

Regionale Schwerefeldmodellierung durch Slepian- und radiale Basisfunktionen

Markus Antoni, Wolfgang Keller und Matthias Weigelt

Zusammenfassung

Zur Erfassung des statischen und des zeitvariablen Gravitationsfeldes der Erde sind seit 2000 bzw. seit 2002 die beiden Satellitenmissionen CHAMP und GRACE im Orbit. Üblicherweise wird das aus den Missionsdaten abgeleitete Gravitationsfeld als Summe von Kugelfunktionen dargestellt. Es zeigt sich aber, dass diese global definierten Basisfunktionen nicht in der Lage sind, alle in den Daten enthaltenen lokalen Details zu repräsentieren. Der Beitrag stellt zwei Methoden dar, das verbleibende Restsignal durch lokale Basisfunktionen aufzulösen. Zum einen werden die radialen Basisfunktionen verwendet, zum anderen die Slepianfunktionen, die durch eine Orthogonalisierung der Kugelflächenfunktionen auf einem vorgegebenen Teil der Kugeloberfläche entstehen. Werden alle Parameter der radialen Basisfunktionen geschätzt, so erfordert das nichtlineare Problem neben den Startwerten auch eine Strahlminimierung, die die Konvergenz der Lösung verbessert, und ein Penalty-Verfahren, damit die geschätzten Parameter in einem sinnvollen Intervall fixiert werden. In einer Simulationsstudie werden die erreichbaren Genauigkeiten und der numerische Aufwand beider Funktionensysteme miteinander verglichen.

Summary

The satellite missions CHAMP and GRACE have been launched in 2000 and 2002 respectively, to determine the static and the time variable gravity field of the Earth. Usually, the gravity field, which is derived from the observations of the missions, is modeled by a sum of spherical harmonics. It turns out, however, that these global base functions are not able to represent all the local details of the data. This study demonstrates two methods of modeling the residual signal by localizing base functions. On the one hand, the radial base functions are used, on the other hand the Slepian functions, which are generated by orthogonalization of spherical harmonics on a subspace of the sphere. If all parameters of the radial base functions are estimated, the non-linear problem demands initial values. Additionally a line search is necessary to improve the convergence of the solution and a penalty method to fix the estimated parameters in a reasonable interval. The achieved accuracy and the numerical efforts are compared in a simulation for both systems of localizing functions.

1 Einführung

Da die Gestalt der Erde annähernd einer Kugel mit dem Radius R entspricht, liegt es nahe, Funktionen auf der Oberfläche und im Außenraum durch ein entsprechendes

Basisssystem von Kugelflächenfunktionen zu approximieren. Diese Darstellung ist für alle außerhalb einer Kugel harmonischen und regulären Funktionen möglich, insbesondere für das harmonische Störpotential des Schwerkraftfeldes der Erde (Moritz 1980):

$$\begin{aligned} T(\lambda, \vartheta, r) &= \frac{GM}{R} \sum_{n=0}^{\infty} \left(\frac{R}{r} \right)^{n+1} \sum_{m=0}^n \bar{P}_n^m(\cos \vartheta) \\ &\times (\Delta \bar{C}_{nm} \cos m\lambda + \Delta \bar{S}_{nm} \sin m\lambda). \end{aligned} \quad (1)$$

Dabei entstehen die Kugelflächenfunktionen vom Grad n und Ordnung m aus dem Produkt der assoziierter Legendrefunktionen $\bar{P}_n^m(\cos \vartheta)$ und den trigonometrischen Funktionen $\{\sin m\lambda, \cos m\lambda\}$. Der Term $\left(\frac{R}{r} \right)^{n+1}$ garantiert die eindeutige, harmonische Lösung im Außenraum $r \geq R$, während der Faktor $\frac{GM}{R}$ die Einheit eines Potentials sicherstellt. Jede harmonische Funktion im Außenraum einer Kugel lässt sich in den Positionen (λ, ϑ, r) durch obige Reihenentwicklung darstellen, wobei lediglich die einheitslosen Stokes-Koeffizienten $\{\Delta \bar{C}_{nm}, \Delta \bar{S}_{nm}\}$ angepasst werden, und gegebenenfalls zusätzliche Faktoren in der Reihe auftreten. Für eine endliche Rechenzeit muss die sphärisch-harmonische Synthese (1) bei einem Entwicklungsgrad N abgebrochen werden, zumal nicht alle Koeffizienten bekannt sind:

$$\begin{aligned} T^N(\lambda, \vartheta, r) &= \frac{GM}{R} \sum_{n=0}^N \left(\frac{R}{r} \right)^{n+1} \sum_{m=0}^n \bar{P}_n^m(\cos \vartheta) \\ &\times (\Delta \bar{C}_{nm} \cos m\lambda + \Delta \bar{S}_{nm} \sin m\lambda). \end{aligned} \quad (2)$$

Dabei entsteht ein Abbruchfehler bzw. »omission error«, der die Informationen aus den höheren Graden und Ordnungen beinhaltet. Der Informationsverlust betrifft vor allem die lokalen Details, während die globalen Signalanteile durch die endliche Synthese im Allgemeinen gut repräsentiert werden.

Die Idee der lokalen Analyse ist es, die regionalen Details im Signal mit einer anderen Repräsentation zu modellieren. Dazu arbeitet man mit Residualsignalen oder Restsignalen, also den Differenzen zwischen den tatsächlichen Beobachtungen $T(\lambda, \vartheta, r)$, die sich aus der unendlichen Reihe ergeben würden, und den synthetischen Beobachtungen aus einem globalen Modell:

$$\Delta T(\lambda, \vartheta, r) = T(\lambda, \vartheta, r) - T^N(\lambda, \vartheta, r). \quad (3)$$

Die synthetischen Beobachtungen beschreiben den langwelligen Anteil $T^N(\lambda, \vartheta, r)$, also die großräumigen Strukturen des Feldes, so dass durch die Differenz vor allem die

kleinen Details im Restsignal verbleiben. Das Residualsignal $\Delta T(\lambda, \vartheta, r)$ enthält folgende Elemente:

- lokale Details, die durch die abgebrochene Reihe $T^N(\lambda, \vartheta, r)$ nicht im globalen Modell enthalten sind,
- numerische Fehler, die durch Rundungsfehler, Integration, Differentiation, etc. entstehen,
- Fehler der Koeffizienten, sobald diese nicht mehr wie in der Simulation als fehlerfrei vorausgesetzt werden,
- Messrauschen und Messfehler, falls die Daten für das Störpotential $T(\lambda, \vartheta, r)$ aus echten Beobachtungen gewonnen werden.

Nimmt man an, dass die lokalen Details im Residualsignal dominieren, so ist es möglich dieses regional auszuwerten. Dazu gehört die Beschränkung auf ein Interessensgebiet und die Verwendung lokaler Basisfunktionen, die in diesem konzentriert sind. Zum anderen können auch die Beobachtungen selektiert werden, indem nur Bahnbögen über dem Interessensgebiet verwendet werden.

Nach einer Simulation von Residualsignalen für verschiedene Testfelder werden diese mit zwei Basisfunktionen analysiert, und die Ergebnisse gegenübergestellt.

2 Lokale Basisfunktionen

Der Abbruchfehler $\Delta T(\lambda, \vartheta, r)$ kann bei einer globalen Betrachtung nicht mehr als Signal interpretiert werden, anderenfalls wäre eine höhere Auflösung durch Kugelflächenfunktionen möglich. Da ein wichtiger Anteil des Fehlers die lokalen Variationen beschreibt, liegt es nahe, diesen als Residualsignal zu betrachten, und regional zu analysieren. Dazu werden die Daten für die Testregion mit einem zweiten, lokalen Basissystem $L_b(\lambda, \vartheta, r, \psi_b)$ modelliert:

$$\Delta T(\lambda, \vartheta, r) = \sum_{b=1}^B \eta_b L_b(\lambda, \vartheta, r, \psi_b) + \delta T(\lambda, \vartheta, r). \quad (4)$$

In dieser Darstellung bezeichnet η_b die Koeffizienten der Reihenentwicklung, die aber je nach Betrachtungsweise auch als ein Teil der Basisfunktionen angesehen werden können. Außer von der Position können die lokalen Basisfunktionen von weiteren Parametern abhängen, was durch das zusätzliche Argument ψ_b symbolisiert wird. Weil auch diese Reihe in der numerischen Berechnung abgebrochen werden muss, was hier nach der Anzahl von B Basisfunktionen geschieht, ergibt sich ein neuer Abbruchfehler $\delta T(\lambda, \vartheta, r)$. Das Ziel der Studie ist es, dass der neue Fehler mindestens um eine Größenordnung kleiner ist als das Residualsignal und damit das Signal zu mehr als 90 % durch das lokale Basissystem approximiert wird.

Grundsätzlich existiert für diese Analyse eine große Anzahl von lokalen Basisfunktionen, z. B. Punktmassen, sphärische Wavelets, Splines oder die Slepianfunktionen.

Während die Kugelflächenfunktionen bei einem globalen Support orthogonal und symmetrisch sind, schließen sich diese Eigenschaften für lokale Basissysteme aus. Im Folgenden werden die beiden Extremfälle gegenübergestellt, zum einen die orthogonalen Slepianfunktionen, die in der Regel asymmetrisch sind und deren Support sich vorab nur schwer abschätzen lässt, zum anderen die radialen Basisfunktionen, die nach ihrer Definition rotationsymmetrisch sind, und deren Support bekannt ist. Die allgemeine Darstellung der radialen Basisfunktionen beinhaltet die sphärischen Wavelets und Splines, sowie eine Approximation der Punktmassen als Spezialfall.

2.1 Radiale Basisfunktionen

Ein weit verbreitetes System lokaler Basen besteht aus den radialen Basisfunktionen $\psi_b(\varpi_b, \lambda_b, \vartheta_b, \sigma_b, \eta_b)$ mit einem zugehörigen Potential $\Psi_b(\lambda, \vartheta, r, \psi_b)$ im Orbit, wobei die Parameter im Folgenden erläutert werden. Anstelle der allgemeinen lokalen Basisfunktionen $L_b(\lambda, \vartheta, r, \psi_b)$ in Gl. (4) werden dann die »Potentialfunktionen« $\Psi_b(\lambda, \vartheta, r, \psi_b)$ verwendet, deren Koeffizienten η_b hier als Skalierungsfaktoren der Basen gedeutet werden. Nach ihrer Definition handelt es sich bei den radialen Basisfunktionen um rotationssymmetrische Basen, wobei die Funktionswerte von einem sphärischen Abstand ϖ_b abhängen. Dieser wird zwischen dem Zentrum der Basis (λ_b, ϑ_b) und dem Berechnungspunkt (λ, ϑ) auf der Einheitskugel gemessen. Aufgrund der Symmetrie ist eine Entwicklung der radialen Basisfunktionen durch zonale Kugelflächenfunktionen $\bar{Y}_{n0}(\lambda, \vartheta)$, oder alternativ durch unnormierte Legendrepolynome $P_n(\cos \varpi_b)$ möglich:

$$\psi_b(\varpi_b, \lambda_b, \vartheta_b, \sigma_b, \eta_b) = \eta_b \sum_{n=n_0}^{\tilde{N}} \sigma_b^n P_n(\cos \varpi_b). \quad (5)$$

Dabei bezeichnet n_0 den minimalen und \tilde{N} den maximalen Entwicklungsgrad der Reihenentwicklung, wobei letzterer nicht mit dem Entwicklungsgrad N des subtrahierten Referenzfeldes übereinstimmen muss. Die vom Grad n abhängigen Koeffizienten σ_b^n verändern das Aussehen der Basis, wobei in dieser Studie die Werte jeweils durch die n -te Potenz eines Formparameters σ_b berechnet werden, wie es die Schreibweise bereits nahelegt. Die Position des Zentrums (λ_b, ϑ_b) erscheint in der Gleichung nur implizit, da der sphärische Abstand

$$\varpi_b = \varpi_b(\lambda_b, \vartheta_b, \lambda, \vartheta) \quad (6)$$

zu diesem Punkt gemessen wird. Im Gegensatz zur sphärisch-harmonischen Synthese, bei der die Koeffizienten $\{\Delta \bar{C}_{nm}, \Delta \bar{S}_{nm}\}$ eine Gewichtung der unveränderten Basen bewirken, bietet es sich hier an, den Skalierungsfaktor η_b als einen Parameter der Basis zu betrachten, da diese so auf eine Punktmasse mit dem Gewicht $(\eta_b M)$ zurückgeführt werden kann (vgl. Abschnitt 4.1). Geht man

von einheitslosen Parametern aus, so muss die Gleichung wieder mit dem Term $\frac{GM}{R}$ multipliziert werden, um Potentialwerte auf der Kugel zu erhalten. Für eine harmonische Fortsetzung in den Außenraum wird analog zur sphärisch-harmonischen Synthese der Term $\left(\frac{R}{r}\right)^{n+1}$ eingeführt, so dass sich das Potential $\Psi_b(\lambda, \vartheta, r, \psi_b)$ einer Basis ψ_b folgendermaßen berechnen lässt:

$$\Psi_b(\lambda, \vartheta, r, \psi_b) = \frac{GM}{R} \eta_b \sum_{n=0}^{\tilde{N}} \left(\frac{R}{r}\right)^{n+1} \sigma_b^n P_n(\cos \varpi_b). \quad (7)$$

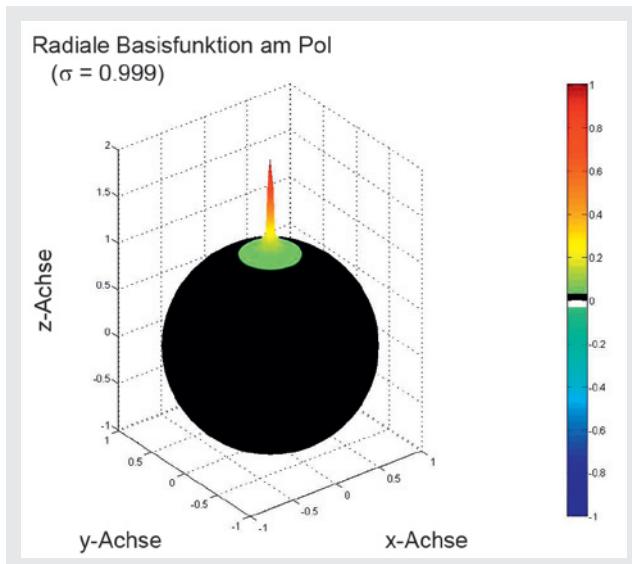


Abb. 1: Radiale Basisfunktion in polarer Lage mit einem »gutartigen« Formparameter $\sigma_b = 0.999$ und einem Entwicklungsgrad bis $\tilde{N} = 120$. Aus dem Farbverlauf ist ersichtlich, dass die Basis außerhalb ihres Zentrums rasch gegen Null konvergiert oder zumindest nur kleine, positive Werte annimmt.

Fixiert man die Entwicklungsgrade n_0 und \tilde{N} , so sind für jede radiale Basisfunktion vier Parameter zu bestimmen, nämlich die Position des Zentrums (λ_b, ϑ_b) , der Skalierungsfaktor η_b und ein Formparameter σ_b . Bei der Analyse der Testfelder wird versucht, alle Parameter aus den simulierten Daten zu schätzen. Während bei idealen Signalen die Bestimmung von Näherungswerten für die Skalierung und die Lageparameter¹ unproblematisch ist, ist dies für den σ_b -Term nicht mehr der Fall. Zum einen hängt die Bestimmung aller Parameter vom Formparameter ab, so dass das Problem nur iterativ und mit Startwerten für σ_b gelöst werden kann. Zum anderen sind nur bestimmte Größen für den Formparameter sinnvoll.

Dazu betrachtet man zunächst eine »gutartige« Basisfunktion mit $\sigma_b = 0.999$ und $\tilde{N} = 120$, wie sie in polarer Lage ($\lambda_b = \vartheta_b = 0$) in der Abb. 1 dargestellt ist. Diese, wie auch die nachfolgenden Basen, ist jeweils so skaliert, dass sie im Basiszentrum den Wert 1 annimmt. Wie man der modifizierten Farbskala entnehmen kann, konvergiert die Basis außerhalb einer kleinen Region rasch gegen Null, im Gebiet selbst ergibt sich eine sehr konzentrierte, lokale Struktur.

Erhält man durch die Ausgleichung einen kleineren Formparameter, zum Beispiel $\sigma_b = 0.888$, so wird die Basis selbst breiter, wie dies in der Abb. 2(a) veranschaulicht ist. Die Werte sind überall positiv, und das Verhalten erinnert an einen langwelligen Anteil eines Referenzfeldes, so dass diese Basisfunktionen in regionalen Analysen im Idealfall nicht zu erwarten sind.

Für die obere Grenze des Formparameters betrachtet man eine Basis mit $\sigma_b > 1$, die in der Abb. 2(b) für

¹ Bisherige Lösungen der Positionssuche gehen von einer an nähernden Kreisbahn aus, für deutlich exzentrische Bahnen muss die Suche noch modifiziert werden!

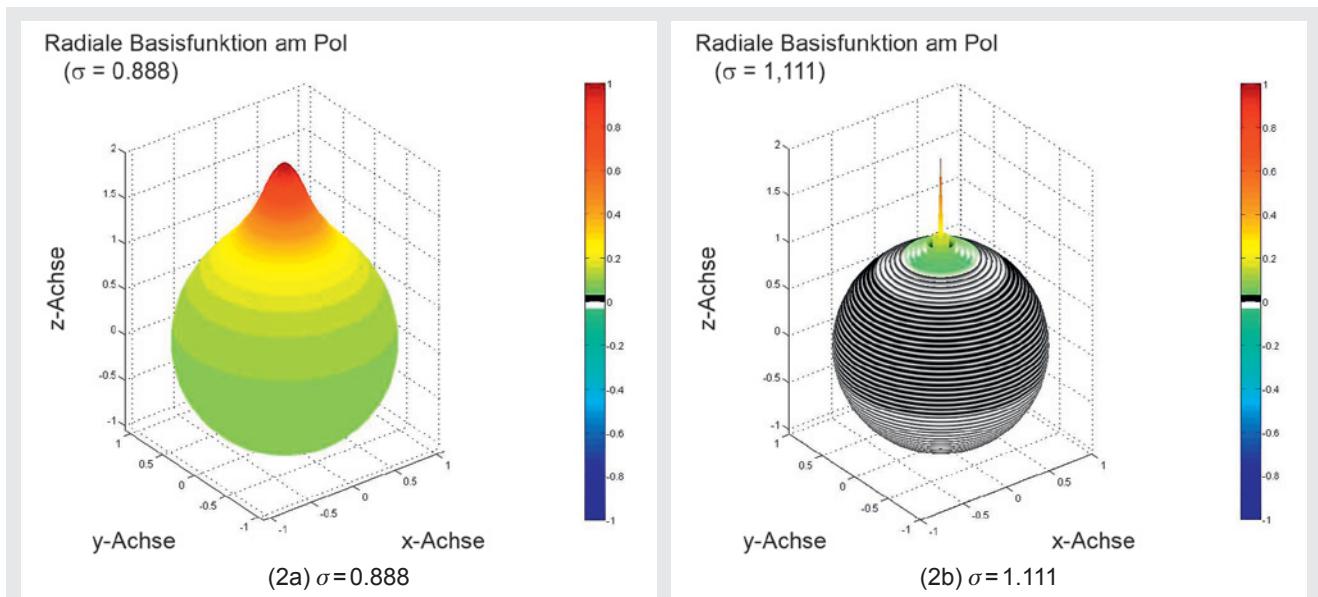


Abb. 2: Zwei radiale Basisfunktionen mit problematischen Formparametern $\sigma_b = 0.888$ bzw. $\sigma_b = 1.111$. Beide Basen wurden bis zum Entwicklungsgrad $\tilde{N} = 120$ in polarer Lage entwickelt und auf die Höhe 1 normiert.

$\sigma_b = 1.111$ gezeigt wird. Wie man erkennen kann, besitzt diese Basisfunktion wieder die sehr konzentrierte, lokale Struktur im Zentrum, die für lokale Analysen geeignet wäre. Außerhalb des Gebietes oszillieren die Funktionswerte jedoch, was für die Darstellung eines Schwerkiefeldes kein realistisches Verhalten beschreibt. Die Oszillationen lassen sich aus den Gleichungen (5) und (7) erklären, die für die Parameter $\sigma_b^n > 1$ bzw. $\frac{R}{r} \left(\frac{R}{r} \sigma_b \right)^n > 1$ keine konvergenten Reihenentwicklungen darstellen. Für einen »unendlichen« Entwicklungsgrad \tilde{N} divergieren diese Darstellungen, während die endliche Summation hier zu den Oszillationen führt.

Dieses Verhalten der Basisfunktionen in Abhängigkeit von den Formparametern σ_b erfordert es, die Schätzung derselben in der Ausgleichung auf ein sinnvolles Intervall zu beschränken. In diesem Fall geschieht dies über ein Penalty-Verfahren, wie es im Abschnitt 3.1 beschriebenen wird.

2.2 Slepianfunktionen

Ein zweites lokales Basissystem auf einer Kugel bilden die Slepianfunktionen. In diesem System wird nicht die Rotationssymmetrie, sondern die Orthogonalität der Basisfunktionen untereinander angestrebt. Im Folgenden soll das prinzipielle Konstruktionsverfahren aufgezeigt werden. Eine Darstellung der Theorie findet sich zum Beispiel in Simons und Dahlen (2006).

Man wählt zunächst ein beliebig geformtes Gebiet Ω' auf der Kugel, das dem Einzugsgebiet von Flüssen, einem Land oder Kontinent entsprechen kann. In der Simulation der Testfelder wird für eine einfachere Selektion der Daten jeweils eine sphärische Kappe um einen Zentralpunkt gewählt, wie sie in der Abb. 3 dargestellt ist.

Sphärische Kappe Ω' um $(45^\circ, 45^\circ)$

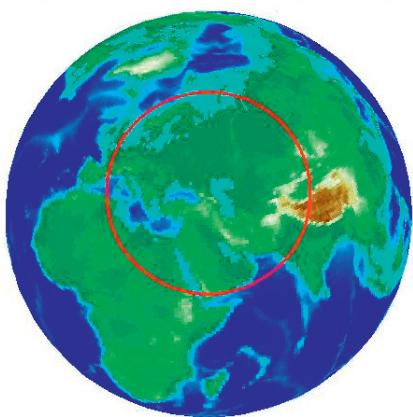


Abb. 3:
Sphärische
Kappe um
den Zen-
tralpunkt
 $(45^\circ, 45^\circ)$
mit dem
sphärischen
Abstand
 $w = 30^\circ$

Für dieses Gebiet werden numerisch die Integrale

$$D_{n,m,n',m'} = \iint_{\Omega'} Y_{nm}(\lambda, \vartheta) Y_{n'm'}(\lambda, \vartheta) d\Omega' \quad (8)$$

über alle Kombinationen von Kugelflächenfunktionen $Y_{nm}(\lambda, \vartheta)$ bis zum maximalen Grad N berechnet. Die Werte der Integrationen ordnet man in einer Matrix D an:

$$D = \begin{pmatrix} D_{0,0,0,0} & \dots & D_{0,0,N,N} \\ \vdots & & \vdots \\ D_{N,N,0,0} & \dots & D_{N,N,N,N} \end{pmatrix}. \quad (9)$$

Diese Matrix ist quadratisch und symmetrisch und besitzt bei einem maximalen Entwicklungsgrad N der Kugelflächenfunktionen insgesamt $(N+1)^2$ Zeilen und Spalten. Im nächsten Schritt wird das Eigenwertproblem

$$D\vec{g} = \Lambda\vec{g} \quad (10)$$

gelöst, was wegen der Größe der Matrix eine Einschränkung bezüglich der möglichen Auflösung darstellt. Um eine Slepianfunktion zu erhalten, wählt man einen der Eigenwerte zum Beispiel den b -ten Wert mit $b = 1, 2, \dots, (N+1)^2$ mit dem dazugehörigen Eigenvektor \vec{g}_b . Dieser liefert bei einer entsprechenden Indizierung der Elemente g_{nm}^b die sphärisch-harmonischen Koeffizienten der Slepianfunktion S_b . Das Potential einer Slepianfunktion im Außenraum $r \geq R$ lässt sich durch

$$S_b(\lambda, \vartheta, r) = \frac{GM}{R} \sum_{n=0}^N \sum_{m=-n}^n \left(\frac{R}{r} \right)^{n+1} g_{nm}^b \bar{Y}_{nm}(\lambda, \vartheta) \quad (11)$$

berechnen, während das (Residual-)Signal analog zu (4) durch die Linearkombination

$$\Delta T(\lambda, \vartheta, r) \approx \sum_{b=1}^B \eta_b S_b(\lambda, \vartheta, r) \quad (12)$$

approximiert wird.

Üblicherweise werden die Slepianfunktionen nicht nach der Auflösung – also Ordnung und Grad – geordnet, sondern nach der Größe der Eigenwerte. Je größer der Eigenwert ist, desto besser ist die Funktion im Gebiet konzentriert, wobei aus theoretischen Gründen die Eigenwerte zwischen Null und Eins liegen. Daraus ergibt sich quasi ein natürliches Kriterium für die Verwendung einer Slepianfunktion im lokalen Basissystem für ein Gebiet, indem eine Schranke für die zulässigen Eigenwerte festgelegt wird. Für die Darstellung des Residualfeldes werden nur Basisfunktionen mit großen Eigenwerten verwendet, zum Beispiel die Slepianfunktion $S_1(\lambda, \vartheta, r)$ aus Abb. 4, während andere Basen wie $S_{3010}(\lambda, \vartheta, r)$ (Abb. 5) in der Entwicklung vernachlässigt werden.

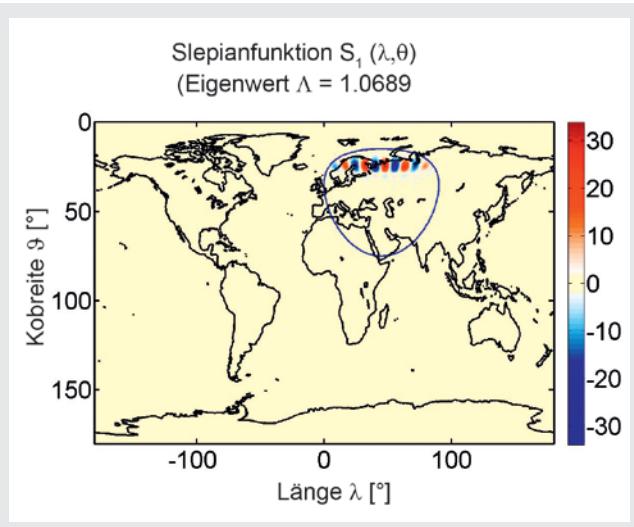


Abb. 4: Gut konzentrierte Slepianfunktion $S_1(\lambda, \vartheta, r)$ für eines der Testgebiete. Die Maxima und Minima der Funktion liegen vollständig innerhalb der sphärischen Kappe um den Zentralpunkt ($45^\circ 45^\circ$), mit dem sphärischen Abstand $w = 30^\circ$.

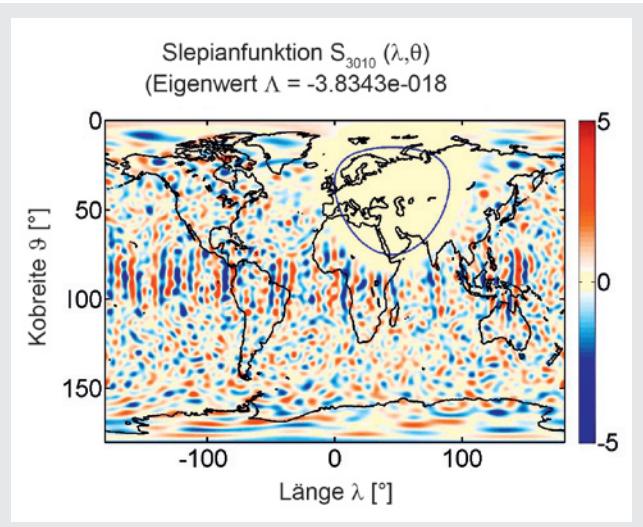


Abb. 5: Schlecht konzentrierte Slepianfunktion $S_{3010}(\lambda, \vartheta, r)$ für das gleiche Testgebiet. Die Maxima und Minima der Funktion liegen vollständig außerhalb des Interessensgebiets.

3 Schätzung der Parameter

In der Analyse werden die freien Parameter oder Koeffizienten des lokalen Basissystems so geschätzt, dass das Residualpotential bestmöglich approximiert wird. Im Fall der Slepianfunktionen $S_b(\lambda, \vartheta, r)$ mit $b = 1, \dots, B$ ergibt sich ein lineares Minimierungsproblem

$$\|\Delta T(\lambda, \vartheta, r) - \sum_{b=1}^B \eta_b S_b(\lambda, \vartheta, r)\|^2 \rightarrow \min \quad (13)$$

für die Koeffizienten η_b . Dabei bezeichnet $\|f\|$ die diskrete Norm der Funktion f

$$\|f\| := \sqrt{\sum_{k=1}^K f^2(\lambda_k, \vartheta_k, r_k)}, \quad (14)$$

mit den selektierten Punkten $(\lambda_k, \vartheta_k, r_k)$ im Interessengebiet. Für die radialen Basisfunktionen erhält man die analoge Optimierungsaufgabe

$$\|\Delta T(\lambda, \vartheta, r) - \sum_{b=1}^B \Psi_b(\lambda, \vartheta, r, \psi_b)\|^2 \rightarrow \min, \quad (15)$$

wobei die Skalierung η_b in den Parametern ψ_b enthalten ist. Üblicherweise fixiert man die Positionen der Basen in einem Gitter und wählt einen konstanten Formparameter, der sich beispielsweise aus der Kaula-Regel und eventuell den Parametern der Satellitenmission festlegen lässt. In diesem Fall ist die Aufgabe ebenfalls linear, es können aber unter Umständen sehr viele Basisfunktionen nötig sein, um eine gute Darstellung zu ermöglichen. Dadurch

verschlechtert sich die Stabilität der Lösung, was oft eine Regularisierung erforderlich macht. Im Gegensatz dazu kann man versuchen alle Parameter, also Skalierung, Formparameter und Positionen für wenige Basisfunktionen zu schätzen, wie es in dieser Studie untersucht wird. Dies reduziert die Anzahl der Parameter und macht damit eine Regularisierung zumindest für simulierte Daten unnötig, andererseits wird das Problem nichtlinear. In der Ausgleichung werden daher Startwerte, Iterationen und Abbruchkriterien benötigt.

3.1 Nichtlineare Optimierung

Die Festlegung der optimalen Positionen und Formparameter der radialen Basisfunktionen ist eine nichtlineare Minimierungsaufgabe:

$$\Phi(x_1, \dots, x_n) = \Phi(\vec{x}) \rightarrow \min. \quad (16)$$

Diese Art der Optimierungsaufgabe wird gewöhnlich durch Abstiegsverfahren gelöst (Großmann und Terno 1993). Sie bestehen darin, dass man ausgehend von einer Startnäherung $\vec{x}^{(0)}$ eine Folge $\{\vec{x}^{(n)}\}$ gemäß folgender Vorschrift erzeugt

$$\vec{x}^{(n+1)} = \vec{x}^{(n)} - \gamma_n \vec{p}^{(n)}. \quad (17)$$

Man bewegt sich also von der aktuellen Näherung $\vec{x}^{(n)}$ in der Richtung $\vec{p}^{(n)}$ zur neuen Näherung $\vec{x}^{(n+1)}$. Die Länge des Schrittes ist dabei $\gamma_n \|\vec{p}^{(n)}\|$. Beim Übergang von $\vec{x}^{(n)}$ zu $\vec{x}^{(n+1)}$ verändert sich der Wert der Zielfunktion gemäß

$$\Phi(\vec{x}^{(n+1)}) \approx \Phi(\vec{x}^{(n)}) - \gamma_n \nabla \Phi(\vec{x}^{(n)})^\top \vec{p}^{(n)}. \quad (18)$$

Er verkleinert sich also, wenn die Abstiegsbedingung

$$\nabla \Phi(\vec{x}^{(n)})^\top \vec{p}^{(n)} > 0 \quad (19)$$

erfüllt ist. Der einfachste, aber nicht unbedingt der beste Weg, dies zu erreichen, ist natürlich

$$\vec{p}^{(n)} = \nabla \Phi(\vec{x}^{(n)}). \quad (20)$$

Der Dämpfungsparameter γ_n wird durch eine Strahlminimierung bestimmt:

$$\Phi(\vec{x}^{(n)} - \gamma_n \vec{p}^{(n)}) = \min\{\Phi(\vec{x}^{(n)} - \gamma \vec{p}^{(n)}) \mid \gamma \geq 0\}. \quad (21)$$

Man berechnet das Minimum der Zielfunktion Φ entlang eines Strahles der von $\vec{x}^{(n)}$ ausgeht und in die Richtung $\vec{p}^{(n)}$ verläuft. Natürlich ist auch die Strahlminimierung ein nicht exakt lösbares Problem und muss näherungsweise durch einen Suchalgorithmus gelöst werden.

Eine Möglichkeit besteht darin, den »vollständigen« Gradienten \vec{p}^n mit den diskretisierten Dämpfungsparametern $\gamma_j = \frac{j}{J}$ mit $j = 0, 1, \dots, J$ in $(J+1)$ Teile aufzuteilen. Anschließend bestimmt man aus diesen diskreten Parametern und der Bedingung (21) den Dämpfungspараметer γ_n mit der kleinsten Residuennorm. Wenn sich dabei der Dämpfungsparameter $\gamma_n = 0$ ergibt, liegt die optimale Lösung näher am Startpunkt $\vec{x}^{(n)}$ als an allen anderen Punkten $\vec{x}^{(n)} - \gamma_j \vec{p}^{(n)}$ mit $j > 0$. In diesem Fall kann man entweder die Unterteilung verfeinern oder man nimmt dies als ein mögliches Abbruchkriterium der Iterationen.

Die bisher beschriebene Minimierungsmethode konvergiert, sobald die Startnäherung nahe genug am Minimum liegt. Die Beschaffung der Startnäherungen ist problemabhängig und muss von Fall zu Fall betrachtet werden.

Allerdings kann der Grenzwert der Minimierungsfolge auch außerhalb des zulässigen Bereiches liegen (Jarre und Stoer 2004). Von einem zulässigem Bereich spricht man, wenn für die Variablen x_i nicht alle Werte erlaubt sind, sondern sie innerhalb der Intervalle $l_j \leq x_{i_j} \leq u_j$, $j = 1, 2, \dots$ liegen müssen. Dies ist bei den radialen Basisfunktionen der Fall, deren Formparameter σ_b innerhalb vorgegebener Grenzen liegen müssen. Die annähernde Einhaltung dieser Grenzen sichert man durch ein Penalty-Verfahren, bei dem die Zielfunktion Φ durch Strafterme erweitert wird, die das Abweichen der Variablen x_i aus ihren vorgeschriebenen Schranken durch ein starkes Anwachsen der Zielfunktion bestrafen:

$$\Psi(\vec{x}) = \Phi(\vec{x}) + \rho \sum_{j=1}^k (l_j - x_{i_j})_+^n + (x_{i_j} - u_j)_+^n. \quad (22)$$

Dabei werden in der Simulation die Intervallgrenzen $l_j = 0.9$ und $u_j = 1$ sowie die Penalty-Parameter $\rho = 10000$ und $n = 20$ gewählt. Die nichtlineare Optimierungsaufgabe mit Nebenbedingungen

$$\Phi(\vec{x}) \rightarrow \min, \quad l_j \leq x_{i_j} \leq u_j \quad (23)$$

kann dann näherungsweise durch die reine Minimierungsaufgabe

$$\Psi(\vec{x}) \rightarrow \min \quad (24)$$

ersetzt werden.

3.2 Bestimmung der Näherungswerte

Um Näherungswerte für die Parameter der radialen Basen zu erhalten, werden zuerst Startwerte für Positionen und dann für die Skalierungen bestimmt. Anschließend werden mit einem Startwert σ_b und einem iterativen Gradientenverfahren alle Parameter zusammen verbessert, wobei verschiedene Abbruchkriterien denkbar sind.

Die Startwerte für die Positionen erhält man, indem man die Daten auf ein $1^\circ \times 1^\circ$ -Gitter interpoliert und mit einem Algorithmus der digitalen Bildverarbeitung lokale Extrema sucht. Durch die Interpolation wird einerseits ein mögliches Rauschen gemindert, andererseits werden durch die Gitterweite auch zu kleine Abstände der Positionen vermieden. Wegen der Exzentrizität der Bahnen werden die Extremwerte bevorzugt in den niedrigeren Orbiten lokalisiert, da diese größere Potentialwerte und -abweichungen aufweisen. Nimmt man an, dass die Bahnbögen annähernd noch kreisförmig sind, so liegt die »wahre« Position in der Nähe dieser ersten Lokalisierung. Man kann die Daten um die gefundenen Extremwerte selektieren und das Potential in einem lokalen, kartesischen Koordinatensystem (X, Y, Z) mit einem harmonischen Polynom zweiten Grades approximieren:

$$\begin{aligned} P(X, Y, Z) = & a_{11}X^2 + a_{22}Y^2 + a_{33}Z^2 + \\ & + a_{12}XY + a_{13}XZ + a_{23}YZ + \\ & + a_{10}X + a_{30}Z + a_{20}Y + a_{00}. \end{aligned} \quad (25)$$

Durch die Polynomkoeffizienten lassen sich die Werte in eine konstante Höhe transformieren. Danach verdichtet man das Koordinatengitter um den ausgewählten Extremwert im globalen Koordinatensystem und transformiert die Punkte ins lokale System. Durch die Suche der Extremwerte des Polynoms kann man die Position lokal verbessern, die man anschließend ins globale Koordinatensystem zurücktransformiert. Nach der Verbesserung aller Positionen der lokalen Extremwerte ordnet man diese nach den Potentialwerten und selektiert die betragsmäßig größten Abweichungen. Dadurch wird die Anzahl der Basen verringert und damit die Stabilität der Ausgleichung verbessert. Im nächsten Schritt wählt man einen konstanten Formparameter, zum Beispiel $\sigma_b = 0.97$, als Startwert für alle Basen und schätzt in einer linearen Ausgleichung bei festgehaltenen Positionen Näherungswerte für die Skalierungen. Da nun Startwerte für alle Unbekannten $(\lambda_b, \vartheta_b, \eta_b, \sigma_b)$ mit $b = 1, \dots, B$ gefunden sind, kann man durch eine iterative Ausgleichung das Minimierungsproblem lösen.

4 Testfelder

4.1 Simulation der Daten

Für die Simulation der Residualfelder wird das EGM96-Modell bis zum Grad $N = 150$ als Referenzfeld gewählt. In diesem Feld werden zusätzliche Punktmassen versteckt, deren Potential durch

$$V(r_Q) = \frac{GM\eta}{||\vec{r}_P - \vec{r}_Q||} = \frac{GM\eta}{r_P} \sum_{n=0}^{\infty} \left(\frac{r_Q}{r_P} \right)^n P_n(\cos \varpi) \quad (26)$$

dargestellt werden kann. Dabei bezeichnet $M\eta$ die Masse im Punkt \vec{r}_Q (mit $r_Q = ||\vec{r}_Q||$), \vec{r}_P die Position des Satelliten und ϖ den Zwischenwinkel zwischen den Vektoren. Nach wenigen Umformungen erhält man eine Formel, die Gl. (7) entspricht, wenn man die Variablen mit $r_P = r$ und $r_Q = \sigma R$ ersetzt und den Entwicklungsgrad durch \tilde{N} beschränkt. Die radialen Basisfunktionen mit dem minimalen Entwicklungsgrad $n_0 = 0$ und den Koeffizienten σ_b^n mit $0 \leq \sigma_b \leq 1$ können daher als eine Approximation der Punktmassen betrachtet werden. Zerlegt man diese in sphärisch-harmonische Koeffizienten bis zum Grad \tilde{N} und addiert diese zum EGM96, so entspricht dies zusätzlichen Testmassen in der Region.

Über diesem modifizierten Schwerefeld wird eine Orbitintegration für einen CHAMP-ähnlichen Satelliten durchgeführt, um Positionen und Geschwindigkeiten für einen Monat zu erhalten. Aus diesen Beobachtungen kann mit dem Energieintegral

$$T(\lambda, \vartheta, r) = E_{kin}(\lambda, \vartheta, r) - U(\lambda, \vartheta, r) - Z(\lambda, \vartheta, r), \quad (27)$$

in dem neben der kinetischen Energie $E_{kin}(\lambda, \vartheta, r)$ auch das Normalpotential $U(\lambda, \vartheta, r)$ und das Zentrifugalpotential $Z(\lambda, \vartheta, r)$ auftritt, das in Abb. 6 dargestellte Potential berechnet werden (Weigelt 2007). Durch eine sphärisch-harmonische Synthese über dem ursprünglichen EGM96-Schwerefeld an den entsprechenden Positionen erhält man ein Referenzpotential $T^N(\lambda, \vartheta, r)$, dessen Potentialwerte sich in den gewählten Beispielen visuell nicht von denen des modifizierten Feldes unterscheiden lassen.

Die Differenz beider Potentiale führt auf das Residualsignal $T(\lambda, \vartheta, r)$, das in der Simulation neben numerischen Fehlern vor allem das Potential der Punktmassen beinhaltet (Abb. 7). Für kleinere Datens Mengen und für eine regionale Analyse werden die Daten selektiert, so dass nur Bahnbögen und »Messdaten« über dem Interessensgebiet verwendet werden (Abb. 8).

4.2 Ergebnisse

Im ersten Testfeld werden drei Punktmassen im EGM96-Modell versteckt, die ein Residualpotential von maximal $|\Delta T(\lambda, \vartheta, r)| = 0.7 \frac{\text{m}^2}{\text{s}^2}$ erzeugen, was in der Abb. 9 dargestellt ist.

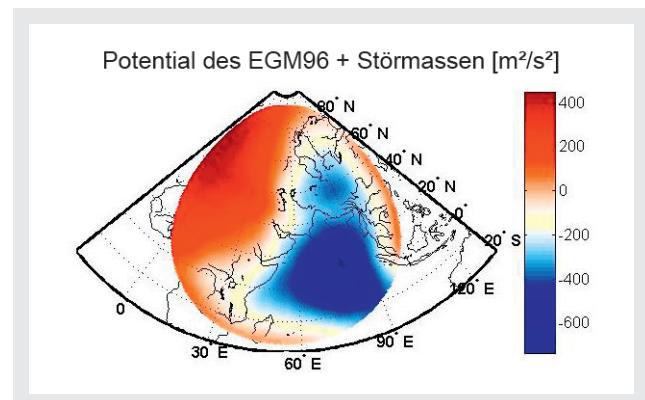


Abb. 6: Selektierte Potentialwerte für eines der Testfelder. Dieses wurde mit dem Energieansatz aus dem EGM96-Modell mit fünf zusätzlichen Punktmassen berechnet. In Abb. 7 ist dann das Residualsignal aus dem Testfeld und dem Referenzfeld dargestellt.

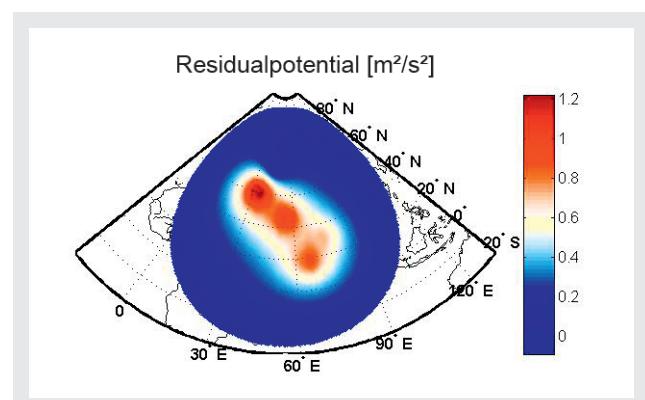


Abb. 7: Residualpotential $\Delta T(\lambda, \vartheta, r)$ aus Beobachtung und Referenzfeld. In dieser Abbildung lassen sich bereits drei der fünf versteckten Basisfunktionen unschwer erkennen.

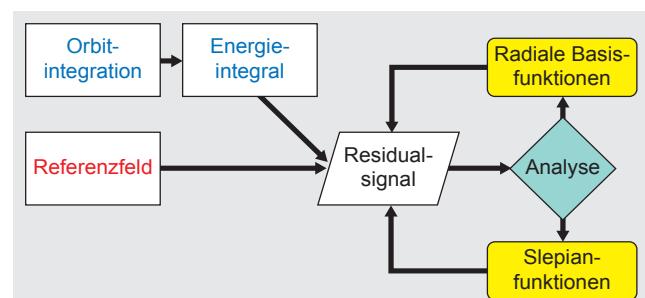
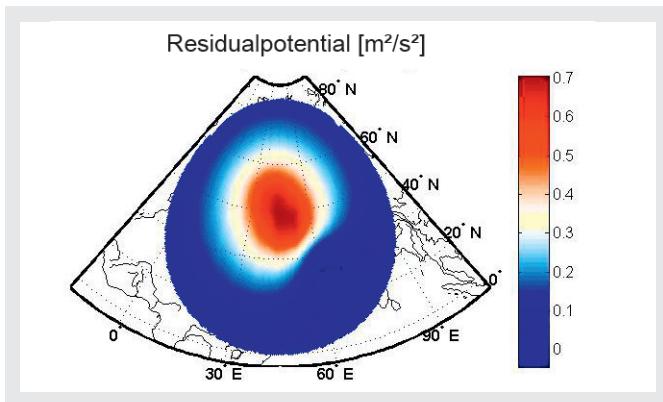
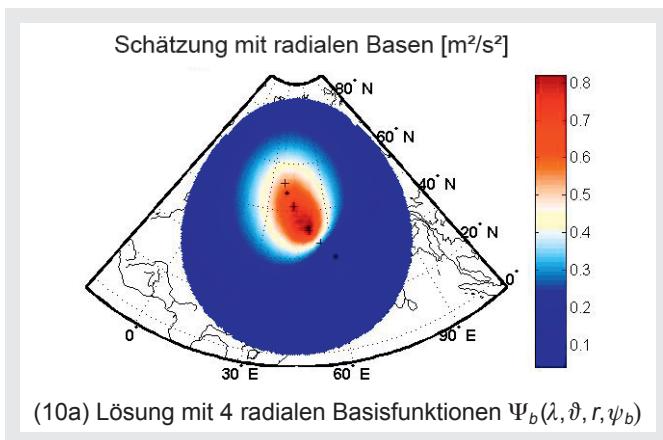


Abb. 8: Simulation der Residualpotentiale $\Delta T(\lambda, \vartheta, r)$. Dabei werden mit einer Orbitintegration und dem Energieintegral Beobachtungen erzeugt, und von diesen ein Referenzsignal aus einem globalen Modell subtrahiert. Für die Analyse werden radiale Basisfunktionen und Slepianfunktionen verwendet und die Schätzungen mit der Beobachtung verglichen.

Die Formparameter der Testmassen variieren zwischen $0.89 \leq \sigma_b \leq 0.94$, was zu einem breiten Verlauf der Potentialwerte der einzelnen Basen führt, die im Orbit zu einer einzigen Anomalie verschmelzen. In der Parameter-

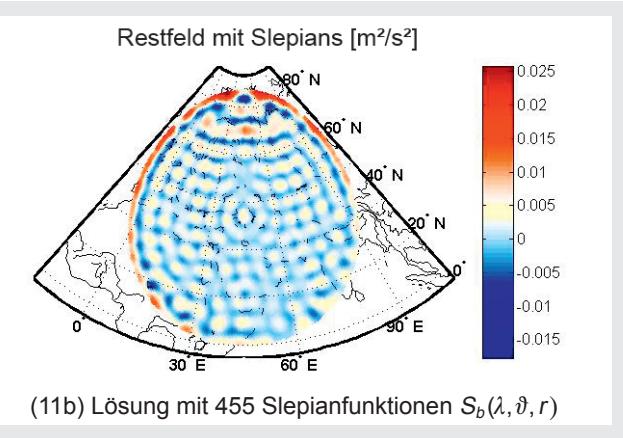
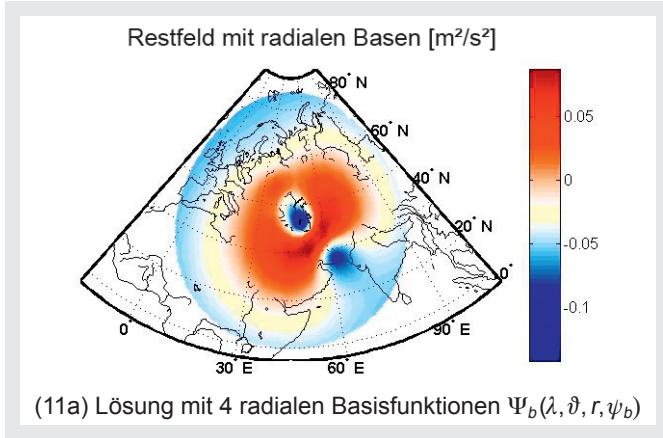
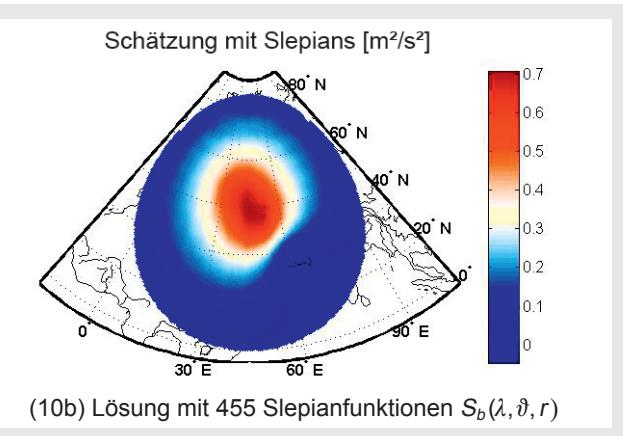


◀ Abb. 9: Darstellung der Residualpotentiale im Orbit für das Testfeld 1 mit drei Punktmasse



▼ Abb. 10: Geschätzte Residualfelder für das erste Testfeld mit drei Punktmasse

▼▼ Abb. 11: Differenzen der geschätzten Residualfelder und der »Beobachtung« für das erste Testfeld. Das Feld konnte mit den radialen Basisfunktionen zu etwa 90 % angenähert werden; der Restfehler mit den Slepianfunktionen beträgt etwa die Hälfte davon.

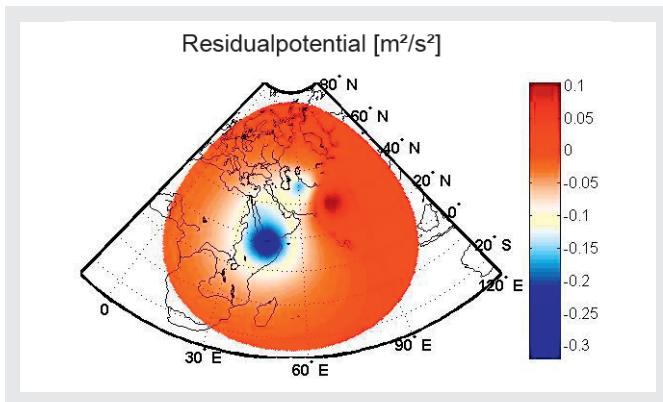


schätzung ergeben sich vier radiale Basisfunktionen, deren Formparameter mit 0.90 bis 0.95 denen der versteckten Basen entsprechen. Das Gesamtpotential wird in den Maxima um fast $0.1 \frac{\text{m}^2}{\text{s}^2}$ überschätzt, auch wenn die Strukturen bereits ähnlich sind (Abb. 10(a)). Bildet man die Differenz zur Beobachtung, sieht man, dass die Abweichungen an den größten Stellen um eine Größenordnung kleiner sind als das Signal, das Feld konnte also zu 90 % approximiert werden (Abb. 11(a)). Die Restfehler sind vor allem in der Nähe der gefundenen Positionen zu finden. Mit den Slepianfunktionen kann das Ergebnis deutlich verbessert werden, hier sieht man in der Abb. 10(b) keinen Unterschied, das Feld wurde wesentlich besser wiedergegeben, allerdings mit 455 verwendeten Basisfunktionen.

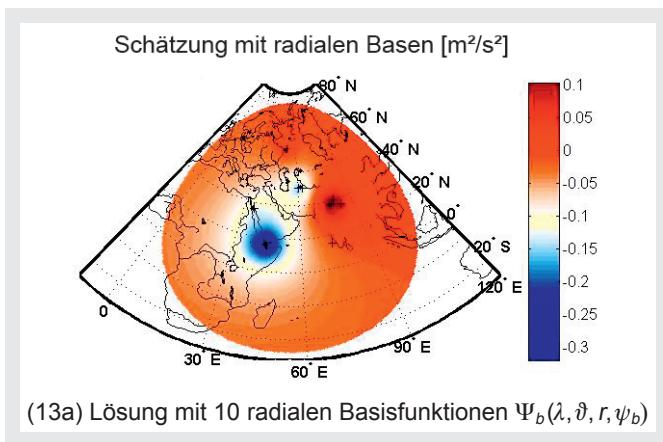
Bildet man mit den Beobachtungen und den Schätzungen die RMS-Werte pro Feld, also

$$RMS = \sqrt{\sum (SOLL - IST)^2}, \quad (28)$$

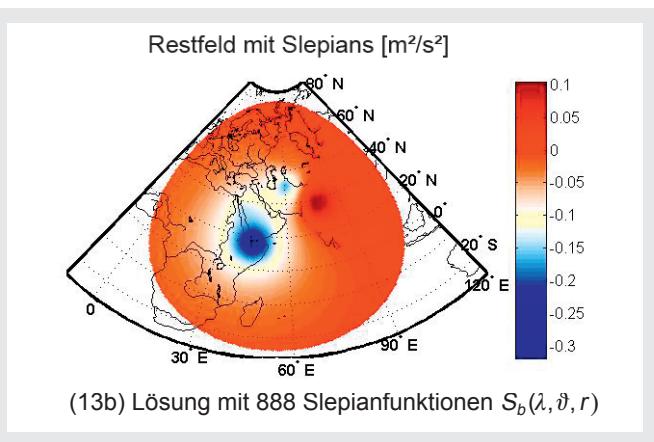
ergibt sich mit den Slepianfunktionen ein RMS-Wert von $0.62 \frac{\text{m}^2}{\text{s}^2}$, während der Wert bei den radialen Basisfunktionen mit $RMS = 6.8 \frac{\text{m}^2}{\text{s}^2}$ um eine Größenordnung größer ist.



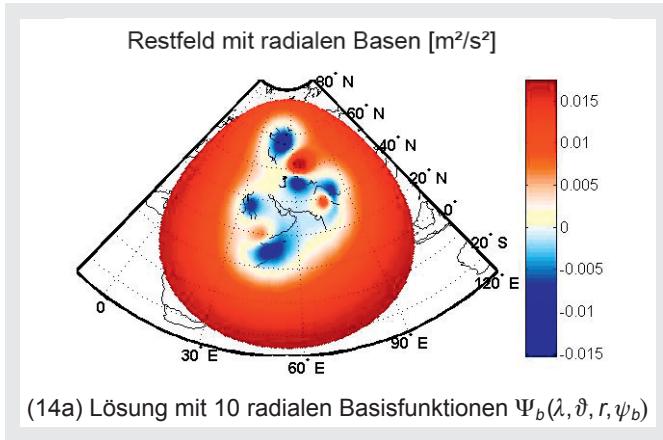
◀ Abb. 12: Darstellung der Residualpotentiale im Orbit für das Testfeld 2 mit sechs Punktmassen



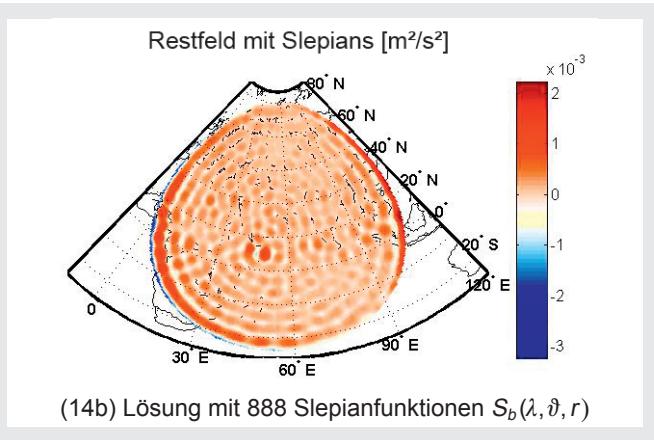
(13a) Lösung mit 10 radialen Basisfunktionen $\Psi_b(\lambda, \vartheta, r, \psi_b)$



(13b) Lösung mit 888 Slepianfunktionen $S_b(\lambda, \vartheta, r)$



(14a) Lösung mit 10 radialen Basisfunktionen $\Psi_b(\lambda, \vartheta, r, \psi_b)$



(14b) Lösung mit 888 Slepianfunktionen $S_b(\lambda, \vartheta, r)$

Ein ähnliches Bild ergibt sich auch für das zweite Testfeld mit 6 versteckten Punktmassen (Abb. 12).

Dabei werden Formparameter von $0.96 \leq \sigma_b \leq 0.98$ verwendet, was zu besser auflösbaren Strukturen für die radialen Basen führt. Der Ansatz der radialen Basisfunktionen liefert 10 Basen, deren Parameter zwischen 0.92 und 0.97 variieren, während der RMS-Wert ($2.1 \frac{\text{m}^2}{\text{s}^2}$) im Vergleich zum ersten Feld etwas abnimmt (Abb. 13(a) und 14(a)). Mit den Slepianfunktionen wird das Feld fast um eine Größenordnung besser approximiert, und man erhält einen RMS von $0.057 \frac{\text{m}^2}{\text{s}^2}$. Dies entspricht einer Verbesserung des RMS-Werts um einen Faktor 36 im Vergleich zu den radialen Basisfunktionen! Allerdings mussten dafür anstelle von 10 radialen Basisfunktionen immerhin 888 Slepianfunktionen verwendet werden, um dieses Ergebnis zu erzielen.

Schlussfolgerungen und Ausblick

Nach den Ergebnissen dieser Studie ist es mit beiden Basisystemen möglich, simulierte Residualfelder des Störpotentials auf mindestens 90 % genau zu approximieren. Die Restfehler und RMS-Werte des Ansatzes mit den Slepianfunktionen sind dabei in allen bisher untersuchten Feldern um eine Größenordnung kleiner als mit den radialen Basisfunktionen. Die Parameterschätzung mit den Slepianfunktionen ist linear, während die Bestimmung der Basisparameter auf ein nichtlineares Problem führt, das nur mit zusätzlichem Aufwand (Strahlminimierung, Penalty-Verfahren, Abbruchkriterien,...) durchgeführt werden kann. Die Vorteile der radialen Basisfunktionen liegen in den kürzeren Rechenzeiten von 5 – 10 min,

anstelle von 2 – 2.5 h, und der geringeren Anzahl von Basisfunktionen.

Grundsätzlich offen bleibt die Frage, welchen Teil der Eingangsdaten man durch ein in Kugelflächenfunktionen dargestelltes Referenzfeld modellieren sollte. Ein bis zu einem zu hohen Grad entwickeltes Referenzfeld könnte für das Residualfeld nur noch regellose Strukturen übrig lassen, für die es zweifelhaft sein kann, ob man sie noch durch lokale Basisfunktionen analysieren sollte. Für die im Beitrag angesprochene lokale Analyse von GRACE- und CHAMP-Daten liegt der Fall klarer: Hier ist das beste, allein aus GRACE- bzw. CHAMP-Daten abgeleitete Kugelflächenmodell als Referenzfeld zu wählen. Für das verbleibende Residualfeld ergeben sich Strukturen, die mit der Topographie korrelieren, also nicht vollständig vom globalen Referenzfeld erfasst werden. Die Analyse dieser Strukturen kann dann sinnvollerweise mit lokalen Basisfunktionen erfolgen.

Literatur

- Großmann, Ch., Terno, J.: Numerik der Optimierung. Teubner, Stuttgart, 1993.
Jarre, F., Stoer, J.: Optimierung. Springer-Verlag, Berlin, Heidelberg, New York, 2004.
Moritz, H.: Advanced physical geodesy. Wichmann Verlag, Karlsruhe, 1980.
Simons, F. J., Dahlen F. A.: Spherical Slepian functions and the polar gap in geodesy. *Geophys. J. Int.*, 166, S. 1039–1061, 2006.
Weigelt, M.: Global and Local Gravity Field Recovery from Satellite-to-Satellite Tracking. PhD thesis, University of Calgary, January 2007.

Anschrift der Autoren

Dipl.-Ing. Markus Antoni
Geodätisches Institut der Universität Stuttgart
Geschwister-Scholl-Straße 24D, 70174 Stuttgart
Fax: +49 711 6858-3285
markus.antoni@gis.uni-stuttgart.de

Prof. Dr. sc. techn. Wolfgang Keller
Geodätisches Institut der Universität Stuttgart
Geschwister-Scholl-Straße 24D, 70174 Stuttgart
Fax: +49 711 6858-3285
wolfgang.keller@gis.uni-stuttgart.de

Dr.-Ing. (Univ. of Calgary) Matthias Weigelt
Geodätisches Institut der Universität Stuttgart
Geschwister-Scholl-Straße 24D, 70174 Stuttgart
Fax: +49 711 6858-3285
matthias.weigelt@gis.uni-stuttgart.de