

Inversion

Fernerkundung für Fortgeschrittene
Wintersemester 2018

Stefan Bühler

Meteorologisches Institut

Universität Hamburg

$$\vec{y} = F(\vec{x})$$

$$\vec{x} = ?$$

Quellen

- ▶ Clive D. Rodgers, Inverse Methods for Atmospheric Sounding, World Scientific, 2000.
- ▶ E. T. Jaynes, Probability Theory, The Logic of Science, Cambridge, 2003.

- ▶ Achtung: In diesem Kapitel wird es etwas mathematisch.

Übersicht

- ▶ Opazitätsregel und Jacobians
- ▶ Das Inverse Problem
- ▶ Kochrezept „Optimal Estimation“
- ▶ Naiver Lösungsversuch
- ▶ Exkurs: Beschreibung der „Varianz“ eines Vektors
- ▶ Weniger naive Lösungsversuche
- ▶ Optimal Estimation (Bayesianische Methoden)
- ▶ Zusammenfassung

Übersicht

- ▶ **Opazitätsregel und Jacobians**
- ▶ Das Inverse Problem
- ▶ Kochrezept „Optimal Estimation“
- ▶ Naiver Lösungsversuch
- ▶ Exkurs: Beschreibung der „Varianz“ eines Vektors
- ▶ Weniger naive Lösungsversuche
- ▶ Optimal Estimation (Bayesianische Methoden)
- ▶ Zusammenfassung

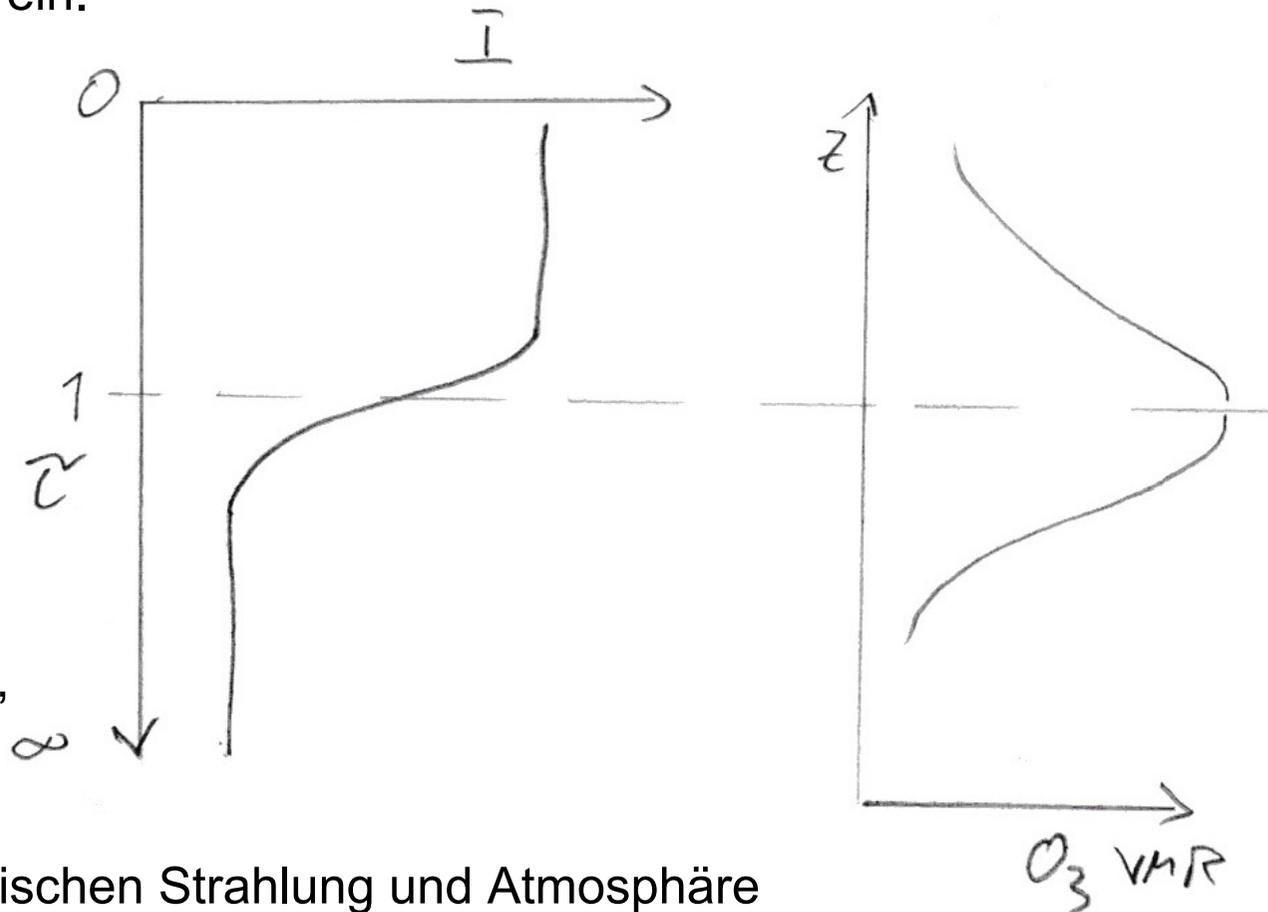
Quelle für Opazitätsregel

- ▶ Thomas & Stamnes, Radiative Transfer in the Atmosphere and Ocean, Cambridge, 1999.
- ▶ Sonst nicht leicht zu finden, wird oft einfach als selbstverständlich angenommen.
- ▶ Auch T&S geben keine formale Herleitung.

Absorption: Beispiel Ozonschicht (nach Chapman)

$$\tau(s_1, s_2) = \int_{s_1}^{s_2} \alpha(s') ds'$$

- ▶ Ultraviolettes Sonnenlicht dringt in die Atmosphäre ein.
 - ▶ Zunächst Atmosphäre dünn:
 - kaum Absorption,
 - I bleibt konstant.
 - ▶ In der Nähe von $\tau=1$ wird viel Energie absorbiert.
 - ▶ Danach ist zwar die Absorption hoch, aber keine Strahlung mehr da, die absorbiert werden könnte.
- Die ganze Interaktion zwischen Strahlung und Atmosphäre findet in einer relativ schmalen Schicht statt.
- Dort ist die Ozonschicht (in dieser Weise hat Chapman argumentiert, um die Lage und Dicke der Ozonschicht herzuleiten).



Emission: Beispiel Sonnenspektrum (nach Thomas&Stamnes)

- ▶ In großer Höhe: Wenig Emission.
- ▶ In geringer Höhe: Viel Emission, kann aber nicht zum Weltraum hin entkommen.
- ▶ Dazwischen gibt es ein Optimum (bei $\tau = 1$).
- ▶ Im Bild des Sonnenspektrums rechts kann man sehen, aus welcher Schicht der Sonnenatmosphäre die Strahlung bei verschiedenen Wellenlängen kommt.
Sichtbar: kühle Photosphäre
Fern UV: höhere und heißere Chromosphäre

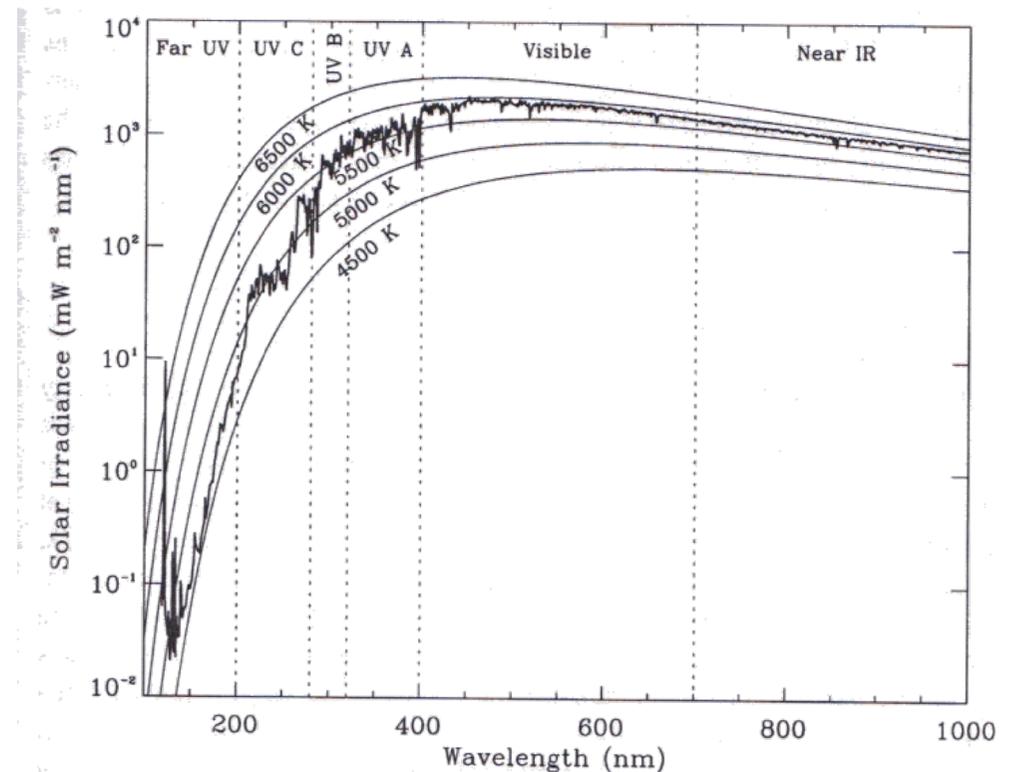
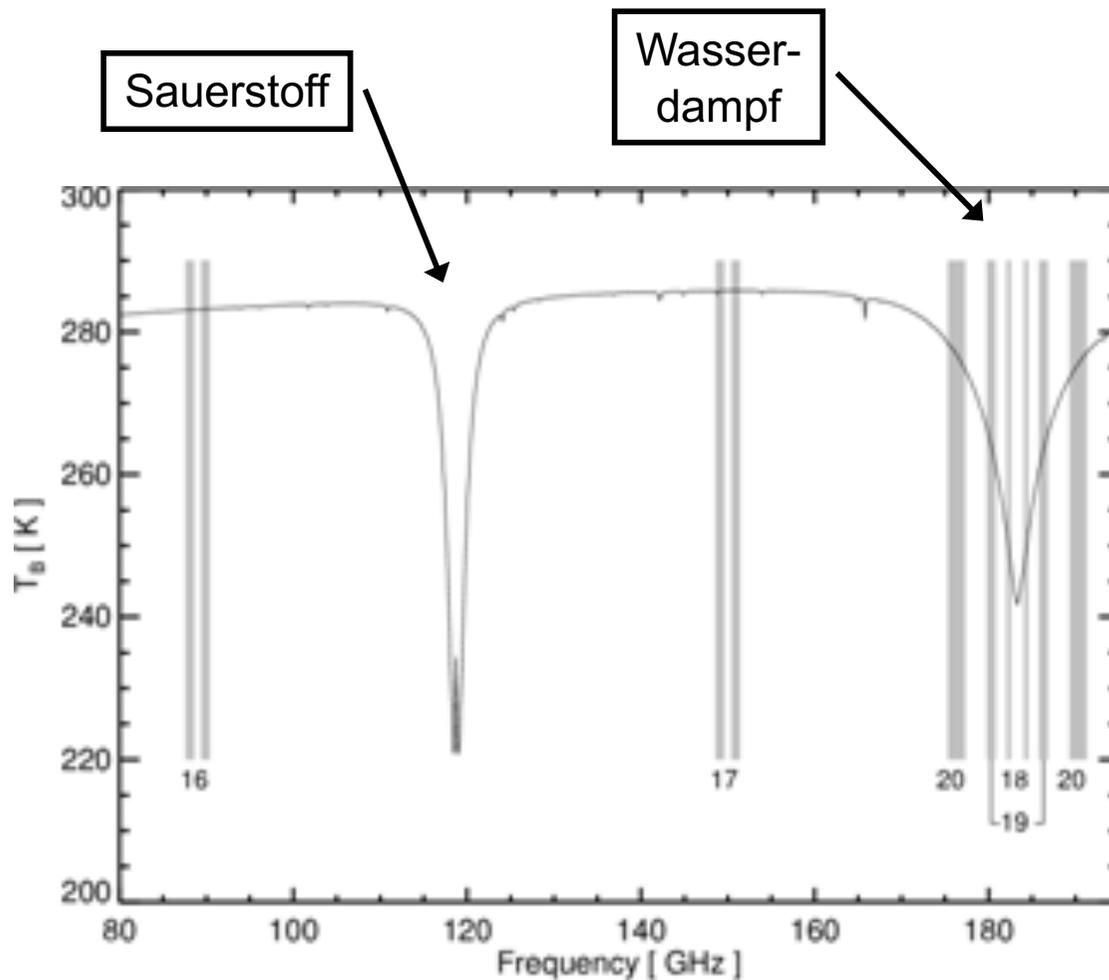


Figure 1.1 Extraterrestrial solar flux, or irradiance, measured by a spectrometer on board an Earth-orbiting satellite. The UV spectrum ($119 < \lambda < 420$ nm) was measured by the SOLSTICE instrument on the UARS satellite (modified from a diagram provided by G. J. Rottmann, private communication, 1995). The vertical lines divide the various spectral subranges defined in Table 1.1. The smooth curves are calculated blackbody spectra for a number of emission temperatures.

Bild: T&S

Beispiel Mikrowellen Strahlungstemperatur der Erde



- ▶ 16+17: Oberfläche
- ▶ 18: Obere Troposphäre ($p < 500$ hPa)
- ▶ 19: Mittlere Troposphäre
- ▶ 20: Untere Troposphäre

- ▶ Grau schattiert: Kanäle des AMSU-B Instruments

AMSU-B Opazität

▶ Passiver Mikrowellen-sensor

▶ Misst Feuchte

16+17:

▶ Fensterkanäle ($\tau_z \leq 1$)

18-20:

▶ „Sounding Channels“, messen Wasserdampf-Profil

▶ $\tau_z > 1$

▶ $z(\tau=1)$ je nach Kanal bei verschiedenen Höhen in der Troposphäre

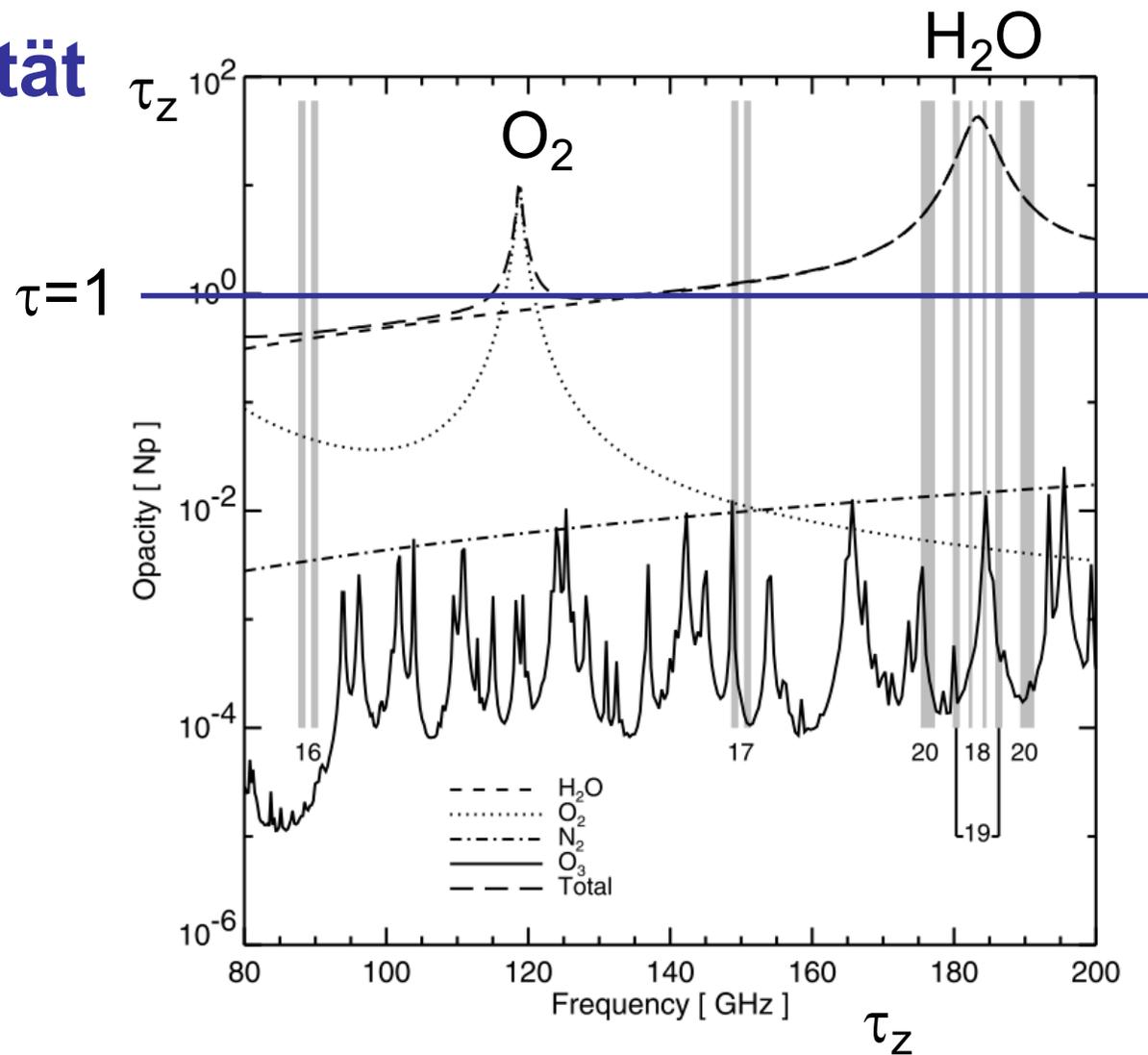
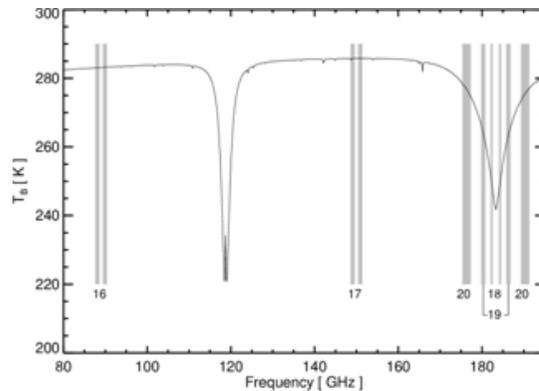


Figure 1. Atmospheric zenith opacity for the FASCOD midlatitude summer scenario for H₂O (dashed), O₂ (dotted), N₂ (dash dotted), and O₃ (solid). The long dashed line represents the total opacity. Shaded regions represent the passband positions of AMSU-B channels. The channel numbers are also printed near the passbands.

AMSU-B Jacobians



- ▶ Ableitung der Intensität (y) nach dem Atmosphärenzustand (x , hier: Wasserdampf-konzentration).

$$\vec{y} = F(\vec{x})$$

$$\mathbf{K} = \frac{dF(\vec{x})}{d\vec{x}} \quad (\text{Jakobi Matrix})$$

$$\vec{y} = \vec{y}_0 + \mathbf{K}(\vec{x} - \vec{x}_0) \quad (\text{Linearisierung})$$

- ▶ Nomenklatur mit x/y aus der Inversionstheorie
- ▶ Besagen, welche **Information** die Messung enthält.
- ▶ Opazitätsregel: Maximum jeweils bei $\sim z(\tau=1)$

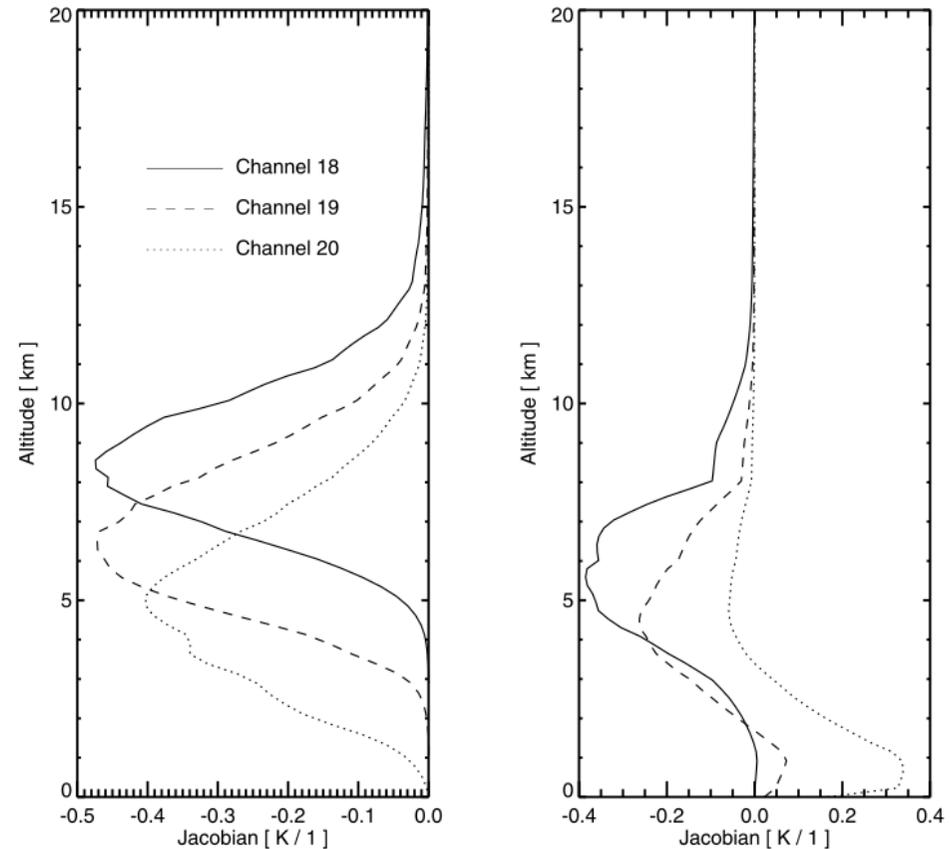
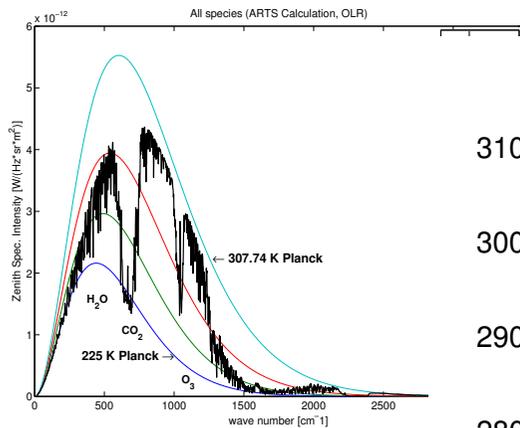
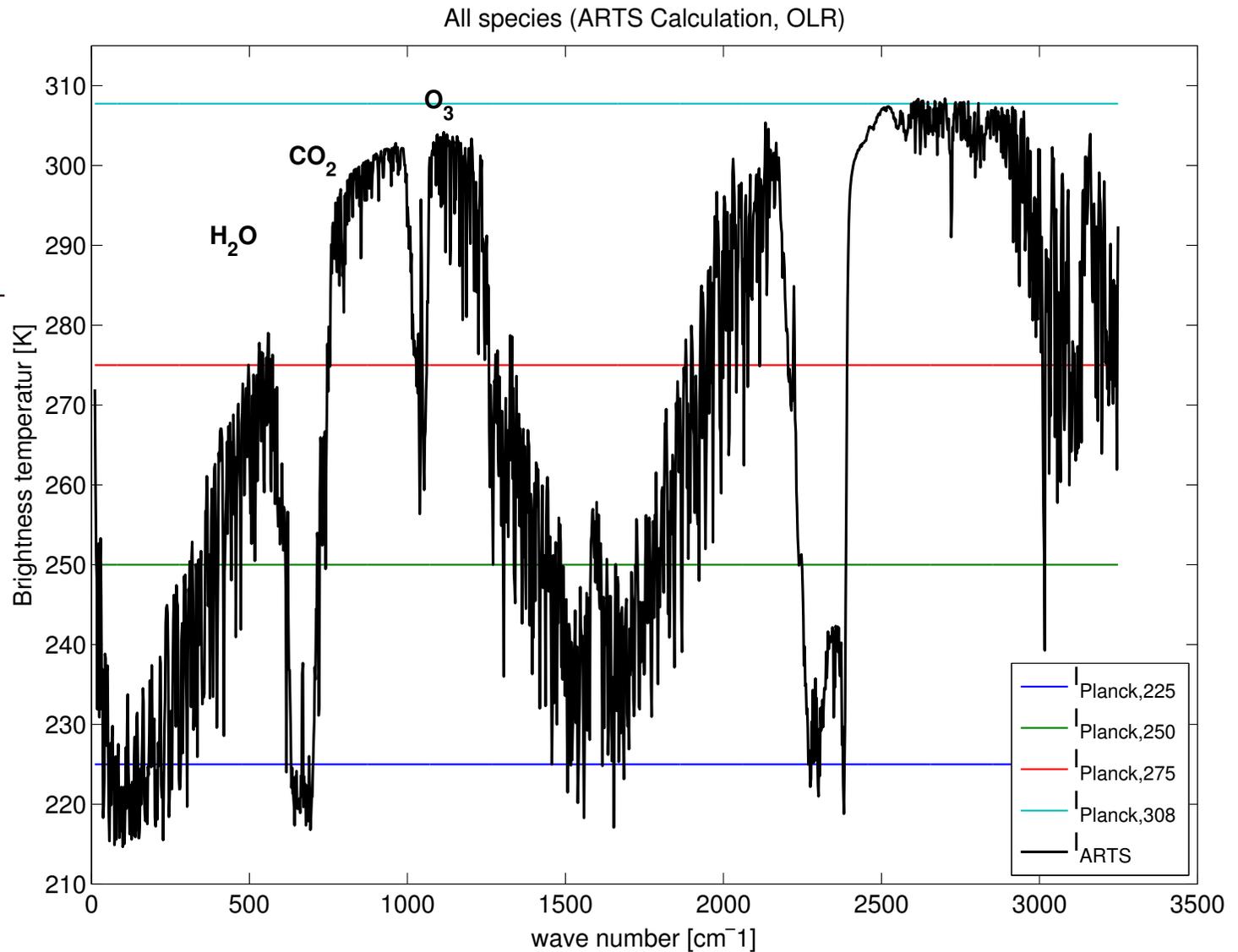


Figure 1. AMSU-B nadir humidity Jacobians for (left) a tropical scenario and (right) a subarctic winter scenario. The scenarios are from *Anderson et al.* [1986]. The Jacobians are in fractional units, so that the values correspond to the change in brightness temperature for a doubling of the mixing ratio at one vertical grid point. Displayed are the Jacobians for channel 18 (solid line), channel 19 (dashed line), and channel 20 (dotted line).

Opazitätsregel am Beispiel OLR

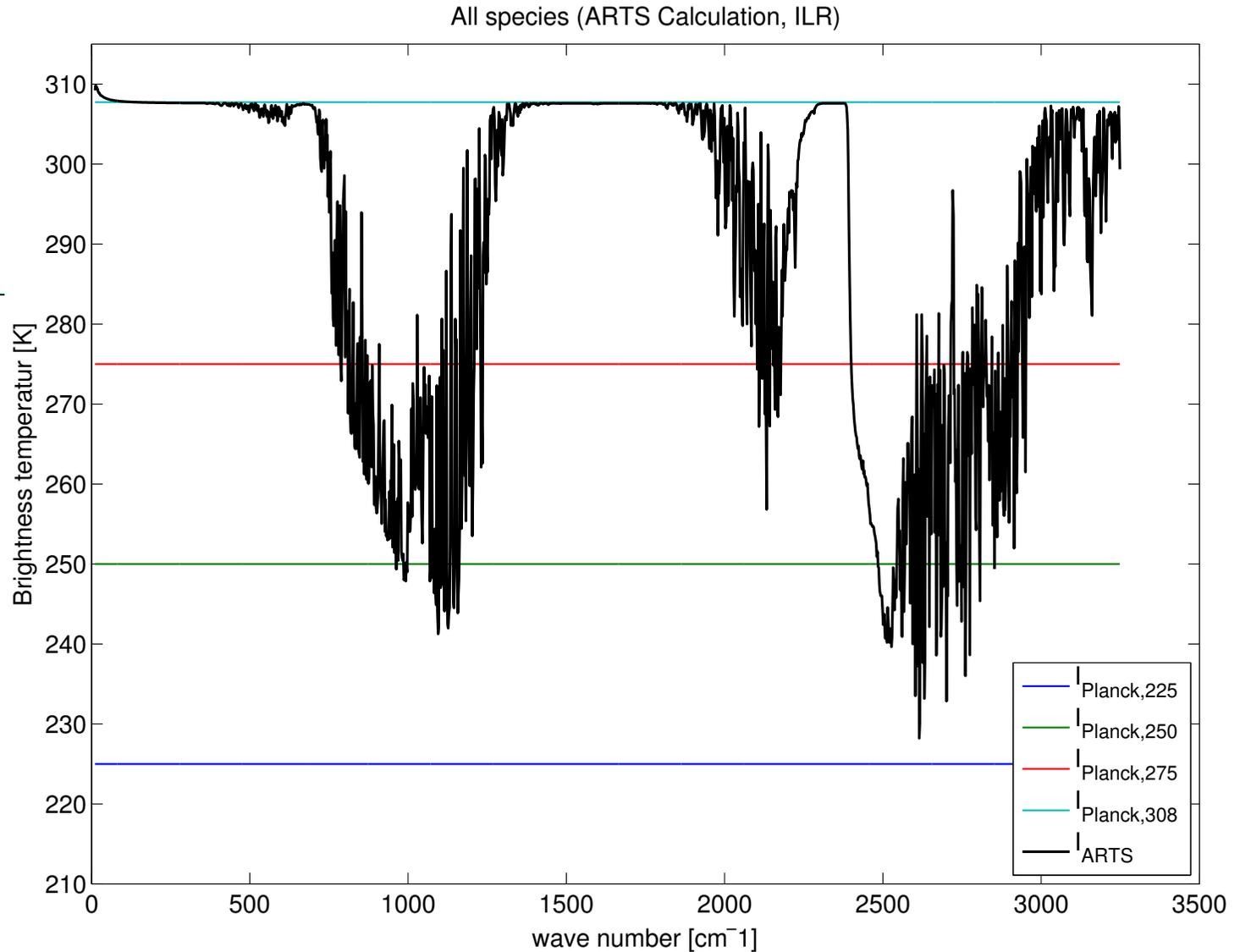
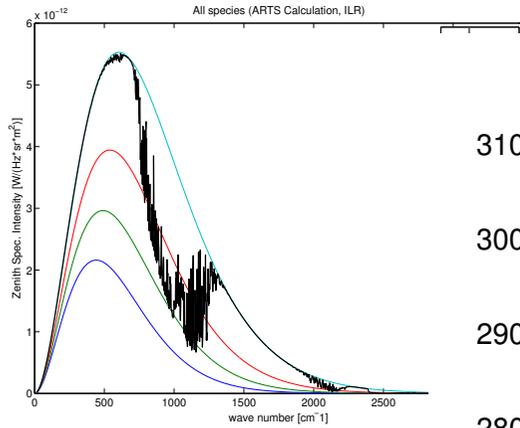


- ▶ Strahlung bei jeder Frequenz vorzugsweise aus einer Höhenschicht.
- ▶ Von da, wo Opazität (von oben) den Wert 1 erreicht.



Rechnung und Abbildung: Manfred Brath

Opazitätsregel am Beispiel der „Gegenstrahlung“ am Boden



In der Gegenstrahlung sieht man gerade da, wo die Atmosphäre optisch dicht ist, die Planck Kurve.

Rechnung und Abbildung: Manfred Brath

Jacobi Matrix für OLR und Wasserdampf

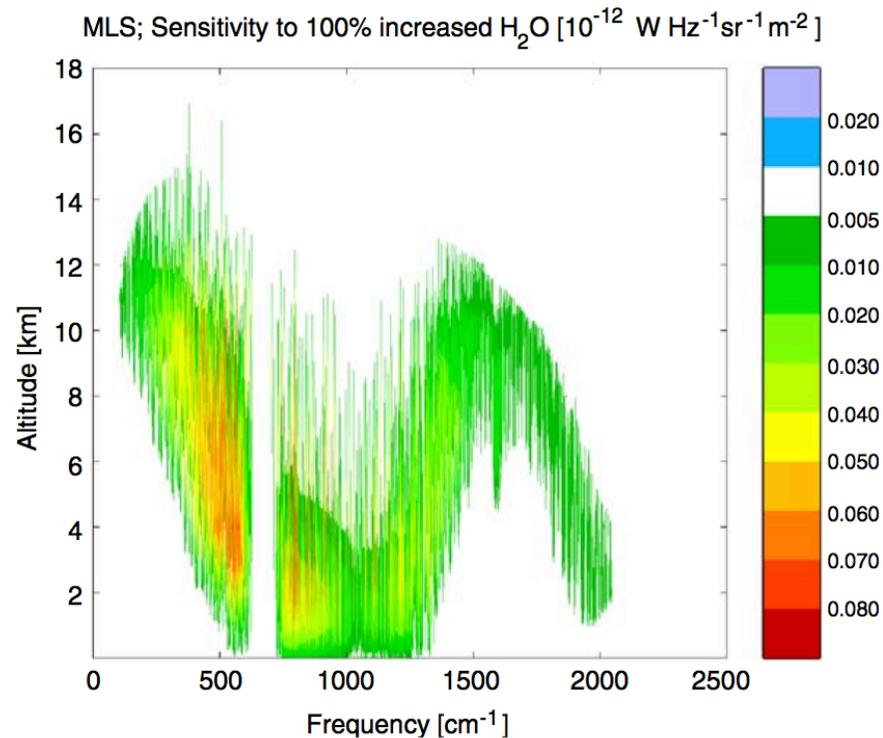


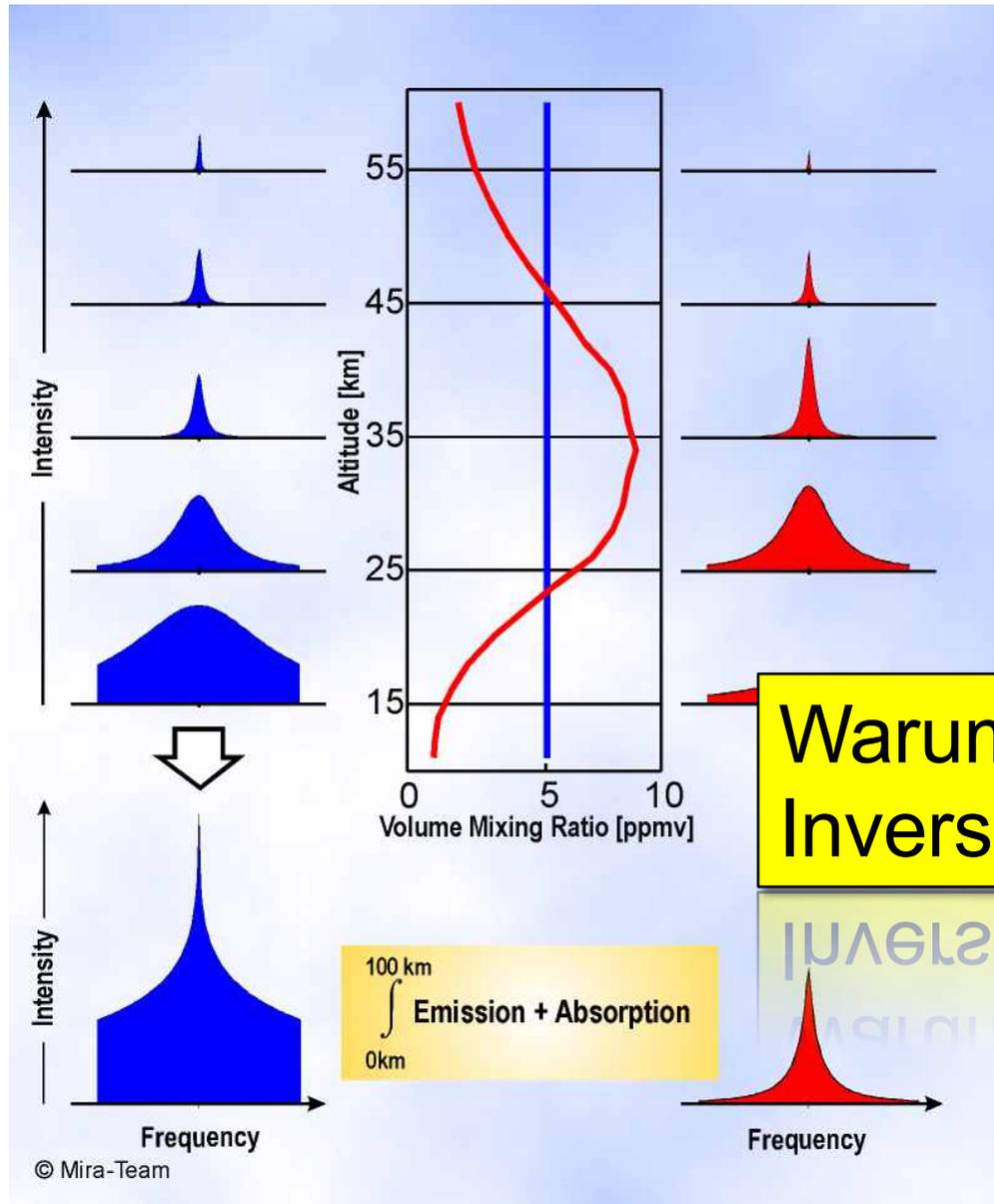
Fig. 2. The Jacobian of TOA zenith monochromatic radiance with respect to humidity in $\text{W Hz}^{-1} \text{sr}^{-1} \text{m}^{-2}$ for a midlatitude summer (MLS) atmosphere. The units correspond to the OLR change for a doubling of the humidity concentration at one altitude, decreasing linearly to zero at the adjacent altitudes above and below (triangular perturbations). The grid spacing is 1 km. The calculation was performed with the model and setup described in Section 2. The spectrum corresponding to this Jacobian is the one displayed in Fig. 1.

Buehler, S. A., A. von Engeln, E. Brocard, V. O. John, T. Kuhn, and P. Eriksson (2006), **Recent developments in the line-by-line modeling of outgoing longwave radiation**, *J. Quant. Spectrosc. Radiat. Transfer*, **98**(3), 446–457, doi:[10.1016/j.jqsrt.2005.11.001](https://doi.org/10.1016/j.jqsrt.2005.11.001).

Übersicht

- ▶ Opazitätsregel und Jacobians
- ▶ **Das Inverse Problem**
- ▶ Kochrezept „Optimal Estimation“
- ▶ Naiver Lösungsversuch
- ▶ Exkurs: Beschreibung der „Varianz“ eines Vektors
- ▶ Weniger naive Lösungsversuche
- ▶ Optimal Estimation (Bayesianische Methoden)
- ▶ Zusammenfassung

Fernerkundungsinstrumente geben indirekte Messung



Warum spricht man von Inversionsproblem?

Figure: University of Bremen

Warum Name Inverses Problem?

- ▶ Atmosphäre → Strahlung:
„Einfach“ (Vorwärtsproblem)
- ▶ Strahlung → Atmosphäre:
„Schwierig“ (Inverses Problem)

- ▶ Die **Vorwärtsrichtung** ist die, in die die Physik funktioniert (mathematisch korrektere Definition kommt später).
- ▶ Die **inverse** Richtung ist die, in die ich (oft) gehen möchte.

Notation

$$\vec{y} = F(\vec{x}) + \vec{\varepsilon}$$

\vec{y} : Messvektor (z.B. Spektrum)

\vec{x} : Zustandsvektor (z.B. Temperaturprofil)

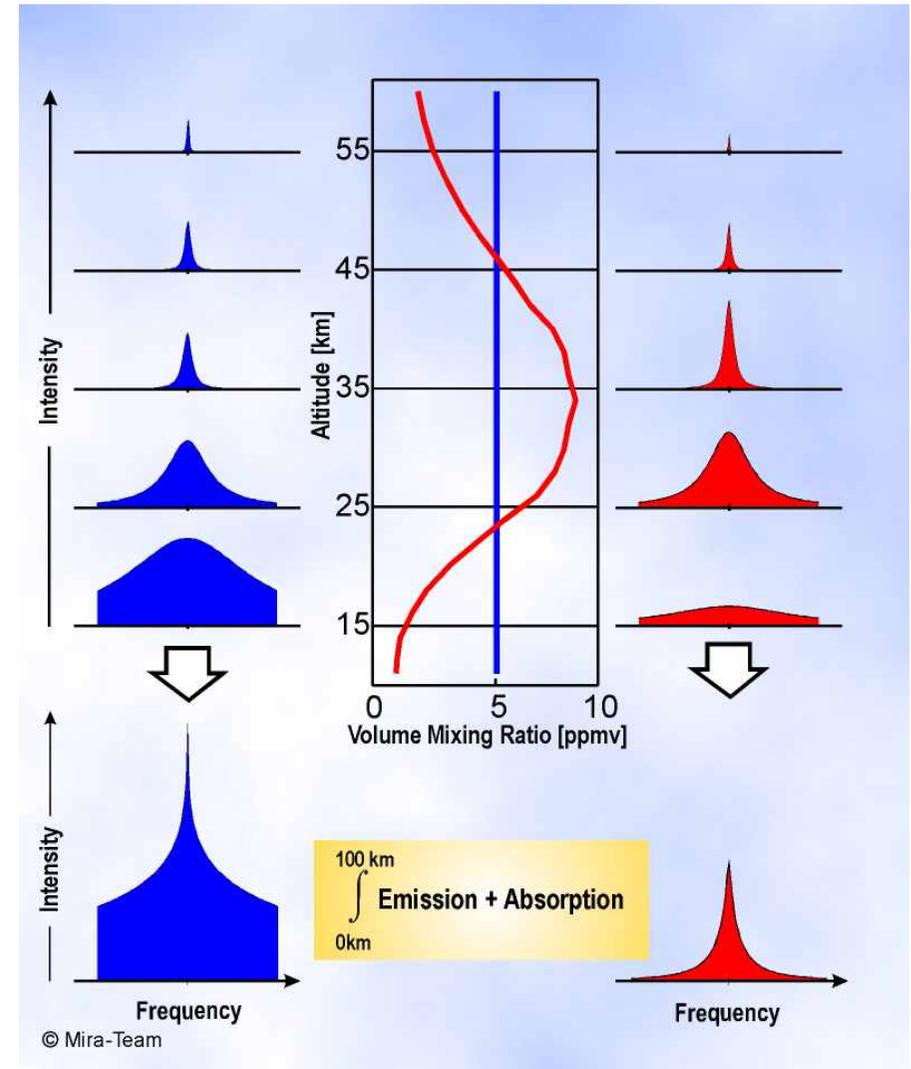
F : Vorwärtsmodell

$\vec{\varepsilon}$: Messfehler

Linearisierung:

$$\vec{y} = \vec{y}_0 + \mathbf{K}(\vec{x} - \vec{x}_0) + \vec{\varepsilon}$$

$$\mathbf{K} = \frac{dF(\vec{x})}{d\vec{x}} \quad (\text{Jakobi Matrix})$$



Messvektor y und Zustandsvektor x

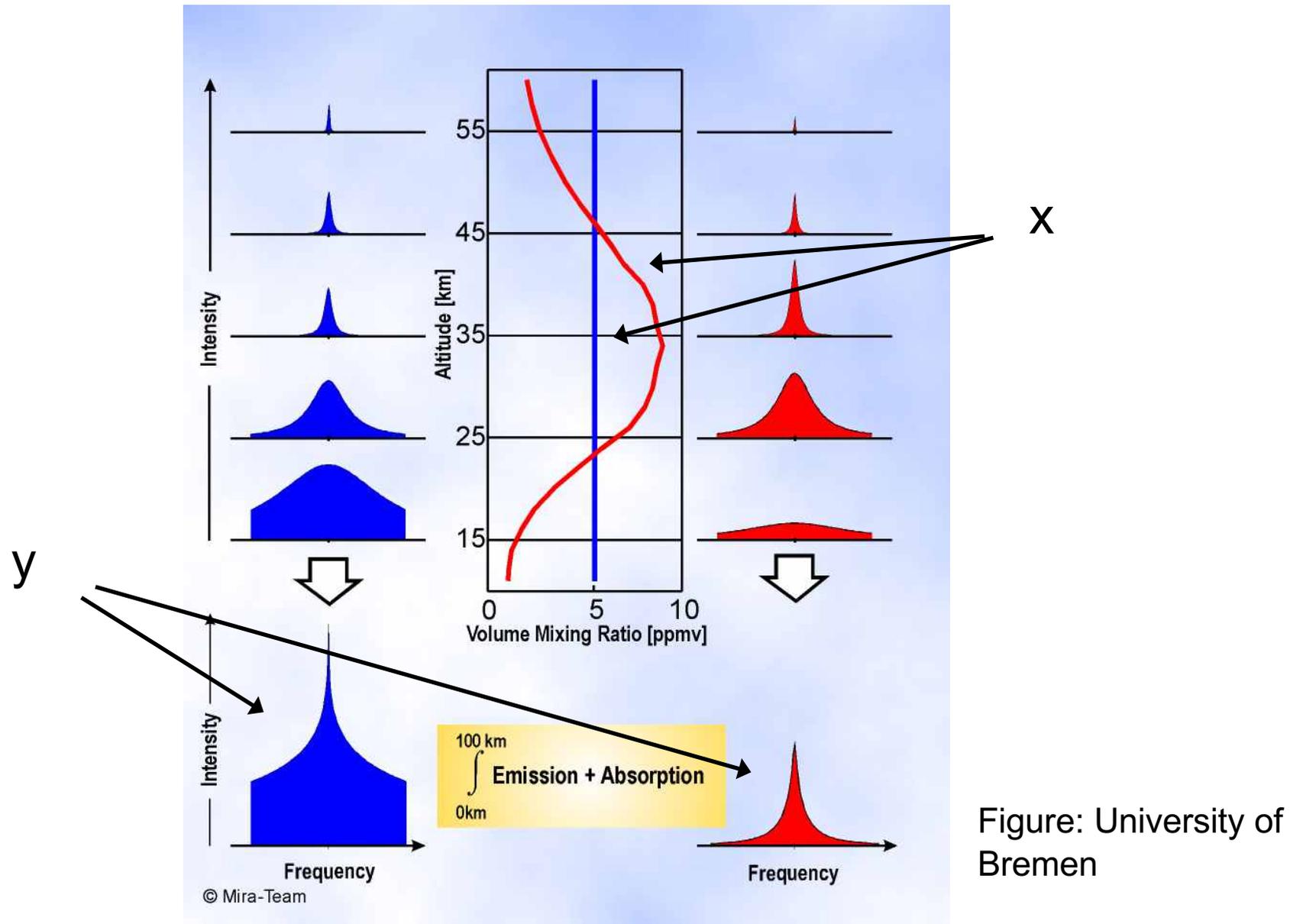


Figure: University of Bremen

Die Jacobi Matrix K

$$\vec{y} = \vec{y}_0 + \mathbf{K}(\vec{x} - \vec{x}_0) + \vec{\varepsilon}$$

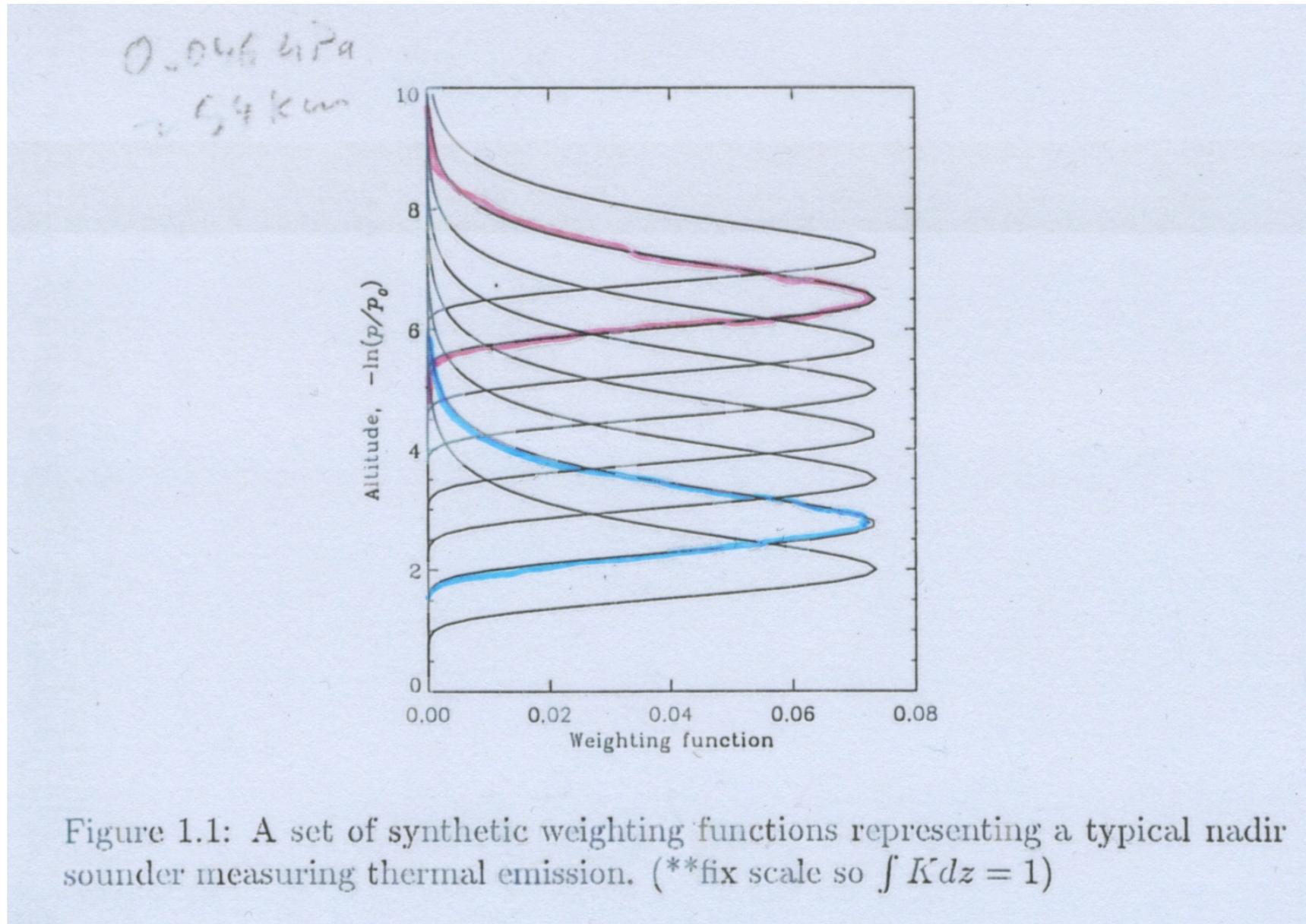


Figure from Clive D. Rodgers, Inverse Methods for Atmospheric Sounding, World Scientific, 2000.

Übersicht

- ▶ Opazitätsregel und Jacobians
- ▶ Das Inverse Problem
- ▶ **Kochrezept „Optimal Estimation“**
- ▶ Naiver Lösungsversuch
- ▶ Exkurs: Beschreibung der „Varianz“ eines Vektors
- ▶ Weniger naive Lösungsversuche
- ▶ Optimal Estimation (Bayesianische Methoden)
- ▶ Zusammenfassung

Optimal Estimation (OEM)

- ▶ Name in der Meteorologie: „**1D-Var**“
- ▶ Kombiniert Information aus der Messung mit Vorwissen („**A Priori**“)
- ▶ Lineare Algebra, Erweiterung des Konzepts der Methode der kleinsten Quadrate
- ▶ Wichtige Matrizen:
 - ▶ Jacobian
 - ▶ Kovarianzmatrizen für
 - Messung (\mathbf{S}_y)
 - A Priori (\mathbf{S}_x)

Optimal Estimation (OEM)

$$\hat{\mathbf{x}} = \vec{\mathbf{x}}_0 + (\mathbf{K}'\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_0^{-1})^{-1}\mathbf{K}'\mathbf{S}_y^{-1}(\vec{\mathbf{y}} - \mathbf{K}\vec{\mathbf{x}}_0)$$

oder

$$\hat{\mathbf{x}} = \vec{\mathbf{x}}_0 + \mathbf{S}_0\mathbf{K}'(\mathbf{K}\mathbf{S}_0\mathbf{K}' + \mathbf{S}_y)^{-1}(\vec{\mathbf{y}} - \mathbf{K}\vec{\mathbf{x}}_0)$$

$\hat{\mathbf{x}}$: Invertierter Atmosphärenzustand $\vec{\mathbf{x}}$

$\vec{\mathbf{x}}_0$: A Priori $\vec{\mathbf{x}}$

$\vec{\mathbf{y}}$: Messung $\vec{\mathbf{y}}$

\mathbf{K} : Jacobi-Matrix des Vorwärtsmodells

\mathbf{S}_0 : A Priori Kovarianzmatrix

\mathbf{S}_y : Kovarianzmatrix der Messung

$()^{-1}$: Matrix-Inversion

$()'$: Matrix-Transponierung

Averaging Kernel Matrix

Jacobi-Matrix: $\mathbf{K} = \frac{d\mathbf{y}}{d\mathbf{x}}$

Contribution Function Matrix: $\mathbf{D} = \frac{d\hat{\mathbf{x}}}{d\mathbf{y}}$

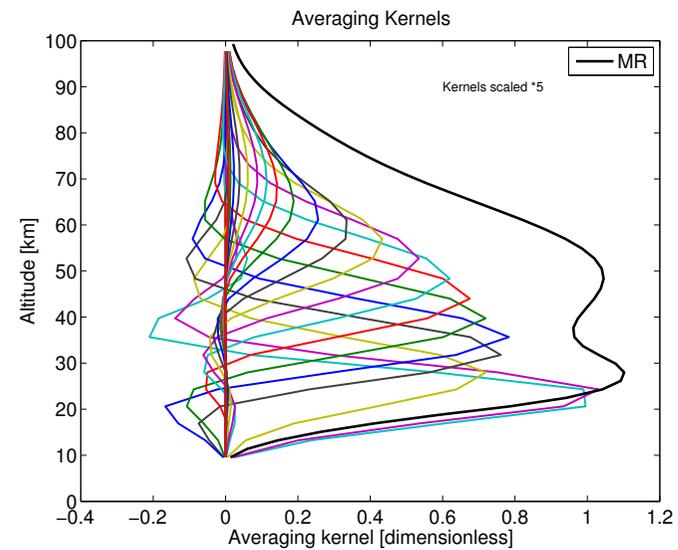
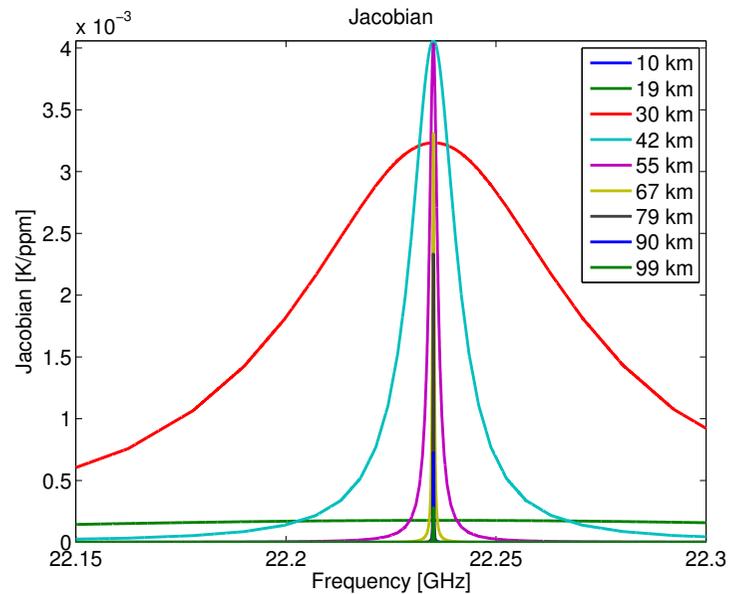
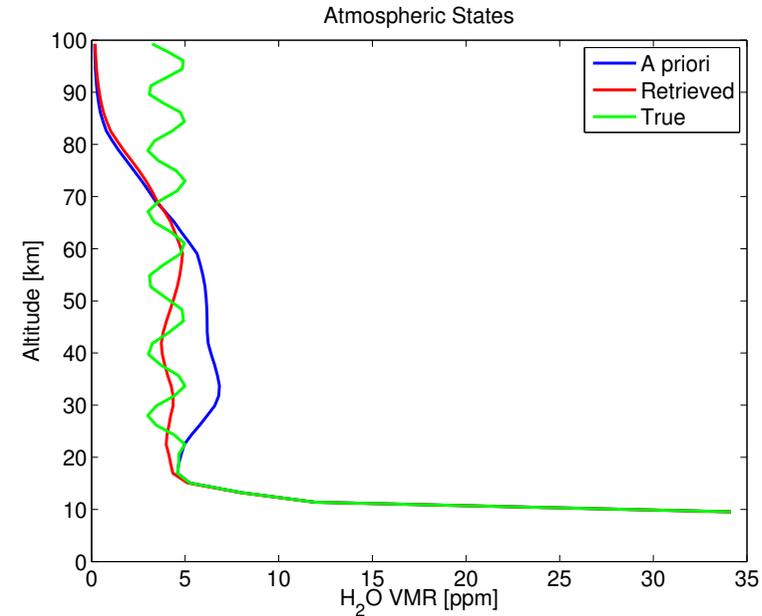
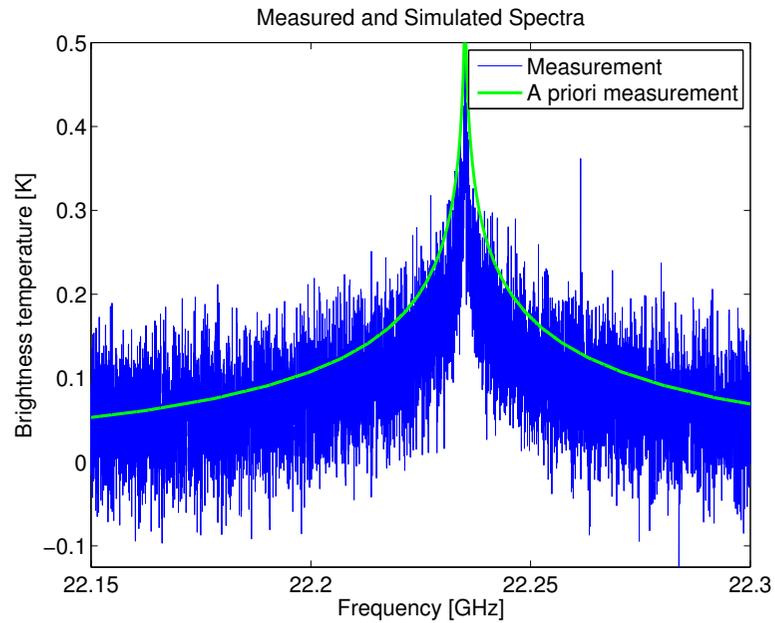
Averaging Kernel Matrix: $\mathbf{A} = \mathbf{D}\mathbf{K}$

$$\mathbf{A}^{\text{OEM}} = (\mathbf{K}'\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_0^{-1})^{-1}\mathbf{K}'\mathbf{S}_y^{-1}\mathbf{K}$$

\mathbf{A} Verbindet invertierten Zustand und wahren Zustand:

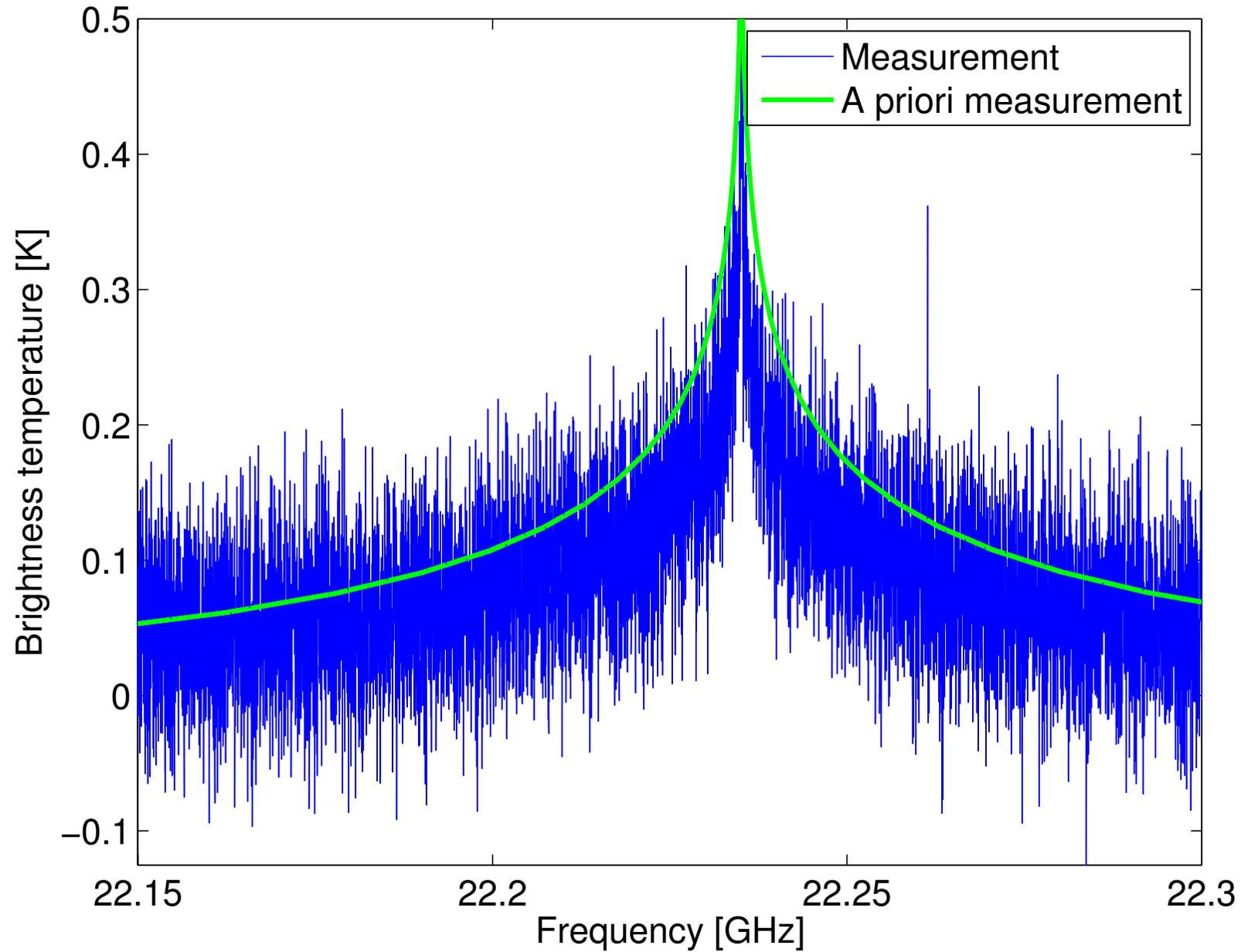
$$\hat{\mathbf{x}} - \mathbf{x}_0 = \mathbf{A}(\mathbf{x} - \mathbf{x}_0) + \mathbf{D}\boldsymbol{\varepsilon}$$

Beispiel: Bodengebundenes Wasserdampf radiometer

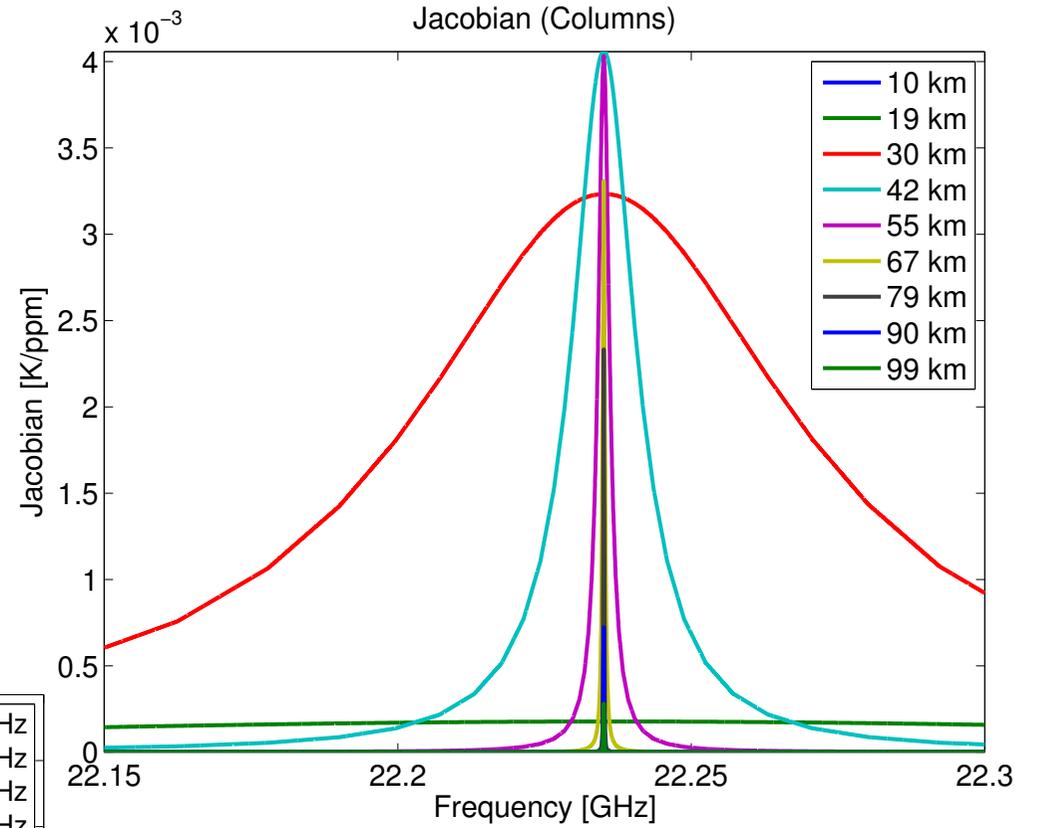
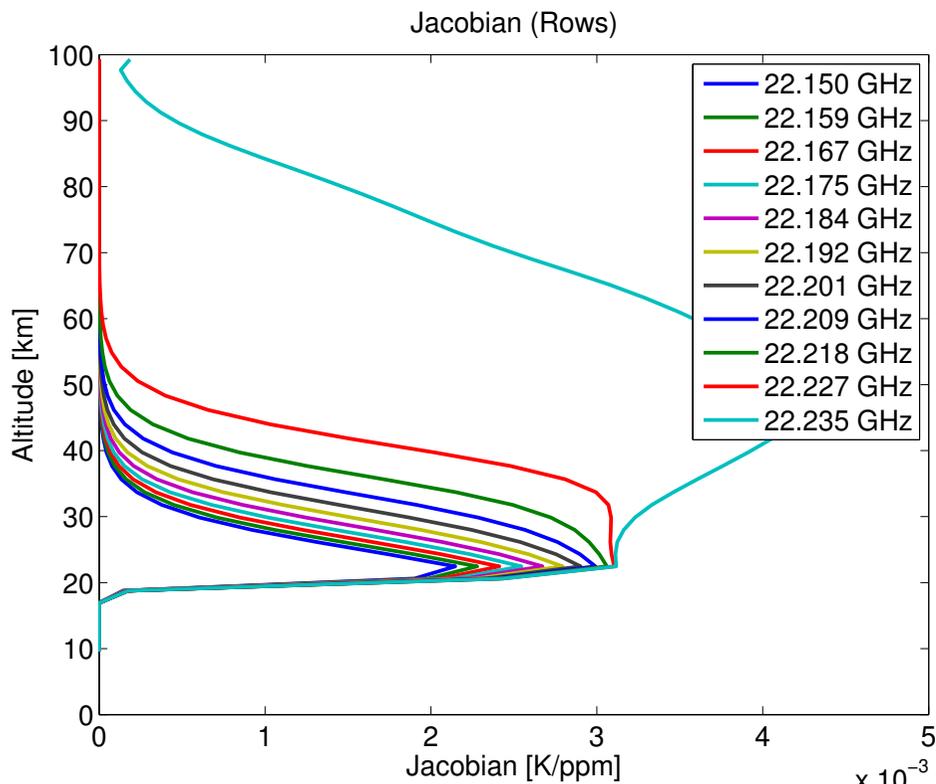


Messung und A Priori

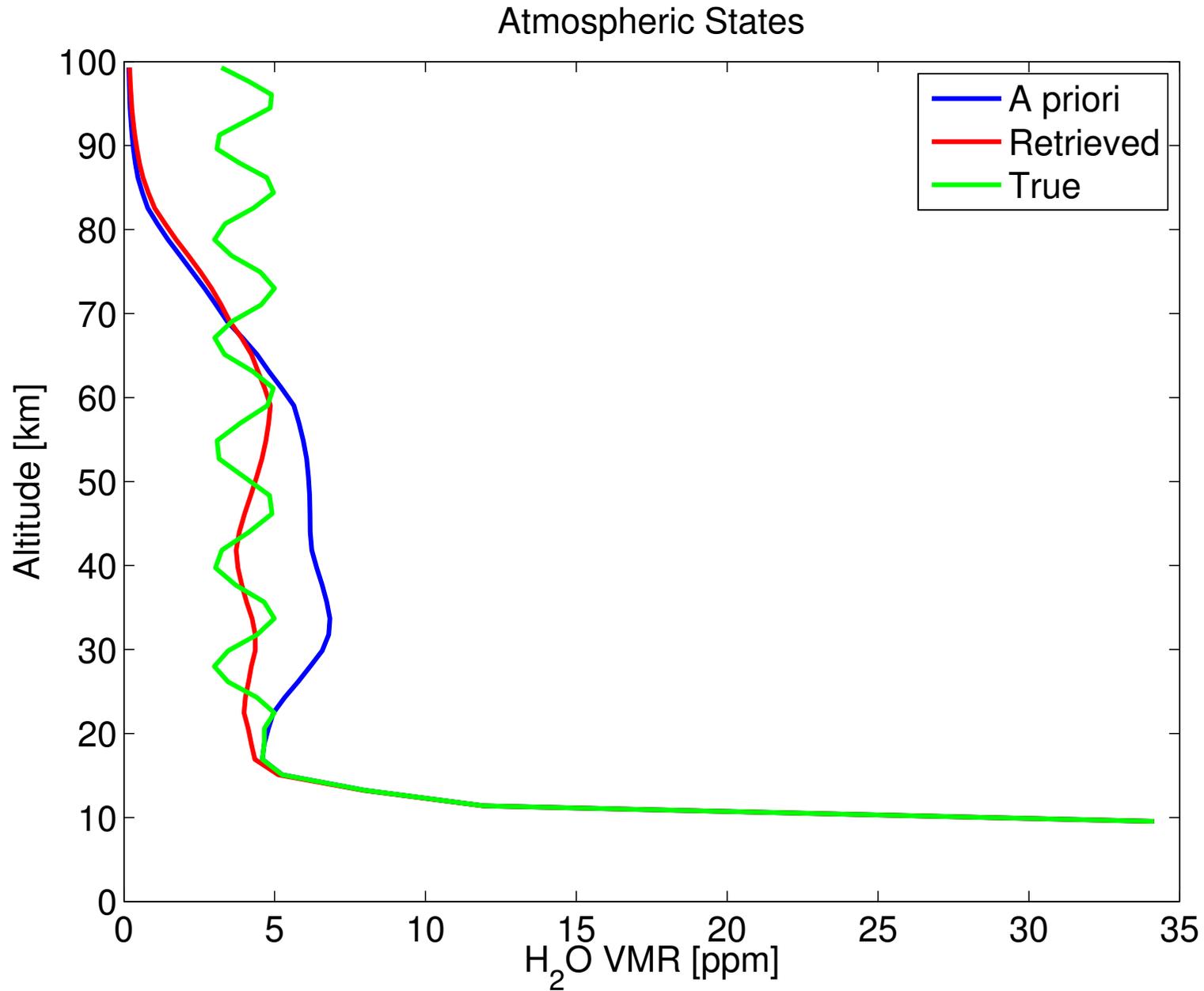
Measured and Simulated Spectra



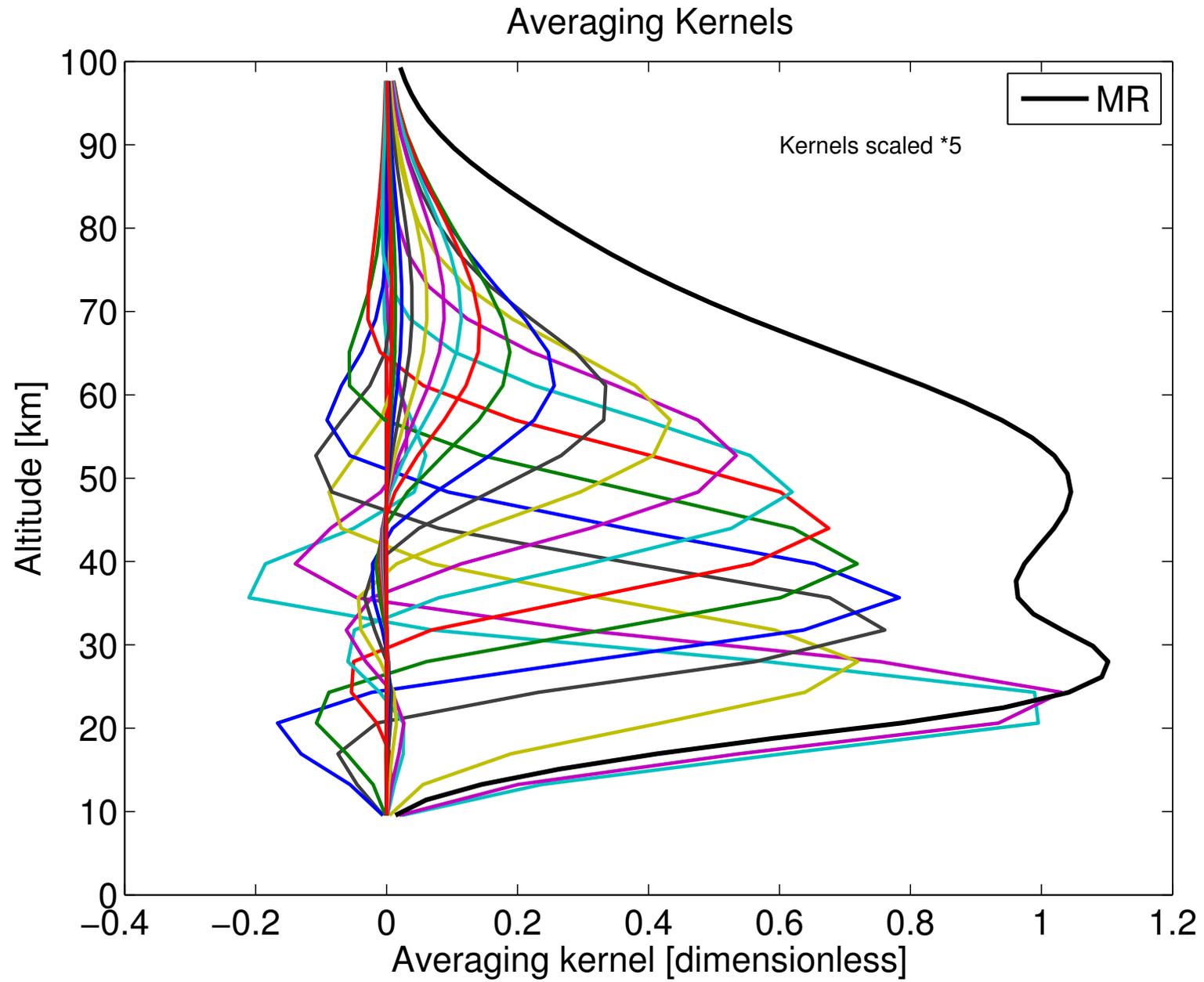
Jacobi-Matrix



Profile



Averaging Kernels



Übersicht

- ▶ Opazitätsregel und Jacobians
- ▶ Das Inverse Problem
- ▶ Kochrezept „Optimal Estimation“
- ▶ **Naiver Lösungsversuch**
- ▶ Exkurs: Beschreibung der „Varianz“ eines Vektors
- ▶ Weniger naive Lösungsversuche
- ▶ Optimal Estimation (Bayesianische Methoden)
- ▶ Zusammenfassung

Ein lineares Gleichungssystem

$$\vec{y} = \vec{y}_0 + K(\vec{x} - \vec{x}_0) + \vec{\varepsilon}$$

$$\vec{y} = \vec{y}_0 + \mathbf{K}(\vec{x} - \vec{x}_0) + \vec{\varepsilon}$$

Macht man eigentlich nicht, hier nur damit es kürzer wird.

▶ Vereinfachung: $\vec{y} = \mathbf{K}\vec{x}$

Ich linearisiere global, und ignoriere den Messfehler.

▶ \mathbf{K} ist eine $n_y * n_x$ Matrix, die Gleichung definiert ein System von n_y Gleichungen für die n_x unbekannt Elemente von x .

▶ Mathematisch, kann dieses System **überbestimmt** oder **unterbestimmt** sein. (Was heißt das?)

▶ Überbestimmt: Mehr Gleichungen als Unbekannte.

▶ Unterbestimmt: Weniger Gleichungen als Unbekannte.

▶ Was ist besser?

▶ Im unterbestimmten Fall zu wenige Gleichungen, um alle Unbekannten zu bestimmen

▶ Im überbestimmten Fall widersprechen sich die Gleichungen (zumindest wenn auch Messfehler im Spiel ist) → ebenfalls keine Lösung.



Matrix-Inversion: Ansatz

$$\vec{y} = \mathbf{K}\vec{x}$$

- ▶ Idee: wähle $n_x = n_y$.
(Ich kann meine Diskretisierung für x wählen, zum Beispiel die Höhenschritte für ein Temperaturprofil)
- ▶ Dann: $\vec{x} = \mathbf{K}^{-1}\vec{y}$
- ▶ Problem gelöst?

Matrix-Inversion: Ergebnis

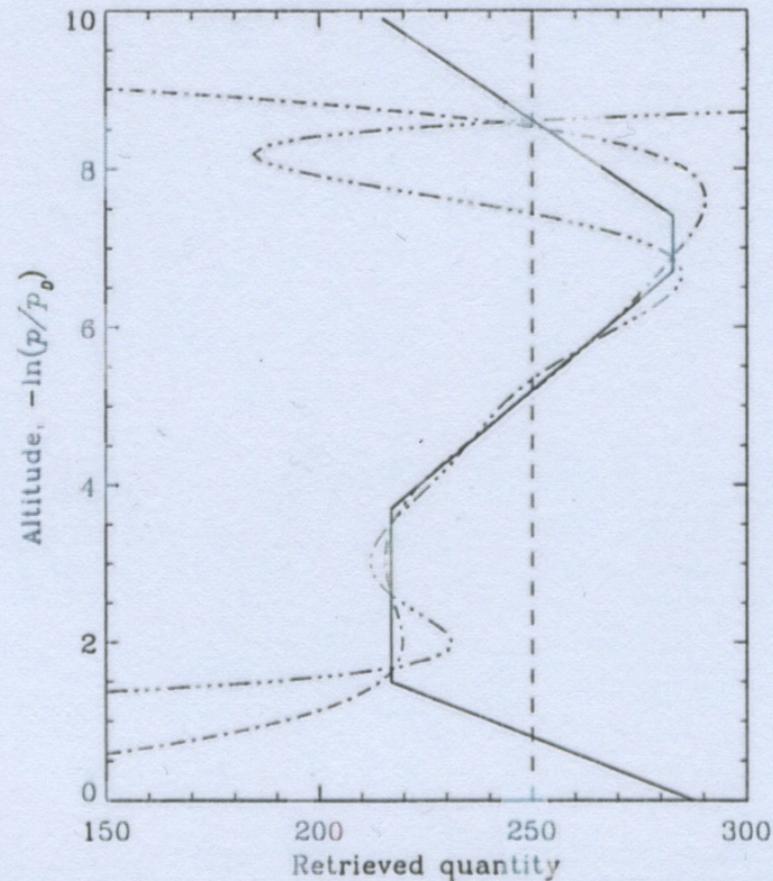
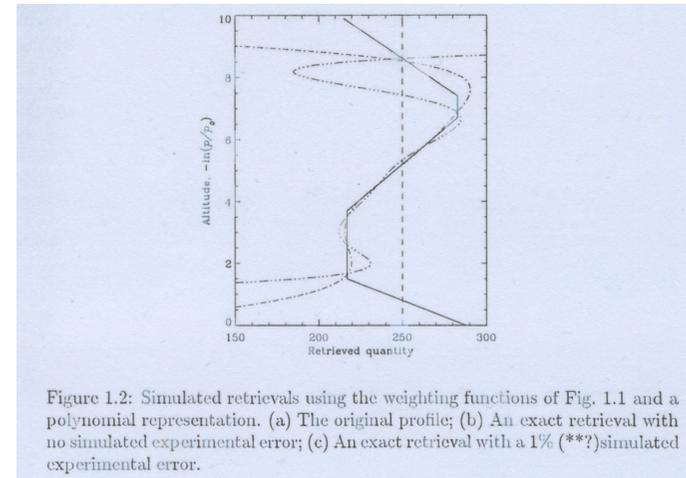


Figure 1.2: Simulated retrievals using the weighting functions of Fig. 1.1 and a polynomial representation. (a) The original profile; (b) An exact retrieval with no simulated experimental error; (c) An exact retrieval with a 1% (**?) simulated experimental error.

Figure from Clive D. Rodgers, Inverse Methods for Atmospheric Sounding, World Scientific, 2000.

Matrix Inversion: Ergebnis

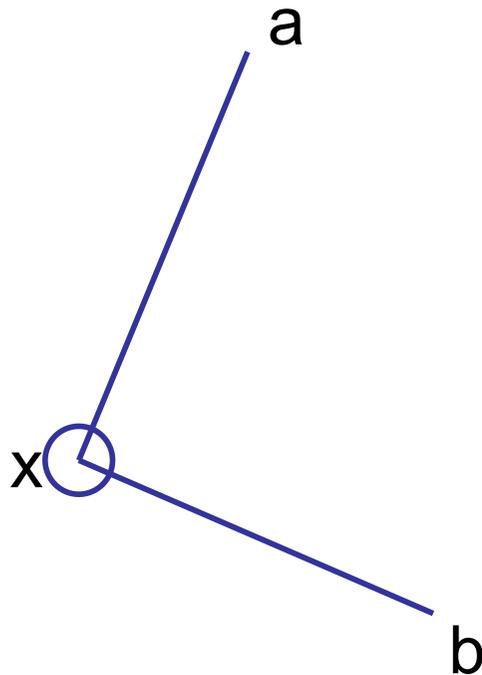
- ▶ Kleine Fehler in y (Messfehler / Rauschen) führen zu großen Unterschieden in x . (**Rauschverstärkung.**)
- ▶ Was ist das Problem mit meinem Gleichungssystem?
- ▶ Die Information der Gleichungen ist nicht unabhängig. (Gleichungen sind linear abhängig, oder beinahe linear abhängig.)
- ▶ Mathematisch sagen wir:
Das Problem ist **schlecht gestellt**. Oder:
Die K -Matrix ist **schlecht konditioniert**.
(Beide Aussagen sind äquivalent.)



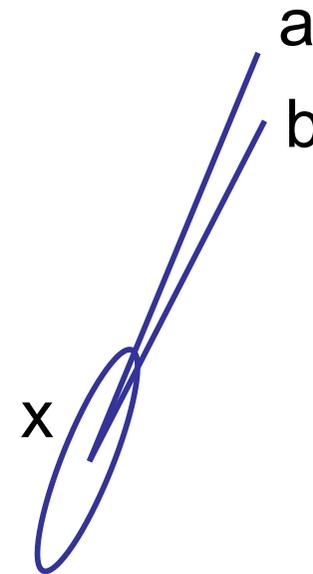
Beispiel Peilung

- ▶ Bestimme Position x eines Schiffes durch Peilung zweier bekannter Landmarken a, b .

Gut gestellt:



Schlecht gestellt:



Ortsungenaugigkeit lässt sich als Fehler-Ellipse beschreiben.

Schlussfolgerung

- ▶ Die naive direkte Matrixinversion ist in der Praxis nicht brauchbar.
- ▶ Leider führt die Strahlungstransfergleichung notorisch zu schlecht gestellten Inversionsproblemen.
(Die Mathematische Klasse, zu der die Integralform der Schwarzschild-Gleichung gehört, heißt **Fredholm-Integralgleichung**. Und schlechte Invertierbarkeit ist leider eine bekannte Eigenschaft dieser Klasse.)
- ▶ Neue Definition der Vorwärtsrichtung: Die Richtung, in der das Problem gut gestellt ist ($y=F(x)$).

Überbestimmte Gleichungssysteme

- ▶ Wir könnten x gröber diskretisieren (z.B. weniger Höhenschritte für das Temperaturprofil), so dass wir mehr Gleichungen als Unbekannt haben.
- ▶ Welche Methode ist dann anwendbar?
 - ▶ Die Methode der kleinsten Quadrate (Least-Squares Fit, lineare Regression)



Die Methode der kleinsten Quadrate (Least Squares Method)

- ▶ Anwendbar bei mehr Gleichungen als Unbekannten.
- ▶ Idee: Finde das x , dass den „Fehler“ zwischen $F(x)$ and y minimiert.
- ▶ Erste Idee: Messe Fehler durch die quadratische Abweichung:

$$\chi^2 = (\vec{y} - F(\vec{x}))' (\vec{y} - F(\vec{x}))$$

- ▶ Lösung:

$$\chi^2 = (\vec{y} - \mathbf{K}\vec{x})' (\vec{y} - \mathbf{K}\vec{x}) \quad (\text{global linearisiert})$$

$$\hat{x} = (\mathbf{K}'\mathbf{K})^{-1} \mathbf{K}' \vec{y} \quad (\text{einfacher "Least-Squares" Fit})$$

$()'$: Transponierte(r) Matrix/Vektor

$()^{-1}$: Matrizeninversion

Probleme mit der Methode der kleinsten Quadrate

- ▶ Typischerweise nur anwendbar wenn y sehr viel mehr Elemente hat als x .
- ▶ Funktioniert am besten, wenn man ein physikalisches Modell für die Lösung hat, und nur einige wenige freie Parameter bestimmen will (gut gestelltes Problem).
- ▶ Funktioniert nicht bei schlecht gestellten Problemen (Rauschverstärkung, gleiches Problem wie bei der naiven direkten Matrizeninversion).
- ▶ Löst also leider nicht unser eigentliches Problem.

Übersicht

- ▶ Opazitätsregel und Jacobians
- ▶ Das Inverse Problem
- ▶ Kochrezept „Optimal Estimation“
- ▶ Naiver Lösungsversuch
- ▶ **Exkurs: Beschreibung der „Varianz“ eines Vektors**
- ▶ Weniger naive Lösungsversuche
- ▶ Optimal Estimation (Bayesianische Methoden)
- ▶ Zusammenfassung

Fehlerbeschreibung für Vektoren

▶ Skalar: $y = \bar{y} \pm \sigma_y$ (Mittelwert und Standardabweichung)

▶ Vektor:

$$\mathbf{S}_y = \begin{pmatrix} \sigma_{11}^2 & \cdots & \sigma_{1n_y}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n_y 1}^2 & \cdots & \sigma_{n_y n_y}^2 \end{pmatrix}$$

$$\sigma_{ij}^2 = \frac{1}{n_s} \sum_s (y_s[i] - \overline{y[i]})(y_s[j] - \overline{y[j]})$$

▶ Die Kovarianzmatrix \mathbf{S} ist die Erweiterung des Konzepts der Varianz σ^2 für Vektoren.

Gauß-Funktion (Normalverteilung) für einen Vektor

Skalar:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)$$

$$\ln(p(x)) = \ln\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)\right]$$

$$\ln(p(x)) = \ln\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right] - \frac{(x-x_0)^2}{2\sigma^2}$$

$$-2\ln(p(x)) = \frac{(x-x_0)^2}{\sigma^2} + \text{const.} = (x-x_0) \sigma^{-2} (x-x_0) + \text{const.}$$

Vektor:

$\sigma^2 \rightarrow \mathbf{S}$

$$p(\vec{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{S}|^{1/2}} \exp\left[-\frac{1}{2} (\vec{x} - \vec{x}_0)' \mathbf{S}^{-1} (\vec{x} - \vec{x}_0)\right]$$

$$-2\ln(p(\vec{x})) = (\vec{x} - \vec{x}_0)' \mathbf{S}^{-1} (\vec{x} - \vec{x}_0) + \text{const.}$$

n : Dimension von \vec{x}

Flächen konstanter Wahrscheinlichkeit werden durch eine quadratische Form beschrieben! → Ellipsen

Kovarianzmatrix

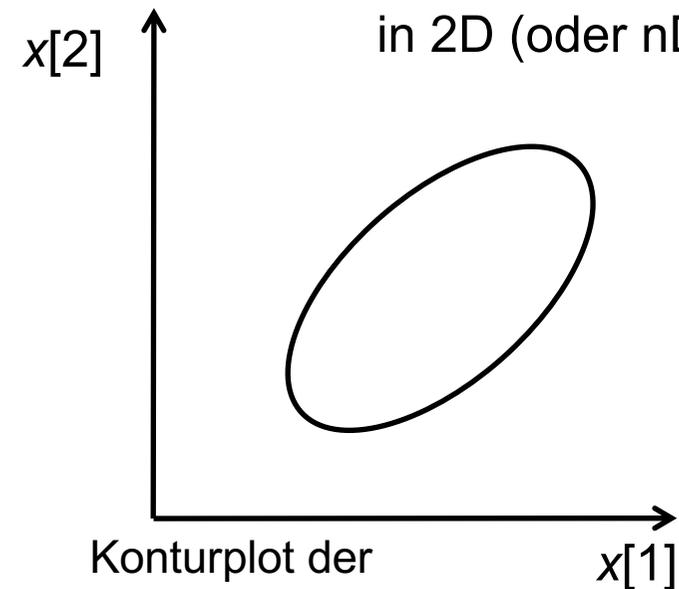
- ▶ Diagonal-Elemente: Varianzen für die einzelnen Vektor-Elemente.

$$\mathbf{S}_y = \begin{pmatrix} \sigma_{11}^2 & \cdots & \sigma_{1n_y}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n_y 1}^2 & \cdots & \sigma_{n_y n_y}^2 \end{pmatrix}$$

- ▶ Nicht-Diagonal-Elemente: **Korrelationen** zwischen den Elementen.

- ▶ Bei geeigneter Koordinaten-Transformation wird die Matrix diagonal.
→ Die Wurzeln der Diagonalelemente sind die Hauptachsen der Fehlerellipse.

Ellipse =
„Fehlerbalken“
in 2D (oder nD)



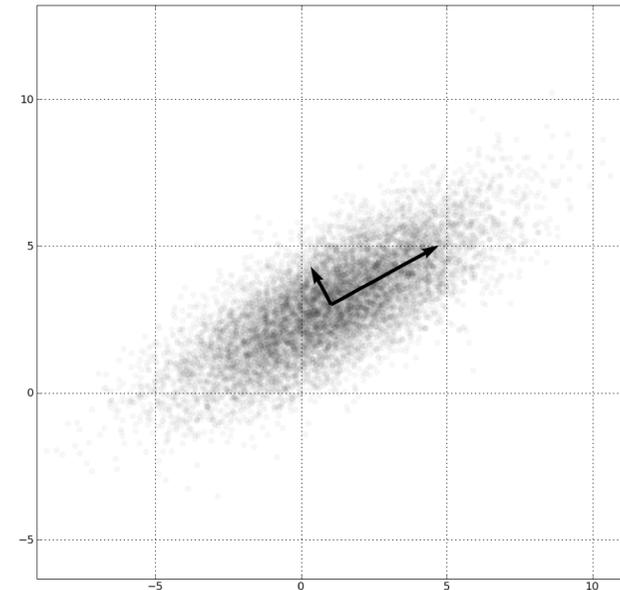
Konturplot der
Wahrscheinlichkeitsdichte
(Probability Density Function, PDF)

Fehler oder Variabilität

- ▶ Die Kovarianzmatrix muss nicht unbedingt einen (Mess-) Fehler repräsentieren, es könnte sich auch um die (natürliche) Variabilität um einen (klimatologischen) Mittelwert handeln.
- ▶ Genau entsprechend zur skalaren Varianz, es kommt auf den Zusammenhang an, wie die Größe zu interpretieren ist.
- ▶ Sie beschreibt aber immer die Variation um einen Mittelwert herum.

Empirical Orthogonal Functions (EOF)

- ▶ Auch bekannt als: Principal Components.
- ▶ Deutsch: Hauptkomponenten (Wikipedia, sagt aber niemand).
- ▶ Die Hauptachsen des Ellipsoids.
- ▶ Mathematisch: man bestimmt **Eigenvektoren** und **Eigenwerte** der Kovarianzmatrix **S**.



PCA of a multivariate Gaussian distribution centered at (1,3) with a standard deviation of 3 in roughly the (0.878, 0.478) direction and of 1 in the orthogonal direction. Source: Ben FrantzDale (Wikipedia)

Statistisch korrektere Methode der kleinsten Quadrate

- ▶ “Unterschied” zwischen Messung und Modell gemessen durch Chi-Quadrat-Test:

$$\chi^2 = (\vec{y} - F(\vec{x}))' \mathbf{S}_y^{-1} (\vec{y} - F(\vec{x}))$$

Sogenannte Quadratische Form.

Bra-Ket Notation aus der Physik:

$$\chi^2 = \langle \vec{y} - F(\vec{x}) | \mathbf{S}_y^{-1} | \vec{y} - F(\vec{x}) \rangle$$

- ▶ \mathbf{S}_y beschreibt den Messfehler (z.B. Rauschen des Instruments).
- ▶ Test misst das Quadrat der Differenz zwischen Messung und Vorwärtsmodell, gewichtet nach Größe des Messfehlers (unsichere Messungen bekommen geringeres Gewicht).

Bringt uns das weiter?

- ▶ Die statistische Verfeinerung bringt uns zunächst überhaupt nicht weiter, bahnt aber den Weg für die Optimal Estimation (OEM, in der Meteorologie auch 1D-Var)
- ▶ Dazu später...

Übersicht

- ▶ Opazitätsregel und Jacobians
- ▶ Das Inverse Problem
- ▶ Kochrezept „Optimal Estimation“
- ▶ Naiver Lösungsversuch
- ▶ Exkurs: Beschreibung der „Varianz“ eines Vektors
- ▶ **Weniger naive Lösungsversuche**
- ▶ Optimal Estimation (Bayesianische Methoden)
- ▶ Zusammenfassung

Singulärwertzerlegung (Singular Value Decomposition, SVD)

- ▶ Hilft uns dabei, das Problem der Rauschverstärkung besser zu verstehen.

$$\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

- ▶ \mathbf{U} und \mathbf{V} sind orthogonale Matrizen, \mathbf{D} ist eine diagonale Matrix.

Achtung: Das \mathbf{D} hier ist nicht die Contribution Function Matrix (Gegenstück zum Jacobian \mathbf{K}). Leider sind beide Namen Konvention.

- ▶ Was bedeutet orthogonale Matrix?

Orthogonale Matrizen

- ▶ Quadratisch, reell.
- ▶ Zeilenvektoren stehen paarweise senkrecht aufeinander, und sind alle auf 1 normiert.
- ▶ Das gleiche gilt dann automatisch für die Spaltenvektoren.
- ▶ Die Transponierte ist dann die Inverse, also:

$$\mathbf{U}\mathbf{U}' = \mathbf{I}$$

- ▶ Und ebenfalls:

$$\mathbf{U}'\mathbf{U} = \mathbf{I}$$

Singulärwertzerlegung (Singular Value Decomposition, SVD)

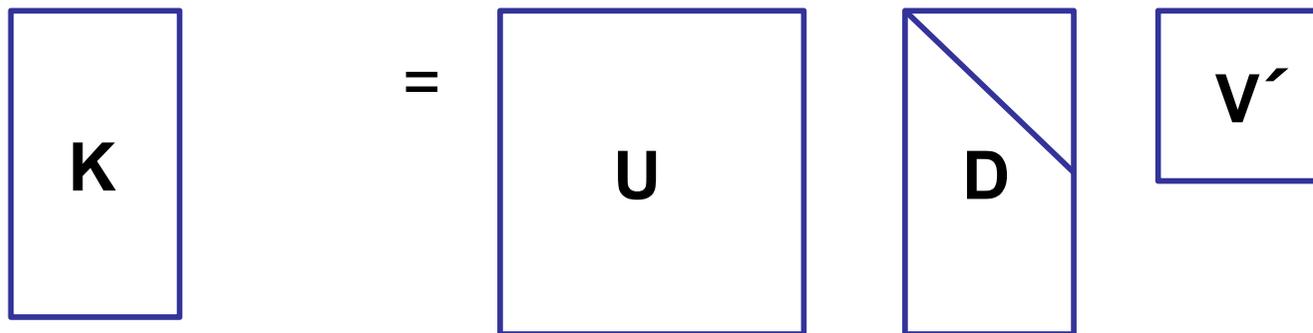
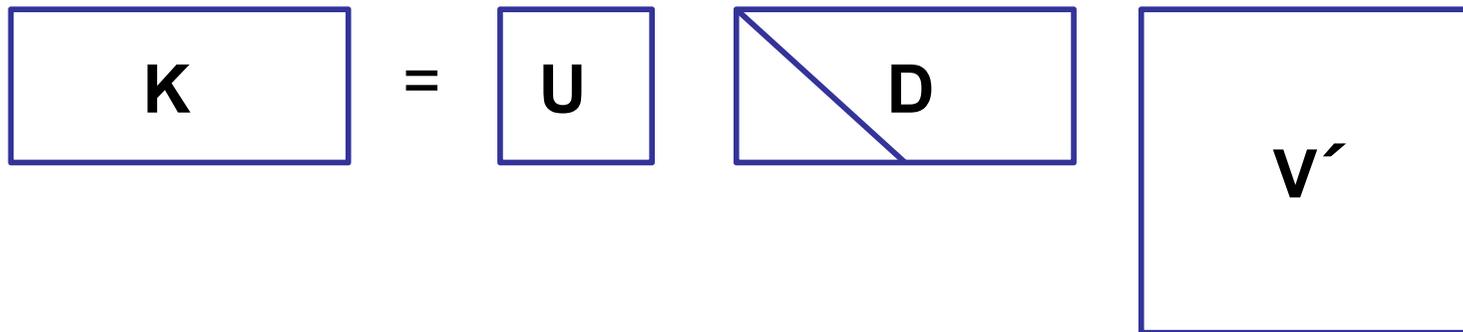
$$\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

- ▶ \mathbf{U} und \mathbf{V} sind orthogonale Matrizen, \mathbf{D} ist eine diagonale Matrix.
- ▶ Wenn man die SVD hat, ist die „Inversion“ von \mathbf{K} trivial:

$$\mathbf{K}^{-1} = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}'$$

- ▶ Nur wenn \mathbf{K} quadratisch ist, ist das die echte Inverse, sonst eine Pseudo-Inverse.
- ▶ The Elemente von \mathbf{D} heißen **Singulärwerte**.

Formfaktor



- ▶ D ist nicht quadratisch, aber diagonal, hat nur so viele Diagonalelemente wie die kleinere Dimension.
- ▶ SVD ist analog zur Bestimmung der Eigenvektoren, nur eben für nicht-quadratische Matrizen.
- ▶ Singulärwerte sind verwandt mit Eigenwerten (identisch für symmetrische (impliziert quadratisch) Matrizen).

Singulärwerte

▶ Beispiel \mathbf{D}^{-1} für $n_y=4$, $n_x=3$:

▶ Für kleine d_{ii} gibt es numerische Probleme (Teilung durch Null).

▶ Wenn alle klein sind, könnten wir einfach die ganze Matrix skalieren, das Problem ist, wenn einige groß sind, und andere klein.

▶ **Kondition** (Condition number):

$$c = \frac{\text{größter Singulärwert}}{\text{kleinster Singulärwert}}$$

$$\begin{pmatrix} d_{11}^{-1} & 0 & 0 \\ 0 & d_{22}^{-1} & 0 \\ 0 & 0 & d_{33}^{-1} \\ 0 & 0 & 0 \end{pmatrix}$$

Schlecht Konditionierte Matrizen

- ▶ Je größer die Kondition c , um so numerisch instabiler die Matrix-Inversion.
- ▶ Man sagt, \mathbf{K} ist **schlecht konditioniert (ill-conditioned)**, wenn $c \gg 1$
- ▶ Wenn \mathbf{K} schlecht konditioniert ist, dann ist das Inversionsproblem $y=\mathbf{K}x$ **schlecht gestellt**.
- ▶ Dann gibt es keine stabile Lösung ohne zusätzliche Annahmen.
- ▶ Verfahren zur Stabilisierung durch zusätzliche Annahmen heißen **Regularisierung**.

Übersicht

- ▶ Opazitätsregel und Jacobians
- ▶ Das Inverse Problem
- ▶ Kochrezept „Optimal Estimation“
- ▶ Naiver Lösungsversuch
- ▶ Exkurs: Beschreibung der „Varianz“ eines Vektors
- ▶ Weniger naive Lösungsversuche
- ▶ **Optimal Estimation (Bayesianische Methoden)**
- ▶ Zusammenfassung

Gedankenexperiment



Bild: <http://www.smosh.com>

Warum pfeift der Polizist?



Bild: <http://2.bp.blogspot.com>

Warum pfeift der Polizist?

- ▶ Weil es wahrscheinlich ist, dass ein Einbrecher dieses Verhalten zeigen würde.
- ▶ Würde er auch pfeifen, wenn gerade Fasching wäre?
 - ▶ Wahrscheinlich nicht. Grund: Mehr harmlose Bürger würden dann so herumlaufen.
- ▶ In Utopia gibt es keine Verbrechen. Würde ein utopianischer Polizist pfeifen?
 - ▶ Wahrscheinlich nicht, er würde nicht auf die Idee kommen, dass es sich um ein Verbrechen handeln könnte.

Das Bayes-Theorem

$$p(x | y) = \frac{p(x)p(y | x)}{p(y)}$$

x : Hypothese (Mann ist Einbrecher)

y : Beobachtung (steigt mit Sack aus dem Fenster)

$p(x|y)$: Gesucht: Wahrscheinlichkeit, dass die Hypothese wahr ist

$p(y|x)$: Wahrscheinlichkeit, dass ein echter Dieb sich so verhalten würde

$p(x)$: „**A Priori**“ (Wahrscheinlichkeit von Dieben)

$p(y)$: Normierung (Wahrscheinlichkeit, dass irgend jemand (Dieb oder harmloser Bürger) sich so verhalten würde)

Noch ein Beispiel

- ▶ Wenn es regnet, wird die Straße nass.
(Vorwärtsrichtung)
- ▶ ~~Wenn die Straße nass ist, hat es geregnet~~
Ist logisch nicht korrekt
- ▶ Wenn die Straße nass ist, ist es wahrscheinlicher, dass es geregnet hat. (Es gibt aber Alternativen, z.B. Straßenreinigung.)
- ▶ Das Quantitative regelt das Bayes-Theorem



$$p(x | y) = \frac{p(x)p(y | x)}{p(y)}$$

Noch ein Beispiel: Psi und so

- ▶ Skeptikerin „Paula Penibel“ testet die übersinnlichen Fähigkeiten des Mediums „Mister Mystique“. Dazu wirft P. verdeckt eine Münze, und M. muss erraten, ob Kopf oder Zahl oben liegt.
- ▶ Von 20 Versuchen liegt M. dabei 20 Mal richtig.
- ▶ Wie hoch ist die Wahrscheinlichkeit, dass dieses Ergebnis zufällig zustande gekommen ist (Nullhypothese)?
 - ▶ $0.5^{20} \approx 10^{-6}$, also ungefähr 1 zu 1 Million. 
- ▶ Ist P. jetzt davon überzeugt, dass M. tatsächlich Übersinnliche Fähigkeiten besitzt?
 - ▶ Nein, aber warum nicht? 

„Resurrection of Hypotheses“

In Wirklichkeit gibt es nicht nur die beiden Alternativhypothesen „Zufall“ oder „Hellseherei“, sondern auch z.B.: „Betrug“.

y : Beobachtung (20 Mal richtig geraten)
 $p(x|y)$: Gesucht: Wahrscheinlichkeit, dass die Hypothese wahr ist
 $p(y)$: Normierung (Wahrscheinlichkeit, dass irgend jemand (Hellseher, Betrüger, oder Zufall) dieses Ergebnis erzielen könnte
→ niedrig, aber vor allem für alle Alternativen gleich

$$p(x | y) = \frac{p(x)p(y | x)}{p(y)}$$

x_0 : **Hypothese 0 (glücklicher Zufall)**

$p(y|x_0)$: Wahrscheinlichkeit, dass Ergebnis zufällig zustande kam

~1e-6

$p(x_0)$: „A Priori“ (Wahrscheinlichkeit von Zufall)

~1



x_1 : **Hypothese 1 (M. kann Hellsehen)**

$p(y|x_1)$: Wahrscheinlichkeit, dass ein echter Hellseher richtig raten würde

~1

$p(x_1)$: „A Priori“ (Wahrscheinlichkeit von Hellsehern)

~1e-5 (jedenfalls niedrig)

x_2 : **Hypothese 2 (Betrug)**

$p(y|x_2)$: Wahrscheinlichkeit, dass ein Betrüger richtig raten würde

~1

$p(x_2)$: „A Priori“ (Wahrscheinlichkeit von Betrügern)

~1e-4 (niedrig aber $>p(x_1)$)

→ Hellseherei ist zwar wahrscheinlicher als Zufall, aber Betrug ist noch wesentlich wahrscheinlicher. Interessanterweise spielt die statistische Signifikanz (wie außergewöhnlich das beobachtete Ergebnis ist) hier also überhaupt keine Rolle.



Schlussfolgerungen aus dem Psi Beispiel



- ▶ Betrug ist insgesamt aus Paulas Sicht wahrscheinlicher als Hellseherei.
- ▶ Andere Personen mit anderem A Priori (Psi-Gläubige) können zum umgekehrten Schluss kommen! (Und sich trotzdem innerhalb ihres Weltbilds logisch verhalten.)
- ▶ Alle Naturwissenschaften zusammen bilden ein konsistentes Gebäude (das A Priori von Paula).
 - ➔ Kein einzelnes Experiment, wie überzeugend auch immer, kann Paula von Mystikes hellseherischen Fähigkeiten überzeugen. (Dafür bräuchte es eine Serie von unabhängig reproduzierbaren Experimenten und einige Jahre Zeit.)

Bayesianische Methoden

- ▶ Sind Methoden, die in Einklang mit der Wahrscheinlichkeitstheorie sind.
- ▶ Das Bayes-Theorem beschreibt, wie verschiedene Informationen kombiniert werden können. (Neue Information wird relativ zum A Priori interpretiert.)
- ▶ Daraus kann einfach und elegant die Methode der Optimalen Schätzung (**Optimal Estimation Method, OEM**) abgeleitet werden.
- ▶ Annahme bei OEM: Sowohl y als auch x (Messung und Atmosphärenzustand) haben Gauß Statistik. (Die durch entsprechende Kovarianzmatrizen beschrieben werden kann.)

Herleitung der Kostenfunktion: Bayes + Gauß

Bayes:

$$p(\vec{x} | \vec{y}) = \frac{p(\vec{x})p(\vec{y} | \vec{x})}{p(\vec{y})}$$

$$-2\ln(p(\vec{x} | \vec{y})) = -2\ln(p(\vec{x})) - 2\ln(p(\vec{y} | \vec{x})) + 2\ln(p(\vec{y}))$$

Gauß-Fehler für Messung:

$$-2\ln(p(\vec{y} | \vec{x})) = (\vec{y} - F(\vec{x}))' \mathbf{S}_y^{-1} (\vec{y} - F(\vec{x})) + \text{const.}$$

Gauß-Fehler für das A Priori:

$$-2\ln(p(\vec{x})) = (\vec{x} - \vec{x}_0)' \mathbf{S}_0^{-1} (\vec{x} - \vec{x}_0) + \text{const.}$$

Der dritte Term ($2\ln(p(\vec{y}))$) hängt nicht von \vec{x} ab.

Bayes wird zu:

$$-2\ln(p(\vec{x} | \vec{y})) = (\vec{y} - F(\vec{x}))' \mathbf{S}_y^{-1} (\vec{y} - F(\vec{x})) + (\vec{x} - \vec{x}_0)' \mathbf{S}_0^{-1} (\vec{x} - \vec{x}_0) + \text{const.}$$

Damit die Wahrscheinlichkeit maximal wird, muss die rechte Seite minimiert werden.

Die Methode der Optimalen Schätzung (Optimal Estimation Method, OEM)

- ▶ A priori: x sollte in der Nähe des klimatologischen Mittelwerts sein.
- ▶ Die **Kostenfunktion** für die Methode der kleinsten Quadrate wird dann:

$$\chi(\vec{x}) = \underbrace{(\vec{y} - F(\vec{x}))' \mathbf{S}_y^{-1} (\vec{y} - F(\vec{x}))}_{\text{Abweichung von der Messung}} + \underbrace{(\vec{x} - \vec{x}_0)' \mathbf{S}_0^{-1} (\vec{x} - \vec{x}_0)}_{\text{Abweichung vom A Priori}}$$

Neu:

- ▶ Das x , das diese Kostenfunktion minimiert, ist die OEM Lösung.
- ▶ Kann für ein lineares Problem durch simple lineare Algebra gefunden werden.

Ganz schön viele x hier

\vec{x} : Irgend ein Atmosphärenzustand

\vec{x}_0 : A Priori Atmosphärenzustand

\hat{x} : Invertierter (=bestimmter = retrievter) Atmosphärenzustand
Meine beste Schätzung, basierend auf Messung und A Priori

\vec{y} : Messung

K: Jacobi Matrix

S₀: Kovarianzmatrix zu \vec{x}_0

S $\hat{}$: Kovarianzmatrix zu \hat{x}

S_y: Kovarianzmatrix zu \vec{y}

(⁻¹): Matrix Inversion

([']): Matrix Transponierung

Lösungsansatz: Gauß-Verteilung auch für das invertierte x

Annahme: Gauß auch für die Lösung $p(\vec{x} | \vec{y})$

$$-2\ln(p(\vec{x} | \vec{y})) = (\vec{x} - \hat{x})' \hat{\mathbf{S}}^{-1} (\vec{x} - \hat{x}) + \text{const.}$$

\hat{x} : Invertiertes (=bestimmtes = retrievedes) \vec{x}

$\hat{\mathbf{S}}$: Kovarianzmatrix des invertierten \vec{x}

Bayes wird zu:

$$(\vec{x} - \hat{x})' \hat{\mathbf{S}}^{-1} (\vec{x} - \hat{x}) = (\vec{y} - \mathbf{K}\vec{x})' \mathbf{S}_y^{-1} (\vec{y} - \mathbf{K}\vec{x}) + (\vec{x} - \vec{x}_0)' \mathbf{S}_0^{-1} (\vec{x} - \vec{x}_0) + \text{const.}$$

Um das nach \hat{x} aufzulösen setzen wir jeweils die quadratischen und linearen Terme in \vec{x} gleich.

Quadratisch:

$$\vec{x}' \hat{\mathbf{S}}^{-1} \vec{x} = \vec{x}' \mathbf{K}' \mathbf{S}_y^{-1} \mathbf{K} \vec{x} + \vec{x}' \mathbf{S}_0^{-1} \vec{x}$$

$$\hat{\mathbf{S}}^{-1} = \mathbf{K}' \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{S}_0^{-1}$$

Linear (in \vec{x}'):

$$\vec{x}' \hat{\mathbf{S}}^{-1} (-\hat{x}) = (-\mathbf{K}' \vec{x})' \mathbf{S}_y^{-1} (\vec{y}) + (\vec{x})' \mathbf{S}_0^{-1} (-\vec{x}_0)$$

Das muss für jeden Wert von \vec{x} gelten, also können wir die \vec{x}' rauskürzen. Dann $\hat{\mathbf{S}}^{-1}$ von oben einsetzen:

$$(\mathbf{K}' \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{S}_0^{-1}) \hat{x} = \mathbf{K}' \mathbf{S}_y^{-1} \vec{y} + \mathbf{S}_0^{-1} \vec{x}_0$$

$$\hat{x} = (\mathbf{K}' \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{S}_0^{-1})^{-1} (\mathbf{K}' \mathbf{S}_y^{-1} \vec{y} + \mathbf{S}_0^{-1} \vec{x}_0)$$

Verwende nochmal $\mathbf{S}_0^{-1} \vec{x}_0$ aus der vorletzten Gleichung:

$$= \vec{x}_0 + (\mathbf{K}' \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{S}_0^{-1})^{-1} \mathbf{K}' \mathbf{S}_y^{-1} (\vec{y} - \mathbf{K} \vec{x}_0)$$

Wieder nur für die, die's wirklich wissen wollen.

Ausführliche Herleitung:
Rodgers Buch,
Abschnitt 2.3.2.2

OEM Lösung für lineares Problem

- ▶ Gegeben durch:

$$\hat{\mathbf{x}} = \vec{\mathbf{x}}_0 + (\mathbf{K}' \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{S}_0^{-1})^{-1} \mathbf{K}' \mathbf{S}_y^{-1} (\vec{\mathbf{y}} - \mathbf{K} \vec{\mathbf{x}}_0)$$

oder

$$\hat{\mathbf{x}} = \vec{\mathbf{x}}_0 + \mathbf{S}_0 \mathbf{K}' (\mathbf{K} \mathbf{S}_0 \mathbf{K}' + \mathbf{S}_y)^{-1} (\vec{\mathbf{y}} - \mathbf{K} \vec{\mathbf{x}}_0)$$

- ▶ Sieht kompliziert aus, aber alle Größen sind bekannt. (Für Definitionen siehe Folie am Anfang)
- ▶ Sehr einfach zu berechnen mit Matlab oder Python.
- ▶ Am teuersten: Eine Matrix-Inversion.

Erinnerung: Averaging Kernel Matrix

Jacobi-Matrix: $\mathbf{K} = \frac{d\mathbf{y}}{d\mathbf{x}}$

Contribution Function Matrix: $\mathbf{D} = \frac{d\hat{\mathbf{x}}}{d\mathbf{y}}$

Averaging Kernel Matrix: $\mathbf{A} = \mathbf{D}\mathbf{K}$

$$\mathbf{A}^{\text{OEM}} = (\mathbf{K}'\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_0^{-1})^{-1}\mathbf{K}'\mathbf{S}_y^{-1}\mathbf{K}$$

\mathbf{A} Verbindet invertierten Zustand und wahren Zustand:

$$\hat{\mathbf{x}} - \mathbf{x}_0 = \mathbf{A}(\mathbf{x} - \mathbf{x}_0) + \mathbf{D}\boldsymbol{\varepsilon}$$

Nicht-Linearität

- ▶ Absolut Linear: $y = \mathbf{K}x$
- ▶ Lokal linear: $y - y_0 = \mathbf{K}(x - x_0)$
- ▶ Gemäßigt nichtlinear (typisch bei Messung von Spurengasen oder Temperatur)
 - Iterative Methoden, speziell **Levenberg-Marquardt Iteration** (Aktualisiere $\mathbf{K}(x_i)$ während x_i iterativ verbessert wird)
- ▶ Stark nichtlinear (typisch bei Messung von Wolkenparametern)
 - Neuronale Netze, Bayesianische Monte Carlo Integration, ...

Übersicht

- ▶ Opazitätsregel und Jacobians
- ▶ Das Inverse Problem
- ▶ Kochrezept „Optimal Estimation“
- ▶ Naiver Lösungsversuch
- ▶ Exkurs: Beschreibung der „Varianz“ eines Vektors
- ▶ Weniger naive Lösungsversuche
- ▶ Optimal Estimation (Bayesianische Methoden)
- ▶ **Zusammenfassung**

Zusammenfassung

- ▶ Opazitätsregel wichtig zum intuitiven Verständnis passiver Fernerkundungsinstrumente.
- ▶ Jacobians: Wo kommt die Information her?
- ▶ Averaging Kernels: Wie gut ist mein Mess-System?
- ▶ Konditionsnummer: Wie „unmöglich“ ist mein Inversionsproblem?
- ▶ Inverse Probleme sind in der Regel schlecht gestellt.
- ▶ Statistische Verfahren
(Bayesianische Verfahren)
notwendig → Optimal Estimation
- ▶ Sehr abstrakt → Übungsaufgabe!

Leseempfehlung

- ▶ Rodgers, Chapter 2