ozon, vergrößert sich trotz eines

verringerten RMSE, der alle Perzentile berücksichtigt, der relative Fehler beim Perzentil des Grenzwertes.

Die Vergrößerung der Fehlermaße bei den Land- und Vorstadt-Stationen von NO_2 ist ein Hinweis darauf, dass räumliche Mittelungen von Messungen nicht ohne Berücksichtigung der Stationscharakteristik durchgeführt werden sollten, da hier offensichtlich durch räumliche Mittelbildung mit benachbarten urbanen oder Verkehrsstationen das Niveau der Landstationen unsachgemäß angehoben wurde. Die starke Verbesserung des relativen maximalen Fehlers bei PM_{10} und SO_2 zeigt weiterhin, dass durch die Glättung Maximalfehler aufgrund extrem hoher Messwerte mit Ausreißer-Charakteristik gedämpft werden.

Es bleibt aber festzuhalten, dass mit Ausnahme von SO_2 eine reine räumliche Mittelung im Mittel zu einer Verbesserung der Fehlermaße führt. Dies gilt insbesondere für Stationen, die durch die Modellskala schlecht wiedergegeben werden können, d.h. Stadt- und Verkehrsregime und die Bergstationen. Ohne zusätzliche Nebenbedingung, wie z.B. die Berücksichtigung des Immissionsregimes oder die Gewichtung von Stationsabständen, treten jedoch höhere relative Modellfehler bei den Land- und Vorstadt-Stationen von NO_2 auf.

Auch die mit der Glättung verbundene Veränderung des normierenden Messwerts führt zu einer Änderung der relativen Fehler. Weiterhin muss immer berücksichtigt werden, dass die Wirkung des Glättungsansatzes immer von der vorgegebenen Stationsdichte beeinflusst ist.

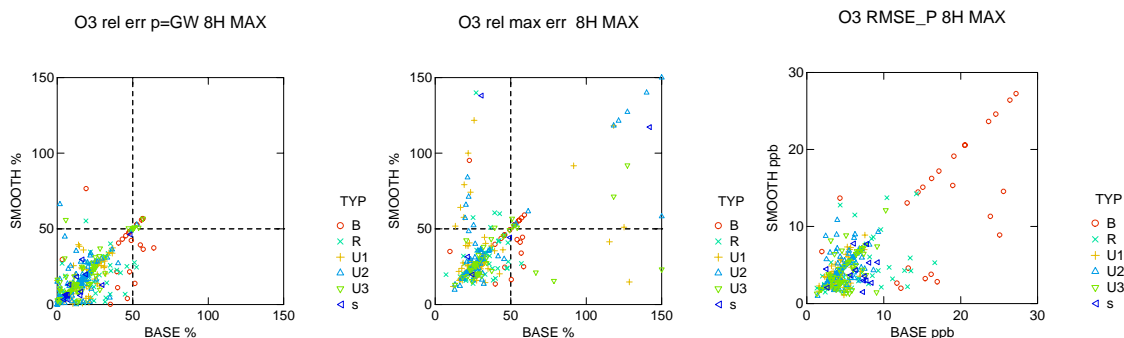


Abbildung 31 Relativer Fehler beim Perzentil des Zielwertes (rel per err_{p_{GW}}, links), relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und RMSE mit Perzentil-Bezug (rechts). Stationsbezogen mit direktem Vergleich (Base) und räumlicher Mittelung der Messung in der Auflösung des Modells (SMOOTH) für die Tagesmaxima des 8 h Mittels von Ozon.

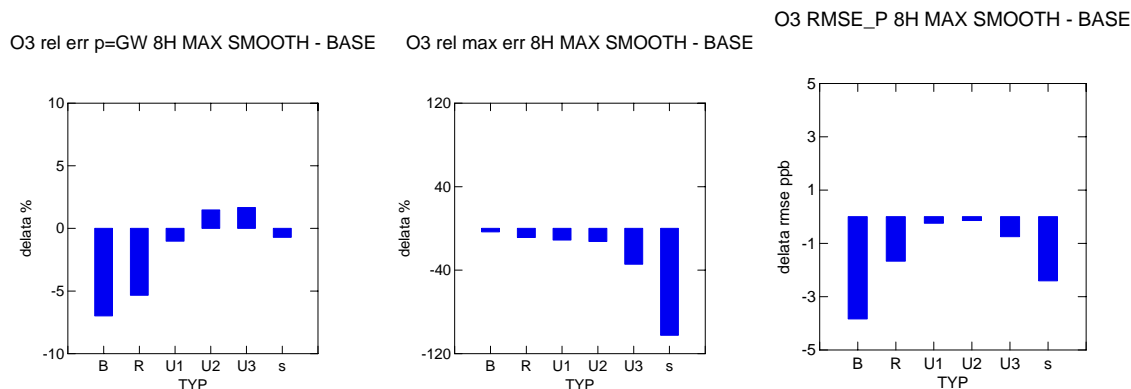


Abbildung 32 Mittlere Veränderung pro Immissionsregime des relativen Fehlers beim Perzentil des Zielwertes (rel per err_p_{GW}, links), des relativen maximalen Fehlers mit Perzentil-Bezug (rel max_err_p, Mitte) und des RMSE mit Perzentil-Bezug (rechts) zwischen direktem Vergleich und räumlicher Mittelung der Messung in der Auflösung des Modells für die Tagesmaxima des 8 h Mittels von Ozon.

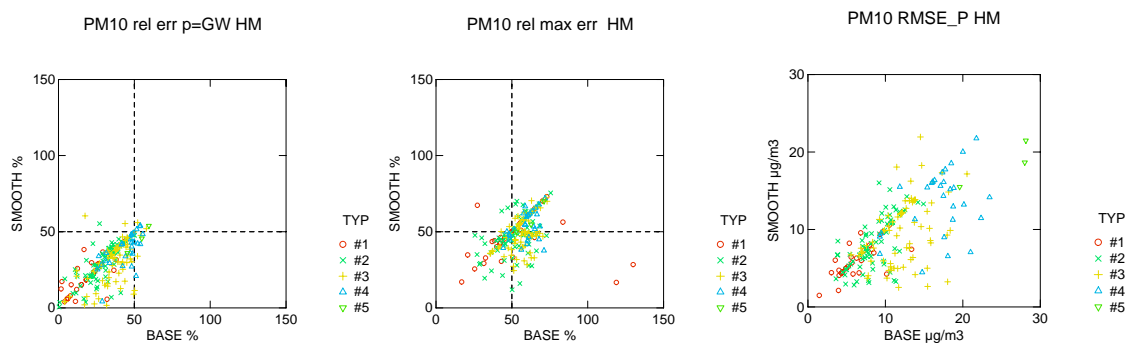


Abbildung 33 Relativer Fehler beim Perzentil des Zielwertes (rel per err_p_{GW}, links), relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und RMSE mit Perzentil-Bezug (rechts). Stationsbezogen mit direktem Vergleich (Base) und räumlicher Mittelung der Messung in der Auflösung des Modells (SMOOTH) für die Tagesmittelwerte von PM10.

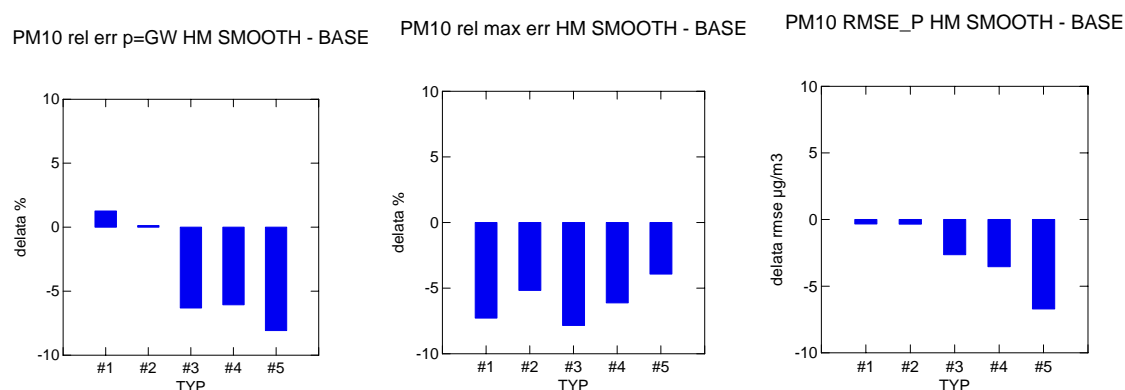


Abbildung 34 Mittlere Veränderung pro Immissionsregime des relativen Fehlers beim Perzentil des Zielwertes (rel per err_p_{GW}, links), des relativen maximalen Fehlers mit Perzentil-Bezug (rel max_err_p, Mitte) und des RMSE mit Perzentil-Bezug (rechts) zwischen direktem Vergleich und räumlicher Mittelung der Messung in der Auflösung des Modells für die Tagesmittel von PM10.

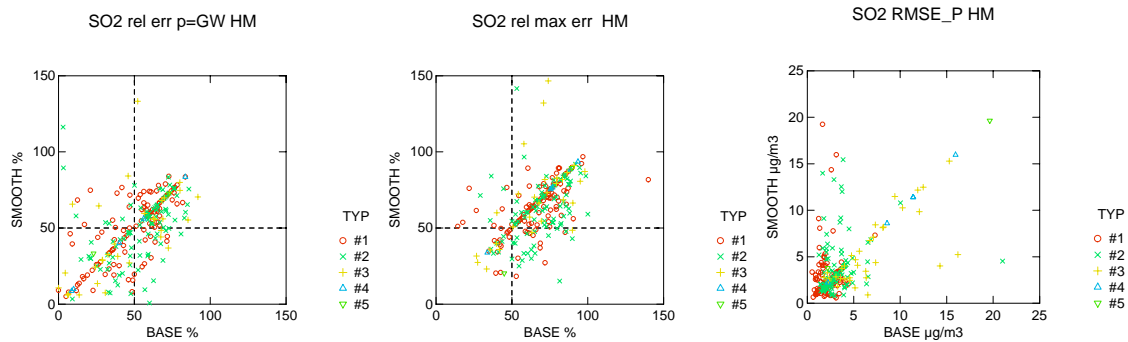


Abbildung 35 Relativer Fehler beim Perzentil des Zielwertes (rel per err_{p_{GW}}, links), relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und RMSE mit Perzentil-Bezug (rechts). Stationsbezogen mit direktem Vergleich (Base) und räumlicher Mittelung der Messung in der Auflösung des Modells (SMOOTH) für die Stundenwerte von SO₂.

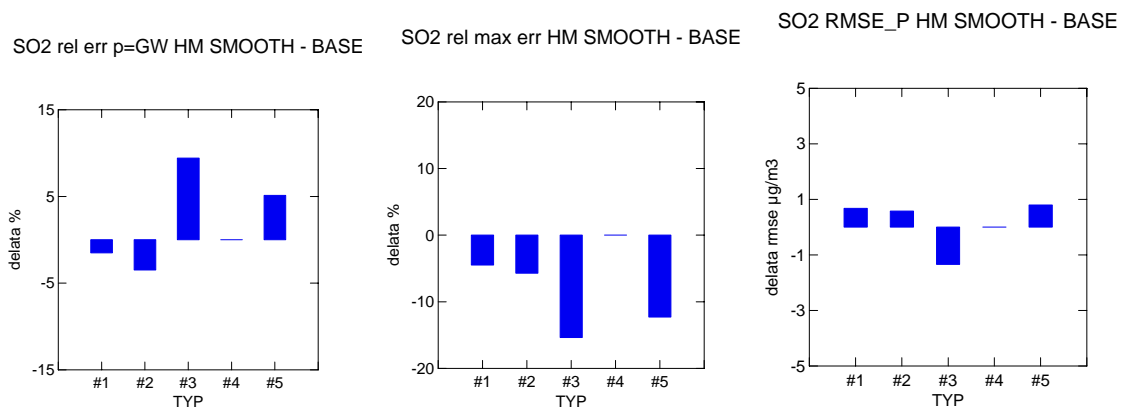


Abbildung 36 Mittlere Veränderung pro Immissionsregime des relativen Fehlers beim Perzentil des Zielwertes (rel per err_{p_{GW}}, links), des relativen maximalen Fehlers mit Perzentil-Bezug (rel max_err_p, Mitte) und des RMSE mit Perzentil-Bezug (rechts) zwischen direktem Vergleich und räumlicher Mittelung der Messung in der Auflösung des Modells für die Stundenwerte von SO₂.

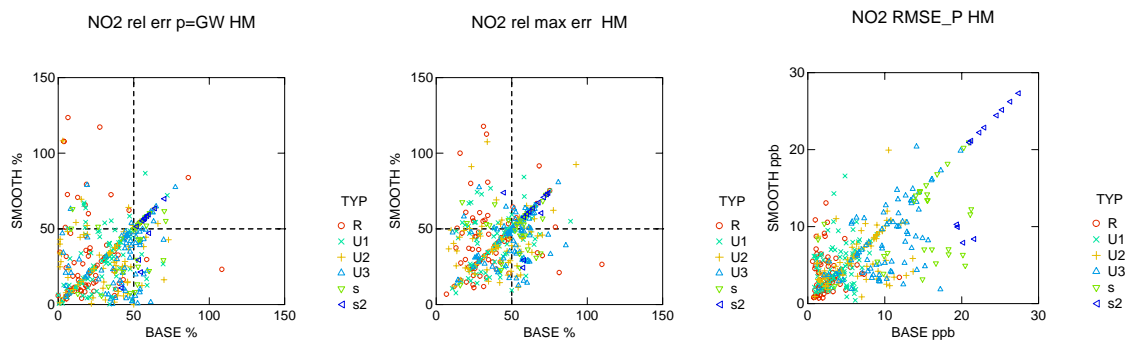


Abbildung 37 Relativer Fehler beim Perzentil des Zielwertes (rel per err_{p_{GW}}, links), relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und RMSE mit Perzentil-Bezug (rechts). Stationsbezogen mit direktem Vergleich (Base) und räumlicher Mittelung der Messung in der Auflösung des Modells (SMOOTH) für die Stundenwerte von NO₂.

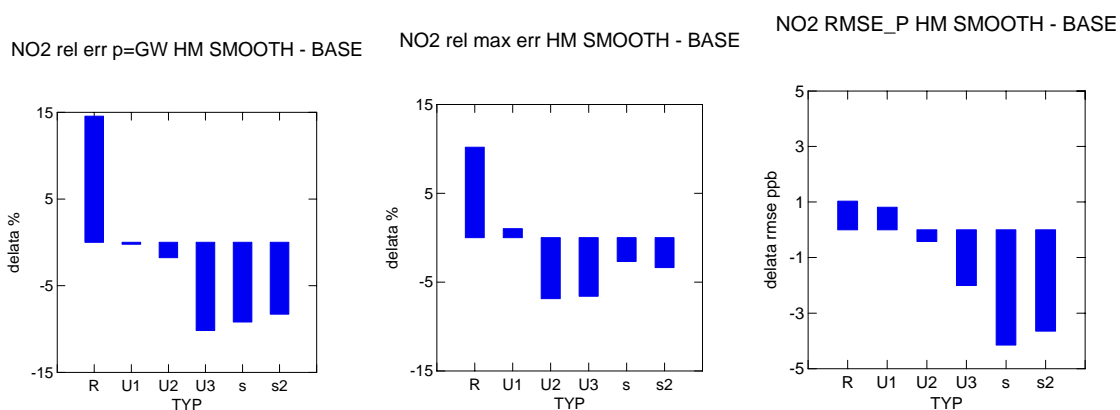


Abbildung 38 Mittlere Veränderung pro Immissionsregime des relativen Fehlers beim Perzentil des Zielwertes (rel per err_p_{GW}, links), des relativen maximalen Fehlers mit Perzentil-Bezug (rel max_err_p, Mitte) und des RMSE mit Perzentil-Bezug (rechts) zwischen direktem Vergleich und räumlicher Mittelung der Messung in der Auflösung des Modells für die Stundenwerte von NO₂.

5 Beobachtungsfehler und der Auflösungsfehler des „perfekten“ Modells

In diesem Kapitel wird der bereits in Kap. 3.4 angesprochene Gedanke über eine Begrenzung der erreichbaren Modellgüte durch zwei Arten von „grundsätzlichen“ Fehlermöglichkeiten für die Nichtübereinstimmung von Modell und Messung aufgegriffen, die auch bei einem hypothetisch perfekten Modell auftreten würden. Sie haben vergleichbare Ursachen und sind auf die Problematik der unterschiedlichen räumlichen Repräsentativität von Beobachtung und Modell zurückzuführen. Es handelt sich dabei um

- den Beobachtungsfehler, und
- den Auflösungsfehler des Modells

Der Beobachtungsfehler ist eine Größe aus der Datenassimilation und geostatistischen Interpolationsverfahren. Er wird als die räumlich nicht korrelierte Komponente der Messung definiert. Im Rahmen dieser Theorie und in den Ansätzen zu seiner Schätzung wird er als ein statistisches Rauschen mit verschwindendem Mittelwert angesehen. In der Praxis ist der Fehler der Konzentrationsmessung allerdings noch zusätzlich systematisch, d.h. mit einem bestimmten von Null verschiedenen Mittelwert versehen. Weiterhin kann mangelnde Vollständigkeit der Messung eine Fehlerursache sein. In Annex VIII der Tochterrichtlinien ist die geforderte Mindestgenauigkeit von Messungen angegeben. Für die kontinuierlichen Messungen von Ozon, NO₂ und SO₂ sind mindestens 15% und für PM₁₀ mindestens 25%-Genauigkeit in der Nähe des jeweiligen Grenzwertes gefordert (siehe auch Kap. 3.5).

Die räumliche Repräsentativität eines Modells kann sehr einfach mit der räumlichen (horizontalen) Auflösung quantifiziert werden. Da Eulersche Chemie-Transport Modelle immer mit Gitterboxmittelwerten arbeiten und auch alle Eingangsdaten nur als Mittelwerte über die Auflösung verwendet werden können, muss es zu Unterschieden hinsichtlich der punktbezogenen Messungen kommen, selbst wenn das Modell in seiner Beschreibung der physikalisch/chemischen Prozesse „perfekt“ wäre und auch die Eingangsdaten fehlerfrei wären. Will man diesen „Auflösungsfehler“ abschätzen, so müssen mehr als eine Station in der Gitterbox liegen. Die Schätzung des Auflösungsfehlers hängt demzufolge von der vorliegenden Stationsdichte ab. Eine geringe Stationsdichte, d.h. z.B. nur eine oder gar keine Station in der Gitterbox, bedeutet jedoch nicht, dass der Auflösungsfehler nicht existiert sondern nur, dass er nicht abgeschätzt werden kann.

Das Problem der Repräsentativität wird häufig diskutiert, eine Quantifizierung des Fehlers wird jedoch meist nicht gegeben. Ansätze wurden im Rahmen einer Anwendung des RCG-Modells für die Datenassimilation entwickelt (Flemming 2003a, 2003c). Diese Ansätze werden hier zusammenfassend wiedergegeben und erweitert.

Die Abschätzung der beiden genannten Fehlergrößen ist für alle Modellevaluierungsvorhaben von enormer Bedeutung, da sie eine untere Schranke für den praktisch ermittelbaren Fehler darstellen, denn auch ein hypothetisch perfektes Modell mit einer gewissen Auflösung würde keine geringeren Fehler erzeugen können. Aus diesem Schwellencharakter heraus können Beobachtungs- und Auflösungsfehler als ein problembezogenes Maß für die praktisch erreichbare Modellgenauigkeit betrachtet werden. Die Forderungen für die Modellgüte sollten mit Hilfe dieses Maßes abgeglichen werden.

5.1 Beobachtungsfehler - Rauschen

5.1.1 Varianz des Beobachtungsfehlers

Der Beobachtungsfehler wird als eine räumlich nicht korrelierte Komponente der Beobachtung mit verschwindendem Erwartungswert angesehen (siehe Flemming 2003a, 2003c), der für die Spezies, Tagesstunde und Halbjahr sowie für jede Station ermittelt wird. Dazu wurde von Flemming (2003a, 2003c) ein homogenes lokales Kovarianzmodell für jede Station aus den Kovarianzen zwischen den betrachteten Stationen und den Nachbarstationen abgeleitet. Durch den Vergleich der totalen empirischen Varianz der Station mit der lokalen, durch das Kovarianzmodell ermittelten Varianz des glatten Feldes kann die Varianz des Beobachtungsfehlers geschätzt werden.

Mit diesem klimatischen Ansatz können stationsspezifische Fehler ermittelt werden. Für jede Station ist aus der Zeitreihenauswertung der Beobachtungssinkremente die Gesamtvarianz bekannt. Die Aufgabe besteht nun in einer möglichst guten Schätzung der Varianz des Modellfehlers bzw. des glatten Feldes am Stationsort. Hierzu wird ein lokales homogenes und isotropes Kovarianzmodell auf Basis der empirischen Kovarianzwerte mit der betrachteten Station in einem Radius von 100 km herangezogen (Abbildung 39).

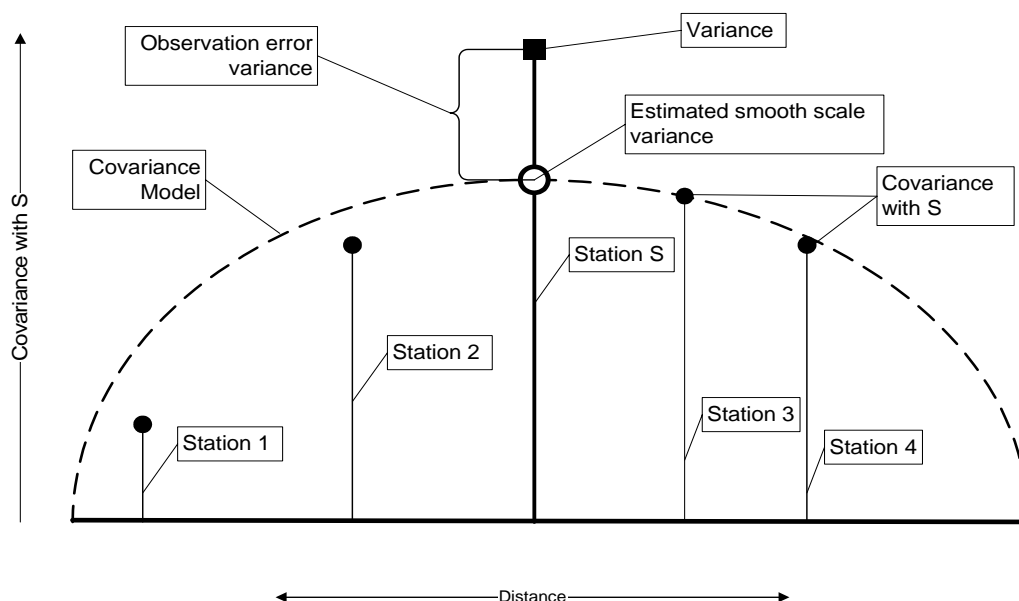


Abbildung 39 Schätzung der Varianz des klimatischen Beobachtungsfehlers (observation error variance) der Station S mit Hilfe eines Kovarianzmodells nach der Beobachtungsmethode.

5.1.2 Die stationsbezogenen Beobachtungsfehlervarianzen für deutsche Messstationen

Die Gesamtheit der stationsbezogen bestimmten Standardabweichungen, d. h. die Wurzel der Varianz, der Beobachtungsfehler aller betrachteten Spurenstoffe, ist in Abbildung 40 zu finden. Um eine relative Angabe zu erhalten, wurden die Standardabweichungen mit dem

Mittelwert normiert. Für PM10 sind an allen Tagesstunden die geschätzten stationsbezogenen Beobachtungsfehler deutlich größer als für NO₂, Ozon oder SO₂. Die stündliche Variabilität von PM10 ist sehr hoch und dementsprechend werden große Fehler geschätzt. Da PM10 keinen ausgeprägten Tagesgang, aber eine große Tagesschwankung besitzt (Flemming, 2003b), ist es im Sinne der räumlichen Repräsentativität empfehlenswert, die stündlichen Werte durch das Tagesmittel zu ersetzen. Die zugehörigen Beobachtungsfehler sind dann deutlich geringer. Dies entspricht dem Ansatz der EU-Richtlinien, die für PM10 als feinste Auflösung den Tagesmittelwert verlangen.

Der Zusammenhang zwischen der Standardabweichung des Beobachtungsfehlers und dem Immissionsregime, d. h. der typischen Belastung, ist in Tabelle 6 dargestellt. Für die primär emittierten Spurenstoffe, d. h. mit Ausnahme von Ozon, steigt die Standardabweichung der Beobachtungsfehler mit zunehmender mittlerer Immission an. Bei PM10, SO₂ und NO bleibt der relative, d. h. auf die mittlere Belastung bezogene Fehler, nahezu konstant in allen Regimes; bei NO₂ sinkt der relative Fehler leicht mit zunehmender Belastung.

Stoff\Regime:	B	R / 1	U1 / 2	U2 / 3	U3 / 4	S / 5	S2	Tagesgang
O ₃ (ppb)	8	6	6	6	6	6		schwach
NO ₂ (ppb)		3	4	5	6	7	8	schwach
PM10(µg/m ³)		8	8	10	12	15		Schwach
SO ₂ (µg/m ³)		2	3	4	6	18		Schwach

Tabelle 6 Geschätzte Standardabweichung des klimatischen Beobachtungsfehlers der stündlichen Messung, dargestellt als Median aller Stationen innerhalb der Immissionsregimes nach Fleming (2003b) (B = Berg, R = Land, U1 = Vorstadt, U2 = Stadt, U3 = belasteter Stadt, S = Straße, S2 = Straße extrem)

Der Ausgangspunkt für die klimatische Kovarianzmodellierung sind separate Kovarianzmodelle für jede Tagesstunde. Betrachtet man die geschätzten Standardabweichungen des Beobachtungsfehlers für jede Tagesstunde, so stellt man bei den primären Stoffen eine Übereinstimmung zwischen dem Tagesgang der Immission und dem des Beobachtungsfehlers fest. Interessant ist jedoch, dass für Ozon in der Zeit der höchsten Konzentrationen eine leichte Abnahme des absoluten und eine große des relativen klimatischen Beobachtungsfehlers festzustellen ist. Offensichtlich ist bei der nachmittäglichen Ozonbildung und guter Durchmischung die räumliche Repräsentativität größer als in den nächtlichen Stunden, wo lokale NO-Emission und Depositionsprozesse die Ozonkonzentration steuern.

5.1.3 Vergleich der Beobachtungsfehlervarianzen mit der Forderung von 15% Messgenauigkeit in den EU-Richtlinien

In der Rahmenrichtlinie werden relative Mindestgenauigkeiten für die Einzelmessung angegeben. Sie betragen für NO₂, SO₂ und O₃ 15% und für PM10 25 % (siehe Tabelle 3). Bei

dem in den Richtlinien angegebenen Vertrauensbereich von 95% entspricht diese Spanne dem zweifachen der Standardabweichung des Messfehlers. Der Messfehler in der Rahmenrichtlinie ist ein biasfreier Instrumentenfehler.

In Abbildung 40 wird die Standardabweichung der geschätzten Beobachtungsfehler in Verhältnis zum Mittelwert gesetzt, um ein relatives Maß zu erhalten. Dieses Maß kann jedoch nur mit Einschränkungen als der relative Beobachtungsfehler angesehen werden, da die Normierung nicht mit dem jeweiligen Messwert, sondern nachträglich mit dem Jahresmittelwert erfolgt. Nur mit der Annahme eines in allen Immissionsbereichen konstanten relativen Beobachtungsfehlers könnte diese Schlussfolgerung gezogen werden. Der Beobachtungsfehler setzt sich aus dem Instrumentenfehler und dem räumlichen Repräsentativitätsfehler zusammen. Seine relativen Werte schwanken im Bereich von 20–50% und sind damit deutlich größer als die geforderte Mindestgenauigkeit, die einer Standardabweichung des Instrumentenfehlers von 7% bzw. 12 % (PM10) entspricht.

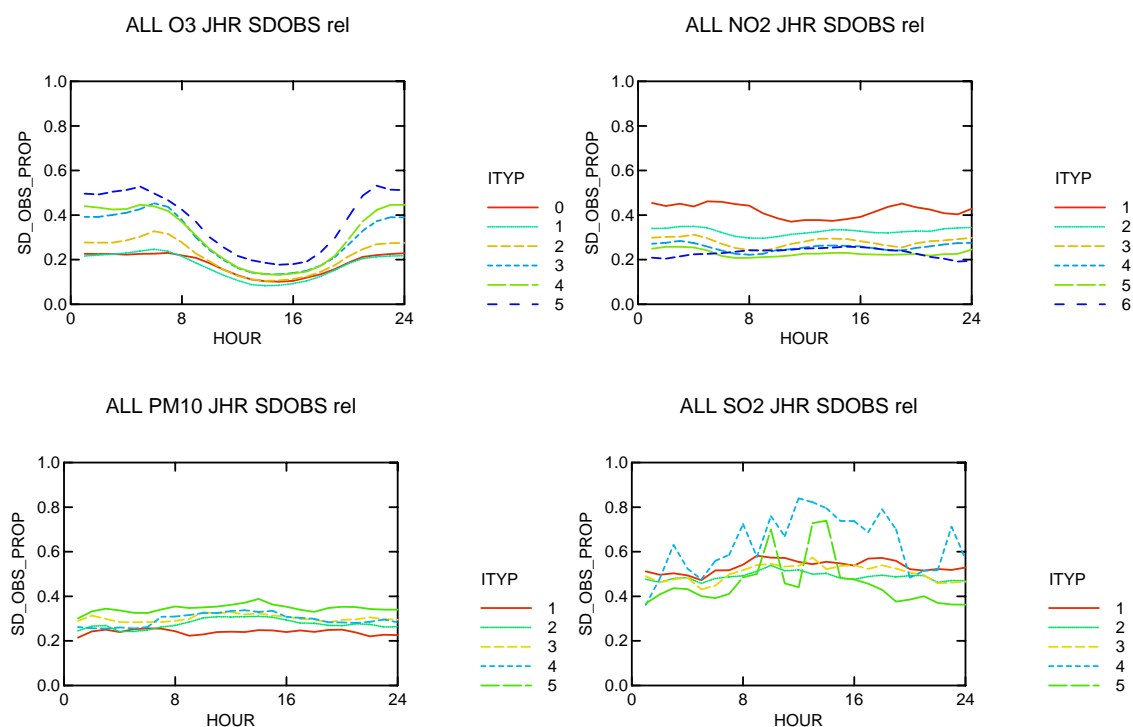


Abbildung 40 Tagesgang der relativen Standardabweichung des Beobachtungsfehlers (normiert mit dem Mittelwert) von O₃, NO₂, PM10 und SO₂ für verschiedene Immissionsregimes im Jahr 2002 in Deutschland in 100% (=1). (O₃ (0 bis 5) und NO₂ (1 bis 6) B: Berg/Küste, U1: Vorstadt U2: Stadt, U3: belastete Stadt, S: Strasse, S2: Strasse extrem, 1-5: Immissionsregime mit wachsender Verschmutzung für SO₂ und PM10)

5.1.4 Einfluss des erlaubten 10% Datenausfalls

Die EU-Richtlinien fordern eine Mindestdatenerfassung von 90% bei der stündlichen Messung. Demzufolge ist ein Datenausfall von mehr als einem Monat gestattet, unabhängig davon, ob der Datenausfall zusammenhängend oder zu verschiedenen Zeitpunkten auftritt. Durch den Datenausfall verringert sich das Perzentil, das einer Überschreitung von einer gewissen Anzahl von Tagen oder Stunden entspricht, da die Gesamtanzahl der Messungen verringert ist.

Für die Modellgütebewertung stellt sich die Frage, wie der Datenausfall behandelt werden soll, da durch die Modellierung im Allgemeinen immer vollständige Zeitreihen geliefert werden. Es ergeben sich die folgenden Möglichkeiten zur Behandlung der Fehlwerte, deren Einfluss auf die ermittelte Modellgüte im Folgenden untersucht wird:

1. Entfernen der entsprechenden Modellwerte bei fehlender Messung mit Zeitbezug
2. keine Ersetzung der Fehlwerte, betrachten der Überschreitungsanzahl und Perzentile
3. Ersetzen der Fehlwerte der Messung mit den Werten der Modellierung

Der 1. Ansatz ist einfach zu realisieren, denn durch die gleiche Anzahl der Modell- und Messwerte sind der Zeitbezug und die Perzentile immer eindeutig vergleichbar. Bei der Modellgütebestimmung im Perzentil-Bezug ist es jedoch fragwürdig, die Modellwerte in den Zeiten fehlender Messung zu entfernen, da gerade diese Modelldaten zu einer besseren Übereinstimmung der Häufigkeitsverteilungen von Messung und Rechnung im Bereich der hohen Werte führen können.

Aus diesem Grund empfiehlt sich für die Modellgütebewertung im Perzentil-Bezug der 2. Ansatz, der die Gesamtheit der vorliegenden Mess- und Modelldaten nutzt. Es ergeben sich jedoch technische Probleme, da der Perzentilbezug nicht mehr mit dem Vergleich von je einem Mess- und Modellwert erfolgen kann. Das Perzentil eines Modellwertes muss dann unter Umständen durch die Interpolation von zwei Messwerten gebildet werden. Gerade im Bereich der hohen Perzentile, wo der Abstand zwischen den Werten schnell anwachsen kann, könnten dadurch bemerkbare Unterschiede bei der Wahl des maximalen Fehlers mit Perzentil-Bezug entstehen. Unproblematisch ist jedoch das Ermitteln des Konzentrationswertes, der der erlaubten Häufigkeit der Grenzwertüberschreitung entspricht. Hierfür werden einfach die um die Datensatzlängen korrigierten Perzentile verwendet, was bedeutet, dass z.B. bei 3 erlaubten Überschreitungen jeweils der 4. höchste Wert miteinander verglichen wird.

Der dritte Ansatz dient der Abschätzung des größtmöglichen Einflusses einer nicht vollständigen Messdatenerfassung auf die Modellgütebewertung. Bei zusammenhängendem Datenausfall ist es durchaus möglich, dass es sich dabei um eine Periode handelt, die vom Modell besser erfasst werden kann bzw. Modellwerte in einem Wertebereich ermittelt, der den Vergleich mit dem Histogramm der Beobachtung verbessert.

Um den Einfluss des Datenausfalls der Messwerte für die Modellgütebestimmung zu ermitteln, werden die relativen maximalen Fehler mit Perzentil-Bezug und die relativen Fehler bei dem Perzentil, das der erlaubten Grenzwertüberschreitungshäufigkeit entspricht, für die Ansätze 1 und 3 verglichen. Ansatz 1 ist bei den bisherigen Ausführungen verwendet worden. Ansatz 3 lässt die stärkste Veränderung der Fehlermaße vermuten. Die Gegenüberstellung erfolgt mit Hilfe von Streudiagrammen für die Tages und Stundenwerte in Abbildung 41 bis Abbildung 44. Für die Untersuchung wurden die Messdaten aus Deutschland im Jahr 2002

verwendet. Wie aus Abbildung 45 zu erkennen ist, ist der Datenausfall einer Überzahl der Stationen sehr gering und liegt meist unter 2%. Es ist deswegen bereits absehbar, dass der Einfluss nicht sehr groß werden kann. Bei Datensätzen mit geringerer Qualität wird eine stärkere Beeinflussung zu vermuten sein. Auf die Konstruktion eines künstlichen Datensatzes mit einem Messdatenausfall von 10% wurde jedoch verzichtet.

Abbildung 41 bis Abbildung 44 zeigen, dass der reale Datenausfall im Jahr 2002 in Deutschland, der nach den EU-Richtlinien bis zu maximal 10% der Werte ausmachen darf, keinen großen Einfluss auf die ermittelten relativen Fehlermaße hat. Für einige Stationen verringert sich für Ansatz 3 der relative Fehler um ca. 5-10%, vereinzelt tritt eine Erhöhung des relativen maximalen Fehlers mit Perzentil-Bezug auf. Da Ansatz 3 bereits die stärkste Veränderung zu Gunsten des Modells darstellt, kann der praktizierte Ansatz 1 auch weiterhin ohne Bedenken verwendet werden.

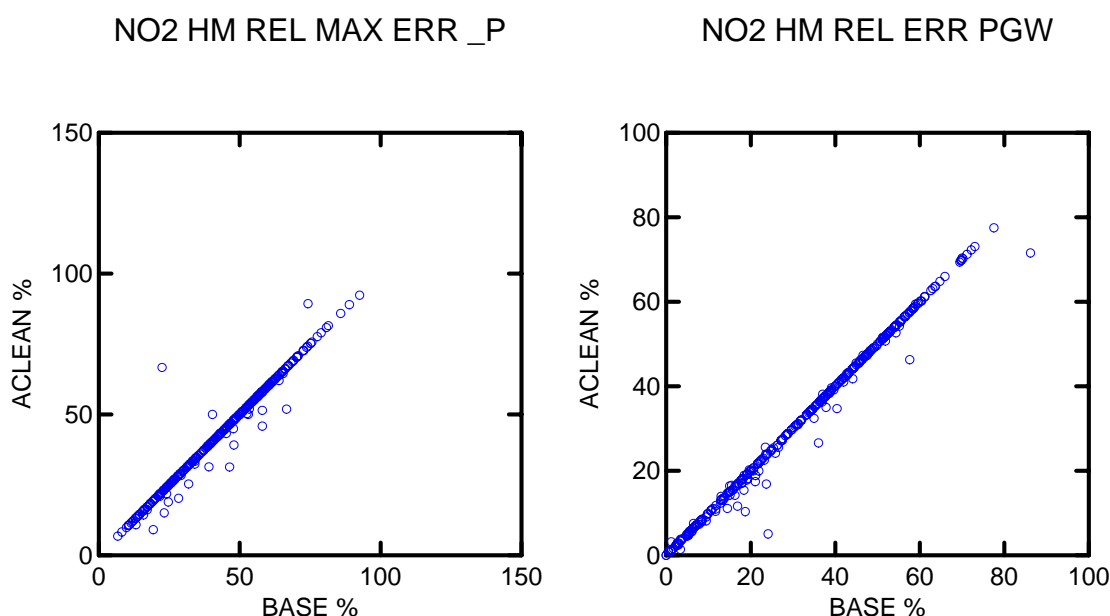


Abbildung 41 Relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, links) und relativer Fehler beim Perzentil der erlaubten Grenzwertüberschreitungshäufigkeit (rel per err_p_{GW}, rechts) bei unterschiedlicher Behandlung des Datenausfalls (< 10%) der Messwerte für die Stundenwerte von NO₂. BASE: Entfernen der Modellwerte mit Zeitbezug, ACLEAN: Ersetzen der fehlenden Messwerte durch die in diesem Zeitraum berechneten Werte.

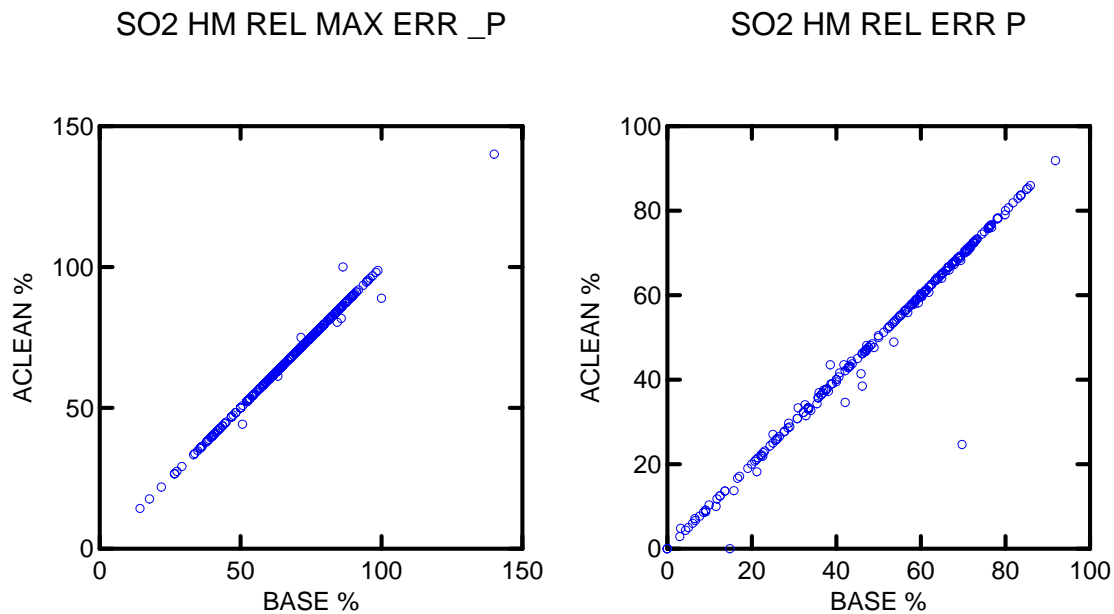


Abbildung 42 Relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, links) und relativer Fehler beim Perzentil der erlaubten Grenzwertüberschreitungshäufigkeit (rel per_err_p_{GW}, rechts) bei unterschiedlicher Behandlung des Datenausfalls (< 10%) der Messwerte für die Stundenwerte von SO₂. BASE: Entfernen der Modellwerte mit Zeitbezug, ACLEAN: Ersetzen der fehlenden Messwerte durch die in diesem Zeitraum berechneten Werte.

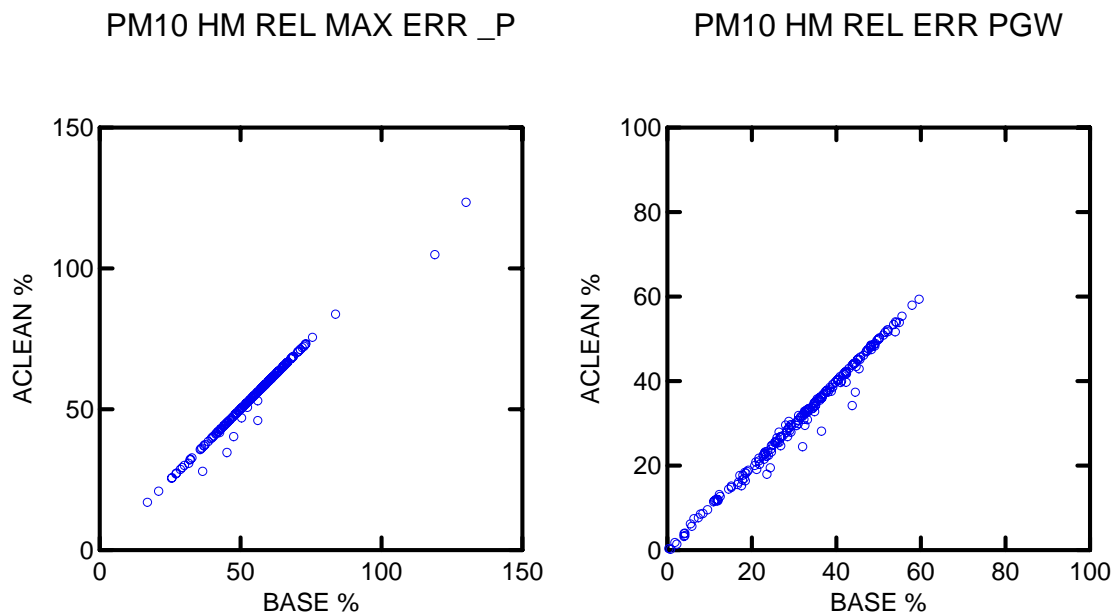


Abbildung 43 Relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, links) und relativer Fehler beim Perzentil der erlaubten Grenzwertüberschreitungshäufigkeit (rel per_err_p_{GW}, rechts) bei unterschiedlicher Behandlung des Datenausfalls (< 10%) der Messwerte für die Tageswerte von PM₁₀. BASE: Entfernen der Modellwerte mit Zeitbezug, ACLEAN: Ersetzen der fehlenden Messwerte durch die in diesem Zeitraum berechneten Werte.

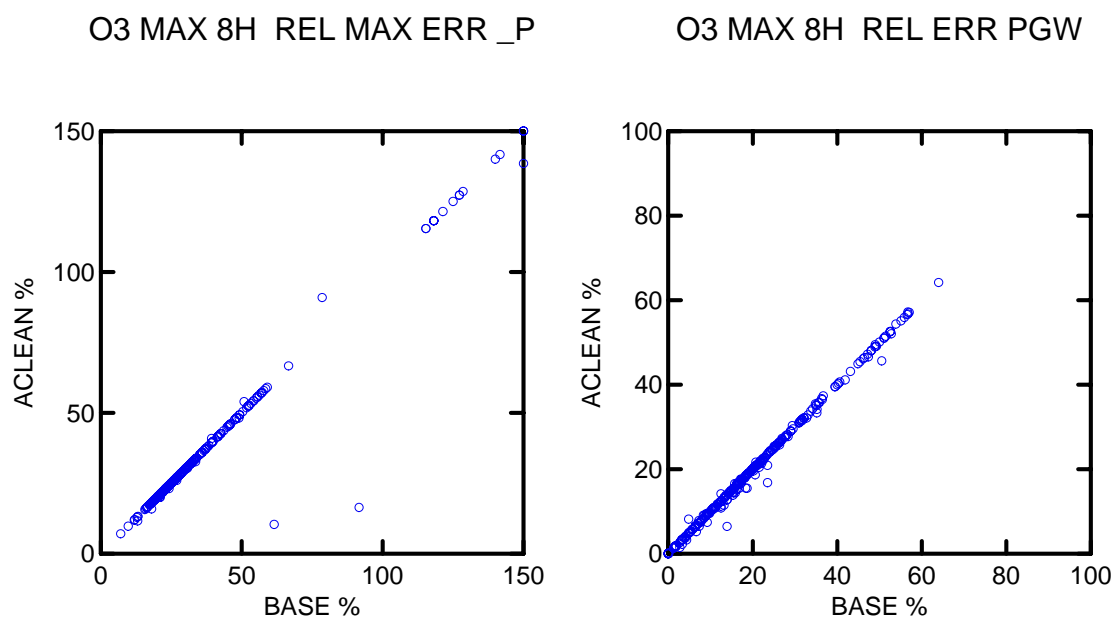


Abbildung 44 Relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, links) und relativer Fehler beim Perzentil der erlaubten Grenzwertüberschreitungshäufigkeit (rel per err_p_{GW}, rechts) bei unterschiedlicher Behandlung des Datenausfalls (< 10%) der Messwerte für die Tagesmaxima des 8 Stundenmittels von Ozon. BASE: Entfernen der Modellwerte mit Zeitbezug, ACLEAN: Ersetzen der fehlenden Messwerte durch die in diesem Zeitraum berechneten Werte.

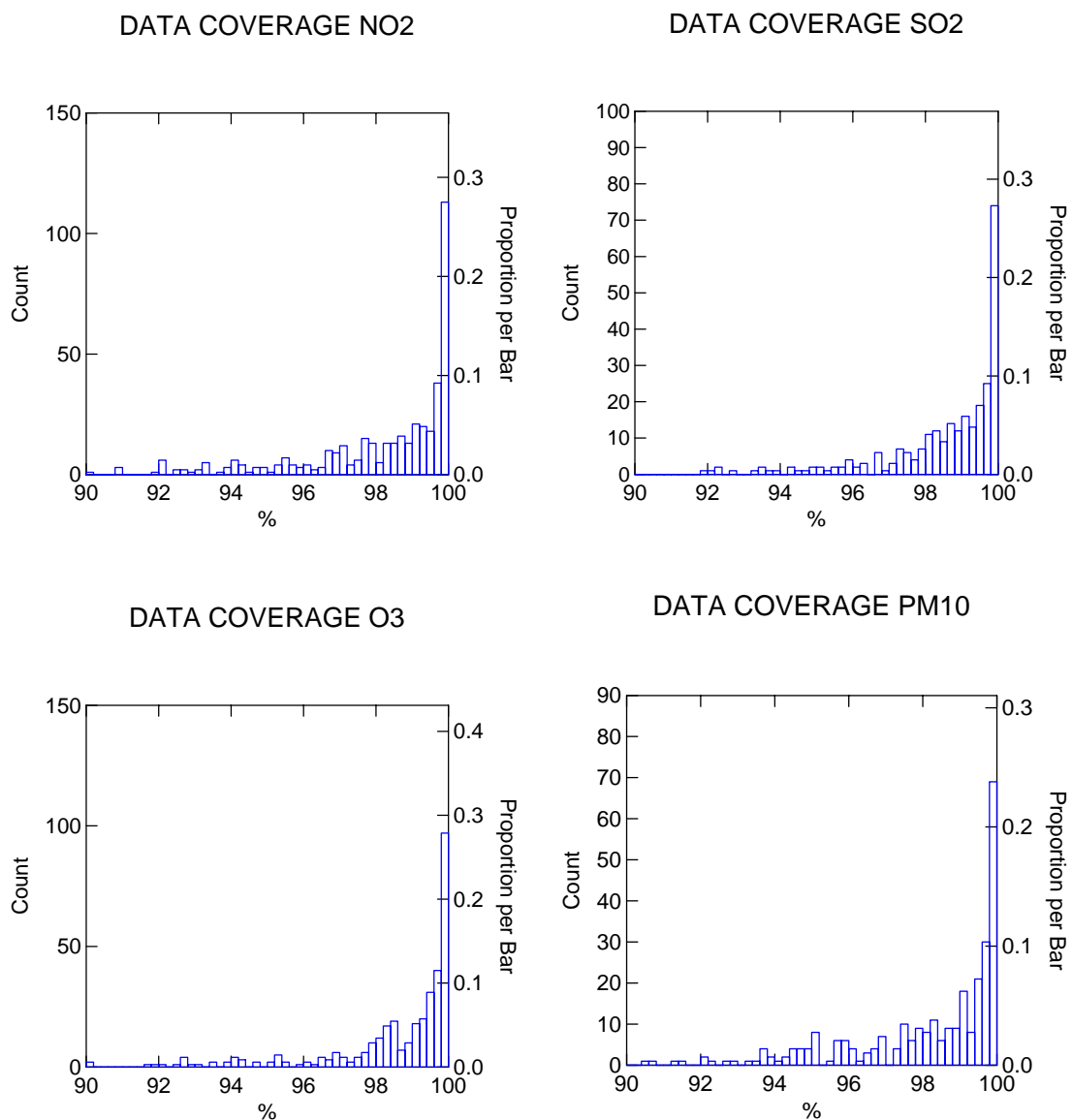


Abbildung 45 Histogramme der Vollständigkeit der stündlichen Beobachtungsreihen von NO₂, SO₂, O₃ und PM₁₀ an den Messstationen in Deutschland im Jahr 2002.

5.2 Modellauflösung

Eulersche Chemie Transport Modelle werden verwendet, um durch die Simulation der relevanten Prozesse (Emission, Ausbreitung, Chemische Umwandlung und Deposition) Felder der Immission zu gewinnen. Diese Felder werden durch ein Netz von Gitterboxen mit einer bestimmten Auflösung repräsentiert. Der modellierte Wert für die Gitterbox wird als ein Volumenmittelwert in der Box verstanden.

Die Auflösung des Modells hängt von dem Rechenaufwand und von der gegebenen Auflösung der Eingabedaten, d.h. der meteorologischen Felder und der Emissionsdaten, ab. Typische Auflösungen von Modellen für die verschiedenen Skalen sind in Tabelle 7 zusammengestellt. Je gröber die Auflösung ist, desto mehr Messorte können sich in einer Gitterbox befinden (siehe Abbildung 46). Es ist unwahrscheinlich, dass zwei Messreihen in einer Gitterbox vollständig übereinstimmen, da jede Station subskaligen Einflüssen unterworfen ist. Damit entsteht in Bezug auf die Messungen ein Modellfehler allein durch den Umstand, dass vom Modell nur ein Flächenmittelwert für das entsprechende Gebiet ermittelt wurde, aber mehrere, unterschiedliche Messreihen vorliegen.

Skala der Modellanwendung	Typische Gitterauflösung
Global	100 bis 300 km
Hemisphärisch	50 bis 100 km
Kontinental	25 bis 50
Ballungsgebiet	1 bis 10 km
Stadtquartier	100 m bis mehrere Dekameter
Straßenschlucht	1 bis mehrere Meter

Tabelle 7 Typische Gitterauflösungen für Modellanwendungen in verschiedenen Skalen

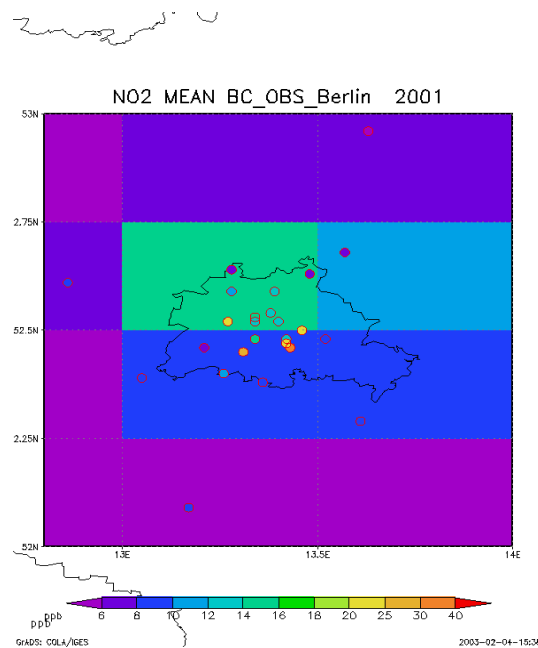


Abbildung 46 Berechnete Jahresmittelwerte von NO₂ (ppb) für die Modellgitterboxen des REM-CALGRID-Modells in und um Berlin (Brandenburg) und die beobachteten Werte an den Messorten. Gitterauflösung circa 25 km.

Im Folgenden wird versucht, den Fehler zu quantifizieren, der allein aufgrund der unterschiedlichen Modellauflösung entsteht. Dazu wird ein „perfektes“ Modell mit einer variablen horizontalen Auflösung konstruiert. Dieses „perfekte“ Modell „berechnet“ für jede Gitterbox eine Zeitreihe, die exakt der mittleren Zeitreihe entspricht, die sich durch die räumliche Mittelung aller Messstationen in der betrachteten Box ergibt. Der Fehler dieses perfekten Modells hängt von den Gradienten der Immissionsfelder ab, seine Größe wird jedoch wesentlich von der Dichte der Messung, d.h. der Auflösung¹⁰ des Messnetzes beeinflusst. Je dichter das Messnetz, desto größer wird der ermittelte Auflösungsfehler des perfekten Modells sein. Für Gitterboxgrößen kleiner als der geringste Abstand zwischen zwei Messstationen kann kein Auflösungsfehler mehr ermittelt werden. Das bedeutet aber nicht, dass das Problem der Modellevaluierung „Punktmessung vs. Modellgitterboxmittel“ dadurch nicht vorhanden wäre.

Das relativ dichte Messnetz der Bundesländer und des UBA bietet eine sehr gute Möglichkeit, den Auflösungsfehler des perfekten Modells abzuschätzen. Die Gebiete höherer Stationsdichte sind in den Ballungsgebieten mit hohen Konzentrationen von primären Spurenstoffen und den Gebieten mit den hohen Gradienten zu finden. Abbildung 47 und Abbildung 48 geben einen Eindruck von der Auflösung des Messnetzes in Deutschland, in dem die maximale Anzahl der in einer Gitterbox vorhandenen Messstationen und die Anzahl der Stationen, die in ihrer Gitterbox noch mindestens eine weitere Station aufweisen, bei verschiedenen Auflösungen gezeigt wird. Letztere Anzahl ist die Anzahl der Stationen, für die ein nichttrivialer, d. h. ein von Null verschiedener Auflösungsfehler des perfekten Modells be-

¹⁰ Die „Auflösung“ eines Messnetzes lässt sich durch die meist irreguläre Anordnung nicht ohne weiteres quantifizieren. Mögliche Lösungsansätze sind eine Histogrammdarstellung der Abstände zu jeweils nächsten Station.

stimmt werden kann. Es wird deutlich, dass sich bei dem derzeitigen Ausbau des deutschen Messnetzes erst ab 6 km Auflösung eine vertretbare Anzahl von auswertbaren Stationen, d.h. Stationen, die mindestens eine Nachbarstation in ihrer Gitterbox aufweisen, ergibt.

#Stationen number > 2

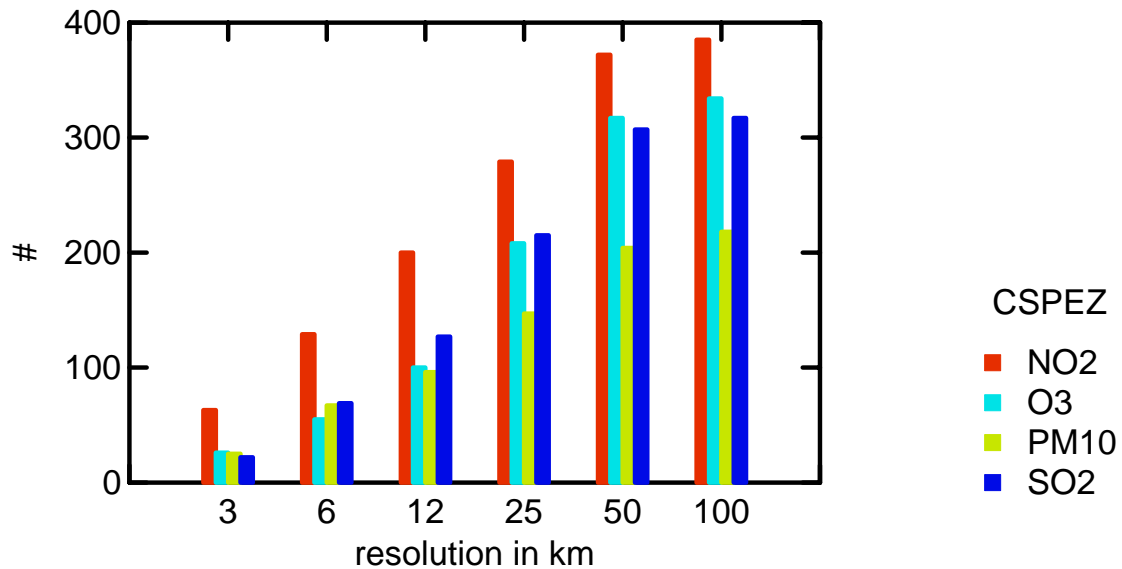


Abbildung 47 Anzahl der Stationen, die bei der gegebenen Auflösung (y-Achse) mindestens noch eine zweite Station in ihrer Gitterbox aufweisen. Beobachtungsnetz in Deutschland.

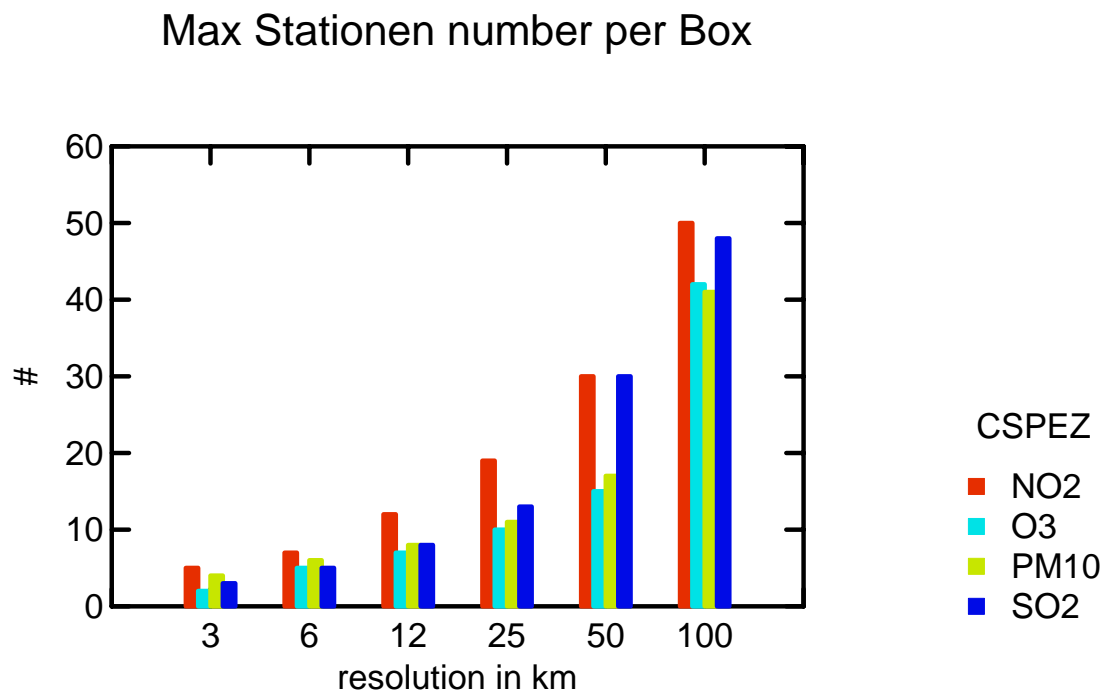


Abbildung 48 Maximale Anzahl von Stationen in Gitterboxen verschiedener Auflösung (y-Achse) . Beobachtungsnetz in Deutschland.

Es muss betont werden, dass die Schätzung des Auflösungsfehlers von der Stationsdichte abhängt und deren Auflösung eine untere Schranke für die Bestimmbarkeit darstellt. Wenn bei einer sehr geringen Messdichte keinen Auflösungsfehler bestimmbar ist, bedeutet dies jedoch keinesfalls, dass er nicht existiert.

Der Auflösungsfehler des perfekten Modells wurde für Ozon, NO₂, PM10 und SO₂ mit den Daten für 2002 für jeden möglichen Termin an allen Stationen bestimmt. Der Fehler wurde stationsbezogen mit den folgenden Kenngrößen für 6 verschiedene Auflösungen zusammengefasst.

1. RMSE mit und ohne Zeitbezug (RMSE, RMSE_p)
2. relativer maximaler Fehler mit Perzentil-Bezug (rel max err_p)
3. relativer Fehler beim Perzentil der erlaubten Grenzwertüberschreitung (rel per err_{GW})

Um den Auflösungsfehler mit dem realen Fehler der Modellrechnung zu vergleichen, werden zusätzlich die entsprechenden Fehlermaße der RCG-Modellrechnung mit ca. 25 km für 2002 dargestellt.

Die Gesamtheit der stationsbezogenen ermittelten Fehlermaße werden in Abbildung 49 bis Abbildung 52 mit Hilfe von Box-Whiskers-Plots dargestellt. Dafür werden jeweils alle verfügbaren Stationen verwendet, auch wenn sich in der umgebenden Gitterbox keine weitere Station befindet. Es ist aus den Abbildungen offensichtlich, dass sich der Auflösungsfehler mit Verkleinerung der Maschenweite von 100 auf 3 km in seiner Gesamtheit verringert, so dass bei 3 km Auflösung an den meisten Stationen kein Fehler mehr feststellbar ist. Der

größte Anteil an dieser Verringerung ist jedoch darauf zurückzuführen, dass bei kleinerer Auflösung immer weniger Stationen mit anderen in einer Gitterbox liegen (siehe Abbildung 47). Untersucht man nur die Stationen, die nicht allein in einer Gitterbox liegen, so ist die Verringerung deutlich geringer und macht bezogen auf die Fehler bei 100 km nur ca. 10-20% bei der Auflösung von 6 bzw. 3 km aus.

Vergleich man die Fehlermaße der Modellrechnung (R25) mit dem Auflösungsfehler, so liegt der Fehler der hier als Beispiel verwendeten RCG-Rechnung meist ungefähr im Größenbereich von 80-150 % des Auflösungsfehlers von 100 km. Der Median des Ensembles der stationsbezogenen Maße des Auflösungsfehlers bei 25 km ist ungefähr halb so groß wie der bei der Modellrechnung. Damit ist eine untere Schranke für den Modellfehler definiert, denn Modellverbesserungen können nur zu Fehlermaßen führen, die größer oder höchstens gleich („perfektes Modell“) dem Auflösungsfehler sind. Legt man die Modellgüteanforderung der EU-Richtlinien zu Grunde (rel max err_p kleiner als 50%), so wird klar, dass bei NO₂ und SO₂ aufgrund des inhärenten Auflösungsfehlers ca. 20 % und bei Ozon und PM10 ca. 5-10% der Messstationen niemals mit der erforderlichen Güte von einem Eulerschen Modell mit 25 km Auflösung simuliert werden können. Das Potential zur Modellverbesserung¹¹ hinsichtlich der untersuchten Fehlermaße, definiert durch den Unterschied zwischen dem Fehlermaß der aktuellen Modellrechnung und dem des Auflösungsfehlers eines „perfekten“ Modells, liegt bei einer Verringerung der derzeitigen Fehler um ca. 30 bis 50%. Diese Aussagen gelten für die hier verwendete Messnetzdicke.

¹¹ Entweder durch eine Verbesserung der Modellphysik und der Modellchemie oder durch eine Verbesserung der Qualität der Eingangsdaten

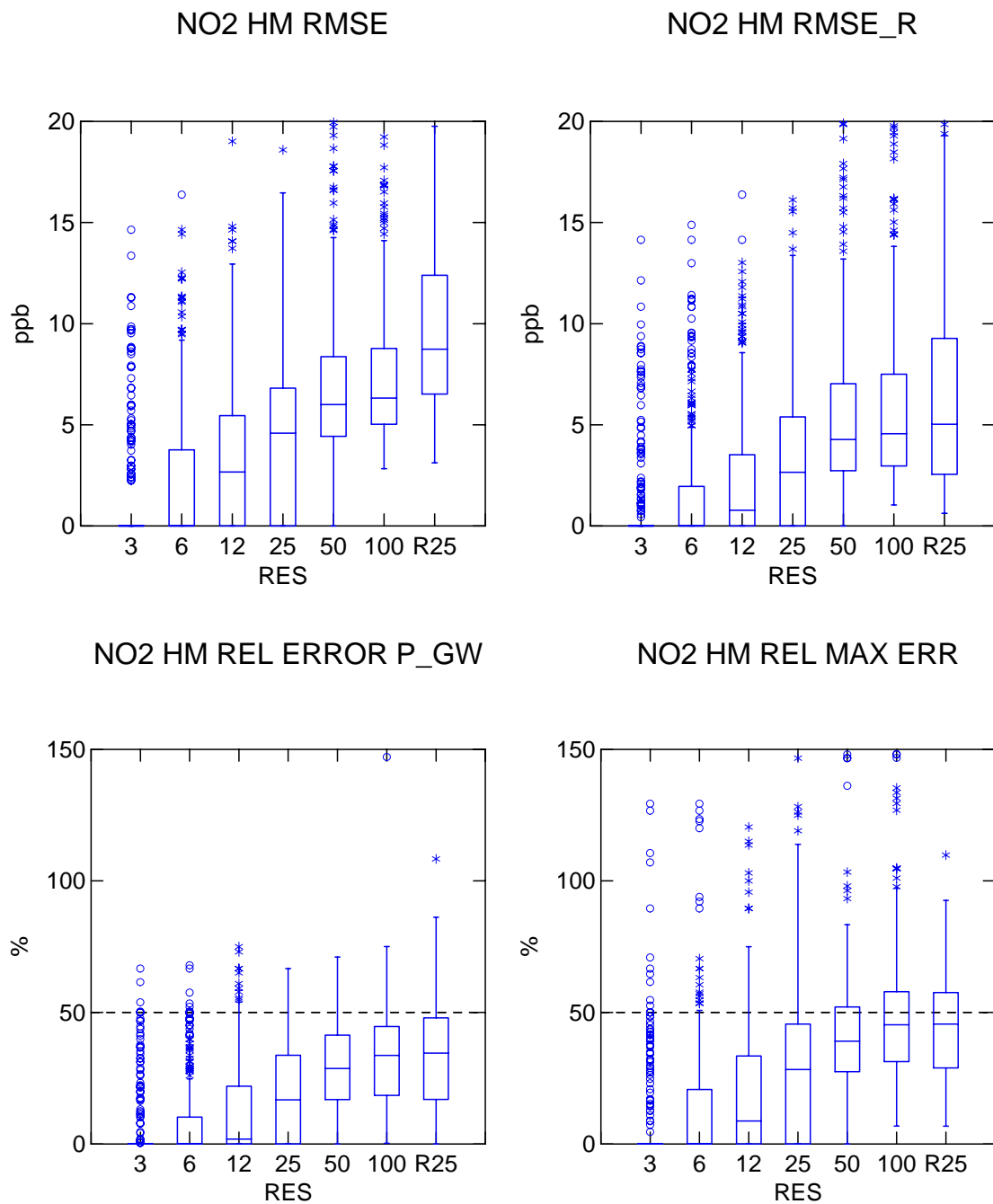


Abbildung 49 Box Whiskers Plots der stationsbezogenen Fehlermaße (RMSE, RMSE-p, rel per err_p_{GW}, rel max err_p) des perfekten Modells bei verschiedenen Auflösungen (RES in km) und der realen RCG-Modellrechnung mit 25 km Auflösung (R25) für NO₂-Stundenmittelwerte im Jahr 2002.

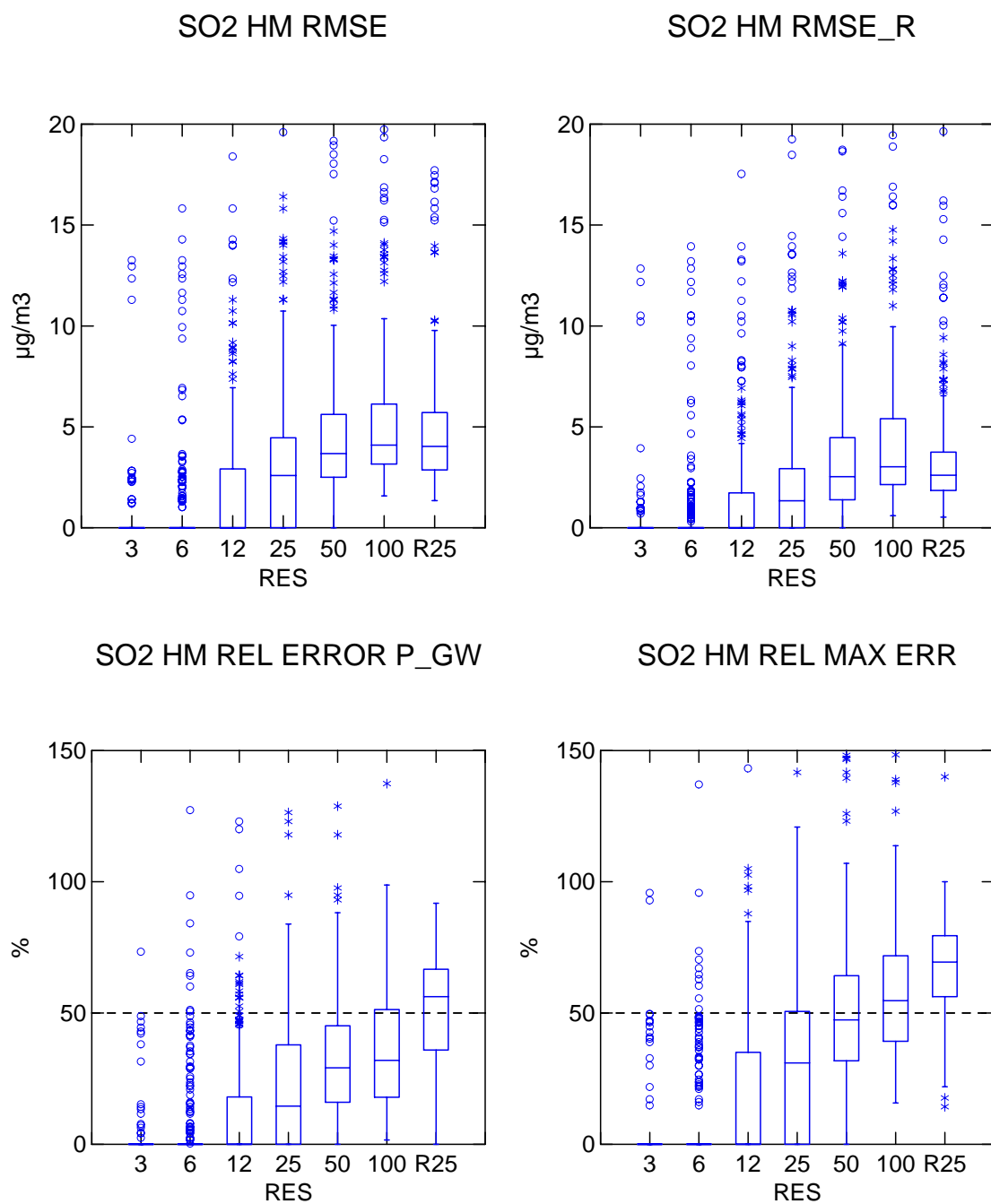


Abbildung 50 Box Whiskers Plots der stationsbezogenen Fehlermaße (RMSE, RMSE-p, rel per err_{pGW}, rel max err_p) des perfekten Modells bei verschiedenen Auflösungen (RES in km) und der realen RCG-Modellrechnung mit 25 km Auflösung (R25) für SO₂ – Stundenmittelwerte im Jahr 2002.

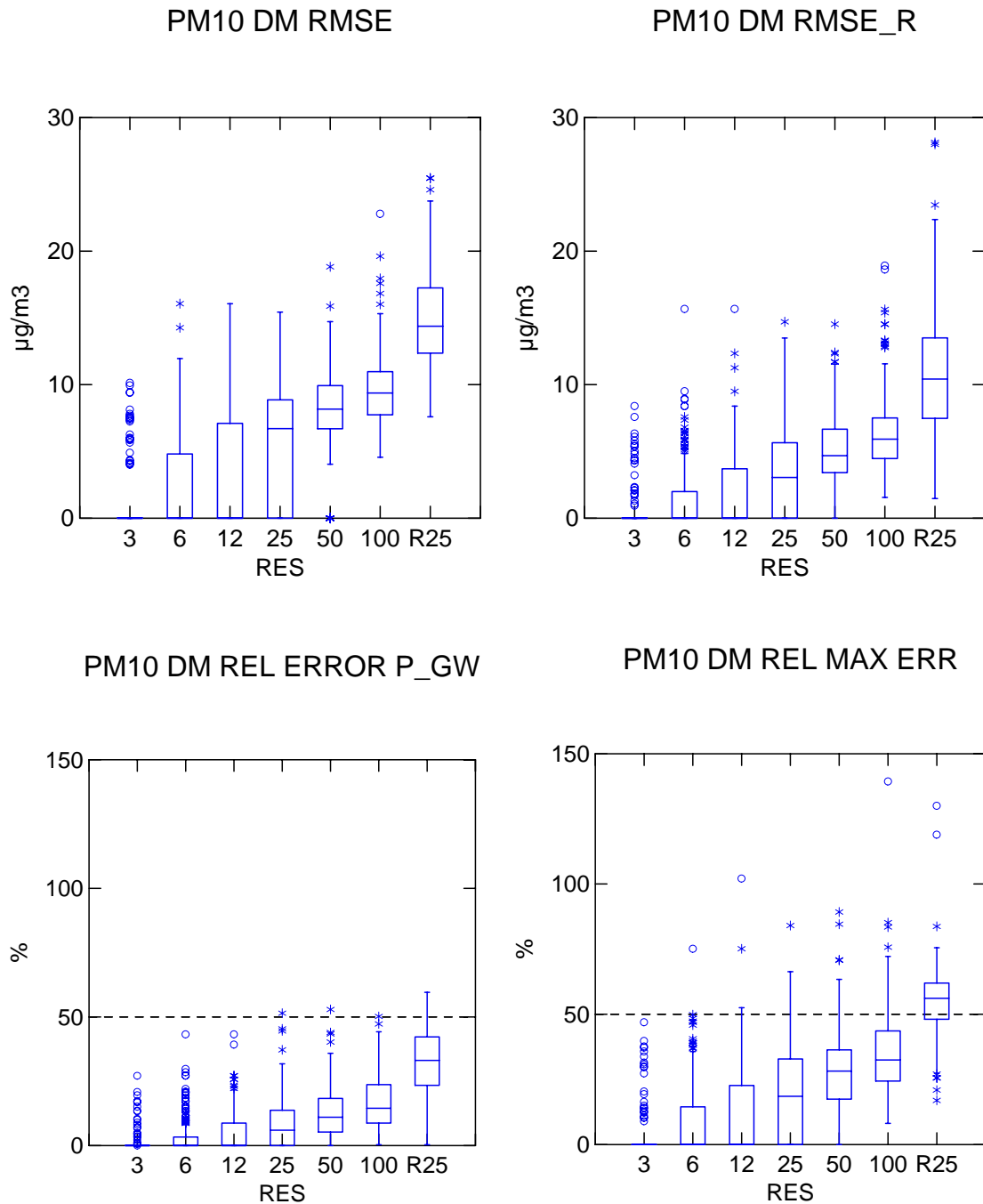


Abbildung 51 Box Whiskers Plots der stationsbezogenen Fehlermaße (RMSE, RMSE-p, rel per err_p_{GW}, rel max err_p) des perfekten Modells bei verschiedenen Auflösungen (RES in km) und der realen RCG-Modellrechnung mit 25 km Auflösung (R25) für PM10-Tagesmittelwerte im Jahr 2002.

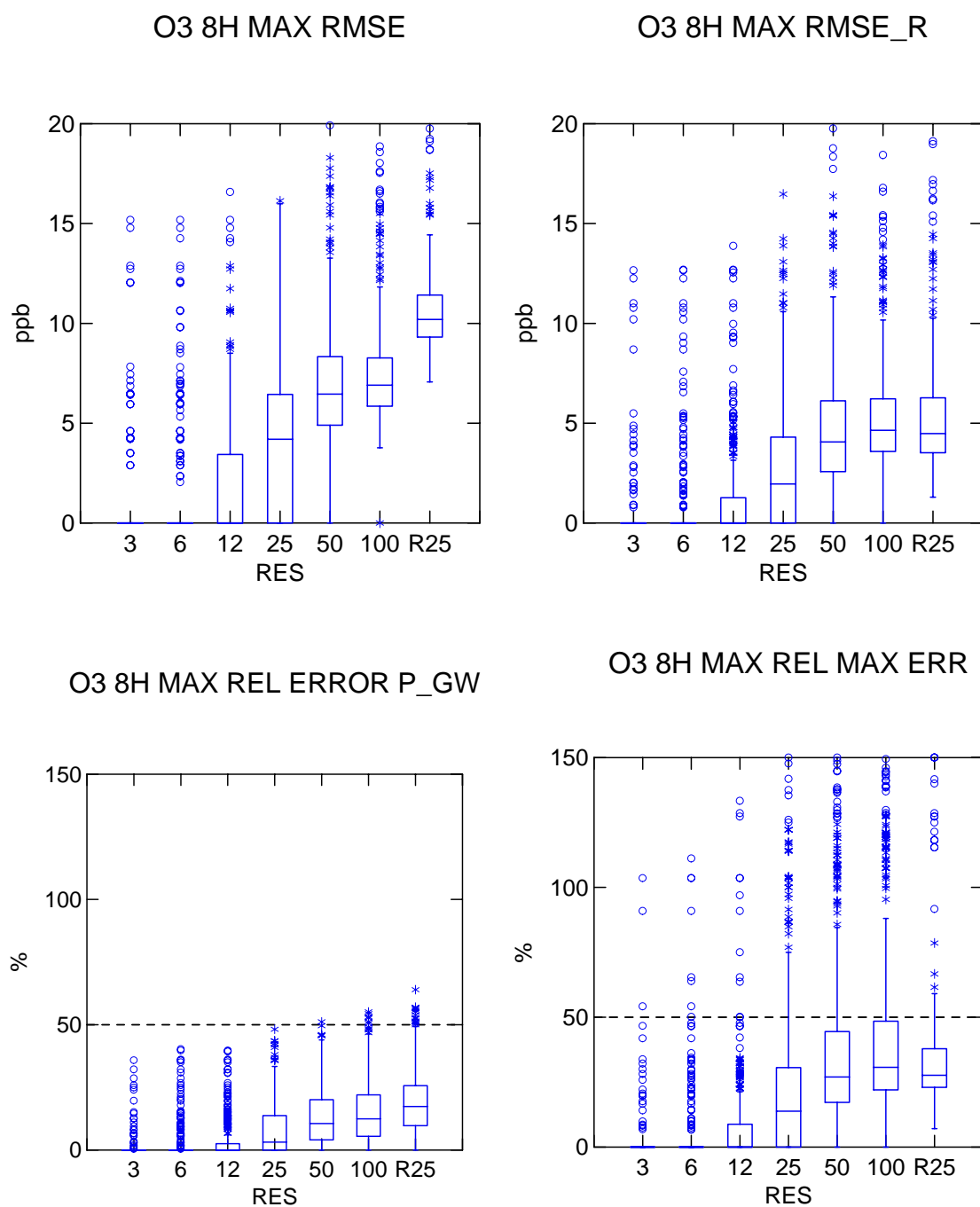


Abbildung 52 Box Whiskers Plots der stationsbezogenen Fehlermaße (RMSE, RMSE-p, rel per err_p_{GW}, rel max err_p) des perfekten Modells bei verschiedenen Auflösungen (RES in km) und der realen RCG-Modellrechnung mit 25 km Auflösung (R25) für Ozon Tagesmaxima des 8h Mittels im Jahr 2002.

6 Zusammenfassung

Die EU-Rahmenrichtlinie 96/62/EG über die Beurteilung und die Kontrolle der Luftqualität (EU, Amtsblatt der Europäischen Gemeinschaften, 1996) gestattet ausdrücklich den Einsatz von Ausbreitungsmodellen zur Beurteilung zur Luftqualität sowie implizit zur Erstellung von Maßnahmenplänen. Danach können Modelle eingesetzt werden

- zur räumlichen Darstellung der Schadstoffverteilung in allen Skalenbereichen (International, national, Ballungsgebiet, Straße), insbesondere in Gebieten, wo die Schadstoffkonzentrationen unter der jeweiligen Oberen Beurteilungsschwelle liegen,
- zur Prognose der Wirksamkeit von Maßnahmen.

Damit ein Modell für die Belange der EU-Richtlinien zur Luftqualität eingesetzt werden kann, muss es den in den Richtlinien definierten Datenqualitätsanforderungen genügen. Analog zur Messung werden Anforderungen an die Mindestgenauigkeit der Modellergebnisse gestellt, wie diese Anforderungen aber erfüllt werden müssen, ist aber nur sehr allgemein und nicht eindeutig formuliert.

In dieser Arbeit wurde daher versucht, die Vorgaben der EU-Richtlinien zu konkretisieren und Kriterien zu formulieren, die zum Nachweis der von den EU-Richtlinien verlangten Genauigkeit von Modellrechnungen benutzt werden können. Es wurden verschiedene Maße vorgeschlagen und anhand einer europaweiten Anwendung des chemischen Transportmodells REM-CALGRID (RCG) auf das gesamte Jahr 2002 und eines Vergleichs der Modellergebnisse mit Messungen aus den deutschen Messnetzen diskutiert. Der Schwerpunkt der Untersuchung lag in der Umsetzung der Genauigkeitsmaße für die Schadstoffe O₃, PM₁₀, SO₂ und NO₂. Das RCG-Modell wurde dabei lediglich zur Demonstration der Bedeutung verschiedener Fehlermaße für die Güte von Modellergebnissen benutzt. Die vorgeschlagenen Maße sind so formuliert, dass sie auf jedes Modell angewendet werden können, das in der Lage ist, die von den EU-Richtlinien geforderte Auflösung von einer Stunde bis zu mindestens einem Jahr zu liefern. Grundsätzlich sind die Gütemaße für Modellrechnungen in allen Skalenbereichen anwendbar, auch wenn sie hier nur für eine großräumige Anwendung diskutiert wurden.

Die Vorgehensweise zur Bestimmung der Modellgüte ist in den EU-Richtlinien in nur einem Satz festgehalten: „Die Genauigkeit von Modellberechnungen und objektiven Schätzungen ist definiert als die größte Abweichung der gemessenen und berechneten Konzentrationswerte über den betreffenden Zeitraum in Bezug auf den Grenzwert, ohne dass die zeitliche Abfolge der Ereignisse berücksichtigt wird“. Die Aufgabe der zeitlichen Zuordnung lässt vermuten, dass damit die Übereinstimmung in den Häufigkeitsverteilungen untersucht werden soll. Ausgehend von dieser Interpretation wurden hauptsächlich folgende Fehlermaße untersucht:

- maximaler relativer Fehler mit Zeitbezug (max rel err_t)
- maximaler relativer Fehler mit Perzentilbezug (max rel err_p)
- Relativer maximaler Fehler mit Zeitbezug (rel max err_t)
- Relativer maximaler Fehler mit Perzentilbezug (rel max err_p)
- Relativer Fehler bei dem Perzentil, das der erlaubten Überschreitungshäufigkeit des Grenzwertes entspricht (rel per err_pgw)

Alle Fehlermaße sind stationsbezogen definiert, d.h. beziehen sich auf beobachtete und berechnete Zeitreihen an demselben Ort. Werden die Abweichungen zwischen Modell und Beobachtung aus Daten des gleichen Zeitpunktes gewonnen, so werden die Fehler mit Zeit-Bezug ermittelt. Ordnet man stattdessen die Werte der Größe nach und bildet die Differenz immer zwischen den Werten gleicher Rangfolge, so spricht man von einem Gütemaß mit Perzentil-Bezug.

Insgesamt zeigt die Auswertung der Gütemaße, dass die Modellfehler mit Perzentil-Bezug deutlich kleiner sind als diejenigen mit Zeitbezug. Entsprechend den Intentionen der EU-Richtlinien wurde der Schwerpunkt der Untersuchung daher auf den Vergleich der Maße ohne Zeitbezug gelegt.

Entsprechend der hier vorgenommenen Interpretation ist das in den Richtlinien 1999/39/EC und 2002/3/EC verlangte Fehlermaß der größte absolute Fehler mit Perzentil-Bezug. Die Zeitreihe von Messung und Rechnung wird der Größe nach geordnet, danach wird die Differenz zwischen den Werten gleicher Rangfolge gebildet. Die größte Differenz ergibt den größten absoluten Fehler $\max \text{err}_p$ und normiert mit der Beobachtung den relativen Fehler rel max err_p .

Die Auswertung dieses Fehlermaßes zeigte, dass der maximale Fehler an vielen Stationen beim 100-Perzentil, d.h. dem jeweils höchsten gemessenen Stunden- oder Tageswert auftritt. Die Beurteilung der Modellgüte hängt damit unter Umständen von einem Extremwert ab, dessen Auftretenswahrscheinlichkeit sehr gering sein kann. Für diesen Wert besteht zudem eine gewisse Wahrscheinlichkeit, dass es sich um einen fehlerhaften Ausreißer handelt. Auf der anderen Seite sind die Güteanforderungen bei kleinen Messwerten, die vom Modell überschätzt werden, durch die Relativierung (Normierung mit Messwert) extrem hoch.

Als weiteres Fehlermaß wurde daher die Konzentrationsdifferenz bei demjenigen Perzentil betrachtet, das der erlaubten Anzahl von Überschreitungen der Kurzzeitgrenzwerte (Stunden- und Tagesmittelwerte) entspricht (Überschreitungsperzentil). Dieses Maß untersucht ebenfalls die Abweichungen der Modellrechnung im Bereich der Maximalwerte ist aber viel robuster als das Maß des größten absoluten Fehlers, da es nicht am höchsten Perzentil ansetzt wie das beim maximalen Fehler oft der Fall ist. Formal kann dieses, an ein festes Perzentil geknüpfte Maß auch auf die Überprüfung der Jahresmittelwertberechnungen angewandt werden, da der Jahresmittelwert in etwa dem Median (50-Perzentil) der Häufigkeitsverteilung entspricht. Damit hat man eine Überprüfung der Modellierung in denjenigen Konzentrationsbereichen, die durch die Ziel- bzw. Grenzwerte festgelegt sind, wodurch ein direkter Bezug zu dem Kernpunkt der EU-Richtlinien hergestellt ist. Im Vergleich zu dem maximalen Fehlermaß der EU-Richtlinien, das aus den oben genannten Gründen wahrscheinlich von keinem Modell vollständig eingehalten werden kann, erhält man mit dem nur für den Mittelwert und das Überschreitungsperzentil ermittelten Fehler eine deutlich bessere und vor allem stabilere Modellgüte, die nicht durch die sehr hohen und sehr niedrigen Werte dominiert wird. Es wird daher vorgeschlagen, die Modellgüte nur am Mittelwert und dem Überschreitungsperzentil zu bestimmen.

Die Richtlinien machen keine Aussagen über eine mögliche Auswahl der zur Modellgütebestimmung heranzuziehenden Stationen und an welchen Stationen das Modell die Gütekriterien einhalten muss. Nimmt man alle verfügbaren Stationen, hat man Probleme mit der Repräsentativität, da das Immissionsregime, das durch die Messung beschrieben wird, auch von der Modellrechnung auflösbar sein muss. So ist es offensichtlich, dass verkehrsnahen Messungen, deren Repräsentanzbereich nur wenige hundert Quadratmeter betragen kann, von

einem Modell mit einer Maschenweite im Dekakilometerbereich nicht simuliert werden kann. Eine erste Hilfe zur Auswahl von Stationen, die zur Gütebestimmung einer bestimmten Modellanwendung herangezogen werden können, bietet die in diesem Bericht angeführte objektive Einordnung der Messungen in Immissionsregimes. Es konnte gezeigt werden, dass die gestellten Güteanforderungen in der hier als Beispiel verwendeten großräumigen Modellanwendung deutlich besser für diejenigen Stationen erfüllt werden, die der Skala des Modells entsprechen.

Insgesamt muss aber festgehalten werden, dass in der hier präsentierten Modellanwendung die gestellten Güteanforderungen nicht an allen Stationen eingehalten werden, selbst wenn man eine skalengerechte Stationsauswahl vornimmt und die Güte nur am Mittelwert und am Überschreitungssperzentil mit den in den Richtlinien vorgegebenen Genauigkeitsanforderungen bestimmt. Dies hat zum einen seine Ursache in der Problematik einer skalengerechten Stationsauswahl. Selbst objektive Klassifizierungsschemata haben ihre Grenzen und können nur Abschätzungen einer „mittleren“ Repräsentativität liefern, die im Einzelfall durch für die Modellauflösung subskalige Einflüsse oder ungewöhnliche Ereignisse wie z. B. den zeitweiligen Ferntransport von Staub, gestört werden kann. Zum anderen stellt jedes Modell nur ein unvollkommenes Abbild der in der Realität stattfindenden Prozesse dar. Dies bezieht sich sowohl auf die physikalisch-chemischen Parametrisierungen im Modell als auch auf die vom Modell verwendeten Eingangsdaten. Weiterhin kann die Heterogenität der Immissionsmessung so groß sein, dass allein durch die verwendete räumliche Modellauflösung beträchtliche Fehler entstehen können. Ein dichtes Messnetz kann so zu größeren Gesamtfehlern führen, da die Beobachtungen verschiedener Stationen in einer Modellbox niemals alle mit derselben Güte modelliert werden können.

Zentraler Punkt bei der Auswahl von Stationen, die zur Modellevaluierung verwendet werden sollen, bleibt die oft unbekannte oder nur unzureichend definierbare Repräsentativität einer Station. Eine räumliche Glättung der Messungen innerhalb einer Gitterbox des Modells verbessert zwar meist die Modellgüte, kann aber in Gebieten, wo durch die Mittelung Messungen von Stationen zusammengefasst werden, die unterschiedlichen Regimen angehören, sogar zu einer Vergrößerung des Modellfehlers führen. In diesem Zusammenhang wären zusätzliche temporäre Messungen im angenommenen Repräsentativitätsbereich einer ortsfesten Probenahmestelle hilfreich. Solche Messungen könnten dazu dienen, den Beobachtungsfehler, d.h. die Unterschiede zwischen den Messungen innerhalb des Gebiets, für das die Messungen der ortsfesten Station repräsentativ sein sollen, zu quantifizieren. Die Abschätzung dieses so genannten Beobachtungsfehlers ist für die Modellevaluierung von großer Bedeutung, da er eine untere Schranke für den praktisch ermittelbaren Modellfehler darstellt.

Die Untersuchungen haben gezeigt, dass es auf Basis der gegenwärtigen EU-Richtlinienbestimmungen nicht möglich ist, eindeutige und objektivierbare Kriterien für eine Modellevaluierung anhand von Messungen zu entwickeln. Hier besteht also noch weiterer Handlungsbedarf.

7 Literatur

EU (1996), Richtlinie 96/62/EG des Rates vom 27.9.1996 über die Beurteilung und die Kontrolle der Luftqualität. Amtsblatt der Europäischen Gemeinschaften Nr. L 296/55 vom 21.11. 1996

EU (1999), Richtlinie 1999/30/EG des Rates vom 22.4.1999 über Grenzwerte für Schwefeldioxid, Stickstoffdioxid und Stickstoffoxide, Partikel und Blei in der Luft. Amtsblatt der Europäischen Gemeinschaften Nr. L 163/41 vom 29.6. 1999.

EU (2000), Richtlinie 2000/69/EG des Europäischen Parlaments und des Rates vom 16.11.2000 über Grenzwerte für Benzol und Kohlenmonoxid in der Luft. Amtsblatt der Europäischen Gemeinschaften Nr. L 313/12 vom 13.12. 2000.

EU (2002), Richtlinie 2002/3/EG des Europäischen Parlaments und des Rates vom 12.2.2002 über den Ozongehalt der Luft. Amtsblatt der Europäischen Gemeinschaften Nr. L 67/14 vom 9.3. 2002.

Flemming, J., Reimer, E. und Stern, R. (1999) Ozonprognose mit REM3, Teil B im Abschlußbericht zum UBA-Forschungsvorhaben FE-Vorhaben 29543817, Freie Universität Berlin, Institut für Meteorologie.

Flemming, J. (2003a) Immissionsfelder aus Beobachtung, Modellierung und deren Kombination, FU-Berlin, <http://www.diss.fu-berlin.de/2003/71>.

Flemming, J. (2003b) Klimatische Auswertung von modellierten und beobachteten Zeitreihen für Ozon, NO₂, NO, SO₂ und PM₁₀ in Deutschland für die Jahre 1995-2001 und deren Charakterisierung nach Immissionsregimes mit einem Clusterverfahren, Abschlussbericht zum Forschungs- und Entwicklungsvorhaben 298 41 252 des Umweltbundesamts "Modellierung und Prüfung von Strategien zur Verminderung der Belastung durch Ozon.

Flemming, J. (2003c) Anwendung von Optimaler Interpolation und Kalman Filter Techniken für die Ausbreitungsrechnung mit dem chemischen Transportmodell REM/CALGRID, Abschlussbericht zum Forschungs- und Entwicklungsvorhaben 298 41 252 des Umweltbundesamts "Modellierung und Prüfung von Strategien zur Verminderung der Belastung durch Ozon

Flemming, J. und Stern, R. (2004), Immissionsklima 1995-2003 aus Beobachtung und Modellrechnung: Untersuchung des Einflusses von Wettervariabilität und Emissionstrends, Abschlussbericht zum Forschungs- und Entwicklungsvorhaben 201 43 250 UBA auf dem Gebiet des Umweltschutzes.

Hogrefe, Christian and Rao, S. Trivikrama (2001), Demonstrating attainment of the air quality standards: integration of observations and model predictions into the probabilistic Framework, "J. Air & Waste Manage. Assoc.", p. 1060 - 1072.

Hollingsworth, A. und Lönnberg, P. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data Part I: The wind field, Tellus, Bd. 38 A, S. 111-136.

Kielland, Gabriel et al. (1998), Guidance report on preliminary assessment under EC air quality directive, EEA Technical report No 11, <http://reports.eea.eu.int/TEC11a/en>.

Roemer, M. et al. (2003), Ozone trends according to ten dispersion models, Eurotrac-2 Special report, ISS-GSF, Munich.

Stedman, John R., Bush, Tony J, Vincent, Keith J and Baggott, Sarah (2003), UK air quality modelling for annual reporting 2002 on ambient air quality assessment under Council Directives 96/62/EC and 1999/30/EC, report AEAT/ENV/R/1564 Issue 1.

Stern, R. (2003) Modellierung und Prüfung von Strategien zur Verminderung der Belastung durch Ozon. Entwicklung und Anwendung des chemischen Transportmodells REM/CALGRID. Zusammenfassender Abschlussbericht zum Forschungs- und Entwicklungsvorhaben 298 41 252 des Umweltbundesamts.

Stern, R., Kerschbaumer, A., Fath, J. (2003) Anwendung des chemischen Transportmodells REM/CALGRID auf die Region Berlin-Brandenburg. Abschlussbericht zum Forschungs- und Entwicklungsvorhaben 299 43 246 des Umweltbundesamts

Stern, R. (2004) Weitere Entwicklung und Anwendung des chemischen Transportmodells REM-CALGRID für die bundeseinheitliche Umsetzung der EU-Rahmenrichtlinie Luftqualität und ihrer Tochterrichtlinien. Abschlussbericht zum Forschungs- und Entwicklungsvorhaben 201 43 250 des Umweltbundesamts.

8 Abbildungsverzeichnis

Abbildung 1 Zeitreihen des PM10 Tagesmittelwertes an der Station BB001 aus der Beobachtung OBS und der Modellrechnung RCG.....	22
Abbildung 2 Zeitreihen der Differenz zwischen Beobachtung und Modellrechnung an der Station BB001, PM10 Tagesmittelwerte.	22
Abbildung 3 Zeitreihen der relativen Differenz zwischen Beobachtung und Modellrechnung an der Station BB001, PM10 Tagesmittelwerte.	23
Abbildung 4 Akkumulierte Häufigkeitsverteilung der PM10 Tagesmittelwerte an der Station BB001 für die Modellrechnung RCG, die Beobachtung OBS und eine zusätzliche Schwankungsbreite von +/- 50%.	23
Abbildung 5 Absolute Fehler mit Perzentil-Bezug der PM10 Tagesmittelwerte an der Station BB001	24
Abbildung 6 Relativer Fehler mit Perzentil-Bezug der PM10 Tagesmittelwerte an der Station BB001	24
Abbildung 7 Differenz zwischen dem Perzentilwert der Modellierung und dem der Beobachtung bei einem festen Konzentrationswert. PM10-Tagesmittelwerte an der Station BB001.	25
Abbildung 8 Häufigkeitsverteilung der PM10 Tagesmittelwerte an der Station BB001 aus der Beobachtung (OBS) und der Modellrechnung (RCG)	25
Abbildung 9 Häufigkeitsverteilung der Modellfehler (RCG-OBS) mit Zeitbezug (ERR) und ohne Zeitbezug (Differenz der entsprechenden Perzentile, ERR_p).....	26
Abbildung 10 Häufigkeitsverteilung der relativen Modellfehler ((RCG-OBS)/OBS) mit Zeitbezug (ERR) und ohne Zeitbezug (Differenz der entsprechenden Perzentile, ERR_p). Die EU-Richtlinien erlauben einen relativen Maximalfehler von 50%.	26
Abbildung 11 Stationsbezogene Jahresmittelwerte von PM10 aus Beobachtung und Modellrechnung mit RCG. Links: Beobachtung vs. Modellrechnung, Mitte: Beobachtung vs. absoluten Fehler, Rechts: Beobachtung vs. relativer Fehler. Die gestrichelten Linien bei Konzentrationsangaben entsprechen der unteren Beurteilungsschwelle bzw. dem Grenzwert. Die 50% Line (rechts) markiert die Genauigkeitsanforderung nach 1999/30/EC.	28
Abbildung 12 Stationsbezogene Jahresmittelwerte von NO ₂ aus Beobachtung und Modellrechnung mit RCG. Links: Beobachtung vs. Modellrechnung, Mitte: Beobachtung vs. Fehler, Rechts: Beobachtung vs. relativer Fehler. Die gestrichelten Linien bei Konzentrationsangaben entsprechen der unteren Beurteilungsschwelle bzw. dem Grenzwert. Die 30% Line (rechts) markiert die Genauigkeitsanforderung nach 1999/30/EC.	28
Abbildung 13 Stationsbezogene Jahresmittelwerte von SO ₂ aus Beobachtung und Modellrechnung mit RCG. Links: Beobachtung vs. Modellrechnung, Mitte: Beobachtung vs. Fehler, Rechts: Beobachtung vs. relativer Fehler. Die gestrichelten Linien bei Konzentrationsangaben entsprechen der unteren Beurteilungsschwelle bzw. dem Grenzwert. Die 30% Line (rechts) markiert die Genauigkeitsanforderung nach 1999/30/EC.	29
Abbildung 14 Stationsbezogene Jahresmittelwerte von Ozon aus Beobachtung und Modellrechnung mit RCG. Links: Beobachtung vs. Modellrechnung, Mitte: Beobachtung vs. absolutem Fehler, Rechts: Beobachtung vs. relativer Fehler.	29
Abbildung 15 Anzahl der Stationen, die die Genauigkeitsanforderung hinsichtlich der Modellierung der Jahresmittelwerte nach 1999/30/EC erfüllen (YES), geordnet nach Regimes. Links: PM10,	

Mitte: NO ₂ , rechts: SO ₂ (Regime R= Land, U1 = Vorstadt, U2 = Stadt, U3 = belastetet Stadt, S = Verkehr, S2 = Verkehr extrem bzw. Belastungsstufen #1 - #5 für PM10 und SO ₂).....	29
Abbildung 16 Häufigkeitsverteilung (Box-Whiskers-Plot) des stationsbezogenen RMSE mit (RMSE) und ohne Zeitbezug (RMSE_P) für die Stundenwerte von NO ₂ (oben rechts), die Stundenwerte von SO ₂ (oben links), die täglichen 8h-Mittelwerte von Ozon (unten links) und die Tagesmittelwerte von PM10 (unten rechts).....	32
Abbildung 17 Konzentrationswerte aus Beobachtung (OBS) und Modellrechnung (RCG) beim maximalen absoluten Fehler für die Stundenwerte von NO ₂ (oben rechts), die Stundenwerte von SO ₂ (oben links), die täglichen 8h-Mittelwerte von Ozon (unten links) und die Tagesmittelwerte von PM10 (unten rechts). Markiert ist das Konzentrationsniveau des Grenzwertes.....	34
Abbildung 18 Konzentrationswerte aus Beobachtung (OBS) und Modellrechnung (RCG) beim Perzentil der erlaubten Überschreitungsanzahl des Grenzwertes für die Stundenwerte von NO ₂ (oben rechts), die Stundenwerte von SO ₂ (oben links), die täglichen 8h-Mittelwerte Ozon (unten links) und die Tagesmittelwerte von PM10 (unten rechts). Markiert ist das Konzentrationsniveau des Grenzwertes.....	36
Abbildung 19 Beobachtete Konzentrationswerte beim maximalen absoluten Fehler (OBS me) und bei dem Perzentilwert der erlaubten Anzahl von Überschreitungen des Grenzwertes (OBS p) für die Stundenwerte von NO ₂ (oben rechts), die Stundenwerte von SO ₂ (oben links), die täglichen 8h-Mittelwerte von Ozon (unten links) und die Tagesmittelwerte von PM10 (unten rechts).	38
Abbildung 20 Histogramm der Perzentile, für die der größte absolute Fehler bestimmt wird (max err_p). Markiert ist das Perzentil, das der erlaubten Häufigkeit von Grenzwertüberschreitungen entspricht. Aus Gründen der Darstellung ist nur der Bereich der Perzentile 90-100 dargestellt, in der fast alle Werte zu finden sind.	39
Abbildung 21 Häufigkeitsverteilung (Box-Whiskers-Plot) von stationsbezogenen relativen Fehlermaßen mit und ohne Zeitbezug (RMSE_P) für die Stundenwerte von NO ₂ (oben rechts), die Stundenwerte von SO ₂ (oben links), die täglichen 8h-Mittelwerte von Ozon (unten links) und die Tagesmittelwerte von PM10 (unten rechts). Maximaler relativer Fehler mit Zeitbezug (max rel err), maximaler relativer Fehler mit Perzentilbezug (max rel err_p), relativer maximaler Fehler mit Zeitbezug (rel max err), relativer maximaler Fehler mit Perzentilbezug (rel max err_p), relativer Fehler bei dem Perzentil, das der erlaubten Überschreitungshäufigkeit des Grenzwertes entspricht (rel err pGW). Alle Stationen in Deutschland.	42
Abbildung 22 Anzahl der Stationen, die die Genauigkeitsanforderung von 50% hinsichtlich der NO ₂ Stundenmittelwerte erfüllen (YES) für den relativen maximalen Fehler (links) und den Fehler am Perzentil der Häufigkeit der erlaubten Grenzwertüberschreitungen (rechts).....	43
Abbildung 23 Anzahl der Stationen, die die Genauigkeitsanforderung von 50% hinsichtlich der täglichen Maxima der 8h-Mittel von Ozon erfüllen (YES) für den relativen maximalen Fehler (links) und den Fehler am Perzentil der Häufigkeit der erlaubten Grenzwertüberschreitungen (rechts).	43
Abbildung 24 Anzahl der Stationen, die die Genauigkeitsanforderung von 50% hinsichtlich der Tagesmittelwerte von PM10 erfüllen (YES) für den relativen maximalen Fehler (links) und den Fehler am Perzentil der Häufigkeit der erlaubten Grenzwertüberschreitungen (rechts).	44
Abbildung 25 Anzahl der Stationen, die die Genauigkeitsanforderung von 50% hinsichtlich der Stundenmittelwerte von SO ₂ erfüllen (YES) für den relativen maximalen Fehler (links) und den Fehler am Perzentil der Häufigkeit der erlaubten Grenzwertüberschreitungen Grenzwertüberschreitung (rechts).	44
Abbildung 26 Perzentile der Tagesmittelwerte von PM10 beim Überschreiten der Grenzwertschwelle von 40 µg/m ³ (links) , der oberen Beurteilungsschwelle von 30 µg/m ³ (Mitte) und der unteren	

Beurteilungsschwelle von $20 \mu\text{g}/\text{m}^3$ (rechts) in Beobachtung (x-Achse) und Modellrechnung (y-Achse) . Die gestrichelte Linie entspricht dem Perzentil der erlaubten Überschreitungsanzahl. Kleinere Perzentilwerte bedeuten eine häufigere Überschreitung.....	45
Abbildung 27 Streudiagramm der relativen Modellfehler der Jahresmittelwerte (x-Achse) und der relativen maximalen Fehler mit Perzentil-Bezug (y-Achse links) bzw. der Fehler beim Perzentil der Überschreitungshäufigkeit des Grenzwerte (y-Achse rechts) der Stundenwerte von NO_2 . Markiert sind die Mindestgenauigkeitsanforderungen für die Modellrechnung nach 1999/30/EC. Die Darstellung ist beschränkt auf Fehler $\leq 150\%$	48
Abbildung 28 Streudiagramm der relativen Modellfehler der Jahresmittelwerte (x-Achse) und der relativen maximalen Fehler mit Perzentil-Bezug (y-Achse links) bzw. der Fehler beim Perzentil der Überschreitungshäufigkeit des Grenzwertes (y-Achse rechts) der Tagesmittelwerte von PM_{10} . Markiert sind die Mindestgenauigkeitsanforderungen für die Modellrechnung nach 1999/30/EC. Die Darstellung ist beschränkt auf Fehler $\leq 150\%$	48
Abbildung 29 Streudiagramm der relativen Modellfehler der Jahresmittelwerte (x-Achse) und der relativen maximalen Fehler mit Perzentil-Bezug (y-Achse links) bzw. der Fehler beim Perzentil der Überschreitungshäufigkeit des Grenzwertes (y-Achse rechts) der Stundenwerte von SO_2 . Markiert sind die Mindestgenauigkeitsanforderungen für die Modellrechnung nach 1999/30/EC. Die Darstellung ist beschränkt auf Fehler $\leq 150\%$	49
Abbildung 30 Streudiagramm der relativen Modellfehler der Jahresmittelwerte (x-Achse) und der relativen maximalen Fehler mit Perzentil-Bezug (y-Achse links) bzw. der Fehler beim Perzentil der Überschreitungshäufigkeit des Zielwertes (y-Achse rechts) der Tagesmaxima des 8h-Mittels von Ozon. Markiert sind die Mindestgenauigkeitsanforderung an die Tageswerte nach 2002/3/EC. Die Darstellung ist beschränkt auf Fehler $\leq 150\%$	49
Abbildung 31 Relativer Fehler beim Perzentil des Zielwertes (rel per err _{p_{GW}} , links), relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und RMSE mit Perzentil-Bezug (rechts). Stationsbezogen mit direktem Vergleich (Base) und räumlicher Mittelung der Messung in der Auflösung des Modells (SMOOTH) für die Tagesmaxima des 8 h Mittels von Ozon. ...	51
Abbildung 32 Mittlere Veränderung pro Immissionsregime des relativen Fehlers beim Perzentil des Zielwertes (rel per err _{p_{GW}} , links), des relativen maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und des RMSE mit Perzentil-Bezug (rechts) zwischen direktem Vergleich und räumlicher Mittelung der Messung in der Auflösung des Modells für die Tagesmaxima des 8 h Mittels von Ozon.	52
Abbildung 33 Relativer Fehler beim Perzentil des Zielwertes (rel per err _{p_{GW}} , links), relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und RMSE mit Perzentil-Bezug (rechts). Stationsbezogen mit direktem Vergleich (Base) und räumlicher Mittelung der Messung in der Auflösung des Modells (SMOOTH) für die Tagesmittelwerte von PM_{10}	52
Abbildung 34 Mittlere Veränderung pro Immissionsregime des relativen Fehlers beim Perzentil des Zielwertes (rel per err _{p_{GW}} , links), des relativen maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und des RMSE mit Perzentil-Bezug (rechts) zwischen direktem Vergleich und räumlicher Mittelung der Messung in der Auflösung des Modells für die Tagesmittel von PM_{10}	52
Abbildung 35 Relativer Fehler beim Perzentil des Zielwertes (rel per err _{p_{GW}} , links), relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und RMSE mit Perzentil-Bezug (rechts). Stationsbezogen mit direktem Vergleich (Base) und räumlicher Mittelung der Messung in der Auflösung des Modells (SMOOTH) für die Stundenwerte von SO_2	53
Abbildung 36 Mittlere Veränderung pro Immissionsregime des relativen Fehlers beim Perzentil des Zielwertes (rel per err _{p_{GW}} , links), des relativen maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und des RMSE mit Perzentil-Bezug (rechts) zwischen direktem Vergleich	

und räumlicher Mittelung der Messung in der Auflösung des Modells für die Stundenwerte von SO ₂	53
Abbildung 37 Relativer Fehler beim Perzentil des Zielwertes (rel per err_p _{GW} , links), relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und RMSE mit Perzentil-Bezug (rechts). Stationsbezogen mit direktem Vergleich (Base) und räumlicher Mittelung der Messung in der Auflösung des Modells (SMOOTH) für die Stundenwerte von NO ₂	53
Abbildung 38 Mittlere Veränderung pro Immissionsregime des relativen Fehlers beim Perzentil des Zielwertes (rel per err_p _{GW} , links), des relativen maximaler Fehler mit Perzentil-Bezug (rel max_err_p, Mitte) und des RMSE mit Perzentil-Bezug (rechts) zwischen direktem Vergleich und räumlicher Mittelung der Messung in der Auflösung des Modells für die Stundenwerte von NO ₂	54
Abbildung 39 Schätzung der Varianz des klimatischen Beobachtungsfehlers (observation error variance) der Station S mit Hilfe eines Kovarianzmodells nach der Beobachtungsmethode.....	56
Abbildung 40 Tagesgang der relativen Standardabweichung des Beobachtungsfehlers (normiert mit dem Mittelwert) von O ₃ , NO ₂ , PM10 und SO ₂ für verschiedene Immissionsregimes im Jahr 2002 in Deutschland in 100% (=1). (O ₃ (0 bis 5) und NO ₂ (1 bis 6) B: Berg/Küste, U1: Vorstadt U2: Stadt, U3: belastete Stadt, S: Strasse, S2: Strasse extrem, 1-5: Immissionsregime mit wachsender Verschmutzung für SO ₂ und PM10).....	58
Abbildung 41 Relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, links) und relativer Fehler beim Perzentil der erlaubten Grenzwertüberschreitungshäufigkeit (rel per err_p _{GW} , rechts) bei unterschiedlicher Behandlung des Datenausfalls (< 10%) der Messwerte für die Stundenwerte von NO ₂ . BASE: Entfernen der Modellwerte mit Zeitbezug, ACLEAN: Ersetzen der fehlenden Messwerte durch die in diesem Zeitraum berechneten Werte.	60
Abbildung 42 Relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, links) und relativer Fehler beim Perzentil der erlaubten Grenzwertüberschreitungshäufigkeit (rel per err_p _{GW} , rechts) bei unterschiedlicher Behandlung des Datenausfalls (< 10%) der Messwerte für die Stundenwerte von SO ₂ . BASE: Entfernen der Modellwerte mit Zeitbezug, ACLEAN: Ersetzen der fehlenden Messwerte durch die in diesem Zeitraum berechneten Werte.	61
Abbildung 43 Relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, links) und relativer Fehler beim Perzentil der erlaubten Grenzwertüberschreitungshäufigkeit (rel per err_p _{GW} , rechts) bei unterschiedlicher Behandlung des Datenausfalls (< 10%) der Messwerte für die Tageswerte von PM10. BASE: Entfernen der Modellwerte mit Zeitbezug, ACLEAN: Ersetzen der fehlenden Messwerte durch die in diesem Zeitraum berechneten Werte.	61
Abbildung 44 Relativer maximaler Fehler mit Perzentil-Bezug (rel max_err_p, links) und relativer Fehler beim Perzentil der erlaubten Grenzwertüberschreitungshäufigkeit (rel per err_p _{GW} , rechts) bei unterschiedlicher Behandlung des Datenausfalls (< 10%) der Messwerte für die Tagesmaxima des 8 Stundenmittels von Ozon. BASE: Entfernen der Modellwerte mit Zeitbezug, ACLEAN: Ersetzen der fehlenden Messwerte durch die in diesem Zeitraum berechneten Werte.	62
Abbildung 45 Histogramme der Vollständigkeit der stündlichen Beobachtungsreihen von NO ₂ , SO ₂ , O ₃ und PM10 an den Messstationen in Deutschland im Jahr 2002.	63
Abbildung 46 Berechnete Jahresmittelwerte von NO ₂ (ppb) für die Modellgitterboxen des REM-CALGRID-Modells in und um Berlin (Brandenburg) und die beobachteten Werte an den Messorten. Gitterauflösung circa 25 km.....	65
Abbildung 47 Anzahl der Stationen, die bei der gegebenen Auflösung (y-Achse) mindestens noch eine zweite Station in ihrer Gitterbox aufweisen. Beobachtungsnetz in Deutschland.....	66

Abbildung 48 Maximale Anzahl von Stationen in Gitterboxen verschiedener Auflösung (y-Achse) . Beobachtungsnetz in Deutschland.....	67
Abbildung 49 Box Whiskers Plots der stationsbezogenen Fehlermaße (RMSE, RMSE-p, rel per err_p _{GW} , rel max err_p) des perfekten Modells bei verschiedenen Auflösungen (RES in km) und der realen RCG-Modellrechnung mit 25 km Auflösung (R25) für NO ₂ -Stundenmittelwerte im Jahr 2002.....	69
Abbildung 50 Box Whiskers Plots der stationsbezogenen Fehlermaße (RMSE, RMSE-p, rel per err_p _{GW} , rel max err_p) des perfekten Modells bei verschiedenen Auflösungen (RES in km) und der realen RCG-Modellrechnung mit 25 km Auflösung (R25) für SO ₂ – Stundenmittelwerte im Jahr 2002.	70
Abbildung 51 Box Whiskers Plots der stationsbezogenen Fehlermaße (RMSE, RMSE-p, rel per err_p _{GW} , rel max err_p) des perfekten Modells bei verschiedenen Auflösungen (RES in km) und der realen RCG-Modellrechnung mit 25 km Auflösung (R25) für PM10-Tagesmittelwerte im Jahr 2002.....	71
Abbildung 52 Box Whiskers Plots der stationsbezogenen Fehlermaße (RMSE, RMSE-p, rel per err_p _{GW} , rel max err_p) des perfekten Modells bei verschiedenen Auflösungen (RES in km) und der realen RCG-Modellrechnung mit 25 km Auflösung (R25) für Ozon Tagesmaxima des 8h Mittels im Jahr 2002.	72

9 Tabellenverzeichnis

Tabelle 1 Grenzwerte für Ozon, PM10, NO ₂ und SO ₂ nach den EU-Richtlinien	4
Tabelle 2 Geforderte Genauigkeit der Modellrechnung (Datenqualitätsziel) für Ozon nach 2003/3/EG Anhang VII und für NO ₂ , PM10 und SO ₂ nach EC 1999/30/EG Anhang VIII. * noch nicht festgelegt.....	7
Tabelle 3 Geforderte Genauigkeit der kontinuierlichen Messungen bezogen auf die Einzelmessung gemittelt über den betreffenden Zeitraum und in der Nähe des jeweiligen Grenz- oder Schwellenwertes für Ozon nach 2003/3/EG Anhang VII und für NO ₂ , PM10 und SO ₂ nach EC 1999/30/EG Anhang VIII.	15
Tabelle 4 Geforderte räumliche Repräsentativität der kontinuierlichen Messungen und die entsprechende Modellauflösung eines Eulerschen CTM für Ozon nach 2003/3/EG Anhang VII und für NO ₂ , PM10 und SO ₂ nach EC 1999/30/EG Anhang VIII.	16
Tabelle 5 Immissionsregimes nach Flemming (2003). Clusterparameter sind: P50ME = Median des Tagesmittelwertes und P50TA = Median der normierten Tagesschwankung, Log = Logarithmische Transformation	18
Tabelle 6 Geschätzte Standardabweichung des klimatischen Beobachtungsfehlers der stündlichen Messung, dargestellt als Median aller Stationen innerhalb der Immissionsregimes nach Flemming (2003b) (B = Berg, R = Land, U1 = Vorstadt, U2 = Stadt, U3 = belasteter Stadt, S = Straße, S2 = Straße extrem).....	57
Tabelle 7 Typische Gitterauflösungen für Modellanwendungen in verschiedenen Skalen	64