

Kriging

Mit dem geostatistischen Interpolationsverfahren Kriging werden Werte an Positionen, an denen keine Messdaten vorliegen, mithilfe der umliegenden Datenpunkte angenähert. Dabei wird die räumliche Varianz berücksichtigt, welche aus einem Semivariogramm ermittelt werden kann. "Für einen gesuchten Wert werden dabei die Gewichte der in die Berechnung einfließenden Messwerte so bestimmt, dass die Schätzfehlervarianz möglichst gering ist."^[1] Mit den Gewichten kann eine Vorhersage über einen nicht gemessenen Ort abgeleitet werden. Dafür wird mit folgender Formel die gewichtete Summe der Daten gebildet:

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i)$$

dabei gilt:^[2]

$Z(s_i)$ - gemessener Wert am i-ten Ort

λ_i - unbekannte Gewichtung für den am i-ten Ort gemessenen Wert

s_0 - vorhergesagter Ort

N - Anzahl der gemessenen Werte

Die Gewichtungen sind abhängig vom an die Messpunkte angepassten Modell, der Entfernung zur vorhergesagten Position und den räumlichen Beziehungen unter den umliegenden Datenpunkten. Mithilfe der Kriging-Interpolationsmethode kann in zwei Schritten eine vorhergesagte Oberflächenkarte erstellt werden. Der erste Schritt ist das Erstellen des Variogrammes, um die räumliche Autokorrelation der Datenpunkte abzuschätzen. Es wird dabei vorausgesetzt, dass die gesamte Oberfläche statistisch homogen ist, sodass an jedem Ort das gleiche Variationsmuster vorliegt. Der zweite Schritt ist das Vorhersagen der unbekannten Werte. Dafür gibt es unterschiedliche Methoden: Beim "Ordinary"-Kriging wird angenommen, dass der konstante Mittelwert unbekannt ist, während beim "Universal"-Kriging von einem vorrangigen Trend in den Daten ausgegangen wird (z.B. eine vorherrschende Windrichtung).

Variogramme

Variogramme werden in der Geostatistik verwendet und sind die Grundlage für die Interpolationsmethode Kriging. Mithilfe von Variogrammen können Aussagen über die räumlichen

Beziehungen von benachbarten Datenpunkten und deren Ähnlichkeit in Abhängigkeit von der Entfernung getroffen werden.

Aus Datenpunkten werden Paare gebildet und deren Werte miteinander verglichen, dafür wird die quadrierte Differenz gebildet, halbiert und in Abhängigkeit von der Distanz h der beiden Punkte dargestellt, damit erhält man eine Variogrammwolke. Es wird Stationarität angenommen, das heißt, dass die Werte nur von der Entfernung der Punkte und nicht von ihrer räumlichen Position abhängen.

Alle diese Werte $\frac{(Z(x+h)-Z(x))^2}{2}$ [3] für den gleichen Abstand werden summiert und durch die Anzahl der Punktpaare dieser Distanz geteilt. Dadurch ergibt sich die Semivarianz der Messwerte in Abhängigkeit der Entfernung:

$$\gamma(h) = \frac{1}{N_h} \sum_{x_i - x_j = h} \frac{(Z(x_i) - Z(x_j))^2}{2}$$

dabei gilt:

N_h - Anzahl der Punktpaare derselben Distanz

h - Distanz der Punkte

(x_i, x_j) - Punktpaare

Z - Messwerte

Dieser Zusammenhang wird im empirischen bzw. experimentellen Variogramm dargestellt. Meist wird dafür die x-Achse zur Vereinfachung in mehrere Intervalle ("Lags") zerlegt, sodass die Mittelwerte innerhalb dieser berechnet werden. Dabei ist das Grundprinzip der Geographie zu erkennen: Nähere Punktpaare sind sich ähnlicher und werden sich mit zunehmender Distanz immer unähnlicher.

Danach wird ein theoretisches Modell erstellt. Dafür wird eine Kurvenfunktion durch das empirische Variogramm gelegt. Dieser Prozess ähnelt einer Regressionsanalyse. Häufig auftretende Formen sind kreisförmige, sphärische, lineare, exponentielle und gaußsche Variogramme. [4] Das ausgewählte Modell hat Einfluss auf die Vorhersage der unbekannten Werte beim Kriging. Je steiler die Kurve in der Nähe des Ursprungs ansteigt, desto stärker ist der Einfluss der nächsten Nachbarn auf die Vorhersage, sodass die ausgegebene Oberfläche weniger glatt ist.

Außer bei linearen Variogrammen, nähert sich die Kurve meist einem Grenzwert an. Dieser wird *Partial Sill* (Schwellenwert) genannt. Er ergibt sich aus *Sill* minus *Nugget*. Die Entfernung zwischen $x = 0$ und dem x -Wert, an dem die y -Werte den *Partial Sill* erreichen, wird als *Range* bezeichnet. Ab dieser Distanz h ist keine autokorrelative Beziehung der Messwerte mehr feststellbar.^[5] Im Idealfall müsste der y -Wert bei $x = 0$ Null sein, da dort der Abstand der zu vergleichenden Punkte Null ist (die zwei Punkte sind identisch) und damit die Messwerte gleich sein müssten. Dies ist im Variogramm oft nicht der Fall, der y -Wert, an dem die Kurve die y -Achse schneidet, wird *Nugget* genannt und gibt an, wie stark die Daten rauschen. Dieser *Nugget-Effekt* beruht auf Fehlern der Messgeräte und räumlichen Schwankungen bei Entfernungen, die kleiner als das Abtastintervall sind.

1. <https://www.geothermie.de/bibliothek/lexikon-der-geothermie/k/kriging.html>, 03.10.2022, 15:02 ↩
2. <https://pro.arcgis.com/de/pro-app/latest/tool-reference/3d-analyst/how-kriging-works.htm>, 03.10.2022, 15:31 ↩
3. http://www.gitta.info/ContiSpatVar/de/html/Interpolatio_learningObject3.html, 01.08.2022 ↩
4. <https://de.wikipedia.org/wiki/Semivariogramm>, 01.08.2022 ↩
5. <https://www.spektrum.de/lexikon/geographie/variogramm/8520>, 01.08.2022 ↩