

4 Methodik der Analyse und Datenassimilation

4.1 Problemstellung

Die Vorstellung eines kontinuierlichen Feldes der atmosphärischen Eigenschaften ist von zentraler Bedeutung in der Meteorologie. Die dreidimensionale Variabilität wird häufig mit der Vorstellung von zweidimensionalen Feldern in mehreren Höhengniveaus ausgedrückt.

Die Kontinuität des Feldes impliziert eine gewisse „Glattheit“ und damit eine untere Schranke für die räumliche **Skala** (Größenmaßstab) der darstellbaren Phänomene. Der Begriff der Skala ist für die Meteorologie entscheidend, da die Prozesse in verschiedenen Skalen häufig methodisch separiert⁵⁹ werden.

Die Eigenschaften des Feldes werden gewöhnlich nur an ausgewählten Orten mit Messinstrumenten beobachtet. Methoden der räumlichen Interpolation, mit denen aus punktbezogenen Messungen eine Felddarstellung an regelmäßig verteilten Gitterpunkten gewonnen wird, werden in der Meteorologie mit dem Begriff **Analyse** zusammengefasst.

Die gewonnene Felddarstellung wird gleichfalls Analyse genannt. Sie hat eine bestimmte räumliche Auflösung, die die untere Schranke für die kleinste darstellbare Skala vorgibt. Die Wahl der Auflösung sollte demzufolge durch die interessierenden bzw. die erfassbaren Skalen bestimmt sein.

Die **Kartierung**⁶⁰ als Grundlage der Visualisierung von atmosphärischen Feldern ist eine wichtige Motivation für die Analyse. Durch die Felddarstellung können räumliche Beziehungen, wie z. B. das geostrophische Gleichgewicht, erkannt werden. Darüber hinaus kann das erzeugte Feld als Zustandsvektor eines dynamischen Modells verwendet werden. Die Analyse ist damit eine Methode, um Beobachtungswerte in Modellrechnungen einfließen zu lassen. Die Analyse ist ein Kernpunkt in allen Verfahren der **Datenassimilation**. Die Modellauflösung stellt hierbei die Begrenzung für die erfassbaren Skalen dar, die sich von denen der Beobachtungen in vielen Fällen unterscheidet. Die Separation der für Analyse und Beobachtungen relevanten Skalen ist ein wichtiger Punkt bei allen Analyseverfahren.

In den letzten 40 Jahren ist in der numerischen Wettervorhersage ein umfangreiches Instrumentarium von Analyse- und Datenassimilationsmethoden entwickelt worden⁶¹.

⁵⁹ Durch die Nichtlinearität der zugrundeliegenden physikalischen und chemischen Gleichungen beeinflussen die Prozesse der nicht erfassten Skalen die größerskaligen Phänomene. Dies führt zum Schließungsproblem, d. h. zur Parametrisierung des Einflusses der nichtaufgelösten Skalen auf der Basis der aufgelösten Skalen.

⁶⁰ Auch bei der Kartierung ist die räumliche Auflösung von Bedeutung; insbesondere für den Algorithmus zur Isolinienfindung.

⁶¹ Ein Standardwerk zu dieser Thematik ist das Buch von Daley (1991).

Die modernen Verfahren beruhen auf der Theorie der Prognose⁶² von **stochastischen Prozessen**, die unter dem Begriff Optimale Interpolation von Gandin (1965) in die Meteorologie eingeführt wurden. Parallel dazu ist der statistische Zugang für die räumliche Interpolation unter dem Begriff Kriging⁶³ in der Geostatistik ausgearbeitet worden. Immissionsdaten werden häufig mit dem Kriging-Ansatz interpoliert.

Die statistischen Methoden haben sich bewährt, da sie ein Instrumentarium bieten, um die räumlichen Beziehungen zu quantifizieren. Die statistisch geschätzte räumliche Kovarianzfunktion ist die Grundlage für die Bestimmung des Einflusses der Beobachtungen bei der räumlichen Interpolation. Gleichzeitig kann ein möglicher Beobachtungsfehler berücksichtigt werden, der sowohl Ausdruck des Messgerätefehlers als auch des Fehlers durch mangelnde Repräsentativität ist. Letzterer ist die Folge mangelnder Übereinstimmung zwischen den erfassten Skalen von Beobachtung und Analyse. Die statistischen Ansätze beinhalten die Möglichkeit zur Bestimmung von Vertrauensbereichen der interpolierten Werte. Damit kann die Güte der Analyse bewertet und in einem weiteren Schritt die Messnetzanordnung optimiert werden. Der erste Punkt ist für die Aufgabenstellung dieser Arbeit von besonderem Interesse, da die Felddarstellungen zur Ableitung von administrativen Luftgütestandards dienen können.

Die Anwendung der modernen Methoden der Analyse und der Datenassimilation für die numerische Wettervorhersage beruht meist auf der Annahme von homogenen statistischen Eigenschaften der Felder und Beobachtungen. Diese Annahmen sind für bodennahe Immissionsfelder häufig nicht gerechtfertigt. Im Folgenden sind Eigenschaften der Immissionsfelder und ihrer Messung aufgelistet, die bei der Anwendung von statistischen Analyseverfahren zu berücksichtigen sind:

1. Die bodennahen Immissionsfelder sind aufgrund der Emissionsverteilung räumlich sehr inhomogen.
2. Die Messstationen liegen gehäuft in den belasteten Gebieten und dienen hauptsächlich dazu, kleinräumige Spitzenkonzentrationen und nicht ein möglichst großes Gebiet zu erfassen.
3. Die räumliche Repräsentativität und die Güte der Messung sind nur in geringem Maße bekannt.
4. Die Häufigkeitsverteilungen der semi-positiv definiten⁶⁴ Immissionswerte sind unsymmetrisch und können mehrere Größenordnungen umfassen.
5. Immissionsdaten zeichnen sich durch eine hohe zeitliche Variabilität aus.
6. Es existieren keine räumlichen Balancebeziehungen⁶⁵, dafür aber lokale chemische Kopplungen⁶⁶.
7. Die Güte der Chemie-Transport-Modelle liegt hinsichtlich der relevanten Variablen unter denen der numerischen Wettervorhersage⁶⁷.

⁶² In Anlehnung an die Statistik von Zeitreihen wird häufig von Prädiktion bei der räumlichen Interpolation gesprochen.

⁶³ Benannt nach einem südafrikanischen Bergbauingenieur D.G. Krige, der das Verfahren zur Interpolation von Probebohrungen entwickelte.

⁶⁴ Stets größer gleich Null

⁶⁵ wie z. B. der geostrophische Wind

⁶⁶ z. B. das fotostationäre Gleichgewicht zwischen Ozon, NO und NO₂

⁶⁷ Da CTM häufig mit modellierten Daten der Wettervorhersage angetrieben werden, ist diese Aussage trivial. Hier ist jedoch gemeint, dass die Fehlermaße für Ozon etc. weit größer sind als die der Temperatur- oder Windprognose.

4.2 Zielstellung

Das Ziel dieser Arbeit ist die Entwicklung, Anwendung und der Vergleich von Verfahren zur statistischen Analyse von bodennahen Immissionsmessungen. Damit sollen Felder von stündlichen Werten von Ozon, NO₂, NO, SO₂ und PM10 für Deutschland bzw. Mitteleuropa erzeugt werden. Die räumliche Auflösung der Analysen beträgt dabei ca. 15 km. Eine Darstellung der Strukturen der urbanen Skala wird damit nicht angestrebt.

Die aufgeführten Stoffe unterscheiden sich hinsichtlich der räumlichen Variabilität aufgrund der Struktur ihrer Quellen und ihrer atmosphärischen Lebensdauer. Aus Gründen der Vereinfachung und der Vergleichbarkeit wird jedoch versucht, eine möglichst einheitliche Methodik für alle Stoffe zu entwickeln.

Der methodische Ansatz der Analyse ist vorrangig die passive Datenassimilation, d. h. die Kombination der Beobachtungen mit den Modellfeldern des Eulerschen Ausbreitungsmodells REM/Calgrid (Stern, 1994). In Anbetracht der gestellten Aufgabe, stündliche Immissionsfelder für mehrere Jahre zu gewinnen, wird auf die Anwendung der numerisch aufwendigen aktiven 4-dimensionalen Datenassimilationsverfahren verzichtet. Es wird ein auf der Optimalen Interpolation beruhendes passives Analyseverfahren⁶⁸ entwickelt, das die in Kapitel 4.1 aufgeführten Besonderheiten der Immissionsdaten berücksichtigt. Zu Vergleichszwecken wird jedoch eine aktive Assimilation auf der Basis der Optimalen Interpolation und die Anwendung eines Kalman-Filter diskutiert. Bei der aktiven Datenassimilation ersetzt die Analyse die Modellfelder und bewirkt damit eine Assimilation der Beobachtungen während des Modellaufes.

In der Arbeit werden den räumlich statistischen Eigenschaften der Beobachtung und ihrer Fehler besondere Aufmerksamkeit geschenkt. Die Ergebnisse dieser Untersuchung sind auch für alle weiteren Datenassimilationsverfahren von Bedeutung. Für das Analyseverfahren werden terminbezogene und klimatische Kovarianzmodelle der Beobachtungsinkremente bzw. Modellfehler abgeleitet. Die terminbezogenen Kovarianzmodelle erlauben eine bessere Wiedergabe der zeitlichen Variabilität; klimatische Kovarianzmodelle sind besser geeignet, die räumliche Inhomogenität abzubilden.

Darüber hinaus werden ausschließlich auf Beobachtungen oder auf Modelldaten beruhende räumliche Kovarianzmodelle gewonnen. Diese können für die Evaluierung von Modellen verwendet werden, die über den Vergleich der ortsbezogenen statistischen Maße, wie Mittelwert und Varianz, hinausgeht.

4.3 Literaturüberblick

Die für diese Arbeit relevante Literatur lässt sich in zwei Gruppen einteilen. Es sind zum einen Arbeiten, die die Interpolation von Immissionsbeobachtungen für deren Kartierung zum Thema haben (Kapitel 4.3.2). Zum anderen werden Arbeiten zur Datenassimilation von vorrangig bodennahen Messungen mit Eulerschen Chemie-Transportmodellen (Kapitel 4.3.3) vorgestellt. Zu Beginn wird ein kurzer Überblick über praktische Anwendung gegeben (Kapitel 4.3.1), da der gegebene Überblick vor-

⁶⁸ Da die Analyse immer ein wesentlicher Bestandteil der 4D-Datenassimilation ist, können die dabei gewonnenen Erkenntnisse zur Verbesserung der Assimilation von Spurenstoffkonzentrationen verwendet werden.

rangig durch die Methodik motiviert ist. Es liegt in der Natur der Sache, dass die vorgestellte Auswahl unvollständig und subjektiv ist.

4.3.1 Anwendungsgebiete

Die meisten Verfahren der Immissionskartierung dienen der räumlichen Darstellung der Luftqualität. Sie beziehen sich auf die wirkungsspezifischen Eigenschaften, d. h. auf abgeleitete klimatische Felder⁶⁹, mit denen Luftreinhaltungsziele formuliert werden. Es sind dies Jahresmittelwerte, Maximalwerte oder dosisbezogene Größen wie AOT40⁷⁰-Werte. Die Darstellung orientiert sich dabei vorrangig an der Überschreitung von umweltpolitisch vorgegebenen Grenzwerten oder ökologischen Toleranzbereichen. Neben der Kartierung sind Untersuchungen zur optimalen Messnetzgestaltung ein weiteres Anwendungsgebiet der Geostatistik. Es beruht auf der Untersuchung der Interpolationsfehler für verschiedene Messnetzkonfigurationen (Nychka und Saltzman, 1998 und Shindo et al., 1990).

Ein weiteres Anwendungsfeld der Immissionskartierung ist die Evaluierung von Chemie-Transport-Modellen. Aufgrund des Skalenunterschiedes der Beobachtungen hinsichtlich der Gitterboxmittelwerte des Modells erscheint ein direkter Vergleich von Rechnung und Messung problematisch (McNair et al., 1996). Dies gilt besonders dann, wenn die räumliche Repräsentativität der Messung gering oder unbekannt ist. Ein möglicher Lösungsansatz ist, die Modellergebnisse mit einer aus den Beobachtungen interpolierten Felddarstellung zu vergleichen. Die Felddarstellung muss dabei der Auflösung bzw. Skala des Modells entsprechen. Beispiele für dieses Vorgehen bei der Evaluierung sind die Arbeiten von Schaug et al. (1993), Davis et al. (2000) und Flemming et al. (2001).

Aufgrund der hohen Komplexität wird die Datenassimilation für Chemie-Transport-Modelle meist nur in Prozess- und Modellstudien mit Episodencharakter angewendet. Neben den bodennahen Immissionsmessungen liegt der Schwerpunkt auf der Assimilation von Satellitendaten für großräumige Modellrechnungen (Jeuken et al., 1999). Die Initialisierung von operationellen Immissionsprognosen mit gemessenen Ozonkonzentrationen ist ein Beispiel für die Anwendungen von einfachen Assimilationstechniken (Flemming, 1996). In einem weiteren Zusammenhang stehen Methoden der inversen Modellierung, bei denen aus Beobachtungsdaten Emissionswerte abgeleitet werden.

4.3.2 Interpolation von Messwerten

Das Standardwerk für Fragen der räumlichen Statistik, mit Beispielen für Immissionsdaten, ist Cressie (1993). Darauf aufbauend werden neue inhomogene Ansätze von Smith (2001) behandelt.

Die Auswahl der nun vorgestellten Arbeiten orientiert sich vorrangig an der Methodik, mit der Lösungsansätze für die in Kapitel 4.1 erwähnten Probleme angeboten werden. Homogenere Felder wie die von Ozon sind mit geringeren Schwierigkeiten zu interpolieren als die stark strukturierten Felder von NO oder Feinstaub (PM10).

⁶⁹ Für die Datenassimilation mit einem dynamischen Modell, aber auch für die Sommersmogproblematik werden Analyseverfahren für aktuelle Immissionsfelder benötigt.

⁷⁰ Akkumulierte Dosis über 40 pbb Belastung.

Obwohl Beobachtungsfehler und die ungleichmäßige Stationsdichte nicht ohne weiteres behandelt werden können, sind für die Interpolation der Messwerte formale nichtstochastische Methoden geläufig, da sie die aufwendige Schätzung der Kovarianzfunktion und der Beobachtungsfehler vermeiden (Wiegand und Dickmann, 2000). Die direkte Berücksichtigung⁷¹ der Stationsdichte muss bei diesen Ansätzen (siehe 4.5.5) zusätzlich eingeführt werden. Falke und Husar (1998) liefert hierfür ein Beispiel bei der Interpolation von Ozonmessungen.

4.3.2.1 Homogene statistische Ansätze

Zier (1976) entwickelte ein verteilungsunabhängiges Interpolationsverfahren auf der Basis von Quantilkorrelationsmaßen für die Staubbelastung in der DDR. Das Interpolationsergebnis ist kein absoluter Wert, sondern die Wahrscheinlichkeit, dass der Wert in einem bestimmten Werteintervall liegt⁷².

Eine grundlegende Anwendung der statistischen Interpolation stündlicher Werte von SO₂ und zusätzlich für NO₂, NO sowie Ozon in den Niederlanden geben van Egmond und Onderdelinden (1981). Die erstellte Analyse hat dabei eine Auflösung von 40 bzw. 28 km. Die Autoren vergleichen drei Interpolationsmethoden mit 1) einem exponentiellen Korrelationsmodell (Optimale Interpolation, OI), 2) einem nichtparametrischen Ansatz (Eigenvektorinterpolation, siehe Kapitel 5.2.5.2) und 3) mit formaler Abstandswichtung. Dabei wird eine leichte Überlegenheit der ersten beiden statistischen Ansätze festgestellt. Die notwendigen empirischen Korrelationswerte werden sowohl terminbezogen (rein räumlich, siehe Kapitel 5.2.1) als auch klimatisch (aus Zeitreihen, siehe Kapitel 5.2.2) ermittelt. Aufgrund mangelnder zeitlicher Filterung ergeben sich dabei größere Unterschiede, die für die Bewertung der relativen Analysefehler und Beobachtungsfehler von Bedeutung sind. Die Beobachtungsfehlervarianz wird für SO₂ auf ca. 5-10 % der terminbezogenen Varianz geschätzt.

Cressie et al. (1999) liefern ein Beispiel für die räumliche Interpolation von PM10 in der urbanen Skala. In dieser Arbeit wird Standard-Kriging für logarithmisch normalverteilte Daten mit einem modernen Markov-Zufallsfeld-Ansatz mit Hilfe von cross validation⁷³ verglichen. Letzteres Verfahren liefert Felder mit stärkeren Gradienten, berücksichtigt jedoch nicht die Wirkung von Beobachtungsfehlern. Weiterhin treten Probleme bei ungleichmäßiger Stationsverteilung auf

4.3.2.2 Raum-Zeit-Zusammenhänge

Die Gesamtheit der Raum-Zeit Variabilität erfasst Bilonick (1985 und 1988) mit der Modellierung von verbundenen Raum-Zeit Kovariogrammen zur Kartierung der monatlichen sauren Deposition in den USA. Weiterentwicklungen der Kriging-Technik werden zur Bildung von Flächenmitteln (Block Kriging) und lokalen Histogrammen (Indikator Kriging) angewandt. Der letzte Ansatz ist eine Form der nichtlinearen statis-

⁷¹ Dies gilt für die Interpolation und nicht für die Kovarianzmodellierung.

⁷² Das Verfahren ähnelt sehr dem sogenannten Indikator-Kriging (siehe Kapitel 4.4.1.3).

⁷³ Die Interpolation für den Stationsort wird mit dem dabei nicht verwendeten Messwert verglichen.

tischen Interpolation, bei der die Interpolationsgewichte vom Messwert selbst abhängen⁷⁴.

Die Spezifikation der zeitlich periodischen Erwartungswerte ist die Grundlage für die Interpolation von Ozon-AOT40-Werten in Norwegen (Host und Follestad (1999)). Die vom Jahres- und Tagesgang befreiten Abweichungen bilden die Grundlage für die Schätzung der Kovarianzstruktur. Für die Gewinnung normalverteilter Ensembles werden die logarithmierten Ozonmessungen betrachtet.

4.3.2.3 Inhomogene Kovarianzmodellierung

Seit Beginn der 90er Jahre sind in der Geostatistik verstärkt Ansätze zur Behandlung inhomogener und anisotroper Zufallfelder entwickelt worden (Meiring et al., 1997), die auch für Luftverschmutzungsprobleme angewendet werden. Smith (2001) gibt eine gute Darstellung und Diskussion dieser Entwicklungen. Es kann zwischen den folgenden Ansätzen⁷⁵ der Kovarianzmodellierung unterschieden werden:

- Moving-Window-Ansatz
- Deformationsansatz
- EOF-Erweiterung

Der Moving-Window-Ansatz (Haas, 1990) spezifiziert eigenständige Kovarianzmodelle für Teile des Untersuchungsgebietes. Der Deformationsansatz (McNair et al., 1996) ist eine vielversprechende Methode, mit der anisotrope Strukturen behandelt werden können. Es wird dabei eine Verzerrung des räumlichen Koordinatensystems vorgenommen, so dass die zugehörige Kovarianzfunktion möglichst homogen und isotrop ist. Die EOF-Ansätze bilden ein nichtparametrisches Kovarianzmodell, das auf der Eigenvektoraufspaltung (SVD) der empirischen Kovarianzmatrix der Messwerte beruht (Obled und Creutin, 1986, siehe auch Kapitel 5.2.5.2).

Die Basis für das Aufspüren der Inhomogenität ist die empirische Kovarianzmatrix für alle Paare von Stationsorten. Sie wird aus Zeitreihendaten ermittelt (siehe Abbildung 5.1, S. 100). Damit ergibt sich für alle Methoden die Notwendigkeit zur Separation der zeitlichen Variabilität, um zeitlich unabhängige Realisierungen eines Zufallfeldes zu erhalten (siehe 5.2.2).

Das Grundproblem, die unbekannte Kovarianz zwischen Beobachtungs- und Interpolationspunkt zu bestimmen, bleibt zunächst bestehen. Es wird gelöst, indem die für die Beobachtungsorte bestimmte Größen mit einfachen Ansätzen auf die Orte ohne Beobachtung übertragen werden. Eine Alternative sind flächendeckende Immissionswerte, die durch Eulersche Modelle simuliert wurden.

Mit dem Moving-Window-Ansatz identifiziert Haas (1995) für die USA 6 verschiedene Gebiete mit einer unterschiedlichen Kovarianzstruktur hinsichtlich der monatlichen Sulfatdeposition. Für jedes Teilgebiet wird ein homogenes Variogramm der logarithmierten Werte geschätzt und für das Kriging des gesamten Gebietes verwendet.

Der Deformationsansatz ist die Grundlage für die Analyse der stündlichen Ozonkonzentration in Gitterpunktdarstellung in Meiring et al. (1998). Die Analyse wird zur

⁷⁴ Die räumliche Kovarianzstruktur ist für höhere bzw. niedrigere Belastung verschiedenartig.

⁷⁵ Darüber hinaus gibt es komplexere Methoden, die mit Bayesschen Ansätzen die Linearität des traditionellen Krigings aufgeben.

Evaluierung des Eulerschen Ausbreitungsmodells SARMAP in North Carolina für eine 2-monatige Ozonepisode verwendet. Diese Arbeit ist von besonderem Interesse, da das Vorgehen bei der Behandlung der Zeit-Raum-Trennung Ähnlichkeiten mit dem hier gewählten aufweist. Aus einer Vorarbeit von Sampson und Guttorp (1998) stammt die Erkenntnis, dass die räumliche Kovarianzstruktur große Unterschiede im Verlauf des Tages aufweist und kein geeigneter Separationsansatz (siehe 4.4.2) für Raum und Zeit zu bilden ist. Aus diesem Grund wurde – wie in dieser Arbeit – ein eigenständiges räumliches Kovarianzmodell für jede Tagesstunde geschätzt. Die Analyse bezog sich dabei auf die logarithmierten Ozondaten, um eine Normalverteilung zu gewährleisten. Die Zeitreihen für jede Tagesstunde wurden mit dem Mittelwert zentriert und die weiterhin vorhandene zeitliche Korrelation mit Hilfe eines AR(2)-Modells⁷⁶ detektiert. Die Residuen zu der mit AR(2) geglätteten Reihe wurden dann als zeitlich unabhängiges Ensemble zur Bildung der paarbezogenen Kovarianzwerte verwendet. Der Deformationsansatz verzerrte nun das Koordinatensystem, um möglichst homogene exponentielle Kovarianzmodelle bilden zu können. Mit Hilfe dieser Modelle wurde dann der Wert des Residuums am Interpolationsort mit Block-Kriging ermittelt, um dem Mittelwertcharakter der Gitterzellendarstellung gerecht zu werden. Zu diesem Wert wurden nun in Umkehrung der Filterung der Wert des AR(2)-Modells und der Mittelwert addiert. Da diese Werte für den Interpolationsort unbekannt sind, wurden die Koeffizienten für AR(2) und der Mittelwert von den Beobachtungsorten mit Hilfe eines einfachen numerischen Interpolators ermittelt.

Eine Anwendung der EOF-Erweiterung ist die Analyse der jahresgangbefreiten Wochenmittel von SO₂ im Osten der USA (Holland et al., 1999). Für diese Arbeit werden drei Kovarianzmodelle mit unterschiedlicher Komplexität angenommen: 1) das traditionelle homogene und isotrope Kovarianzmodell, 2) ein homogenes und isotropes Korrelationsvarianzmodell bei räumlich variabler Varianz (siehe Gleichung {4.12}) und 3) ein Kovarianzmodell analog zu 2) mit einer zusätzlichen Erweiterung auf Basis der EOF der Kovarianzmatrix der Abweichung der empirischen Kovarianzmatrix vom Kovarianzmodell 2. Durch diese Erweiterung werden die großräumige Inhomogenität und Isotropie erfasst, der kleinräumigen Variabilität wird eine isotrope Struktur zugeschrieben. Median und 98%-Perzentil des Interpolationsfehlers lagen bei dem Kovarianzmodell 3) deutlich unter denen der anderen Ansätze.

4.3.3 Datenassimilation

Daley (1991) liefert die Grundlagen zur Analyse für die numerische Wettervorhersage. Einen guten Überblick über die modernen Methoden der Datenassimilation enthält die Darstellung von Bouttier und Courtier (1999).

Die komplexen vier-dimensionalen Verfahren (Kalman-Filter, 4D-VAR) bedeuten einen hohen mathematisch-numerischen Aufwand, dessen Bewältigung auch häufig der Schwerpunkt der entsprechenden Publikationen ist. Dieser Überblick beleuchtet hauptsächlich die Frage, wie die besondere Spezifik (siehe Kapitel 4.1) der Immissionsbeobachtungen und ihrer Modellierung behandelt wurden.

Es wird vorrangig auf die Assimilation der bodennahen Messungen aus den Luftgütemessnetzen für Eulersche Chemie-Transport-Modelle eingegangen; Arbeiten mit Satellitendaten oder Lagrangesche Modelle sind nur kurz aufgeführt. Weiterhin werden

⁷⁶ Autoregressives Modell auf Basis der beiden zurückliegenden Datenwerte

Arbeiten erwähnt, die sich mit der Analyse auf Basis von Beobachtungen und Modellbetrachtungen befassen und damit als Vorstufen der Datenassimilation dienen können.

4.3.3.1 Einfache Ansätze

Einfache Ansätze für die Verbindung von Modellierung und Beobachtung stellt die Arbeit von Venkatram (1988) zur Kartierung der Sulfatdeposition dar. Hier wird ein Gauß-Modell verwendet, um einen inhomogenen Erwartungswert für die anschließende Interpolation mit normalem Kriging zu gewinnen.

Stedman et al. (1997) gehen den umgekehrten Weg für die hochaufgelöste Kartierung der NO_x-Jahresmittelwerte in Großbritannien. Sie verwenden die Jahresmittelwerte von Stationen mit ländlicher Charakteristik für die Interpolation eines Hintergrundfeldes mit Hilfe von Kriging. Die Darstellung der Immission in belasteten Gebieten beruht auf einem hochaufgelösten Emissionskataster. Mit Hilfe eines einfachen Regressionsmodells wird ein Zusammenhang zwischen der Jahresemission und -immission gewonnen und für die Kartierung verwendet.

Eine methodische Ähnlichkeit zum Kalman-Filter ist in der Arbeit von Nychka und Saltzman (1998) für die räumliche Interpolation von Ozonmessungen im Mittelwesten der USA zu finden. Hier werden die Ergebnisse des Eulerschen Modells ROM zur Schätzung eines inhomogenen Kovarianzmodells herangezogen. Das Modell liefert somit nicht die aktuellen Werte, sondern nur die räumlich statistischen Eigenschaften. Es wird dabei der Umstand ausgenutzt, dass die Modellrechnungen für jeden beliebigen Ort vorliegen und somit Inhomogenitäten erfassbar sind.

Initialisierung von Ozonprognosen

Für ein UBA-Projekt zur operationellen Ozonprognose mit REM3 wurde ein Initialisierungsverfahren mit Ozonbeobachtungen entwickelt (Flemming et al., 1999a,b). Für den Termin des Modellstarts um 15 UTC wird aus dem prognostizierten Ozonfeld und vorliegenden Messungen aus Deutschland eine Analyse mit Hilfe der optimalen Interpolation gewonnen (Gandin, 1965). Die Korrektur des Modellfehlers ist dabei auf die Ozonkonzentrationen in Deutschland beschränkt. Die Varianz des Beobachtungsfehlers wird mit der Beobachtungsmethode nach Hollingsworth und Lönnberg (1986) mit Hilfe eines terminbezogenen und homogenen Kovarianzmodells bestimmt.

Diese Analyse dient als Startfeld für den nächsten Prognoselauf. Aufgrund des beschränkten Repräsentationsgebietes der Messungen erfolgte eine Änderung des Ozonfeldes nur im deutschen Gebiet. Untersuchungen zeigten, dass eine alleinige Änderung der Ozonkonzentration der Bodenschicht des Modells keine Auswirkungen auf die prognostizierten Ozonkonzentrationen des nächsten Tages hat (Flemming, 1996).

Mit der Annahme guter vertikaler Durchmischung der Grenzschicht zum Zeitpunkt des Modellstarts wurden die Bodenmessungen auch zur Analyse der zweiten Modellschicht herangezogen. Gleichzeitig wurden die NO_x-Konzentrationen mit Hilfe des fotostationären Gleichgewichtes an die veränderten Ozonwerte angepasst. In einer Fallstudie für eine typische Sommersmog-Periode (20.7.-30.7.1994) reagierte das Modell auf diese Initialisierung mit besseren Ergebnissen, insbesondere bei hohen Ozonkonzentrationen. Für die Untersuchung wurden diagnostische Daten verwendet. Bei der Auswertung über den gesamten Zeitraum der Prognoseanwendung verringerte sich die festgestellte mittlere Beeinflussung des nächsten Tages durch die Initialisierung. Der Einfluss des zunehmenden Prognosefehlers der meteorologischen Daten überlagert den Effekt der

Initialisierung. Deutlich erkennbar ist jedoch der „Export“ der Initialisierung mit der Hauptwindrichtung nach Polen. Dies macht deutlich, dass Beobachtungen aus ganz Mitteleuropa für eine Initialisierung notwendig sind. Der Umstand, dass die Anfangsbedingungen des Ozons für die Rechnung mit troposphärischen Chemie-Transport-Modellen keinen großen Einfluss haben, wird auch durch anderen Autoren bestätigt (Petry, 1993).

4.3.3.2 Variationsansätze

Assimilation von stratosphärischem Ozon

Eine Hauptanwendung für die Datenassimilation von Spurengasen ist die Gewinnung von Anfangszuständen für globale numerische Wettervorhersagemodelle, in denen die atmosphärische Ozonkonzentration eine prognostische Variable ist (z. B. in Modellen am ECMWF, UK Met Office und NCEP). Dabei stehen die Dynamik und die Wechselwirkung mit der Strahlung des stratosphärischen Ozons im Mittelpunkt des Interesses. Chemische Reaktionen werden von diesen Modellen nur in einfacher Form behandelt.

Die Beobachtungen werden mit Hilfe satellitengestützter Fernerkundungsmethoden gewonnen. Um aus den Strahlungsdaten Konzentrationen zu gewinnen, müssen inverse Strahlungsrechnungen unter Vorgabe von Druck- und Temperaturprofilen durchgeführt werden. Variationsmethoden eignen sich besonders, diese Daten mit ihren komplizierten Beobachtungsoperatoren in Modellzustände zu assimilieren bzw. zu analysieren (Anderson, 1992).

Die europäischen Aktivitäten auf diesem Gebiet sind im SODA⁷⁷-Projekt zusammengefasst. Variationsmethoden (4DVAR) wurden z. B. am ECMWF und am UK Met Office für die Satellitenmessungen von Ozon und weiteren Spurenstoffen entwickelt. Durch die Assimilation von stratosphärischem Ozon erhofft man sich eine Verbesserung der Strahlungsrechnung, der UV-Strahlungsvorhersage am Boden und eine Überprüfung von Strömungsfeldern in der Stratosphäre (Fisher und Lary, 1995).

4DVAR für das EURAD System

In den letzten Jahren wurde von Elbern und Schmidt (2001) ein 4DVAR System für das Ausbreitungsmodell EURAD CTM2 entwickelt. Das adjoint model zur Minimierung der Kostenfunktion ist für den Transportteil und den RADM2-Gasphasen-Chemismus formuliert.

Das Verhalten dieses Systems wurde mit „identical twin“-Experimenten untersucht. Dazu werden synthetische „Beobachtungen“ aus einem Modellauf mit bestmöglicher Konfiguration (Auflösung, Parameter etc.) entnommen. Das Assimilationsverhalten einer anderen Modellkonfiguration bezüglich dieser Beobachtungen ist dann Gegenstand der Untersuchung.

Experimente mit 70 realen Ozonmessungen aus ganz Europa wurden für den August 1997 durchgeführt. Die Assimilation erfolgte in einem 6-stündigen Intervall von 6 bis 12 UTC. Von diesen Analysen ausgehend, konnten deutlich verbesserte Modellsimulationen der Ozonkonzentration bis in die Abendstunden, d. h. mit einem Zeithorizont von 6 Stunden, verzeichnet werden.

⁷⁷ European Union Satellite Ozone Data Assimilation

Der 4DVAR Ansatz wurde auch zur Analyse und Optimierung von Emissionsraten herangezogen. Eine genaue Einschätzung des praktischen Erfolgs dieser Optimierung war aus den beiden vorliegenden Publikationen nicht zu entnehmen.

4.3.3.3 Kalman-Filter-Ansätze

Ausbreitungsrechnung von SO₂

Erste Anwendungen vom Kalman-Filter für die Luftreinhaltung stammen aus den siebziger Jahren (Fronza et al., 1979).

Für die SO₂-Ausbreitungsrechnung wurden Kalman-Filter-Ansätze im Rahmen des UBA-Projektes "Anwendung statistisch-systemdynamischer Methoden in der Ausbreitungsrechnung von Luftbeimengungen" untersucht (Dlabka et al., 1986). Mit dem Kalman-Filter wurde die Schätzung der optimalen Anfangskonzentrationsfelder für die SO₂-Prognose des TUB-Modells durchgeführt. Das TUB-Modell ist ein Eulersches Ausbreitungsmodell ohne chemische Prozesse mit einer horizontalen Gitterweite von 3 km und 10 Modellschichten. Es überdeckt das Stadtgebiet von Berlin. SO₂-Messungen des Berliner Stadtmessnetzes BLUME wurden für die Assimilation in einer Winter-Smog Episode vom 20.2.-28.2.1982 verwendet. Der Kalman-Filter wurde mit einem reduced-rank-Ansatz implementiert. Die Größe der Kovarianzmatrix des Modellzustandsvektors ist mit einer Aufspaltung in Modellteilgebiete verringert worden.

Die mit dem Kalman-Filter erzeugten Anfangsfelder bewirkten eine verbesserte Prognoserechnung für einen Zeitraum von drei Stunden. Diese relativ geringe Zeitspanne ist mit der geringen räumlichen Ausbreitung des Modellgebiets zu erklären. Die Randbedingungen haben so einen besonders großen Einfluss. Die Untersuchung widmete sich darum dem Vergleich von Randbedingungen aus Messungen und aus einer Rechnung eines großräumigen Ausbreitungsmodells.

Ozonmodellierung mit LOTOS

Im Umfeld der Abteilung für Technische Mathematik und Informatik der TU Delft wurden numerische Verfahren für die Implementierung von Kalman-Filter mit dem RRSQRT-Ansatz für nichtlineare Modelle entwickelt (Heemink et al., 1999). Die Anwendungen betreffen sowohl Gezeitevorschagemodelle als auch verschiedene Chemie-Transport-Modelle.

Im Rahmen des RITFOZ (Regional differences in tropospheric Ozone) Projektes wurde zusammen mit der TNO ein Programmpaket für den KF im RRSQRT-Ansatz in verschiedenen Ausbaustufen für das Modell LOTOS entwickelt (Segers et al., 1998 und van Loon et al., 1998). Für die Anwendung des Programmpaketes sind keine wesentlichen Modifikationen des Ausbreitungsmodells notwendig, so dass es relativ leicht auf andere Modelle übertragbar ist. Diese Implementierung wird auch im Rahmen dieser Arbeit angewendet werden.

Um die Modellunsicherheiten für den Kalman-Filter zu beschreiben, wurde ein Set von länder- und speziesspezifischen Emissionsfaktoren für die noise-Parameter verwendet. Ihre Aufgabe ist es, eine stochastische Schwankung der Emissionen mit einer Standardabweichung von 25% zu simulieren. Die Schwankungen werden jedoch zeitlich korreliert, so dass eine allzu schnelle Variation der Emission vermieden wird. Mit der Aufnahme der noise-Parameter in den Zustandsvektor kann deren Verhalten untersucht werden. Die Analyse der noise-Parameter, d. h. die Betrachtung der Korrelationen zwischen diesen Parametern und den Abweichungen zwischen Modell und Beobach-

tung, kann systematische Abweichungen aufdecken. Diese geben Hinweise auf mögliche Unter- oder Überschätzung der verwendeten Emissionsdaten.

Die Simulation galt einer Periode im August 1997, in der bodennahe Ozonbeobachtungen in Deutschland und den Niederlanden sowie drei Ozonsondenmessungen assimiliert wurden. Für alle Messungen ist eine Fehlerstandardabweichung von 10 % angenommen worden. Die Satellitenmessungen der Ozonsäule durch den GOME-Satelliten konnten nicht direkt assimiliert werden; sie dienten als die obere Randbedingung des Modells.

Das Assimilationsverhalten wurde mit Beobachtungen, die nicht in die Assimilation einbezogen wurden, bestimmt. Es konnte gezeigt werden, dass für diese Stationen eine deutliche Verbesserung der Modellrechnung erreicht wird (van Loon et al., 1998).

Es kam jedoch auch zum Ausdruck, dass die durch die Emission eingeführten Modellfehler nur einen Teil der Unsicherheiten erfassten. Die Analyse der noise-Parameter ergab bisher keine belastbaren Aussagen über systematische Abweichungen der Emission. Ein wesentlicher Einfluss der GOME-Daten auf die Ozonkonzentration am Erdboden konnte nicht nachgewiesen werden (van Loon et al., 1999).

4.4 Räumliche Zufallsprozesse

Die atmosphärischen Eigenschaften sind in Raum und Zeit variable Größen. Deren Untersuchung im Rahmen der Theorie der Zufallsprozesse beschränkt sich meist auf eine rein zeitliche oder rein räumliche Herangehensweise. Im ersten Fall handelt es sich um die Betrachtung von Zeitreihen, im zweiten um jene von Feldern, die häufig „Geostatistik“ genannt wird (Cressie, 1993).

Es gibt keine theoretischen Unterschiede bei der Behandlung von zeitlichen und räumlichen Zufallsprozessen. Dies führt dazu, dass in Lehrbüchern häufig nur Zeitreihen behandelt werden mit dem Hinweis, dass die Theorie auf räumliche Prozesse übertragbar sei. Es gilt jedoch, die folgenden Besonderheiten der Zufallsfelder zu beachten:

- Die unabhängige Variable der Zufallsfelder ist der mehrdimensionale Ortsvektor \mathbf{r} (2D oder 3D).
- Die Beobachtungen von Zeitreihen liegen fast immer in äquidistanten Abständen vor, während die räumliche Anordnung der Messstationen im Allgemeinen irregulär⁷⁸ ist.
- Begriffe wie Vergangenheit und Zukunft können für Felder nicht angewandt werden.

Die statistische Analyse ist eine Anwendung der Geostatistik und beruht damit vorrangig auf der Theorie räumlicher Zufallsprozesse. Für die Schätzung der Momente der räumlichen Prozesse wird jedoch ein Ensemble benötigt, das aus den Beobachtungen des Feldes zu verschiedenen Zeitpunkten gewonnen werden kann⁷⁹. Aufgrund des zeitlichen Zusammenhangs ist jedoch die notwendige Unabhängigkeit der Realisierungen nicht gewährleistet. Die Beschäftigung mit der zeitlichen Komponente des Zufalls-

⁷⁸ Dadurch ist die Anwendung der spektralen Betrachtungsweise in der Praxis eingeschränkt. Daten eines Eulerschen Modells bilden dafür eine geeignete Grundlage.

⁷⁹ Es handelt sich dann um ein klimatologisches Kovarianzmodell, mit dem eine höhere räumliche Spezifizierung möglich ist (siehe Kapitel 5.2.2)

prozesses ist deshalb notwendig, um die zeitliche und räumliche Variabilität angemessen zu trennen. Eine zusammenhängende Behandlung der Raum-Zeit-Prozesse bieten 4D-Datenassimilationsverfahren.

4.4.1 Beschreibung räumlicher Zufallsprozesse

Der räumliche Zufallsprozess $X(\mathbf{r})$ kann methodisch in eine mittlere Komponente $\mu(\mathbf{r})$, gegeben durch den Erwartungswert $E(X(\mathbf{r}))$, und eine stochastische Komponente $e(\mathbf{r})$ mit verschwindendem Erwartungswert zerlegt werden:

$$X(\mathbf{r}) = \mu(\mathbf{r}) + e(\mathbf{r}) \quad \{4.1\}$$

$$E(X(\mathbf{r})) = \mu(\mathbf{r}) \quad \text{und} \quad E(e(\mathbf{r})) = 0 \quad \{4.2\}$$

Für die Beschreibung der statistischen räumlichen Beziehungen⁸⁰ wird auf die Momente zweiter Ordnung zurückgegriffen. Handelt es sich um einen Gaußschen Zufallsprozess, so ist er mit den Momenten erster und zweiter Ordnung ausreichend charakterisiert. Es bieten sich hierfür zwei Größen an:

1. Die Autokovarianzfunktion $C = f(r_i, r_j)$ beruht auf den Produkten des stochastischen Anteils e :

$$C(\mathbf{r}_1, \mathbf{r}_2) = \text{Cov}(X(\mathbf{r}_1), X(\mathbf{r}_2)) = E(e(\mathbf{r}_1)e(\mathbf{r}_2)) \quad \{4.3\}$$

2. Die Strukturfunktion $2\gamma = f(r_i, r_j)$ beschreibt den Erwartungswert der quadratischen Differenz zwischen zwei Orten des Feldes \mathbf{r}_i , und \mathbf{r}_j . (Gandin, 1965):

$$2\gamma(\mathbf{r}_1, \mathbf{r}_2) = \text{Var}(X(\mathbf{r}_1) - X(\mathbf{r}_2)) = E((e(\mathbf{r}_1) - e(\mathbf{r}_2))^2) \quad \{4.4\}$$

Die Kovarianzfunktion $C(\mathbf{r}_1, \mathbf{r}_2)$ muß positiv definit sein, d. h. es gilt für jede Menge von Orten \mathbf{r}_i und willkürlichen Koeffizienten a_i die folgende Gleichung:

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{r}_i, \mathbf{r}_j) \geq 0 \quad \{4.5\}$$

Ein positive Spektraldichte ist dafür eine hinreichende Voraussetzung (Bochner-Theorem). Die stetige Differenzierbarkeit der Kovarianzfunktion für den Abstand Null ist Ausdruck der Differenzierbarkeit des zugrundeliegenden Prozesses (Schweschnikow, 1965).

⁸⁰ Persistenz

4.4.1.1 Statistisch motivierte Skalentrennung

Der stochastische Anteil $e(\mathbf{r})$ kann in unabhängige Komponenten aufgeteilt werden (Cressie, 1993), die im Sinne des meteorologischen Skalenbegriffs interpretierbar sind⁸¹. Die Kovarianzfunktionen der einzelnen Subprozesse addieren sich wegen der gegenseitigen Unabhängigkeit zu der des gesamten Prozesses:

$$\begin{aligned} e(\mathbf{r}) &= e_s(\mathbf{r}) + e_{ss}(\mathbf{r}) + e_\varepsilon(\mathbf{r}) \\ \text{Var}(e(\mathbf{r})) &= \sigma^2(\mathbf{r}) = \sigma_s^2(\mathbf{r}) + \sigma_{ss}^2(\mathbf{r}) + \sigma_\varepsilon^2(\mathbf{r}) \end{aligned} \quad \{4.6\}$$

Die Komponente $e_s(\mathbf{r})$ ist die räumliche Variabilität⁸², für die eine von Null verschiedene kontinuierliche Kovarianzfunktion bzw. Strukturfunktion existiert. Die subskalige räumliche Variabilität⁸³ $e_{ss}(\mathbf{r})$ besitzt eine Kovarianz- bzw. Strukturfunktion nur im Bereich der nicht aufgelösten räumlichen Unterschiede, d. h. für Distanzen, die kleiner als der minimale Abstand der Beobachtungen bzw. Modellgitterpunkte sind. Der Beitrag $e_\varepsilon(\mathbf{r})$ weist keine räumliche Korrelation auf. Er entspricht einem weißen Rauschen und wird häufig als Meßgerätefehler interpretiert. Zusammen mit dem subskaligen Anteil $e_{ss}(\mathbf{r})$ bildet er den sogenannten Beobachtungsfehler $e_o(\mathbf{r})$ (siehe Kapitel 4.5.2). Der Zufallsprozess $X_s(\mathbf{r})$ ohne den Beobachtungsfehler:

$$X_s(\mathbf{r}) = \mu(\mathbf{r}) + e_s(\mathbf{r}) \quad \{4.7\}$$

wird als der „glatte“ bzw. „rauschfreie“ Prozess von $X(\mathbf{r})$ bezeichnet. Er ist das eigentliche Ziel der Interpolationsbemühungen (siehe Kapitel 4.4.1.3).

Die konkrete Form der Zerlegung eines räumlichen Zufallsprozesses ist nicht eindeutig und hängt von der Problemstellung, der Datenverfügbarkeit und von der subjektiven Einschätzung ab.

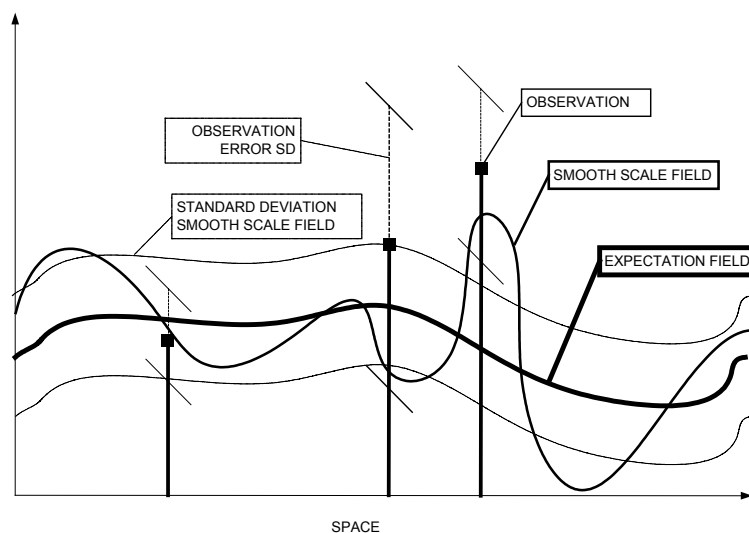


Abbildung 4.1 Schema eines räumlichen Zufallsprozesses und seiner Messung

⁸¹ Dieser Skalenbegriff ist mit dem Problem der Auflösung verbunden und geht nicht auf die physikalisch motivierten Skalen ein.

⁸² Smooth scale variation

⁸³ sub scale variation

4.4.1.2 Homogenität und Isotropie

Durch die Annahme der Homogenität⁸⁴ und Isotropie hinsichtlich der 1. und 2. Momente ist es möglich, die Komplexität der statistischen Eigenschaften des Zufallsfeldes $X(\mathbf{r})$ zu verringern. Homogenität hinsichtlich des Erwartungswertes bedeutet, dass er für das gesamte Feld einen konstanten Wert μ besitzt. Ist dieser Umstand nicht gegeben, so wird vor der weiteren statistischen Behandlung versucht, den mittleren Anteil $\mu(\mathbf{r})$ zu bestimmen. Die Abweichungen $X'(\mathbf{r}) = X(\mathbf{r}) - \mu(\mathbf{r})$ bilden dann ein Zufallsfeld mit homogenem mittleren Anteil.

Homogenität hinsichtlich der Kovarianzfunktion bzw. der Strukturfunktion bewirkt, dass diese nicht mehr von der Lage im Raum, sondern nur noch von dem Differenzvektor $\Delta\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ abhängt. Durch die Annahme von **Isotropie** wird die Richtungsabhängigkeit aufgegeben. Die Annahme der Homogenität und Isotropie führt dazu, dass die Kovarianz- und Strukturfunktion allein vom skalaren Abstand $|\Delta\mathbf{r}| = r$ der beiden Orte bestimmt ist. In diesem Fall, d. h. bei Homogenität und Isotropie, ist der statistische Apparat für Zeitreihen und räumliche Prozesse gleichartig.

Homogenität für den Erwartungswert und die Kovarianzfunktion, d. h. das erste und zweite Moment, wird als Homogenität 2. Ordnung⁸⁵ bezeichnet.

$$\begin{aligned} &\text{Homogenität 2. Ordnung} \\ &E(X(\mathbf{r})) = \mu \quad \text{und} \quad \text{Cov}(X(\mathbf{r}_1), X(\mathbf{r}_2)) = C(\Delta\mathbf{r}) \end{aligned} \quad \{4.8\}$$

Homogenität hinsichtlich der Strukturfunktion und ein konstanter Erwartungswert wird intrinsische Homogenität (intrinsic stationarity) genannt. Diese Annahme ist etwas schwächer, da die Erwartungswerte homogen, aber nicht bekannt sein müssen. Die Strukturfunktion für einen homogenen (stationären) Prozess wird als Variogramm⁸⁶ 2γ bezeichnet (Cressie, 1993).

$$\begin{aligned} &\text{"Intrinsische Homogenität"} \\ &E(X(\mathbf{r}_1) - X(\mathbf{r}_2)) = 0 \quad \text{und} \quad \text{Var}(X(\mathbf{r}_1) - X(\mathbf{r}_2)) = 2\gamma(\Delta\mathbf{r}) \end{aligned} \quad \{4.9\}$$

Unter der Annahme der Homogenität 2. Ordnung lässt sich jedoch eine einfache Beziehung zwischen Strukturfunktion und räumlicher Autokovarianzfunktion finden.

$$2\gamma(\Delta\mathbf{r}) = 2C(0) - 2C(\Delta\mathbf{r}) \quad \{4.10\}$$

Autokovarianzfunktion und Strukturfunktion sind bei Homogenität 2. Ordnung gleichwertig.⁸⁷ Variogramme können jedoch robuster geschätzt werden (Gneiting et al., 2000) und mit Gleichung {4.10} die Kovarianzfunktion festlegen (siehe 5.2.1). Der gegenläufige Charakter von Kovarianzfunktion und Variogramm (Strukturfunktion) wird bei großen Abständen deutlich. Geht man von einer verschwindenden Kovarianz-

⁸⁴ In Anlehnung an Zeitreihen auch häufig Stationarität (stationarity) genannt; Homogenität bezeichnet dann Stationarität und Isotropie.

⁸⁵ auch Stationarität im weiteren Sinne (weak stationarity) genannt.

⁸⁶ γ heißt Semivariogramm

⁸⁷ mit Einschränkung der schwächeren Forderung der „intrinsic stationarity“

funktion aus, so entspricht das Variogramm bei sehr großen Abständen der homogenen Varianz σ_s^2 des glatten Feldes.

Aufgrund der Wirkung der räumlich nichtkorrelierten Schwankung e_{ss} und der nicht auflösbaren Anteile e_ε (Gleichung {4.6}) ist die Varianz des glatten Prozesses $\sigma_s^2(\mathbf{r})$ um den Beitrag der Varianz des Beobachtungsfehlers $\sigma_o^2(\mathbf{r}) = \sigma_{ss}^2(\mathbf{r}) + \sigma_\varepsilon^2(\mathbf{r})$ erhöht:

$$C(r) = \begin{cases} \sigma_s^2 + \sigma_o^2 & r = 0 \\ \sigma_s^2 \rho(r) & r > 0 \end{cases} \quad \{4.11\}$$

Die Kovarianzfunktion des „kontaminierten“ Prozesse weist also einen Sprung am Abstand Null im Vergleich zum „glatten“ Prozess auf. Dieser Umstand wird genutzt, um mit Hilfe eines Kovarianzmodells die Varianz der Beobachtungsfehler zu schätzen (siehe Kapitel 5.2.6).

Eine komplexere Form der Kovarianzfunktion ist gegeben, wenn sich Homogenität und Isotropie nur auf die Korrelationsfunktion⁸⁸ $\rho(r)$ beziehen. Die räumlich variable Varianz $\sigma^2(\mathbf{r})$ wird zur Normierung der Kovarianzfunktion verwendet (Daley, 1991).

$$C(\mathbf{r}_1, \mathbf{r}_2) = \begin{cases} \sigma_s^2(\mathbf{r}) + \sigma_o^2(\mathbf{r}) & r = 0 \\ \sigma_s(\mathbf{r}_1)\sigma_s(\mathbf{r}_2)\rho(r) & r > 0 \end{cases} \quad \{4.12\}$$

4.4.1.3 Interpolation

Die räumliche Interpolation, d. h. die Prädiktion des unbekanntes Wertes $X(\mathbf{r}_0)$ aus $i=1, n$ bekannten Werten $X(\mathbf{r}_i)$, ist eine der Hauptanwendungen der räumlichen Statistik. Sie ist eine objektive Methode der Geowissenschaften, mit der aus Beobachtungswerten ein zusammenhängendes Feld gewonnen wird. Die Interpolation bewirkt jedoch immer eine Glättung, da nur der Prozess ohne den räumlich unkorrelierten Anteil $\varepsilon(\mathbf{r})$ prognostiziert werden soll.

Der verallgemeinerte Ansatz für die statistische Interpolation besteht darin, den Erwartungswert (Minimale Varianz) oder den wahrscheinlichsten Wert (maximum likelihood) der verbundenen Wahrscheinlichkeit $p(X(\mathbf{r}_0) | X(\mathbf{r}_i), i = 1, n)$ zu finden.

Daraus leiten sich zwei allgemeine Teilaufgaben ab: 1. Die Schätzung der statistischen Eigenschaften des Zufallsfeldes $X(\mathbf{r})$ und 2. die Bestimmung des konkreten Feldwertes $X(\mathbf{r}_0)$ mit Hilfe der bekannten Beobachtung $X(\mathbf{r}_i)$ unter Verwendung der Ergebnisse von 1. Der erste Punkt umfasst meist die Schätzung von Erwartungswert und Autokovarianzfunktion. Der zweite Schritt wird zur Unterscheidung vom ersten auch im geostatistischen Kontext häufig Prädiktion genannt. Die Qualität der Interpolation ist wesentlich von der Richtigkeit der Schätzung der Momente bestimmt.

Der optimale Prädiktor minimiert eine Kostenfunktion, die den Interpolationsfehler, also die Abweichung des interpolierten Werts vom wahren Wert bemißt. In dieser Arbeit wird die Schätzung über die Minimierung der Varianz gewonnen:

⁸⁸ In den meisten Darstellungen der Optimalen Interpolation (Daley, 1991 und Gandin, 1965) werden Homogenität und Isotropie angenommen. Die Interpolationsgleichungen werden nur in Form von Korrelationen entwickelt. Die allgemeinere Darstellung mit Kovarianzen in dieser Arbeit wird verwendet, um die Zusammenhänge zwischen OI 3/4D-VAR und Kalman-Filter herauszuarbeiten.

$$E\left(\left(X_p(\mathbf{r}) - X(\mathbf{r})\right)^2\right) \Rightarrow \min \quad \{4.13\}$$

Diese Prädiktion⁸⁹ ist biasfrei und linear. Der gesuchte Wert ergibt sich aus dem Mittelwert $\mu(\mathbf{r})$ und einer linearen Kombination der beobachteten Werte $X(\mathbf{r}_i)$, wobei für X Gaußscher Zufallsprozess⁹⁰ angenommen wird. $X_p(\mathbf{r}_0)$ repräsentiert den „glatten“ Prozess“ (siehe 4.4). Die Gewichte k_i hängen nicht von den aktuellen Beobachtungswerten ab. Aus der Minimierung des quadratischen Interpolationsfehlers von $X_p(\mathbf{r}_0)$ werden die Koeffizienten k_i aus der Kovarianzfunktion oder dem Variogramm gewonnen. Die Lösung für die k_i für die statistische Analyse ist Inhalt des Kapitels 4.5.4.

In der statistischen Analyse und der Datenassimilation wird vorrangig mit Autokovarianzen gearbeitet. Sind Variogramme die Grundlage für die räumliche Interpolation, dann spricht man von der Kriging-Methode. Beruht die Interpolation auf Kovarianzen, dann muss nicht von Homogenität ausgegangen werden. Sie ist jedoch häufig die Voraussetzung, um die Kovarianzfunktion aus vorgegebenem Datenmaterial abzuleiten (siehe Kapitel 5).

Es existieren verschiedene Kriging-Entwicklungen (Cressie, 1993). Im Falle von „Einfachem Kriging“ ist $\mu(\mathbf{r})$ bekannt, beim „Gewöhnlichen Kriging“ ist $\mu(\mathbf{r})$ unbekannt, aber konstant. Man spricht von „Universellem Kriging“, wenn der Mittelwert räumlich variabel ist und durch eine Funktionsfolge parametrisiert wird. „Block-Kriging“ ist ein Verfahren, um aus Beobachtungen Flächenmittelwerte oder Gitterzellenmittelwerte zu interpolieren. „Indikator-Kriging“ ist eine nichtlineare Interpolation, mit der lokale Histogramme geschätzt werden.

4.4.2 Raum-Zeit-Prozesse

Für die zusammenhängende Bearbeitung der Raum-Zeit-Zufallsprozesse gibt es nur wenige Beispiele (Bilonick, 1985). Zur Vereinfachung geht man von einer Separation in eine räumliche und zeitliche Komponente aus (Gneiting und Schlather, 2001). Damit werden die Raum- und Zeitkomponente als unabhängig voneinander angesehen:

$$\begin{aligned} \mu(\mathbf{r}, t) &= \mu_r(\mathbf{r})\mu_t(t) \\ e(\mathbf{r}, t) &= e_r(\mathbf{r})e_t(t) \\ &\text{mit} \\ \text{Cov}(X(\mathbf{r}_1, t_1)X(\mathbf{r}_2, t_2)) &= C_r(\mathbf{r}_1, \mathbf{r}_2)C_t(t_1, t_2) \end{aligned} \quad \{4.14\}$$

oder

$$\begin{aligned} \mu(\mathbf{r}, t) &= \mu_r(\mathbf{r}) + \mu_t(t) \\ e(\mathbf{r}, t) &= e_r(\mathbf{r}) + e_t(t) \\ &\text{mit} \\ \text{Cov}(X(\mathbf{r}_1, t_1)X(\mathbf{r}_2, t_2)) &= C_r(\mathbf{r}_1, \mathbf{r}_2) + C_t(t_1, t_2) \end{aligned} \quad \{4.15\}$$

⁸⁹ auch BLUE („best linear unbiased estimate“) genannt

⁹⁰ Bei der Varianzminimierung für nicht-normalverteilte Prozesse ist diese Schätzung die optimalste in der Klasse der linearen Schätzungen.

Dieses Vorgehen erscheint geeignet, wenn eine ortsunabhängige zeitliche Periodizität, wie z. B. der Tagesgang, erfasst werden soll. Die Wechselbeziehung der raum- und zeitabhängigen Prozesse kann damit nicht wiedergegeben werden.

Ansätze ohne Separation entstehen z. B. durch die Bildung einer einheitlichen Raum-Zeitmetrik. Dabei wird zu dem räumlichen Abstand der zeitliche Abstand in skaliert Form addiert. Die Skalierungsgröße der Zeit kann als eine konstante Driftgeschwindigkeit physikalisch interpretiert werden.

Auch wenn man sich auf Zeitreihenanalyse oder Geostatistik beschränkt, kann die fehlende Betrachtung der Raum-Zeit-Persistenz Fehlinterpretationen von statistischen Tests zur Folge haben, wenn die Unabhängigkeit der Stichprobe vorausgesetzt wurde. Aus dem Blickwinkel der Geostatistik führen weiterhin die Nichtbeachtung von zeitlicher Persistenz oder die Periodizität zu Fehlern in der Schätzung der Momente des räumlichen Prozesses.

4.4.2.1 Schätzung der räumlichen Momente aus Zeitreihen

In der statistischen Interpolation bzw. Analyse bestimmt sich der unbekannte Wert des Zufallsfeldes aus den bekannten Beobachtungen des gleichen Termins. Dafür wird nur die räumliche Struktur von Erwartungswert und Kovarianzfunktion benötigt (siehe 4.4.1.3).

Der zeitliche Zusammenhang ist jedoch bei der Bestimmung der Kovarianzfunktion von Bedeutung, wenn dies mit Hilfe eines klimatischen Kovarianzmodells (siehe 5.2.2) erfolgt. Es werden dafür Felder zu verschiedenen Zeitpunkten als Realisierungen des selben räumlichen Prozesses aufgefasst. Die Zeitreihen der Beobachtungen dienen der Schätzung von stationspaarbezogenen räumlichen Kovarianzwerten. Das klimatische Kovarianzmodell erlaubt aufgrund des erweiterten Ensembles eine höhere räumliche Strukturierung⁹¹.

Aufgrund der zeitlichen Persistenz können die Felder zu verschiedenen Zeiten nicht ohne weiteres als unabhängige Realisierung des räumlichen Prozesses angesehen werden (Zwiers und von Storch, 1995). Es bedarf einer Trennung der zeitlichen und räumlichen Kovarianzanteile durch Filterung der Daten (siehe Kapitel 5.2.2.1). Sie entspricht einer zeitlichen Hochpassfilterung, die durch Differenzbildung mit einer geglätteten, die zeitliche Variation widerspiegelnden Datenreihe erfolgt. Das Ensemble der Residuen wird dann zur Schätzung der räumlichen Kovarianz herangezogen. Die Zeitreihe der Residuen muss dazu ergodisch sein.

Im Gegensatz zum klimatischen wird das terminbezogene Kovarianzmodell nur aus den Daten der vorliegenden Realisierung unter der Annahme von räumlicher⁹² Ergodizität⁹³ gewonnen. Das Konzept der Ergodizität für räumliche Zufallsprozesse ist nur unvollständig ausgearbeitet⁹⁴ (Cressie, 1993). Analog zur Stationarität ergodischer Zeitreihen ist dabei von einer räumlichen Homogenität auszugehen. Weiterhin gilt die physikalisch sinnvolle Forderung der verschwindenden Kovarianzfunktion für große

⁹¹ Dieser Ansatz ist die Grundlage für die Entwicklung von nichtstationären Kovarianzmodellen (Sampson et al., 2001).

⁹² Das Konzept der räumlichen Ergodizität ist nicht vollständig ausgearbeitet.

⁹³ Die Folge von Ergodizität ist, dass die Parameter des Prozesses aus einer Realisierung gewonnen werden können.

⁹⁴ Insbesondere für die Turbulenzforschung ist der Zusammenhang zwischen Eulerscher (Zeitmittel) und Lagrangescher (Ensemblemittel) Kovarianzfunktion interessant (Roedel, 1992, S.246).

Abstände. Die Plausibilität der klimatischen Kovarianzmodelle kann mit Hilfe der terminbezogenen überprüft werden.

4.4.2.2 Ergodizität von Zeitreihen

Um für die klimatischen Kovarianzmodelle die Erwartungswerte aus zeitlichen Mittelwerten zu gewinnen, müssen die entsprechenden Zeitreihen ergodisch sein (Schweschnikow, 1965). Der Erwartungswert einer ergodischen und damit stationären Gaußschen Zeitreihe $X(\mathbf{r},t)$ ergibt sich für den Ort \mathbf{r} aus dem zeitlichen Mittelwert:

$$E(X(\mathbf{r})) = \lim_{t \rightarrow \infty} \frac{1}{2t} \int_{-t}^t (X(\mathbf{r},t)) dt \quad \{4.16\}$$

Die räumliche Kovarianz $Cov(X(\mathbf{r}_1), X(\mathbf{r}_2))$ für zwei Orte \mathbf{r}_1 und \mathbf{r}_2 kann aus den zwei ergodischen Gaußschen Zeitreihen $X(\mathbf{r}_1, t=t_1 \dots t_2)$ und $X(\mathbf{r}_2, t=t_1 \dots t_2)$ über deren Kreuzvarianz gebildet werden:

$$Cov(X(\mathbf{r}_1), X(\mathbf{r}_2)) = \lim_{t \rightarrow \infty} \frac{1}{2t} \int_{t_2}^{t_1} (X(\mathbf{r}_1, t) - E(X(\mathbf{r}_1, t)))(X(\mathbf{r}_2, t) - E(X(\mathbf{r}_2, t))) dt \quad \{4.17\}$$

Für die Ergodizität einer stationären Zeitreihe lässt sich die hinreichende Bedingung angeben, dass für große Zeitdifferenzen Δt die zeitliche Kovarianzfunktion C_t gegen Null geht (Schweschnikow, 1965, S. 167):

$$\lim_{\Delta t \rightarrow \infty} C_t(\Delta t) = 0 \quad \{4.18\}$$

Taubenheim (1969) erläutert die Bedingung für die zeitliche Ergodizität in der spektralen Betrachtungsweise. Danach darf das Spektrum der ergodischen Zeitreihe keine diskreten Spektrallinien enthalten, denn diese führen zu einer Schwingung der Autokovarianzfunktion, die auch für große Zeitabstände erhalten bleibt⁹⁵.

Für die Verwendung von Zeitreihen zur Berechnung der räumlichen Kovarianz für zwei Orte ist die Folge mangelnder Ergodizität sofort einsehbar. Weist das Feld an zwei weit entfernten Orten eine periodische Schwingung⁹⁶ bzw. einen linearen Trend auf und werden die lokalen Erwartungswerte durch zeitliche Mittelbildung gewonnen, so wird eine empirische Kovarianz geschätzt, die keine Entsprechung in der räumlichen Dimension hat. Um dies zu vermeiden, muss der periodische Anteil entfernt bzw. dem mittleren Anteil $\mu(\mathbf{r},t)$ zugeordnet werden.

Trägt man die zeitlichen Kreuzvarianzwerte als Schätzung der räumlichen Kovarianz in Abhängigkeit vom Abstand auf, so sollten die Werte für große Abstände gegen Null streben. Damit besteht die Möglichkeit, die Ergodizität der Zeitreihen $X(\mathbf{r},t)$ zu testen.

⁹⁵ Die Kovarianzfunktion ist die Fouriertransformierte des Energiespektrums (Wiener-Chinchin-Theorem).

⁹⁶ z. B. der durch den Verkehr bedingte Tagesgang der Immission

4.5 Theorie der statistischen Analyse und Datenassimilation

4.5.1 Grundprinzip

Die Analyse stellt die bestmögliche Wiedergabe des Feldes von atmosphärischen Variablen in einer regelmäßigen räumlichen Auflösung dar. Der Begriff Analyse wird dabei auch für die Methode zur Gewinnung der Analyse aus Beobachtungsdaten verwendet. Man spricht von objektiver Analyse, wenn numerische Verfahren für die Analyse angewendet werden (Gandin, 1965).

Die Auflösung der Analyse führt zu einer unteren Schranke für die kleinste darstellbare Skala. Weil die Beobachtungen auch durch nichtaufgelöste Prozesse beeinflusst werden, ist die angemessene Separation⁹⁷ der relevanten Skalen ein wichtiges Problem im Rahmen der Analyseverfahren.

Die modernen Analyseverfahren beruhen auf statistischen Ansätzen zur Interpolation (siehe Kapitel 4.4.1, Interpolation). Sie zeichnen sich dadurch aus, dass sie neben den Beobachtungen \mathbf{y} auf einer unabhängig vorgegebenen Feldverteilung \mathbf{x}_b basieren. Diese wird als Background⁹⁸ bezeichnet und kann als eine erste Vorgabe für die Analyse betrachtet werden. Es handelt sich dabei meist um ein von einem Eulerschen Modell berechnetes Feld. Ein klimatologisch bestimmtes Feld oder die Analyse eines vorangegangenen Zeitpunkts sind weitere Möglichkeiten für den Background. Die räumliche Auflösung⁹⁹ stimmen in Background und Analyse überein.

Durch die Methodik, einen vorgegebenen Background durch Messungen zu verändern, werden neben der Information aus den Beobachtungen zusätzliche Informationsquellen¹⁰⁰ für die zu erstellende Analyse herangezogen. Dies ist von Vorteil, wenn die Messungen aufgrund ihrer begrenzten Repräsentativität keine befriedigenden Rückschlüsse auf die gesamte Feldverteilung zulassen, wie es in beobachtungsarmen Gebieten der Fall ist. Dies gilt auch lokal, falls die Messungen Skalen wiedergeben, die in der Analyse nicht enthalten sein sollen.

Im Fall der Datenassimilation ist der Background das von einem Modell berechnete Feld. Der Background \mathbf{x}_b ist somit ein Teil oder der gesamte Zustandsvektor des dynamischen Modells. Die Beobachtungen fließen bei der Datenassimilation in die Modellrechnung ein, indem die Analyse den berechneten Zustandsvektor des Modells ersetzt. Für die statistische Analyse ist die verbundene Wahrscheinlichkeit von \mathbf{x} bei gegebenen \mathbf{y} und \mathbf{x}_b der theoretische Ausgangspunkt:

$$p(\mathbf{x} | \mathbf{y}, \mathbf{x}_b) \quad \{4.19\}$$

⁹⁷ Filterung

⁹⁸ Der Background darf im Zusammenhang mit der Immissionsanalyse nicht mit einem Hintergrund im Sinne einer unbelasteten bzw. natürlichen Immissionssituation verwechselt werden.

⁹⁹ Falls in der Analyse vom Modell nicht wiedergegebene Skalen dargestellt werden sollen, so wird formal die Auflösung des Backgrounds erhöht. Die Auflösung ist eine technische, die kleinste wiedergegebene Skala eine inhaltliche Fragestellung. Die Auflösung stellt nur die untere Schranke für die kleinste wiedergegebene Skala dar.

¹⁰⁰ Da die Anzahl der Messwerte im Allgemeinen wesentlich geringer als die der Analysepunkte ist, stellt die Analyse im mathematischen Sinne ein unterbestimmtes Problem dar. Zu dessen Lösung werden weitere Informationsquellen (d. h. der Background) benötigt.

Die Analyse \mathbf{x}_A kann sich aus dem Erwartungswert {4.20} oder dem Modalwert {4.21} ergeben. Der erste Fall entspricht einer Schätzung nach dem Kriterium der minimalen Varianz, der zweite einem modifizierten¹⁰¹ Maximum likelihood Ansatz (Lorenc, 1986):

$$\mathbf{x}_A^{mv} = \int \mathbf{x} p(\mathbf{x} | \mathbf{y}, \mathbf{x}_B) d\mathbf{x} \quad \{4.20\}$$

$$\mathbf{x}_A^{ml} = \mathbf{x} \quad \text{für} \quad p(\mathbf{x} | \mathbf{y}, \mathbf{x}_B) = \max \quad \{4.21\}$$

Datenassimilaton

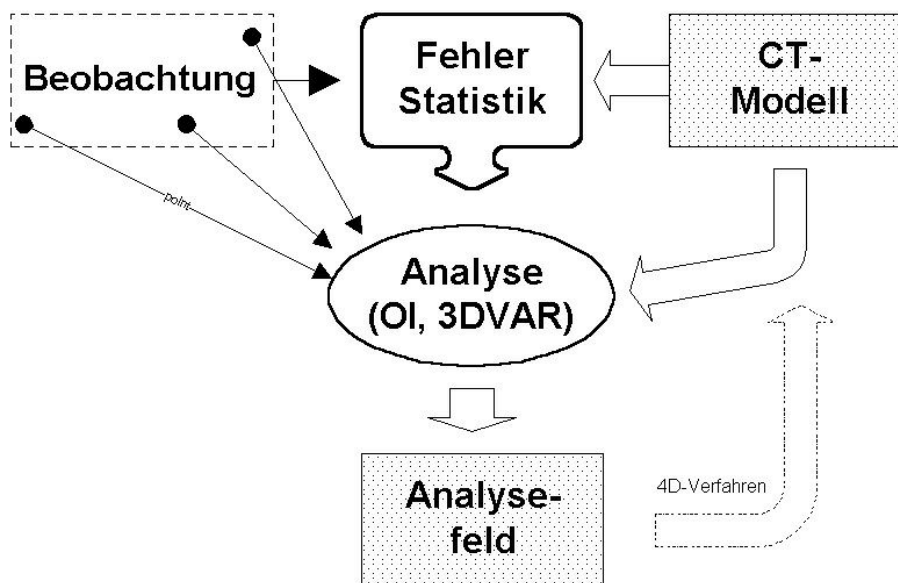


Abbildung 4.2 Schematisches Grundprinzip der Analyse und Datenassimilation

4.5.2 Definition der Fehlermaße

Die Zustandsvektoren bzw. die für Gitterpunkte I gegebenen Felder werden mit dem Vektor \mathbf{x} bezeichnet. Es sind dies der Background \mathbf{x}_B , die Analyse \mathbf{x}_A und das nur methodisch eingeführte „wahre“ Feld \mathbf{x}_{true} . Die Analyse wird durch eine Korrektur des Backgrounds gewonnen. Sie zielt damit auf die Abweichungen¹⁰² des Backgrounds hinsichtlich des „wahren“ Feldes \mathbf{x}_{true} , das durch \mathbf{x}_A geschätzt wird.

Die Abweichung zwischen der Analyse und dem wahren Wert ist der zu minimierende Analysefehler \mathbf{e}_A :

$$\mathbf{e}_A = \mathbf{x}_A - \mathbf{x}_{true} \quad \mathbf{x}_A \in R^N \quad \{4.22\}$$

¹⁰¹ “Maximum a posteriori probability estimation” Todling R. (1999)

¹⁰² Damit ergibt sich ein enger Zusammenhang zwischen Analyse und der Modellevaluierung.

Der Fehlervektor des Backgrounds \mathbf{e}_B stellt die Abweichung des Modellzustandsvektors \mathbf{x}_B von dem hypothetischen wahren Zustandsvektor \mathbf{x}_{true} dar:

$$\mathbf{e}_B = \mathbf{x}_B - \mathbf{x}_{true} \quad \mathbf{x}_B, \mathbf{x}_{true} \in R^N, (x_{iB} = 1, N) \quad \{4.23\}$$

Die unregelmäßig verteilten Beobachtungen i bilden den Vektor \mathbf{y} . Die Fehler der Beobachtungen \mathbf{e}_O sind die Differenzen zwischen den Beobachtungen \mathbf{y} und dem mit dem Beobachtungsoperator H transformierten Zustandsvektor \mathbf{x}_{true} .

$$\begin{aligned} \mathbf{e}_O &= \mathbf{y} - H(\mathbf{x}_{true}) && \text{Beobachtungsfehler} \\ \mathbf{y} &\in R^m, (y_i, i = 1, m) && \\ H &\in R^{N \times m} && \text{Beobachtungsoperator} \end{aligned} \quad \{4.24\}$$

Der Beobachtungsfehler besteht aus dem Messgerätefehler ($\mathbf{y} - \mathbf{y}_{true}$) und einem Fehler, der durch die Ungenauigkeit des Beobachtungsoperators H hervorgerufen wird. Dieser Anteil kann mit dem Fehler in der Repräsentativität der Messung für die Modellvariable interpretiert werden (siehe Kapitel 4.5.3)

Die aktuellen Werte von Modell- und Beobachtungsfehler sind unbekannt, da die „wahren“ Werte naturgemäß nicht bekannt sind. Die statistische Analyse beruht jedoch auf der Annahme, dass probabilistische Angaben über beide Fehler gemacht werden können. Es wird davon ausgegangen, dass die Fehler von Background und Beobachtung als Gaußsche Zufallsprozesse behandelbar und damit durch den Erwartungswert und die räumliche Kovarianzfunktion ausreichend beschrieben sind.

Die Schätzung dieser Größen aus Modell- und Beobachtungsdaten ist der Inhalt von Kapitel 5. Es wird dabei auf die Beobachtungsinkremente \mathbf{e}_{OB} zurückgegriffen, die im Gegensatz zu den Fehlern direkt bestimmbar sind:

$$\mathbf{e}_{OB} = \mathbf{y} - H(\mathbf{x}_B) \quad \text{Beobachtungsinkrement} \quad \{4.25\}$$

Die räumliche Beziehung der Fehler zwischen den einzelnen Elementen des Modellzustandsvektors bzw. zwischen den einzelnen Beobachtungen wird durch die Kovarianzmatrizen \mathbf{B} und \mathbf{R} erfasst.

Genauso wie der Zustandsvektor \mathbf{x} die Diskretisierung eines kontinuierlichen Feldes darstellt, so ist die Matrix \mathbf{B} die Diskretisierung der räumlichen Kovarianzfunktion C (siehe Kapitel 4.4). Die Beobachtungsfehler werden als unkorreliert angenommen, und so ergibt sich \mathbf{R} als Diagonalmatrix:

$$\mathbf{B} = E(\mathbf{e}_B \mathbf{e}_B^T) = E((\mathbf{x}_B - \mathbf{x}_{true})(\mathbf{x}_B - \mathbf{x}_{true})^T) \quad \{4.26\}$$

$$\mathbf{R} = E(\mathbf{e}_O \mathbf{e}_O^T) = E((\mathbf{y} - H\mathbf{x}_{true})(\mathbf{y} - H\mathbf{x}_{true})^T) = \mathbf{I}E(\mathbf{e}_O^2) \quad \{4.27\}$$

Beide Matrizen sind symmetrisch und positiv definit¹⁰³. Das Quadrat¹⁰⁴ der Anzahl der Elemente des Modellzustandsvektors bzw. der Beobachtungen ist die Anzahl der Elemente in den entsprechenden Matrizen.

Die Hauptdiagonalelemente von \mathbf{B} und \mathbf{R} sind die Varianzen der entsprechenden Fehler σ_{BI}^2 und σ_{Oj}^2 für jeden Gitterpunkt I bzw. jede Beobachtung j .

Analog zu den Matrizen \mathbf{B} und \mathbf{R} kann die Kovarianzmatrix \mathbf{A} des Analysefehlers definiert werden:

$$\mathbf{A} = \mathbf{E}(\mathbf{e}_A \mathbf{e}_A^T) = \mathbf{E}((\mathbf{x}_A - \mathbf{x}_{\text{true}})(\mathbf{x}_A - \mathbf{x}_{\text{true}})^T) \quad \{4.28\}$$

4.5.3 Beobachtungsoperator

Im Allgemeinen entsprechen die analysierten Modellvariablen \mathbf{x} nicht vollkommen den Beobachtungen \mathbf{y} : Die Beobachtungsorte stimmen meist nicht mit den Gitterpunkten überein bzw. die Modellvariablen besitzen einen anderen räumlichen und zeitlichen Mittelungscharakter als die Beobachtungen. Darüber hinaus können die Modellvariablen und Beobachtungen keine direkte Entsprechung haben, wie dies bei Satellitenbeobachtungen oder bei der hier vorgestellten aktiven Assimilation für PM10 (siehe Kapitel 6.1.7) der Fall ist. Die ursprünglichen Strahlungsmesswerte müssen erst in die prognostische Modellvariable umgerechnet werden bzw. die PM10-Masse muss auf die modellierten PM10-Bestandteile verteilt werden (siehe Kapitel 1.4.2).

Der Übergang¹⁰⁵ aus dem Raum der Zustandsvektoren \mathbf{x} (Modell oder Analysefeld) in den der Beobachtungen wird formal durch den Beobachtungsoperator \mathbf{H} geleistet (Bouttier und Courtier, 1999):

$$\mathbf{y} \leftrightarrow H(\mathbf{x}) \quad \{4.29\}$$

Für die Ableitung der Gleichungen der Gewichte \mathbf{K} durch eine Minimierung der Varianz des Analysefehlers (siehe Kapitel 4.5.4) wird die Existenz eines linearisierten Beobachtungsoperators \mathbf{H} gefordert¹⁰⁶. Das bedeutet, dass dieser Beobachtungsoperator auch auf Differenzen bzw. Fehler des Zustandvektors \mathbf{x}_B anwendbar ist. \mathbf{H} ergibt sich aus einer Taylor-Reihenentwicklung von H

:

$$H(\mathbf{x}_B + \Delta\mathbf{x}) \approx H(\mathbf{x}_B) + \mathbf{H}\Delta\mathbf{x} \quad \text{mit} \quad \mathbf{H} = \left. \frac{\partial H}{\partial \mathbf{x}} \right|_{\mathbf{x}_B} \quad \{4.30\}$$

¹⁰³ Diese aus der Autokorrelation über das Bochner-Chinchin Theorem (eine Funktion ist positiv definit, wenn ihr Spektrum positiv ist) ableitbare Eigenschaft muss insbesondere bei der numerischen Behandlung der Backgroundfehlermatrix beachtet werden.

¹⁰⁴ Im Falle eines REM/Calgrid-Feldes (80*96 Gitterpunkte) hat sie 58 982 400 Elemente.

¹⁰⁵ z. B. falls Modellvariablen und Beobachtungen keine direkte Entsprechung haben, wie es bei aus Satellitenbeobachtungen abgeleiteten Temperaturprofilen der Fall ist.

¹⁰⁶ Im Rahmen der Datenassimilation mit 3/4DVAR ist diese Forderung nicht nötig. Darum eignet sich dieses Verfahren besonders für komplexe Beobachtungsoperatoren, wie es bei Satellitendaten der Fall ist.

Durch die Linearität ist \mathbf{H} mit der Erwartungswertbildung vertauschbar, so dass \mathbf{H} auf die Kovarianzmatrix \mathbf{B} angewendet werden kann. Die entsprechenden Terme lauten \mathbf{HBH}^T und \mathbf{HB} .

Die elementarste Anwendung des Beobachtungsoperators ist die räumliche Verschiebung: Mit Hilfe des Beobachtungsoperators H oder \mathbf{H} wird der Backgroundwert oder seine Varianz am Gitterpunkt J auf den Ort j der vorliegenden Beobachtung transformiert, um die Beobachtungsinckremente bilden zu können. Die Beobachtung j liegt dabei in der zugehörigen Gitterbox J . Die \mathbf{HBH}^T ist damit die Kovarianzmatrix zwischen den Orten der Beobachtung, \mathbf{HB} zwischen den Beobachtungsorten und den Gitterpunkten. $\mathbf{HBH}^T + \mathbf{R}$ entspricht demzufolge der Kovarianzmatrix der Beobachtungsinckremente \mathbf{e}_{OB} .

Eine numerische Formulierung des Beobachtungsoperators H für den Skalenübergang ist meist nicht bekannt. Eine statistische Beschreibung des Skalenunterschieds kann jedoch mit dem Ansatz {4.6} erfolgen. Mit ihm wird die subskalige Komponente e_{ss} mit dem unkorrelierten Messfehler ε zum Beobachtungsfehler zusammengefasst. Der Beobachtungsfehler ist damit auch Ausdruck der Unzulänglichkeit von H . Die Berücksichtigung des Skalenunterschiedes erfolgt somit durch die Erhöhung der Varianz des Beobachtungsfehlers, der Bestandteil der Matrix \mathbf{R} ist.

Der statistische Ansatz für die Skalentrennung geht von einer „Kontaminierung“ des „glatten“ Prozesses durch eine subskalige Komponente mit verschwindendem Mittelwert aus. Dieses Vorgehen ist für Immissionsbeobachtungen nur teilweise gerechtfertigt, da die subskalige Komponente meist ein zusätzlicher und systematischer Aufschlag aufgrund kleinräumiger Emissionen ist. Der auf die Kovarianzmatrix angewendete linearisierte Beobachtungsoperator \mathbf{H} enthält diesen Aufschlag nicht. Die entsprechende Emissionssituation kann aus den in Kapitel 2 bestimmten Immissionsregimes geschlussfolgert werden. Sie werden im folgenden verwendet, um eine sinnvolle Klassifizierung innerhalb des Kovarianzmodells einzuführen (siehe 5.1.3).

4.5.4 Analyse durch Varianzminimierung – Optimale Interpolation

Für die genaue Herleitung der Analysegleichung durch eine Minimum-Varianz-Schätzung mit Hilfe der in Kapitel 4.5.1 angedeuteten Prinzipien wird auf Todling (1999) oder Lorenc (1986) verwiesen. Die Herleitung wird vereinfacht, wenn man von folgendem linearen Ansatz ausgeht:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H(\mathbf{x}_b)) \quad \{4.31\}$$

Die Analyse wird dabei aus dem Background \mathbf{x}_b und den Beobachtungsinckrementen $(\mathbf{y} - H\mathbf{x}_b)$ gebildet. Die Matrix \mathbf{K} legt die Gewichte für den Einfluss der Beobachtungsinckremente fest.

Die Gleichung für die Gewichte {4.35} wird aus der Minimierung¹⁰⁷ der Gesamtvarianz des Analysefehlers σ^2_A gewonnen. Die Gesamtvarianz σ^2_A ist die Spur Tr der Kovarianzmatrix des Analysefehlers \mathbf{A} {4.28}.

$$Tr(\mathbf{A}) = \sum_{i=1}^N \sigma_{iA}^2 = \sum_{i=1}^N E((x_{iA} - x_{itrue})^2) \rightarrow \min \quad \{4.32\}$$

Bei der Minimierung werden folgende Voraussetzungen gemacht:

1. Kein Bias zwischen den Beobachtungen und dem Modellfeld:

$$E(\mathbf{e}_B) = E(\mathbf{e}_O) = \mathbf{0} \quad \{4.33\}$$

2. Keine Korrelation (Kovarianz) zwischen den Fehlern von Modell und Beobachtungen:

$$E(\mathbf{e}_O \mathbf{e}_B^T) = \mathbf{0} \quad \{4.34\}$$

3. Existenz eines linearisierten Beobachtungsoperators, der auf die Kovarianzmatrizen des Backgroundfehlers \mathbf{B} anwendbar ist.

Unter diesen Voraussetzungen ergibt sich für \mathbf{K} das folgende Gleichungssystem aus den Kovarianzmatrizen von Background \mathbf{B} und Beobachtung \mathbf{R} .

$$\begin{aligned} (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}) \mathbf{K}^T &= \mathbf{H}\mathbf{B} \\ \text{bzw. Invertierbarkeit} & \\ \mathbf{K} &= \mathbf{H}\mathbf{B}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \end{aligned} \quad \{4.35\}$$

Setzt man Gleichung {4.31} in Gleichung {4.28} ein, so ergibt sich die Matrixgleichung für die Kovarianzmatrix des Analysefehlers \mathbf{A} :

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \quad \{4.36\}$$

Mit den optimalen Gewichten \mathbf{K} vereinfacht sich der Ausdruck für den Analysefehler:

$$\mathbf{A} = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B} \quad \{4.37\}$$

4.5.5 Vereinfachung und Diskussion der Analysegleichungen

Die Struktur der Analysegleichungen {4.31} und {4.35} kann durchschaubarer gemacht werden, wenn die Analyse nur für einen Wert x_{AJ} am Punkt J formuliert wird. Eine weitere Vereinfachung ergibt sich, wenn nur eine Auswahl von $i=1, n_J$ Stationen zur Interpolation am Punkt J beiträgt. Der Backgroundwert x_{Bi} an den Beobachtungsorten

¹⁰⁷ Gleichwertig ist eine maximum-likelihood-Schätzung der Wahrscheinlichkeitsfunktion bei Gaußschen pdfs (Lorenz, 1986).

ten i wird durch den Beobachtungsoperator¹⁰⁸ $H(x_{Bi})$ aus dem zugehörigen Gitterpunktswert gewonnen:

$$x_{AJ} = x_{BJ} + \sum_{i=1}^{n_J} k_{Ji} (y_i - x_{Bi}) \quad \{4.38\}$$

Diese Form der Analyse ist von Gandin (1965) unter dem Begriff **Optimale Interpolation** in die Meteorologie eingeführt worden. Die Schätzung der räumlichen Kovarianzfunktionen beruht bei den meisten weiteren Anwendungen auf homogenen und meist isotropen Verhältnissen.

Es erscheint sinnvoll, nur Stationen in der Nähe des Analysepunktes zu verwenden. Das Verschwinden der Kovarianz **HB** kann als ein formales Kriterium für die Auswahl gelten. Es zeigt sich jedoch häufig, dass eine viel geringere Anzahl¹⁰⁹ für die Analyse ausreicht, d. h. keine wesentlichen Veränderungen hinzukommen¹¹⁰. Durch die Daten-selektion wird der numerische Aufwand für die Analyse stark reduziert, da für jeden Gitterpunkt ein Gleichungssystem von deutlich geringerer Größe gelöst werden muss. Die Analyse zeichnet sich durch folgende Eigenschaften aus, die aus der Diskussion von Gleichung {4.39} begründet werden:

- Hohe Kovarianz des Backgrounds zwischen Beobachtungsort und Analysepunkt führen zu hohem Einfluss des zugehörigen Beobachtungsincrements.
- Hohe Kovarianz des Backgrounds zwischen zwei Beobachtungsorten verringert den Einfluss der zugehörigen Beobachtungsincremente.
- Hohe Varianz des Beobachtungsfehlers einer Messung führt zu einem geringen Einfluss des zugehörigen Beobachtungsincrements.

Durch die Stationsauswahl vereinfacht sich die Matrixgleichung {4.35} zu einem inhomogenen linearen Gleichungssystem n_J -ter Ordnung für die Gewichte k_{Ji} :

$$\begin{pmatrix} b_{11} + r_{11} & \cdots & b_{n1} \\ \vdots & \ddots & \vdots \\ b_{1n} & \cdots & b_{nn} + r_{nn} \end{pmatrix} \begin{pmatrix} k_{J1} \\ \vdots \\ k_{Jn} \end{pmatrix} = \begin{pmatrix} b_{J1} \\ \vdots \\ b_{Jn} \end{pmatrix} \quad \{4.39\}$$

Die k_{Ji} werden hauptsächlich durch die rechte Seite, d. h. die Kovarianzen des Backgroundfehlers b_{ji} zwischen den Beobachtungsorten i und dem Analysepunkt J bestimmt. Je ausgeprägter die Beziehung, ausgedrückt durch die Kovarianzen, zwischen Beobachtungsort und Interpolationspunkt ist, desto wichtiger wird das ermittelte Beobachtungsincrement für die Analyse.

Die Matrix der linken Seite beschreibt die Eigenschaften und Beziehungen zwischen den Beobachtungen. Dabei werden die Beobachtungsfehler und die Beziehung des

¹⁰⁸ Der Beobachtungsoperator ist hier eine bilineare Interpolation von den Gitterpunktswerten am Stationort.

¹⁰⁹ Gandin zeigt für das 500 hPa Geopotential, dass ca. 10 Stationen ausreichend sind.

¹¹⁰ Dies gilt nicht, wenn Gradienten etc. des Feldes betrachtet werden, da hier die von Punkt zu Punkt veränderliche Stationsauswahl zu Sprüngen führen kann.

Backgroundfeldes an den Beobachtungsorten berücksichtigt. Eine hohe Varianz des Beobachtungsfehlers $r_{ii} = \sigma_{O_i}^2$ der Messung i verringert den Einfluss der Beobachtung i für die Analyse. Die Stärke dieser Verringerung wird durch Varianz des Backgroundfehlers $b_{ii} = \sigma_{B_i}^2$ relativiert. Es lässt sich zeigen, dass das Verhältnis σ_{O_i} und σ_{B_i} für die Bewertung des Beobachtungsinkrements am Ort i entscheidend ist (siehe Kapitel 5.3.3.2).

Weiterhin trägt die Beobachtungsfehlervarianz zur numerischen Stabilität des Gleichungssystems bei. Ohne ihn könnte die Matrix \mathbf{HBH}^T durch zwei nahezu identische Beobachtungsorte singulär werden.

Die Nichtdiagonalelemente b_{ij} ($i \neq j$) beschreiben die Beziehungen¹¹¹ der Backgroundfehler an verschiedenen Beobachtungsorten zueinander. Hohe Werte der Kovarianz für zwei Orte reduzieren die individuellen Gewichte beider Beobachtungsinkremente. Die Beobachtungsinkremente sind in diesem Fall nicht unabhängig. Ihr individueller Einfluss auf den Analysewert sollte individuell kleiner sein als der eines unabhängigeren Beobachtungsinkrements. Das Berücksichtigen der Beziehungen zwischen den Beobachtungsorten wird als „declusterung“ bezeichnet und ist ein Vorteil der statistischen Analyse gegenüber anderen Verfahren. Durch ihn wird die variable Stationsdichte¹¹² berücksichtigt: Geht man von einem homogenen und isotropen Kovarianzmodell aus¹¹³, so erhalten alle Beobachtungen mit dem gleichen Abstand zum Analysepunkt das gleiche Gewicht. Befinden sich nun viele dieser Stationen dicht gedrängt, z. B. in einer Stadt, so würde der Analysewert stark durch die Beobachtungsinkremente aus diesem Gebiet beeinflusst werden. Die Berücksichtigung der Kovarianzen des Backgroundfehlers führt jedoch dazu, dass der ungerechtfertigt hohe Einfluss dieses Gebietes verringert wird.

Aus der Gleichung {4.36} ergibt sich die Varianz des Analysefehlers im Falle optimal spezifizierter Gewichte k_{ji} aus:

$$\sigma_{JA}^2 = \sigma_{JB}^2 - \sum_{i=1}^{n_i} k_{ji} b_{ji} \leq \hat{\sigma}_{JA}^2$$

Die Varianz des Fehlers der optimalen Analyse σ_{JA}^2 liegt unter dem des Backgrounds σ_{JA}^2 . Der Wert stellt die theoretische untere Schranke für die tatsächliche Varianz des Analysefehlers $\hat{\sigma}_{JA}^2$ dar, bei nicht optimalen Koeffizienten aufgrund unzureichender Kenntnis der Matrizen \mathbf{B} und \mathbf{R} . Für eine suboptimale aber sinnvolle Interpolation sollte $\hat{\sigma}_{JA}^2$ kleiner als $\sigma_{B_i}^2$ sein.

Eine Möglichkeit, den tatsächlichen Analysefehler abzuschätzen, ergibt sich durch einen cross-validation-Ansatz für die Stationsorte. Dabei werden die Stationswerte prognostiziert, ohne sie bei der Interpolation zu verwenden (siehe Kapitel 6.2).

4.5.6 Variationsansätze

Neben der Minimierung der Varianz des Analysefehlers¹¹⁴ (siehe Kapitel 4.5.1) kann die Analyse als Variationsproblem formuliert werden. Der Ausgangspunkt ist die maximum-likelihood-Schätzung für die Analyse, die sich aus den mehrdimensionalen

¹¹¹ Da keine Korrelation der Beobachtungsfehler angenommen wird, gilt $r_{ij} = 0$ für $i \neq j$.

¹¹² Die Immissionsbeobachtungen zeichnen sich durch große Unterschiede in der Stationsdichte aus.

¹¹³ gleichbedeutend mit einer reinen Abstandsabhängigkeit

¹¹⁴ entspricht der Bestimmung des Erwartungswertes

Wahrscheinlichkeitsfunktionen für Background und Beobachtung ergibt (siehe Gleichung {4.19}). Gehorchen diese einer Gauß-Verteilung, so ist das Ergebnis des Ansatzes identisch mit dem der Varianzminimierung (Lorenz, 1986).

Das Ergebnis des Maximum-Likelihood-Ansatzes ist eine skalare Kostenfunktion J {4.40}, die ihr Minimum bei dem wahrscheinlichsten Wert für $\mathbf{x} = \mathbf{x}_A$ hat. Die Kostenfunktion besteht aus zwei Teilen, die die Abweichungen zum Background und zu den Beobachtungen bewerten. Die inversen Kovarianzmatrizen der Fehler von Background \mathbf{B} und Beobachtung \mathbf{R} dienen als verallgemeinerte Gewichte. Je höher der Fehler, umso geringer wird die Abweichung bewertet.

Die Analyse wird durch die Minimierung der Kostenfunktion gefunden, die auf dem Gradient von J beruht. Ein Vorteil des Variationsansatzes ist, dass der Beobachtungsoperator nicht linearisiert werden muss. Weiterhin können leicht Zwangsbedingungen eingeführt werden, denen der Zustandsvektor exakt oder in möglichst hohem Maße gehorchen soll (Sasaki, 1970). Dafür bieten sich physikalische und chemische Gleichgewichte¹¹⁵ oder eine Modellrechnung an. In der 4-dimensionalen Variationsrechnung (4D-VAR) werden die Abweichungen zu den Beobachtungen für mehrere Zeitpunkte einbezogen. Dabei gilt die starke Zwangsbedingung, dass die Modellzustandsvektoren \mathbf{x} für verschiedene Zeiten durch eine Modellintegration gebildet werden. Für die iterative Bestimmung des Minimums der Kostenfunktion wird der Gradient der Kostenfunktion verwendet. Für seine Berechnung wird die Adjungierte des linearisierten¹¹⁶ Modelloperators (adjoint model) benötigt (Bouttier und Courtier, 1999). Sie entspricht der transponierten Linearisierung des ursprünglichen Modells. Für die mit zeitlichen Differenzenschemata formulierten Modellteile (Advektion, Chemie etc.) ist die Programmierung des adjungiert-linearen Modells möglich (Talagrand, 1997). Im Rahmen der Chemie-Transport-Modellierung wird 4D-VAR für das EURAD-Modell verwendet (Elbern und Schmidt, 1997, 2001):

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_B)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_B) + \sum_{i=1}^{n_t} (\mathbf{y}^i - H\mathbf{x}^i)^T \mathbf{R}^{-1} (\mathbf{y}^i - H\mathbf{x}^i) \quad \{4.40\}$$

$$\begin{aligned} J(\mathbf{x}_A) = \min & \quad \text{mit Zwangsbedingung } \mathbf{x}^{i+1} = M\mathbf{x}^i \\ M & \quad \dots \quad \text{Modelloperator} \\ \text{Zeitpunkte } i=1 \dots n_t & \end{aligned} \quad \{4.41\}$$

4.5.7 Kalman-Filter

Der Kalman-Filter¹¹⁷ kann als ein vierdimensionales Verfahren zur Datenassimilation verwendet werden. Sein Grundprinzip ist der Wechsel von Prognose und Analyse des Modellzustandsvektors \mathbf{x} und seiner Fehlerkovarianzmatrix \mathbf{P} ¹¹⁸ (Abbildung 4.3). Der

¹¹⁵ z. B. das geostrophische Gleichgewicht für Wind- und Geopotentialfelder bzw. das fotostationäre Gleichgewicht.

¹¹⁶ Die Linearisierung komplexer Modelle ist, wenn überhaupt, nur für kleine Zeitschritte möglich. Man spricht dann von dem tangential-linearisierten Modell.

¹¹⁷ Kalman-Filter sind eine große Gruppe von optimalen rekursiven Methoden zur Prognose. Sie müssen nicht wie hier mit einem dynamischen Modell gekoppelt sein.

¹¹⁸ Da diese sich ständig verändert, wird sie in Abgrenzung zur empirischen Kovarianzmatrix \mathbf{B} mit \mathbf{P} bezeichnet.

Kalman-Filter ist für lineare Modelle formuliert. Er wird als Kalman Bucy Filter für nichtlineare Modelle erweitert, wenn das Modell M durch eine lineare Approximation \mathbf{M} für die Prognose der Kovarianzmatrix der Fehler \mathbf{P} verwendet werden kann¹¹⁹ {4.43}. Bei dieser Prognose wird der Fehlerhaftigkeit des Modells durch die Modellprognosefehlermatrix \mathbf{Q} Rechnung getragen. Liegen Messungen \mathbf{y} vor, so erfolgt eine Analyse \mathbf{x}_A , bei der der prognostizierte Zustandsvektor an die Messungen angepasst wird {4.45}. Die prognostizierte Kovarianzmatrix \mathbf{P}^{t+1} und die Fehlervarianzen der Messungen bestimmen dabei die Analysegewichte \mathbf{K} {4.44}. Gleichzeitig wird die Kovarianzmatrix für den analysierten Zustand \mathbf{P}_A berechnet {4.46}. Die Analysen des Modellzustandes und der Kovarianzmatrix werden nun als Ausgangspunkt für den nächsten Prognoseschritt verwendet.

$$\text{Prognose des Zustandsvektors:} \quad \mathbf{x}^{t+1} = \mathbf{M}\mathbf{x}_A^t \quad \{4.42\}$$

$$\text{Prognose der Fehlerkovarianzmatrix:} \quad \mathbf{P}_B^{t+1} = \mathbf{M}\mathbf{P}_A^t\mathbf{M}^T + \mathbf{Q} \quad \{4.43\}$$

$$\text{Bestimmung der Analysegewichte:} \quad \mathbf{K}^t = \mathbf{H}\mathbf{P}_B^t(\mathbf{H}\mathbf{P}_B^t\mathbf{H}^T + \mathbf{R})^{-1} \quad \{4.44\}$$

$$\text{Analyse des Zustandsvektors:} \quad \mathbf{x}_A^t = \mathbf{x}_B^t + \mathbf{K}^t(\mathbf{y}^t - \mathbf{H}\mathbf{x}_B^t) \quad \{4.45\}$$

$$\text{Bestimmen der Fehlerkovarianzmatrix der Analyse:} \quad \mathbf{P}_A^t = (\mathbf{I} - \mathbf{K}^t\mathbf{H})\mathbf{P}_B^t \quad \{4.46\}$$

Der Analyseschritt gleicht dem nach dem Minimum-Varianz-Ansatz {4.31}. Es gelten auch für den Kalman-Filter die gleichen Voraussetzungen, d. h. kein Bias und keine Korrelation zwischen den Modell- und Beobachtungsfehlern. Für die Berechnung der Fehlerkovarianzen wird von deren Normalverteilung ausgegangen.

Die Besonderheit des Kalman-Filters besteht in der Prognose und Analyse der Kovarianzmatrix \mathbf{P} . Die Prognose ist dynamisch und durch die zugrundeliegenden Modellgleichungen bestimmt. Dadurch können Zusammenhänge innerhalb des Zustandsvektors (z. B. Advektion oder chemische Kopplung) erfasst und bei der Analyse berücksichtigt werden: Ist eine entsprechende statistische Beziehung in \mathbf{P} entwickelt worden, so erfolgt eine Korrektur auch in Teilen des Zustandsvektors, für die keine direkten Beobachtungen vorliegen. Durch das mehrmalige Durchlaufen des Zyklus wird der Analysefehler schrittweise bis zu einem Konvergenzniveau verringert. Für die Initialisierung des Kalman-Filters muss eine Kovarianzmatrix aus anderen Quellen bereitgestellt werden.

Im einfachsten Fall ist die Kovarianzmatrix \mathbf{Q} des Modellprognosefehlers eine reine Diagonalmatrix, die den Elementen des Zustandsvektors eine Varianz (Unsicherheit) zuordnet. Nichtprognostische Modellparameter können in gewissem Sinne überprüft werden, indem man sie in den Zustandsvektor und die Analyse mit einbezieht. Bewirkt die Analyse zu verschiedenen Zeitpunkten eine gleichgerichtete Korrektur dieser Werte, so deutet dies auf einen systematischen Fehler hin, der im Modell korrigiert werden kann. Aufgrund des rein statistischen Zusammenhanges muss jedoch sichergestellt

¹¹⁹ Vereinfachend wird bei der Datenassimilation jedoch nur vom Kalman-Filter gesprochen.

werden, dass die richtige Ursache für die Modellfehler gefunden wurde und kein unphysikalisches und rein episodisches „Modelltuning“ vorgenommen wird. Van Loon et al. (1999) verwenden diesen Ansatz, um Emissionsangaben in dem Modell LOTOS für eine bessere Modelleistung zu verändern.

Der numerische Aufwand für die Prognose der Kovarianzmatrix ist sehr groß und würde N Modellintegrationen entsprechen, wobei N die Anzahl der Elemente des Modellzustandsvektors ist¹²⁰. Aus diesem Grund sind Vereinfachungen nötig. Mit dem RRSQRT-Ansatz (reduced rank square root, siehe Kapitel 4.5.8) wird die Kovarianzmatrix auf die wichtigen Moden der Variabilität beschränkt, so dass der Rechenzeitbedarf nur ca. 50-100 Modellintegrationen entspricht (Heemink et al., 1999). Für die Chemie-Transport-Modellierung werden Kalman-Filter-Ansätze z. B. für das LOTOS-Modell angewendet (van Loon und Heemink, 1997):

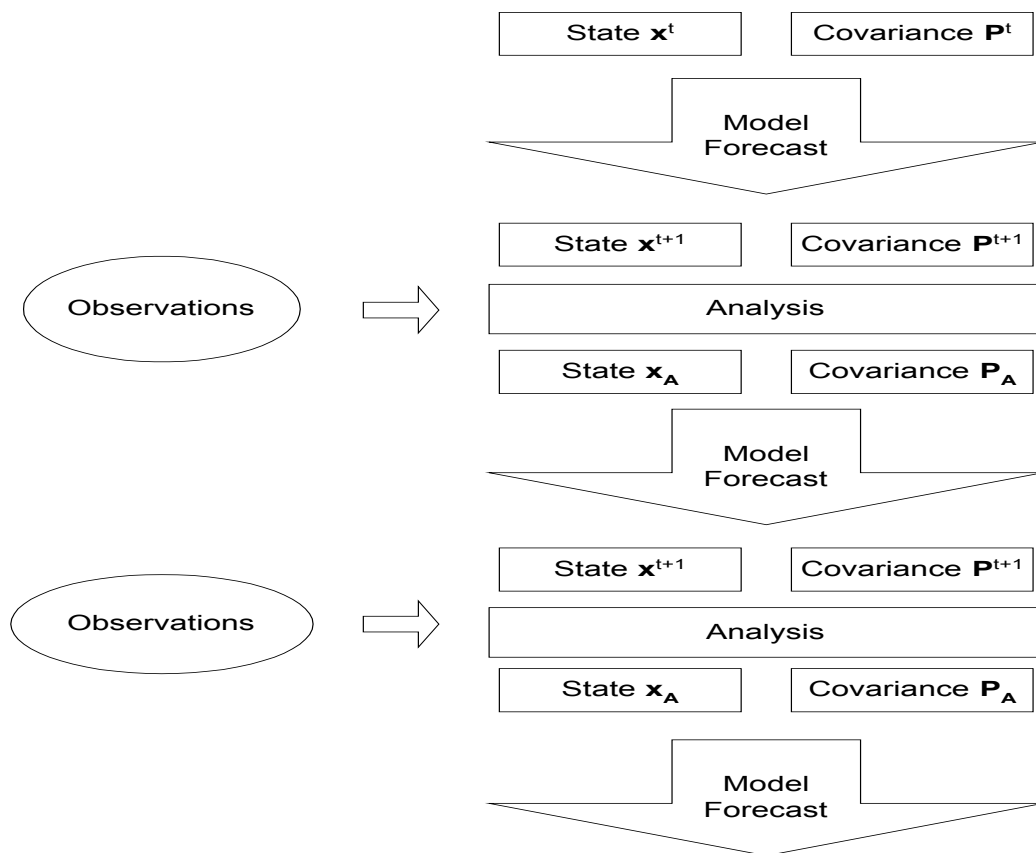


Abbildung 4.3 Funktionsprinzip des Kalman-Filters

4.5.8 RRSQRT-Ansatz für Kalman-Filter

Die Abkürzung RRSQRT steht für reduced rank square root. Es ist eine Methode, die Prognose der Kovarianzmatrix \mathbf{P} {4.43} für nichtlineare Modelle M zu realisieren, ohne explizit eine linearisierte Form \mathbf{M} angeben zu müssen (van Loon und Heemink, 1997). Zunächst wird die originale Kovarianzmatrix \mathbf{P} durch die square root Matrix \mathbf{S} („Wurzelmatrix“) ausgedrückt:

¹²⁰ REM/Calgrid besitzt einen Zustandsvektor mit ca. 10^6 Elementen, der Rechenaufwand beträgt damit für den vollständigen KBF das 10^6 -fache der Modellrechenzeit.

$$\mathbf{P} = \mathbf{S}\mathbf{S}^T \quad \{4.47\}$$

Dadurch wird auch erreicht, dass die Matrix \mathbf{P} immer positiv definit ist, was bei einer direkten Behandlung von \mathbf{P} aufgrund von numerischen Ungenauigkeiten nicht immer der Fall sein muss. Die square root Matrix \mathbf{S} ist eine Matrix aus m Reihen von Vektoren der Größe des Modellzustandsvektors \mathbf{x} :

$$\mathbf{S} = \left[\mathbf{x}'_1, \mathbf{x}'_2, \mathbf{x}'_3 \cdots \mathbf{x}'_m \right] \quad \{4.48\}$$

Die Anzahl m der Spalten bestimmt den Rang der Matrix \mathbf{P} und damit deren statistische Güte. Für eine approximative Darstellung von \mathbf{P} kann \mathbf{S} jedoch eine bedeutend geringere Anzahl von Spalten als die Dimension von \mathbf{x} haben (Seghers, 2002). Selbst mit einer einspaltigen Matrix \mathbf{S} lässt sich eine gültige Kovarianzmatrix bilden.

Anschaulich kann man sich die Spalten von \mathbf{S} als ein Ensemble von $j = 1, m$ verschiedenen, d. h. gestörten Modellzuständen \mathbf{x}'_j , von denen der jeweilige Mittelwert abgezogen wurde, vorstellen. Je größer die Anzahl m , um so mehr Ensembleelemente, d. h. Realisierungen von Modellzuständen, wurden verwendet, um die Varianzen bzw. Kovarianzen in der Matrix \mathbf{P} zu berechnen.

Die Formulierung der Kalman-Filter-Gleichungen erfolgt mit der Matrix \mathbf{S} und die Prognose von \mathbf{P} wird mit einer Prognose von \mathbf{S} ersetzt {4.43}. Dafür wird zu jeder Spalte von \mathbf{S} ein Grundzustand addiert, so dass m Realisierungen eines Modellzustandsvektors entstehen. Diese können mit Hilfe des Computercodes des Modells M prognostiziert werden. Anschließend wird durch Abzug des prognostizierten Grundzustandes die prognostizierte Matrix \mathbf{S}^{t+1} gebildet.

Um den Modellfehler, ausgedrückt durch die Matrix \mathbf{Q} , in die Prognose der Matrix \mathbf{S} einfließen zu lassen, werden q zusätzliche Modellzustände mit einer Störung zeitlich integriert. Die Störung wird gebildet, indem ausgewählte Modellparameter wie die Emission oder die Stärke des vertikalen Austauschs verändert werden. Sie werden mit angenommenen Unsicherheiten der Modellierung quantifiziert und als q Rausch-Parameter (noise) formuliert. Die Differenz der so gestörten Zustände \mathbf{x}^* vom Grundzustand werden als zusätzliche Spalten in die square root Matrix \mathbf{S} aufgenommen. Mit jedem Prognoseschritt erhöht sich demzufolge die Anzahl der Spalten in \mathbf{S} um q :

$$\mathbf{S}^{t+1} = \left[\mathbf{x}'_1^{t+1}, \mathbf{x}'_2^{t+1}, \mathbf{x}'_3^{t+1} \cdots \mathbf{x}'_m^{t+1}, \mathbf{x}^*_1 \cdots \mathbf{x}^*_q \right] \quad \{4.49\}$$

Um die Anzahl der Spalten von \mathbf{S} und damit die Anzahl der notwendigen Modellschritte konstant zu halten, wird eine Eigenvektoraufspaltung der Matrix \mathbf{P} vorgenommen. Für den nächsten Prognoseschritt werden dann nur die m größten Eigenvektoren beibehalten. Die Matrix \mathbf{S} wird damit durch Modellzustände repräsentiert, die einen wesentlichen Anteil an der Varianz bilden. Sie heißen darum auch die **Moden** von \mathbf{S} . Im übertragenen Sinn bedeutet die Eigenvektoraufspaltung, dass gestörte Modellzustände konstruiert werden, die ein bedeutsames Anwachsen der Varianz vermuten lassen. Ohne die Eigenvektoraufspaltung wäre ein großes Ensemble von Modellzuständen nötig, um \mathbf{P} in ausreichender Güte zu gewinnen. Durch die Einführung von Moden reicht eine geringere Anzahl von Spalten und damit von notwendigen Modellprognosen aus, um \mathbf{P} gleichwertig zu repräsentieren.

Zum Beginn der Kalman-Filter Rechnung sind die Matrizen \mathbf{S} bzw. \mathbf{P} unbekannt. Man nimmt für sie Null Matrizen an. Mit jedem Prognoseschritt werden q neue Spalten in

die Matrix \mathbf{S} eingebettet, um den stochastischen Modellfehler zu berücksichtigen. Übersteigt die Anzahl der Spalten die Anzahl m von handhabbaren Moden, so wird deren Anzahl mit der Eigenvektoraufspaltung auf die m größten beschränkt. Somit wird kontinuierlich die Matrix \mathbf{S} bzw. \mathbf{P} entwickelt. Dabei ist zu beachten, dass nur Kovarianzen modelliert werden, die durch die Wahl der Störung, z. B. Emissionsvariationen, im Modell entstehen. Die maximale Anzahl der notwendigen Moden muss durch Konvergenzkriterien und die numerische Kapazität festgelegt werden. In den zitierten Anwendungen von LOTOS werden ca. 30-50 Moden verwendet.

Es ist günstig, den konkreten Wert der q Rausch-Parameter (noise), z. B. eine Erhöhung der NO_x -Emission, in den Modellzustand einzubeziehen. Jeder Analyseschritt verändert diese Werte im Lichte der vorliegenden Beobachtungen aufgrund der aufgebauten Kovarianzen. Die Auswirkung der gebildeten noise-Parameter auf die Modellrechnung ist dadurch erkennbar.