Wie bewältigt man Stationaritätsannahmen in der Geostatistik?

Alexander Brenning[1] & K. Gerald van den Boogaart[2]

Zusammenfassung

In der Geostatistik ist die Stationarität des Zufallsfeldes eine zentrale Annahme. Die räumliche Variabilität vieler Phänomene in unserer Umwelt hängt stark von den Verhältnissen in einer lokalen Umgebung ab, die meist aber instationär sind. Um damit umgehen zu können, wird das Konzept der Stationarität des Zufallsfeldes ersetzt durch eine Stationarität des Einflusses der lokalen Umweltverhältnisse, wie sie in einem GIS gespeichert sind, auf das lokale Kovariogramm. Es wird eine Konstruktionsmethode benutzt, die auf sinnvolle Art räumliche Informationen aus dem GIS in Kovariogramm-Modelle einbinden kann, etwa im Untersuchungsgebiet variierende Umweltverhältnisse, sich räumlich verändernde Anisotropie im Gebirgsrelief oder geologische Störungen, die die Kontinuität unterbrechen.

1 Einführung

Die in Geo-Informationssystemen vorhandene Menge räumlicher Daten hat in den vergangenen Jahrzehnten nicht zuletzt durch die Nutzung moderner Fernerkundungsmethoden sehr schnell zugenommen. Diese Daten eignen sich häufig für die Modellierung lokaler Anisotropien, die beispielsweise tektonischen Strukturen oder dem Relief folgen.

Infolgedessen entsteht ein Bedarf an instationären geostatistischen Methoden, die diese Informationen auf sinnvolle Weise berücksichtigen, so dass komplexe Umweltkonfigurationen modelliert werden können. Zu diesen Umweltbedingungen zählen etwa tektonische Verwerfungen, welche zu Diskontinuitäten führen, oder hydrographische Einzugsgebiete, die hierarchisch verschachtelte Anisotropiestrukturen nach sich ziehen. Methoden der Geostatistik ohne Stationaritätsannahmen müssen mit Geo-Informationssystemen verknüpft werden, um entsprechende Umweltdaten nutzen zu können.

2 Grundlagen

Die Geostatistik stellt Methoden zur Analyse (skalarer) räumlicher Daten zur Verfügung. Diese Daten (Beobachtungen) werden als Realisierungen eines räumlichen stochastischen Prozesses oder Zufallsfeldes $(Z(s))_{s \in D}$, hier mit $D \subset \mathbb{R}^2$, angesehen, also als Menge von Zufallsvariablen auf einer räumlichen Domäne.

Räumliche Abhängigkeiten werden stochastisch durch Kovarianzstrukturen, sogenannte Kovariogramme, beschrieben:

$$C(\mathrm{s}, \mathrm{t}) = \mathrm{Cov}(Z(s), Z(t)) = \mathrm{E}[(Z(s) - m(s)) \cdot (Z(t) - m(t))],$$

wobei $m(s) = \mathrm{E}(Z(s))$ den Erwartungswert einer Zufallsvariable bezeichnet – vgl. z.B. GOOVAERTS (1997).

[1] Humboldt-Universität zu Berlin, Geographisches Institut, Unter den Linden 6, 10099 Berlin, E-Mail: alexander.brenning@rz.hu-berlin.de
[2] TUFreiberg, Institut für Geologie, B.-v.-Cotta-Str.2, 09596 Freiberg, E-Mail: boogaart@iname.com

2.1 Stationarität

In Theorie und Praxis werden in aller Regel für Erwartungswerte und Kovariogramme Stationaritätsannahmen getroffen; dabei handelt es sich um räumliche Invarianzforderungen bezüglich Mittelwert und Kovariogramm. Die meist verwendete Stationarität 2. Ordnung fordert:

i) Der Erwartungswert m ist konstant auf D;

ii) das Kovarogramm C hängt nur von der Differenz $t - s$ der Beobachtungsstellen ab (und nicht von ihrer tatsächlichen Lage).

(Zur technischen Vereinfachung gehen wir hier und im folgenden nicht auf Semivariogramme und intrinsische Stationarität ein; vgl. hierzu BRENNING 2001 oder GOOVAERTS 1997.)

2.2 Zwei Fälle von Instationarität

In folgenden für die Praxis wichtigen Fällen erfüllen Umweltvariablen die Stationaritätsbedingungen nicht – vgl. BRENNING & VAN DEN BOOGAART (2001), CRESSIE (1993):

i) Vorhandensein eines Trends: Es existiert eine an räumliche Kovariablen geknüpfte systematische „Tendenz“ – beispielsweise in einem Gebirge die lineare Abnahme der Temperatur mit zunehmender Höhe über dem Meer. In diesem Fall ist der Erwartungswert nicht auf D konstant, sondern eine Funktion von räumlichen Kovariablen.

ii) Lokale Anisotropien und Diskontinuitäten: Das Streichen geologischer Schichten kann dazu führen, dass Korrelationen quer zur Vorzugsrichtung bei gleicher Entfernung kleiner sind als parallel zu ihr; ferner können Beobachtungen derselben Umweltvariable in Gebieten, die durch eine geologische Störung getrennt sind, unkorreliert sein. In diesen Fällen hängt das Kovariogramm nicht nur von der Distanz zwischen Beobachtungen ab, sondern von lokalen Veränderungen der Kovariablen.

Zwar sind in beiden genannten Fällen Stationaritätsannahmen verletzt. Dennoch entstehen die jeweiligen Instationaritäten auf der ganzen Domäne auf dieselbe Weise: Lediglich die Kenntnis bestimmter Kovariablen ist nötig, um diese Instationaritäten zu erklären. Daher wird im folgenden das Konzept der Stationarität des Zufallsfeldes ersetzt durch eine Stationarität des Einflusses der lokalen Umweltverhältnisse auf das lokale Kovariogramm (VAN DEN BOOGAART 1999).

Fortan wird lediglich der Fall instationärer Kovariogramme (bei Fehlen eines Trends) betrachtet. Geostatistische Schätzung mit linearem Trend wird von GOOVAERTS (1997), VAN DEN BOOGAART & BRENNING (2001) sowie BRENNING (2001) behandelt.

3 Zur Konstruktion instationärer Kovariogramme

Das Konzept der Stationarität des Einflusses der lokalen Umweltverhältnisse auf das lokale Kovariogramm wird nun in Form eines generischen Stationaritätsbegriffes eingeführt und anschließend mit Hilfe einer Konstruktionsmethode operationalisiert. Relevante Kovariablen (die „lokalen Umweltverhältnisse“) werden im folgenden durch eine Abbildung g auf der Domäne D repräsentiert.

3.1 Generische Stationarität

Ein Kovariogramm C heißt *generisch stationär* bezüglich g, falls es eine Funktion C_g gibt, so dass für alle s,t,h gilt:

$$\mathrm{C}(s,t) = C_g(s,t;\, g),$$
$$\mathrm{C}(s+h,\, t+h) = C_g(s,t;\, g(\cdot+h)).$$

Die Kovarianz von Umweltvariablen an den Stellen s und t hängt somit von den lokalen Umweltverhältnissen ab; und wenn an den Stellen $s+h$, $t+h$ dieselben lokalen Umweltverhältnisse vorliegen wie bei s und t, werden auch dieselben Kovarianzen auftreten. An jedem Punktepaar ($s+h$, $t+h$) wirkt daher das gleiche Einflussgesetz $C_g(s,t;\, \cdot)$, jedoch bezüglich „verschobener" Umweltverhältnisse $g(\cdot+h)$. Auf diese Weise kommt eine Stationarität des Einflusses der lokalen Umweltverhältnisse auf das Kovariogramm zustande.

Je nachdem, welche Kovariablen g und welches Einflussgesetz C_g zum Modellieren verwendet werden, ist das resultierende Kovariogramm mehr oder weniger stationär oder instationär – es handelt sich sozusagen um eine „skalierbare" Instationarität.

3.2 Eine Konstruktionsmethode

Allgemein lässt sich zeigen (VAN DEN BOOGAART 1999, BRENNING 2001), dass sich eine hinreichend große Klasse von Kovariogrammen konstruktiv durch eine Faltung erzeugt werden kann:

$$C_w(s,t) := \int_E w(s,p)w(t,p)\,\mathrm{d}p,$$

wobei $w : D \times E \to \mathbb{R}$ eine geeignete Gewichtsfunktion auf einer Domäne $D \times E$ sei.

Werden radialsymmetrische Gewichtsfunktionen verwendet, so können isotrope Kovariogramme ertzeugt werden. Dies ist etwa für die Indikatorfunktion auf einer Kreisscheibe der Fall, während die Indikatorfunktion einer Ellipse mit lokal variierender Hauptachsenorientierung ein lokal anisotropes Kovariogramm induziert.

Die Unabhängigkeit der beobachteten Umweltvariable auf verschiedenen Teilgebieten der Domäne lässt sich durch heranmultiplizieren weiterer Indikatorfunktionen erzielen. Durch geeignete Wahl von Gewichtsfunktionen $w(s,t;\, g)$ können somit die lokalen Umweltverhältnisse g auf ein sie berücksichtigendes generisch stationäres Kovariogramm abgebildet werden. Das hier vorgestellte Konstruktionsprinzip repräsentiert daher einen Modellierungsansatz, dessen Ziel darin besteht, Kovarianzmodelle baukastenartig zusammenzusetzen, wobei Kenntnisse über Strukturen und Prozessmuster in unserer Umwelt abgebildet werden. Das folgende Beispiel soll dies verdeutlichen.

3.3 Eine hydrologisch motivierte Anwendung

Bei der Untersuchung bestimmter hydrologischer oder hydrogeologischer Variablen kann angenommen werden, dass Paare von Punkten, die in unterschiedlichen Einzugsgebieten liegen, unkorreliert sind; dagegen sind Punkte mit sich überschneidenden Teileinzugsgebieten umso stärker korreliert, je größer deren Schnittfläche ist.

Bezeichnet nun $A(s)$ das zu einem Punkt s gehörige hangaufwärts gelegene Teileinzugsgebiet, so lässt sich vermittels einer Gewichtsfunktion mit Träger $A(s)$ – z.B. einer Indikatorfunktion – ein Kovariogramm induzieren, welches unser qualitatives Prozesswissen repräsentiert. Dieses Kovariogramm ist generisch stationär, jedoch in hohem Maße instationär im klassischen Sinne.

Das Teileinzugsgebiet zu einem Punkt kann mit Hilfe von digitalen Geländemodellen ermittelt werden.

3.4 Implementationsansätze

Ausgewählte generisch stationäre Modelle wurden von BRENNING (2001) implementiert. Dabei kam die open-source Datenanalyse-Umgebung und Programmiersprache R (www.r-project.org) zum Einsatz. Die für die Berechnung von Kovariogrammen (genauer: Semivariogrammen) aufgrund des beschriebenen Konstruktionsschemas benötigte numerische Integration wurde als Quasi-Monte-Carlo-Verfahren in C implementiert. Ferner wurde eine einfache GIS-Schnittstelle zur kommerziellen Software ArcView 3.1 in der Programmiersprache AVENUE realisiert, um den Geodatenaustausch zu erleichtern.

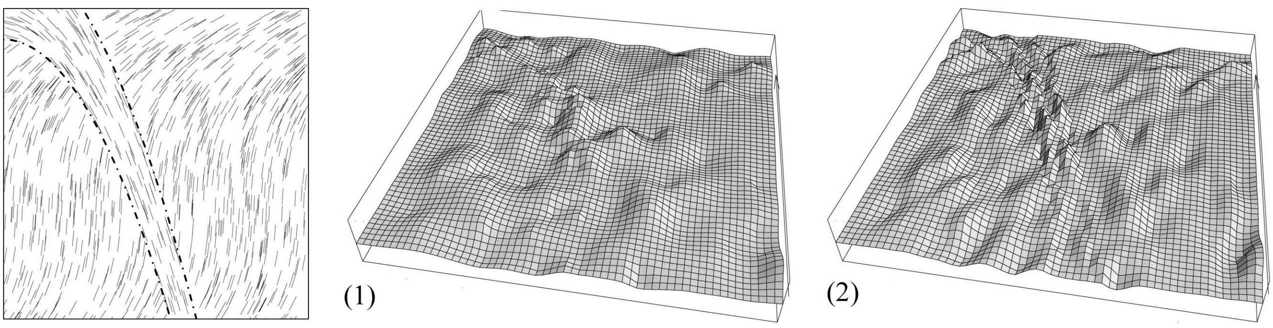

Abbildung 1: Links: Die für die Simulation eines generisch stationären Datensatzes angenommenen Umweltverhältnisse: Es werden zwei unabhängige Teil-Zufallsfelder mit lokal variierenden Anisotropierichtungen (durch Linien angedeutet) betrachtet. Mitte und rechts: Kriging-Oberflächen, die unter Verwendung geeigneter stationärer (Mitte) bzw. generisch stationärer Modelle (rechts) berechnet wurden.

4 Ein simulierter Beispieldatensatz in komplexen Umweltverhältnissen

Als Beispiel für die Anwendung generisch stationärer Kovarianzstrukturen wurde ein zu einem instationären (generisch stationären) Gaußschen Zufallsfeld gehörender Datensatz aus 259 unregelmäßig verteilten Werten mit Hilfe von Pseudozufallszahlen simuliert. Hierbei wurden lokale Instationaritäten auf voneinander unabhängigen Teilgebieten angenommen, wie sie durch häufig auftretende Umweltverhältnisse induziert werden können (Abb. 1). An diesen Datensatz wurden geostatistische Modelle angepasst, welche (1) stationär bzw. (2) generisch stationär unter Berücksichtigung der zugrunde liegenden (bekannten) Anisotropiestruktur sind.

Die Ergebnisse der geostatistischen Schätzung mit Hilfe von Kriging sind in Abbildung 1 dargetellt. Wie deutlich zu erkennen ist, werden die durch die anisotropen Kovariablen vorgegebenen Richtungsabhängigkeiten und Unstetigkeiten unter Stationaritätsannahmen nicht reproduziert. Lediglich in der Mitte des dargestellten Gebietes ist eine etwas ausgeprägtere Variabilität der interpolierten Werte zu beobachten, welche jedoch in alle Richtungen unscharf abnimmt. Dagegen zeichnet die mit einem generisch stationären Modell erzeugte Interpolation relativ scharf die Grenze zwischen den beiden unabhängigen Gebieten sowie die Richtung der lokalen Anisotropien (langgestreckte „Hügel“) nach. Das Kriging-Ergebnis unter der Annahme generischer Stationarität kommt somit den in unserer meist instationären Umwelt näher.

5 Herausforderungen für die Zukunft

Bei den vorgestellten generisch stationären Modellen handelt es sich um eine Auswahl, welche das angegebene Konstruktionsprinzip veranschaulichen und wichtige Anwendungsfelder aufweisen soll. Eine Herausforderung für zukünftige Arbeiten stellt nun die Bereitstellung einer Reihe von weiteren einfachen, generisch stationären Kovarianzmodellen dar, welche wichtige Umweltprozesse und -strukturen – etwa in Abhängigkeit von Relief oder Hydrographie – berücksichtigen. Aufbauend auf diesen einfachen Modellen können dann nach dem Baukastenprinzip komplexere Umweltverhältnisse wiedergegeben werden. Empirische Analysewerkzeuge müssen geschaffen werden, um in der Praxis die Anwendbarkeit solcher Modelle auf einen konkreten Datensätz zu prüfen. In Verbindung mit dem GIS-gestützten Einsatz sollte ein Expertensystem dem Anwender die Verwendung generisch stationärer Modelle erleichtern.

Literatur

BRENNING, A. (2001): Geostatistics without stationarity assumptions within Geographical Information Systems, Freiberg Online Geosciences 6.

BRENNING, A. & VAN DEN BOOGAART, K. G. (2001): Geostatistics without stationarity assumptions within GIS, Proc. 7th Annual Conference of IAMG, Cancún.

CRESSIE, N. A. C. (1993): Statistics for spatial data, Wiley, New York.

GOOVAERTS, P. (1997): Geostatistics for natural resources evaluation, Oxford University Press, New York et al.

VAN DEN BOOGAART, K. G. (1999): A new possibility for modelling variograms in complex geology, Proc. of StatGIS Klagenfurt 1999.

VAN DEN BOOGAART, K. G. & BRENNING, A. (2001): Why is Universal Kriging better than IRFk-Kriging: Estimation of variograms in the presence of trend, Proc. 7th Annual Conference of IAMG, Cancún.