

A guide to the geostatistical analysis of compositional data

Leitfaden zur geostatistischen Analyse von Daten in geschlossenen Systemen

Una guia per a l'anàlisi geostadística de dades composicionals

RAIMON TOLOSANA-DELGADO (Barcelona) & K. GERALD VAN DEN BOOGAART (Freiberg)

Key words: composition, covariance function, Aitchison geometry, simplex.

Abstract

Dealing with spatially-dependent compositional databases (including proportions, data in percentages, concentrations, etc) should honor the mathematical properties of these kinds of data: a valid composition must have positive components summing at most up to a constant (1, 100%, etc.). Generally speaking, that is easily done by working on a set of log-ratios of components, instead of using the raw data. To study the spatial variability of these databases, it is recommended to estimate and model the lr-variograms: the set of variograms of all possible pairwise log-ratios of components in the composition. Such lr-variograms contain all necessary information to deal with intrinsic stationary compositions, and may be modelled with standard geostatistical tools, like the linear model of coregionalization. Moreover, we can study the model properties, and extract from it some relationships between components and link them to processes operating at some given spatial scales. Finally, component-by-component interpolation and mapping is straightforward with existing kriging and simulation techniques: one must just apply these tools to a set of invertible log-ratios of components, i.e. some log-ratio transformations such that the original composition can be recovered from the transformed data and vice versa.

Zusammenfassung

Wenn man Kompositionsdaten, (z.B. Prozent-Daten wie sie typischerweise bei chemischen Analysen oder Proportionen vorliegen) statistisch auswertet, muss man die besonderen mathematischen Eigenschaften dieser Datenstruktur beachten: Kompositionsdaten haben nur positive Komponenten, die sich zu einer Konstanten aufsummieren (1, 100%, usw.). Am einfachsten geschieht dies, indem man anstatt mit den Rohdaten zu arbeiten eine Log-ratio-Transformation durchführt. Um die räumliche Variabilität dieser Daten zu analysieren, wird empfohlen, mit lr-Variogrammen zu arbeiten, d.h. mit den Variogrammen der möglichen paarweisen Log-Ratios der Komposition. Diese lr-Variogramme enthalten die notwendige Information um intrinsisch-stationäre Kompositionen mit den klassischen geostatistischen Methoden (z.B. lineares Modell der Koregionalisierung) zu bearbeiten. Ferner kann man aus den Modelleigenschaften Beziehungen zwischen den Einzelkomponenten ableiten und sie mit räumlichen Prozessen (Sedimentation etc.) in verschiedenen Skalen in Verbindung bringen. Schließlich kann man die geostatistischen Interpolations- und Simulationsmethoden (Kriging) auf die einzelnen Log-Ratio-Variablen der Komponentenpaare anwenden und die Ergebnisse zur Interpretation in die originale Komponentenstruktur rücktransformieren.

Das hier beschriebene Verfahren wird anhand von geochemischen (Kompositions-)Daten von Flusssedimenten aus dem Grazer Paläozoikum vorgestellt (Abb. 1). Die räumliche Strukturanalyse der Log-ratios der verschiedenen Komponenten-

paare ist in den Abb. 2 und 3 dargestellt. Der Zusammenhang mit räumlichen Prozessen auf der Mikro- und Mesoskala kann aus den Dreiecksdiagrammen abgeleitet werden (Abb. 4); die Kartendarstellung der Ergebnisse ist in Abb. 5 zu sehen.

Resum

El tractament de dades composicionals amb dependència espacial (bases de dades en proporcions, en %, en concentracions, etc) hauria de preservar les característiques fonamentals de la definició matemàtica d'aquest tipus de dades: una composició ha de contenir components positives que sumin igual o menys que una constant de referència (1, 100%, etc). En termes generals, això s'aconsegueix de forma automàtica treballant sobre les dades transformades mitjançant un log-quocient, enlloc de les dades originals en brut. Per a l'estudi de la dependència espacial d'aquesta mena de dades, es recomana l'ús dels variaciogrames, o lr-variogrames: el conjunt de variogrames directes de tots els possibles log-quocient de dos components. Els variaciogrames contenen tota la informació necessària per a tractar dades composicionals amb dependència espacial intrínseca, i poden ser modelats amb les tècniques geostatístiques habituals, com e.g. el model lineal de coregionalització. A més, les propietats del model mateix poden ser estudiades, i hom pot obtenir certes relacions entre components potencialment degudes a processos operant a certes escales espacials concretes. Finalment, un mapa de cada component o de qualsevol log-quocient es pot obtenir mitjançant les tècniques habituals d'interpolació per cokriging i simulació geoestadística aplicades a qualsevol conjunt de log-quocients bijectiu, i.e. un conjunt de transformacions log-quocient tals que la composició original es pot recuperar de les dades transformades i viceversa.