

Kapitel 5

Interpolationsmethoden

2-dimensionale Interpolation ist ein Problem das für Geowissenschaftler schon lange existiert, aber dem auch heute noch eine große Aufmerksamkeit zukommt (Burrough 1986). Ursprünglich wurden die Methoden der Interpolation zur Bestimmung der Erzstufen in Erzminen entwickelt, heute finden sie auf allen Gebieten der Geowissenschaften und auch darüber hinaus Verwendung (Bonham-Carter & Greame 1994).

Topographie kann zusätzlich zu der klassischen Konturkarte durch verschiedene analytische, numerische und digitale Methoden dargestellt werden. Da bei der Aufnahme eines Geländes nur einzelne Punkte aufgenommen werden können, ist der zu untersuchende Bereich nicht an allen Stellen definiert. Die Werte der dazwischenliegenden Punkte müssen deshalb über eine angemessene Methode interpoliert werden. Es gibt sehr viele unterschiedliche Methoden, darum ist es wichtig im Vorfeld das für die Fragestellung richtige Modell zu wählen, denn auch die Ergebnisse der verschiedenen Methoden, sind sehr unterschiedlich. Manche sind sehr generalisiert - manche sehr detailliert (Burrough 1986).

Eines der größten Probleme ist, kontinuierliche topographische Informationen einer bestimmten Region in einen angemessenen Satz von diskreten Daten zu überführen. Dabei müssen die Daten so in einer kontinuierlichen Funktion zusammengestellt werden, dass sie die topographische Oberfläche wiedergeben. Eine solche Gleichung der Oberfläche kann digital oder analytisch ausgewertet werden, ohne vorher in die geographische Form reduziert worden zu sein.

Die Interpolation dient zur Datenverdichtung (Bill 1996). Die bekannten Punkte können dabei regelmäßig oder unregelmäßig verteilt sein. Die Qualität der interpolierten Ergebnisse hängt von der Genauigkeit, Anzahl und Verteilung der der Berechnung zugrunde liegenden bekannten Punkte ab, aber auch davon wie gut die mathematische Funktion das darzustellende Phänomen modelliert. Die mathematische Funktion muß sich so stark wie möglich an das zu beschreibende Phänomen anpassen. Bei einer Temperaturmessung über einer einheitlichen Landform ist nicht zu erwarten, daß die Werte sprunghaft wechseln, sondern eher, daß sie ganz allmählich ansteigen oder abfallen. Ein Algorithmus, der die Temperaturwerte über die durchschnittliche Abweichung und glättend über die Distanz berechnet, ist in diesem Fall angemessen. Für eine rauhe Landschaft mit vielen Steigungen würde so eine glättende Methode natürlich nicht angemessen sein. Hier müssen komplexere Operationen angewandt werden. Die einfachen Methoden benutzen zur Berechnung den Durchschnitt der benachbarten Werte oder einen gewichteten Durchschnitt, wobei die Gewichtung umgekehrt

proportional zu der Entfernung des zu bestimmenden Punktes ist (Aronoff 1991). Dahinter steht die häufige Beobachtung, daß im Durchschnitt Punkte die eng beieinander liegen, eher gleiche Werte haben als solche, die weit voneinander entfernt liegen (Burrough 1986). Durch die Approximation (Glättung) werden zufällige Meßfehler beseitigt (Bill 1996). In dieser Arbeit wurde die Genauigkeit der 5 verschiedenen Interpolationsmethoden(nearest Station Methode, Inverse-Distance Methode, Kriging and upscaling) für die Interpolation der beobachteten Niederschlagsdaten vom Iran überprüft. In diesem Kapitel werden die fünf o. g. Methoden erklärt.

5.1 Inverse-Distance Methode

Eine Vielzahl der Methoden benutzt Weighting Moving Averages (Gewichtete gleitende Durchschnittswerte) von Punkten innerhalb von Einflußzonen, die meist kreisförmig sind. Eine der einfachsten Methoden basiert auf Gewichten, die umgekehrt proportional zu dem Quadrat der Distanz sind (Bonham & Graeme 1994). Die Originale Inverse Distance Weighted Methode hat Shepard entwickelt. Alle Methoden dieses Typs sind von der Shepards Methode abgeleitet. Shepards Methode basiert auf dem Ausdruck:

$$F(x, y) = \frac{\sum_{k=1}^N w_k(x, y) f_k}{\sum_{k=1}^N w_k(x, y)}$$

wobei

$$w_k(x, y) = \frac{1}{d_k^m}$$

$$d_k^m = \left((x - x_k)^2 + (y - y_k)^2 \right)^{m/2}$$

und normalerweise $m = 2$ ist, obwohl auch andere Werte eingesetzt werden können. m kann auch durch m_k ersetzt werden und möglicherweise für jedes k unterschiedlich sein (Franke 1982).

Durch die umgekehrt proportionale Gewichtung der Distanz bekommen die Datenpunkte einen, je nach Entfernung gewichteten Einfluß, auf den zu schätzenden Punkt: Datenpunkte die nah an dem zu schätzenden Punkt liegen haben einen großen Einfluß, solche die weit entfernt liegen einen geringen Einfluß. Anstatt des Quadrates der Distanz können auch andere Exponenten gewählt werden. Es können auch unterschiedliche Parameter benutzt werden, die

dann verschiedene Gebiete mit z.B. hoher, mittlerer und geringer Gewichtung definieren (Bonham-Carter & Graeme 1994) . Der Benutzer dieser Interpolationsmethode legt fest, wie viel Nachbarpunkte er in die Interpolation einbeziehen möchte. Der Einfluß der Nachbarpunkte auf einen unbeprobten Ort wird durch deren Abstand ausgedrückt. Meistens wird die Distanz jedoch potenziert ($1/d^2$ oder $1/d^3$), was von den jeweiligen Eigenschaften der zugrundeliegenden wahren Oberfläche abhängt. Je höher der Exponent, desto stärker der Einfluß naher Punkte, desto geringer der Einfluß weiter entfernter Punkte und desto unruhiger die Oberfläche. Bei einem niedrigen Exponenten kommt es hingegen zu einer Glättung. Um all diese interaktiven Einflußmöglichkeiten des Anwenders sinnvoll verwenden zu können, ist die genaue Kenntnis der spezifischen Eigenschaften der zu interpolierenden Oberfläche erforderlich.

Die Shepard Methode ist eine globale Methode, es wurden aber auch schon Schemen zur Lokalisierung vorgeschlagen (Franke 1982). Shepards Methode kann sowohl eine exakte als auch eine glättende Interpolation sein (Surfer Help 1995).

5.2 The Nearest Station Methode

In dieser Methode wird der Wert der Beobachtungsvariablen an einem Ort ohne Meßdaten durch den Wert des nächsten benachbarten gemessenen Werts geschätzt.

Der Benutzer dieser Interpolationsmethode legt fest, welchen Abstand er in die Interpolation einbeziehen möchte. Der Einfluß der Nachbarpunkte auf einen Ort ohne Meßdaten wird durch deren Abstand ausgedrückt. Z. B. betrachtet man die nächste Station unter dieser Voraussetzung, dass der Abstand zwischen dem Ort ohne Meßdaten und der Station kleiner als 2 km ist, sonst gibt es keine Auswertung mit dieser Methode.

5.3 Kriging

5.3.1 Semivariogramme

Typische regionalisierte Variablen sind Funktionen, die über den Raum verteilte, natürliche Phänomene wie z.B. Höhendaten, Niederschlagsverteilungen oder Grundwasserflurabstände beschreiben. Im Gegensatz zu Zufallsvariablen sind regionalisierte Variablen von Punkt zu Punkt über den Raum kontinuierlich. Allerdings erweisen sich die Änderungen der Variable über den Raum als so komplex, daß sie mit einer mathematischen Funktion kaum beschrieben werden können, wie das bei einer deterministischen Variablen der Fall ist.

Eigenschaften von Variablen werden als Größe, Form, Orientierung und relative räumliche Position beschrieben. Geostatistik umfaßt die Abschätzung einer regionalisierten Variablen in 1 - 3 Dimensionen.

Eine grundlegende statistische Größe ist die Semivarianz. Diese Größe definiert die gerichtete, quantitative Änderung einer Variablen. Die Semivarianz ist der Grad der

räumlichen Abhängigkeit zwischen den Proben einer bestimmten Serie. Die Semivarianz wird aus den quadrierten Differenzen zwischen einzelnen Punktepaaren, die innerhalb einer Entfernungsklasse h liegen, ermittelt.

$$\gamma_h = \sum_i^{n-h} (X_i - X_{i+h})^2 / 2n$$

wobei:

gamma = Semivarianz

X_i = ist die Größe der regionalisierten Variablen am Ort i

X_{i+h} = ist eine andere Probe h Intervalle von i entfernt

n = die Anzahl der Probepunkte

n-h = die Anzahl der Vergleiche zwischen Punktepaaren

Werden die Semivarianzen für verschiedene h ermittelt, können diese als ein Semivariogramm dargestellt werden.

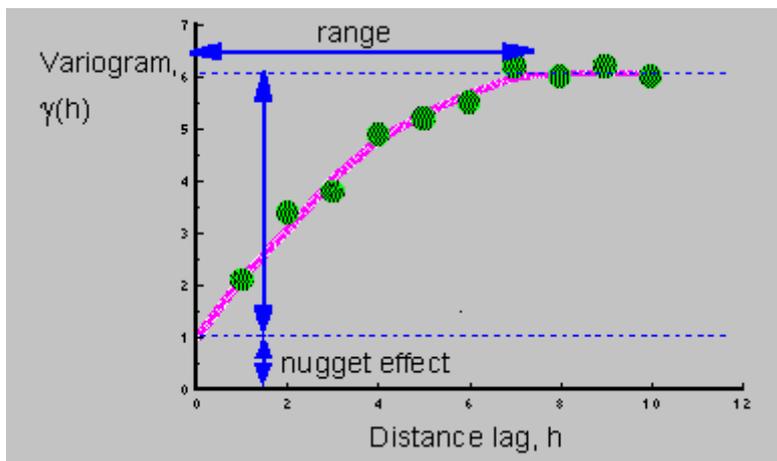


Abb. 5.1 experimentelles Semivariogramm

5.3.2 Kriging

Kriging hat im Gegensatz zu herkömmlichen Interpolationsmethoden die Fähigkeit, den Grad der räumlichen Verteilung einer Variable mit in das Verfahren einzubeziehen. Mit anderen Worten, die Schätzung wird optimiert. Informationen über diese räumliche Verteilung liefert das Semivariogramm. Da das Semivariogramm eine Entfernungsfunktion ist, ändern sich die Gewichtungen der Variablen mit der Distanz zu anderen Probepunkten. Außerdem bietet das Verfahren eine Fehlerschätzung und -optimierung.

5.3.3 Punctual Kriging

Punctual Kriging wird an punktuellen, dimensionslosen Meßpunkten zur Erzeugung eines Gitters von ebenso dimensionslosen Punkten angewendet. Es darf angenommen werden, daß die gemessene Variable stationär und demnach frei von Trends ist. Die Größe an einem nichtgemessenen Punkt wird als gewichtetes Mittel der der Meßpunkte geschätzt:

$$\hat{Y}_p = \sum W_i Y_i$$

Es wird erwartet, daß der Schätzwert vom wahren Wert abweicht:

$$\varepsilon_p = (\hat{Y}_p - Y_p)$$

Wenn die Gewichtungen der Schätzung W_i sich zu '1' aufsummieren, dann sind die resultierenden Schätzungen unbeeinflusst (unbiased), und es ist kein Trend festzustellen. Das heißt, wenn die Durchschnittsabweichung über die Gesamtprobe gleich null ist, dann löschen sich alle Über- bzw. Unterschätzungen aus. Aber auch in diesem Falle können die Schätzungen weit gestreut sein. Diese Streuung kann mit Hilfe der Fehlervarianz bzw. deren Standardabweichung beschrieben werden:

$$s_s^2 = \frac{\sum (\hat{Y}_p - Y_p)^2}{n} \quad s_s = \sqrt{s_s^2}$$

Da bei diesem Verfahren davon ausgegangen werden kann, daß die näherliegenden Punkte einen größeren Einfluß auf die Schätzung der Variablen an einem Ort ohne Messdaten haben als weiter entfernt liegende, liegt es nahe, sowohl die Gewichte als auch die Fehler auf ein vorher erstelltes Variogramm desselben Datensatzes zu beziehen. Eine Schätzung aus den Werten dreier benachbarter Punkte sähe folgendermaßen aus:

$$\hat{Y}_p = W_1 Y_1 + W_2 Y_2 + W_3 Y_3$$

Wenn die Summe aller drei Gewichtungen gleich 1 ist, dann ist die Probe unbeeinflusst und frei von Trend. Angenommen die Gewichtung W_1 ist eins und die anderen Gewichte sind gleich null, dann gilt:

$$\hat{Y}_p = Y_1$$

Der Fehler berechnet sich einfach aus der Differenz dieser beiden Werte, da Y_1 die Schätzung von \hat{Y} ist. Wenn nun viele andere, zu Y_1 nahe Punkte zu schätzen sind, kann die Varianz der Schätzung als durchschnittliche, quadrierte Differenz dieser Punktpaare betrachtet werden:

$$S_s^2 = \frac{1}{n} \sum_{i=1}^n (Y_{pi} - \hat{Y}_{li})^2$$

Die Varianz der Schätzung ist, vorausgesetzt eine identische Anzahl von Punktpaaren in einer Entfernungsklasse, doppelt so groß wie die Semivarianz. Es gibt nun eine ganze Anzahl von Gewichtungskombinationen, die alle eine unterschiedliche Schätzung sowie unterschiedliche Schätzungsfehler errechnen. Allerdings ergibt nur eine Kombination den minimalen Schätzungsfehler, die durch Kriging ermittelt werden soll. Die optimalen Gewichtungen können durch das Lösen eines Gleichungssystems ermittelt werden. Die aus dem Semivariogram gewonnenen Semivarianzen werden in dieses Gleichungssystem eingebaut. Die so errechneten Gewichtungen sind deshalb optimal, weil die Summe aller Gewichtungen gleich 1 also unbeeinflusst ist und die Fehlervarianz auf ein Minimum reduziert wurde. Keine andere lineare Kombination von Meßpunkten kann dann Ergebnisse erreichen, die eine kleinere Streuung um ihre Meßwerte haben. Wird ein Schätzwert z. B. aus drei Meßwerten einer Variablen ermittelt, kann die Lösung aus folgendem Gleichungssystem ermittelt werden:

$$W_1\gamma(h_{11}) + W_2\gamma(h_{12}) + W_3\gamma(h_{13}) = \gamma(h_{1p})$$

$$W_1\gamma(h_{12}) + W_2\gamma(h_{22}) + W_3\gamma(h_{23}) = \gamma(h_{2p})$$

$$W_1\gamma(h_{13}) + W_2\gamma(h_{23}) + W_3\gamma(h_{33}) = \gamma(h_{3p})$$

Die Indizes der Entfernungsklassen sagen aus, welche Semivarianz eingesetzt werden muß. Die Matrix auf der linken Seite ist symmetrisch, weil $h_{ij} = h_{ji}$. An der Diagonale sind die Entfernungen gleich Null, weil die Distanz von einem Punkt zu sich selbst eingesetzt wird. Geht das Variogramm durch den Nullpunkt des Diagramms, ist die Semivarianz an diesen Stellen auch gleich null. Die Semivarianzen werden dem vorher erstellten Semivariogramm entnommen.

Um allerdings die Summe der Gewichte auf '1' zu bringen, muß eine vierte Gleichung in das System eingesetzt werden:

$$W_1 + W_2 + W_3 = 1.0$$

Das ergibt aber ein System von vier Gleichungen mit nur drei Unbekannten. Da mehr Gleichungen als Unbekannte vorhanden sind, kann eine vierte Unbekannte, der sogenannte 'Lagrange-Koeffizient' eingesetzt werden. Diese Unbekannte ermöglicht eine Lösung mit dem kleinsten möglichen Schätzfehler. Das Gleichungssystem sieht dann folgendermaßen aus:

$$W_1\gamma(h_{11}) + W_2\gamma(h_{12}) + W_3\gamma(h_{13}) + \lambda = \gamma(h_{1p})$$

$$W_1\gamma(h_{12}) + W_2\gamma(h_{22}) + W_3\gamma(h_{23}) + \lambda = \gamma(h_{2p})$$

$$W_1\gamma(h_{13}) + W_2\gamma(h_{23}) + W_3\gamma(h_{33}) + \lambda = \gamma(h_{3p})$$

$$W_1 + W_2 + W_3 + 0 = 1$$

Die Matrix kann wie folgt umgestellt werden:

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \gamma(h_{13}) & 1 \\ \gamma(h_{12}) & \gamma(h_{22}) & \gamma(h_{23}) & 1 \\ \gamma(h_{13}) & \gamma(h_{23}) & \gamma(h_{33}) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ \gamma(h_{3p}) \\ 1 \end{bmatrix}$$

oder in der allgemeinen Form:

$$[A] * [W] = [B]$$

Die Werte für Matrix [A] und Vektor [B] können direkt dem Variogramm entnommen werden. Sind die Werte des Vektors [W] ermittelt, können diese direkt in die Gleichung:

$$\hat{Y}_p = W_1 Y_1 + W_2 Y_2 + W_3 Y_3$$

eingesetzt werden. Die Varianz der Schätzung ist demnach:

$$s_s^2 = W_1 \gamma(h_{1p}) + W_2 \gamma(h_{2p}) + W_3 \gamma(h_{3p}) + \lambda$$

Die Varianz ist demnach die Summe der gewichteten Semivarianzen plus Lagrange-Koeffizienten. Aus diesem Beispiel ist ersichtlich, dass Kriging zwei große methodische Vorteile besitzt:

Schätzungen mit dem kleinsten möglichen Fehler werden produziert

Kriging quantifiziert diesen Fehler mit Hilfe der Varianz der Schätzung.

5.3.4 Universal Kriging

Wenn die regionalisierte Variable nicht mehr stationär ist, dann kann 'punctual kriging' nicht mehr angewendet werden. Sobald der Trend sich ändert, oder es einen geringen Unterschied im Durchschnittswert gibt, dann ist der lineare Schätzer nicht mehr unbeeinflusst. Die errechneten Schätzungen nehmen systematisch Steigerungen auf, die von der Lage der Messpunkte abhängt, und in der generellen Fallrichtung einer Fläche folgt. Die nichtstationäre Variable wird in zwei Komponenten aufgeteilt:

- Der Trend besteht aus dem Durchschnitt oder dem Erwartungswert einer regionalisierten Variablen innerhalb einer Nachbarschaft und ist der nichtstationäre, variierende Teil der Variablen
- Das Residual ist die Differenz zwischen dem gemessenen Wert und dem Trend.

Wenn also der Trend von der regionalisierten Variablen entfernt wird, sollten sich die Residuen stationär verhalten, was ein Kriging ermöglichen würde. Der Ablauf des 'Universal kriging' sieht wie folgt aus:

1. Schätzung und Beseitigung des Trends
2. Kriging der stationären Residuen um die benötigten Schätzwerte zu erhalten
3. Die geschätzten Residuen werden mit dem Trend kombiniert, um eine Schätzfläche zu erhalten

Der Trend ist also ein Analogon zur Trendfläche, außer dass die Schätzung des Trends ausschließlich auf Messpunktwerte der umgebenden Nachbarschaft basiert.

5.4 Upscaling

In dieser Methode wird der Mittelwert der gemessenen Variablen, die in einer Modell-Gitterzelle liegen, als Gitterpunkt betrachtet. Der Benutzer dieser Interpolationsmethode muss nur die Anzahl der Stationen und die dazugehörigen Werte festlegen. Wenn man die Koordinaten des betrachteten Modell-Abschnitts und die Auflösung des Modells weiß, berechnet man dann den Mittelwert der Variablen in jeder Gitterzelle. Diese Methode wurde als Interpolationsmethode für die Verifikation des Modells beim EZMW-Zentrum verwendet.

5.5 Crossvalidation

Um die beste Interpolationsmethode für den täglichen Niederschlag über den Iran herauszufinden wurde die Crossvalidation-Technik verwendet. Die Auswahl der Interpolationstechnik erfolgte über zahlreichen Tests in Hinblick auf die Plausibilität der Ergebnisse und mittels des Verfahrens der Crossvalidation. Crossvalidation testet die prädiktiven Fähigkeiten der Interpolationsmethode und ihrer unterschiedlichen Parametereinstellungen: aus der Menge aller Stützpunkte der Interpolation, den Messdaten, wird ein Stützpunkt entfernt, anschließend mit allen anderen Stützpunkten mit einer bestimmten Methode interpoliert. Der nicht in die Interpolation einfließende Stützpunkt dient als Prüfgröße. Er enthält den wahren Niederschlagswert an einem bestimmten Ort. Derart wird nun jeweils ein Punkt, und zwar immer ein anderer, aus der Menge der Stützpunkte entfernt und mit der einmal gewählten Methode interpoliert. Am Ende liegt für jede Station des Messnetzes neben dem gemessenen Wert auch ein Schätzwert vor. Aus den sich daraus ergebenden Residuen (Differenzen zwischen Mess- und Schätzwerten) werden verschiedene statistische Kennzahlen gewonnen und mit jenen Kennzahlen verglichen, die mit anderen Interpolationsmethoden erzielt wurden. Die Crossvalidation erlaubt eine begründete Auswahl jener Interpolationsmethode, die geeignet erscheint, das Phänomen Niederschlag im Iran bestmöglich zu beschreiben.