Simplicial Indicator Kriging

Dr. Raimon Tolosana-Delgado
raimon.tolosana@geo.uni-goettingen.de

Department of Sedimentology and Environmental Geology
University of Göttingen, Germany

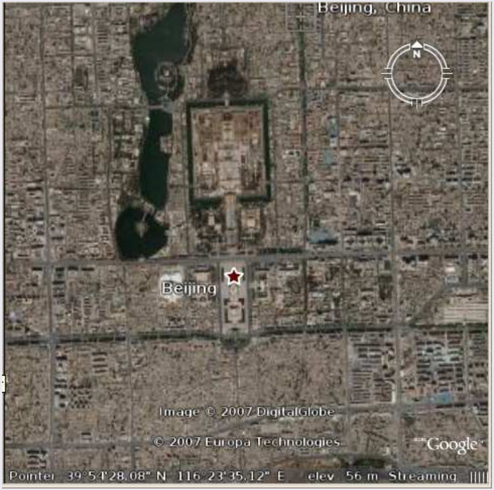
China University of Geosciences
Wuhan, China, September 5-6, 2007

water quality assessment: an online control system

XACQA: on-line water quality control system

- basin NE Barcelona (eastern Spain)
- Mediterranean climate
- main river < 5 m^2/s, 55km long, 0-1000 m above sea level
- an online station, to control Waste-Water Treating Plant effluent (dumps into a *riera*)
- 17000 inhabitants
- chemical industry

location



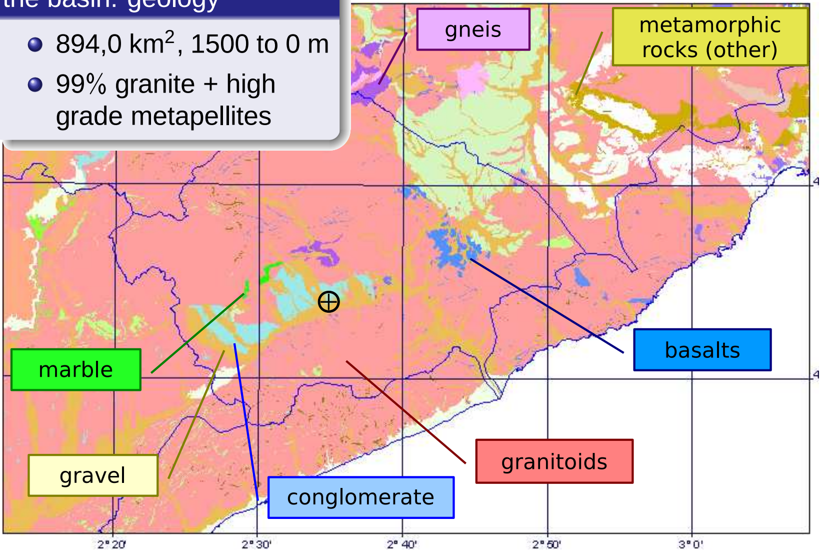
water quality assessment: a particular case

the Gualba riera: the sampled tributary

water quality assessment: a particular case

the basin: geology

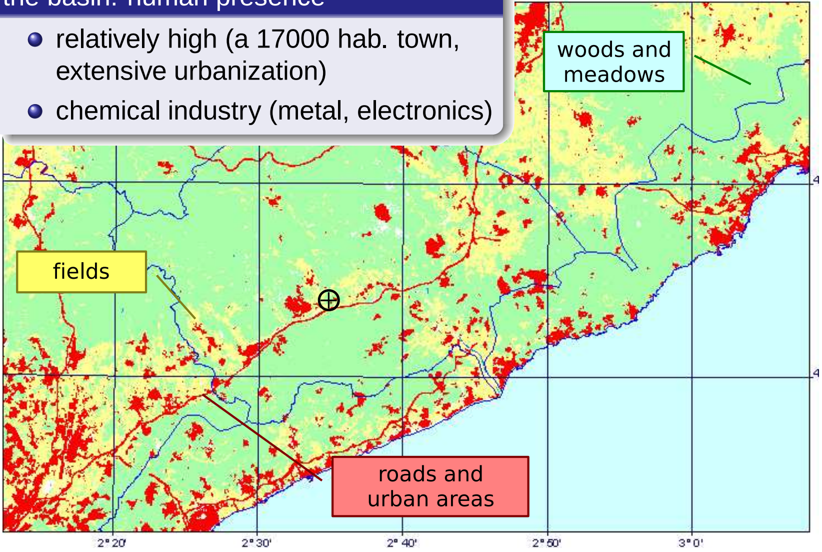
- 894,0 km^2, 1500 to 0 m
- 99% granite + high grade metapellites



water quality assessment: a particular case

the basin: human presence

- relatively high (a 17000 hab. town, extensive urbanization)
- chemical industry (metal, electronics)



water quality assessment: a particular case

measured variables

- conductivity, pH, ammonium, (temperature, $O_2, \ldots$)
- main interest: potential of ammonia production
- ammonia (NH_3): lethal (fishes, macroinvertebrates), but volatile
- ammonium (NH_4^+): much less dangerous on itself, but

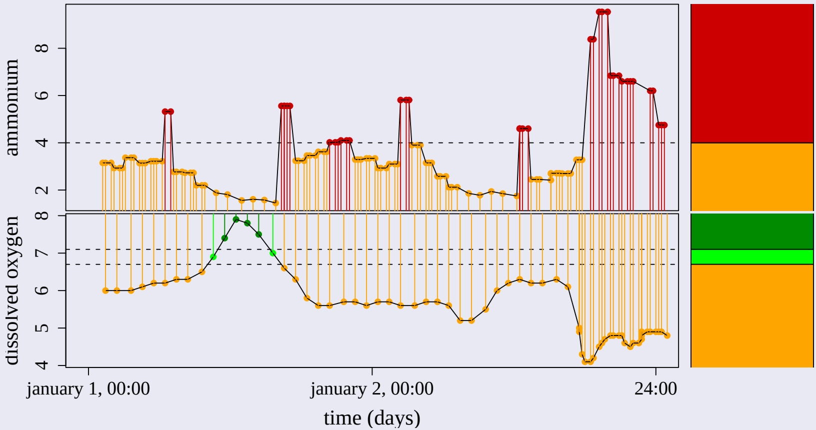
$$NH_4^+ + H_2O \rightleftharpoons NH_3 + H_3O^+ \qquad K_a = \frac{[NH_3] \cdot [H_3O^+]}{[NH_4^+]} = f(T_w)$$

$$\begin{aligned} HCO_3^- + H_2O &\rightleftharpoons CO_3^= + H_3O^+ \\ HCO_3^- + H_2O &\rightleftharpoons H_2CO_3 + OH^- \end{aligned}$$

- basin rich in $HCO_3 \Longrightarrow$ pH buffering $\Longrightarrow NH_3$ controlled

water quality assessment: a particular case

uncertain category assessment

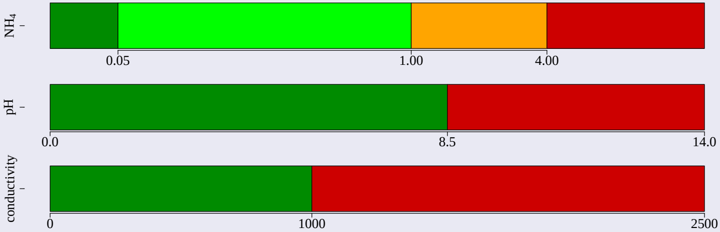


- *which is the distribution of the water quality at a given moment?*

data set

obtaining the data set of water quality categories

1. data available: ammonium concentration, pH, conductivity
2. regularization: 12h geometric averages (pH arithmetic)
3. thresholding



4. final quality category: the worse

data set

categories from 12h averaged values of NH_4^+, pH and conductivity

geostatistics for categorical variables

treatment: Indicator Kriging (IK; Journel, 1983)

1. (re)define the categories as indicator functions

$$I_i(x) = \begin{cases} 1 & Z(x) < z_i \\ 0 & \texttt{otherwise} \end{cases} \qquad J_i(x) = \begin{cases} 1 & Z(x) \in A_i \\ 0 & \texttt{otherwise} \end{cases}$$

2. compute variograms, fit models, interpolate
3. interpret results as probabilities:
$\hat{I}_i(x_0) \Rightarrow \mathrm{Pr}[Z(x_0) < z_i]$ or $\hat{J}_i(x_0) \Rightarrow \mathrm{Pr}[Z(x_0) \in A_i]$

problems: practical and theoretical

- interpolations $\hat{I}_i$ are not ordered ($\Rightarrow$ *ad-hoc* corrections)
- $\hat{J}_i$ are negative, or sum $\neq$ one
- variogram/covariance systems are difficult to model

geostatistics for categorical variables

treatment: Indicator Kriging (IK; Journel, 1983)

1. (re)define the categories as indicator functions

$$I_i(x) = \begin{cases} 1 & Z(x) < z_i \\ 0 & \texttt{otherwise} \end{cases} \qquad J_i(x) = \begin{cases} 1 & Z(x) \in A_i \\ 0 & \texttt{otherwise} \end{cases}$$

2. compute variograms, fit models, interpolate
3. interpret results as probabilities:
   $\hat{I}_i(x_0) \Rightarrow \texttt{Pr}[Z(x_0) < z_i]$ or $\hat{J}_i(x_0) \Rightarrow \texttt{Pr}[Z(x_0) \in A_i]$

problems: practical and theoretical

- interpolations $\hat{I}_i$ are not ordered ($\Rightarrow$ *ad-hoc* corrections)
- $\hat{J}_i$ are negative, or sum $\neq$ one
- variogram/covariance systems are difficult to model
- the scale of I (or J) is NOT the scale of $\texttt{Pr}[Z(x_0) \in A_i]$

geostatistics for categorical variables

treatment: Indicator Kriging (IK; Journel, 1983)

1. (re)define the categories as indicator functions

$$I_i(x) = \begin{cases} 1 & Z(x) < z_i \\ 0 & \texttt{otherwise} \end{cases} \qquad J_i(x) = \begin{cases} 1 & Z(x) \in A_i \\ 0 & \texttt{otherwise} \end{cases}$$

2. compute variograms, fit models, interpolate
3. interpret results as probabilities:
   $\hat{I}_i(x_0) \Rightarrow \mathtt{Pr}[Z(x_0) < z_i]$ or $\hat{J}_i(x_0) \Rightarrow \mathtt{Pr}[Z(x_0) \in A_i]$

problems: practical and theoretical

- interpolations $\hat{I}_i$ are not ordered ($\Rightarrow$ *ad-hoc* corrections)
- $\hat{J}_i$ are **negative, or sum $\neq$ one**
- variogram/covariance systems are difficult to model
- the scale of I (or J) is NOT the scale of $\mathtt{Pr}[Z(x_0) \in A_i]$

geostatistics for categorical variables

treatment: Indicator Kriging (IK; Journel, 1983)

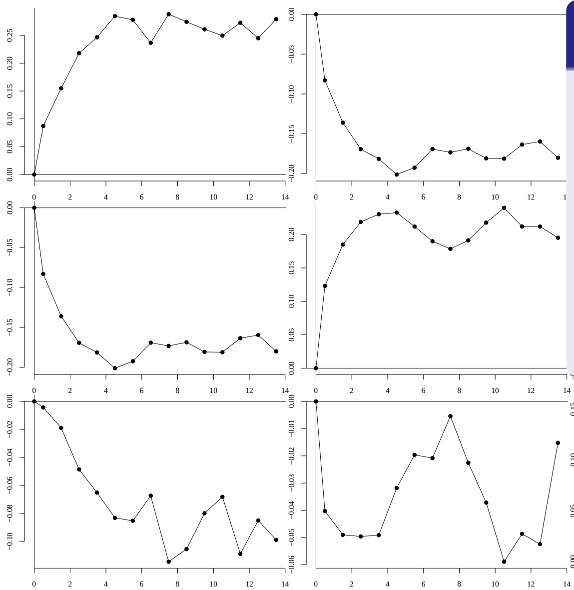
1. (re)define the categories as indicator functions

$$I_i(x) = \begin{cases} 1 & Z(x) < z_i \\ 0 & \texttt{otherwise} \end{cases} \qquad J_i(x) = \begin{cases} 1 & Z(x) \in A_i \\ 0 & \texttt{otherwise} \end{cases}$$

2. compute variograms, fit models, interpolate
3. interpret results as probabilities:
   $\hat{I}_i(x_0) \Rightarrow \Pr[Z(x_0) < z_i]$ or $\hat{J}_i(x_0) \Rightarrow \Pr[Z(x_0) \in A_i]$

problems: practical and theoretical

- interpolations $\hat{I}_i$ are not ordered ($\Rightarrow$ *ad-hoc* corrections)
- $\hat{J}_i$ are **negative, or sum $\neq$ one**
- variogram/covariance systems are difficult to model
- the scale of I (or J) is NOT the scale of $\Pr[Z(x_0) \in A_i]$

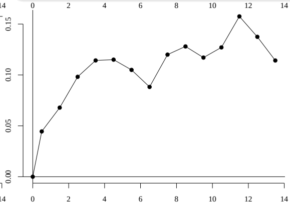


variograms are difficult to model

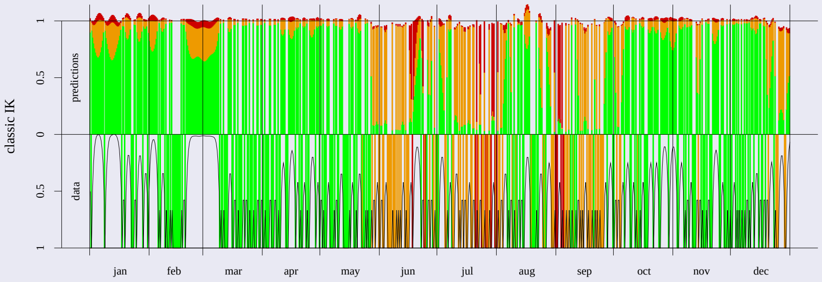
variograms of $\boldsymbol{J}$ bound to:

- sum to 0 by rows
- sum to 0 by columns
- sill condition, $c_{ij} = \bar{p}_i \delta_{ij} - \bar{p}_i \bar{p}_j$
- positive definite

conjecture on variograms of $\boldsymbol{I}$ (Matheron, 1971)

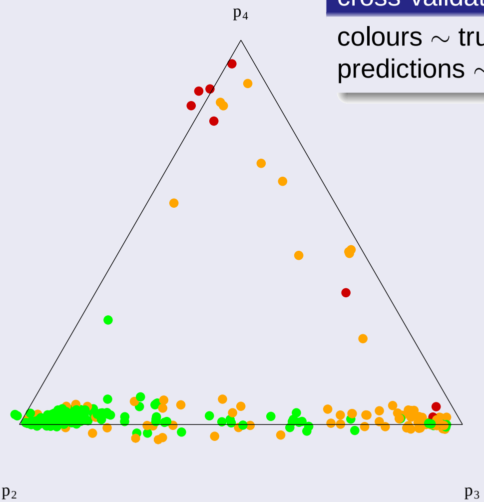


indicator kriging results do not sum up to one



- only a full cokriging ensures predictions summing up to 1

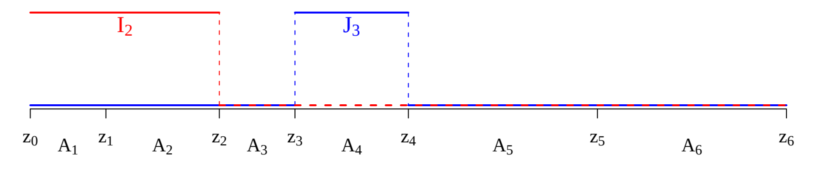
indicator kriging results might be negative



cross-validation excercise

colours ∼ true category
predictions ∼ edges

relation between cumulative and disjoint indicators



$$\mathbf{I} = \mathbf{L} \cdot \mathbf{J} \qquad \mathbf{L} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

- no difference between cokriging **I** and **J**
- $\hat{\mathbf{I}}$ has order violations $\Longleftrightarrow$ $\hat{\mathbf{J}}$ has negative values
- $\hat{\mathbf{I}}$ does not take any profit of the ordering information!

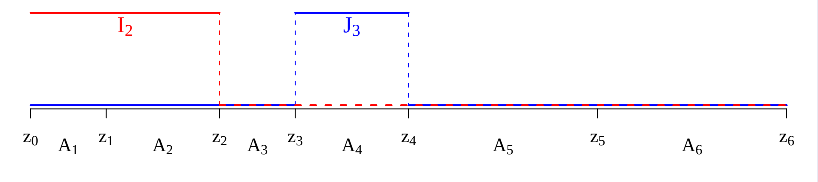
relation between cumulative and disjoint indicators

Proposition:
kriging transformed vectors is transforming kriged vectors

- IF: vector random functions: $\mathbf{Z}$ and $\mathbf{Y}$ (dim. P), with $\mathbf{Z} = \mathbf{T} \cdot \mathbf{Y}$
- transformation: $\mathbf{T}$ a (P, P)-full rank matrix (linear transformation)
- covariance models $\mathbf{C}^z$, $\mathbf{C}^y$, *consistent* if $\mathbf{C}^z(h) = \mathbf{T} \cdot \mathbf{C}^y(h) \cdot \mathbf{T}^t$
- THEN: cokriging predictors also fulfill $\hat{\mathbf{z}}_0 = \mathbf{T} \cdot \hat{\mathbf{y}}_0$
- linear operators commute; Myers (1982-84, *Math. Geol.*)

$$\begin{array}{ccc}
\mathbf{Y} & \longleftarrow T \longrightarrow & \mathbf{Z} \\
\downarrow & & \downarrow \\
\textit{cokriging} & (C_{ij}\ \textit{consistent}) & \textit{cokriging} \\
\downarrow & & \downarrow \\
\hat{\mathbf{y}} & \longleftarrow T \longrightarrow & \hat{\mathbf{z}}
\end{array}$$

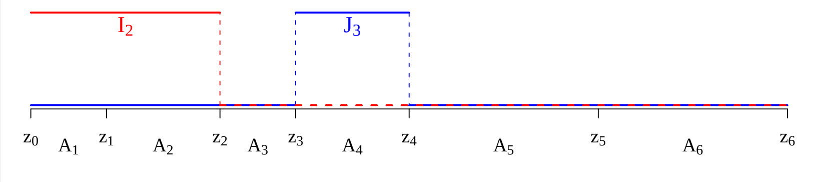
relation between cumulative and disjoint indicators



$$\mathbf{I} = \mathbf{L} \cdot \mathbf{J} \qquad \mathbf{L} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

- no difference between cokriging **I** and **J**
- $\hat{\mathbf{I}}$ has order violations $\Longleftrightarrow$ $\hat{\mathbf{J}}$ has negative values
- $\hat{\mathbf{I}}$ does not take any profit of the ordering information!
- order corrections do not symmetrically treat classes

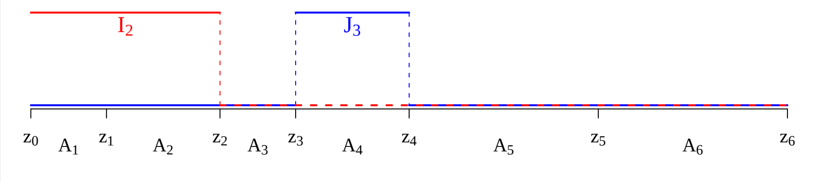
relation between cumulative and disjoint indicators



$$\mathbf{I} = \mathbf{L} \cdot \mathbf{J} \qquad \mathbf{L} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

- no difference between cokriging **I** and **J**
- $\hat{\mathbf{I}}$ has order violations $\Longleftrightarrow$ $\hat{\mathbf{J}}$ has negative values
- $\hat{\mathbf{I}}$ does not take any profit of the ordering information!
- order corrections do not symmetrically treat classes

relation between cumulative and disjoint indicators



$$\mathbf{I} = \mathbf{L} \cdot \mathbf{J} \qquad \mathbf{L} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

- no difference between cokriging **I** and **J**
- $\hat{\mathbf{I}}$ has order violations $\Longleftrightarrow$ $\hat{\mathbf{J}}$ has negative values
- $\hat{\mathbf{I}}$ does not take any profit of the ordering information!
- order corrections do not symmetrically treat classes

summary

1. goal: assess water quality
   - lots of variables, irregular time series
   - several chemical equilibria involved
   - NH_3 from sewers, controlled by pH (not buffered, lack of carbonates)
   - problem simplified to 4 water quality categories (ordered)
2. classical method: Indicator Kriging
   - variogram/covariance functions difficult to model
   - very often negative interpolations
   - without cokriging, almost never summing up to 1
   - can we trust the apparently valid results? and the corrected results?

summary

1. goal: assess water quality
   - lots of variables, irregular time series
   - several chemical equilibria involved
   - NH_3 from sewers, controlled by pH (not buffered, lack of carbonates)
   - problem simplified to 4 water quality categories (ordered)
2. classical method: Indicator Kriging
   - variogram/covariance functions difficult to model
   - very often negative interpolations
   - without cokriging, almost never summing up to 1
   - can we trust the apparently valid results? and the corrected results?

simplicial IK in a nut

two basic principles

- $\mathbf{J} = [J_1, \ldots J_D]$: multinomial variable; interest in its parameter $\mathbf{p}$
- respect the scale of the interpolated object (compositional scale)

five-step algorithm

1. first look at $\mathbf{J}$ structure (variogram: nugget, sill, range)
2. estimate $p_i(x_n)$ at sampled locations: $\hat{\mathbf{p}}(x_n) = \mathbf{A} \cdot \mathbf{J}(x_n)$
3. represent $\mathbf{p}(x_n) = [p_1, p_2, \ldots p_D]$ adequately in its scale (apply log-ratio transformations)
4. compute variograms, fit models, interpolate, in transformed scale
5. extract desired probabilities from interpolations
   multinomial! $\Pr[Z(x_n) \in A_i] = \hat{p}_i(x_0)$

simplicial IK in a nut

two basic principles

- $\mathbf{J} = [J_1, \ldots J_D]$: multinomial variable; interest in its parameter $\mathbf{p}$
- respect the scale of the interpolated object (compositional scale)

five-step algorithm

1. first look at $\mathbf{J}$ structure (variogram: nugget, sill, range)
2. estimate $p_i(x_n)$ at sampled locations: $\hat{\mathbf{p}}(x_n) = \mathbf{A} \cdot \mathbf{J}(x_n)$
3. represent $\mathbf{p}(x_n) = [p_1, p_2, \ldots p_D]$ adequately in its scale (apply log-ratio transformations)
4. compute variograms, fit models, interpolate, in transformed scale
5. extract desired probabilities from interpolations
   multinomial! $\Pr[Z(x_n) \in A_i] = \hat{p}_i(x_0)$

simplicial IK in a nut

two basic principles

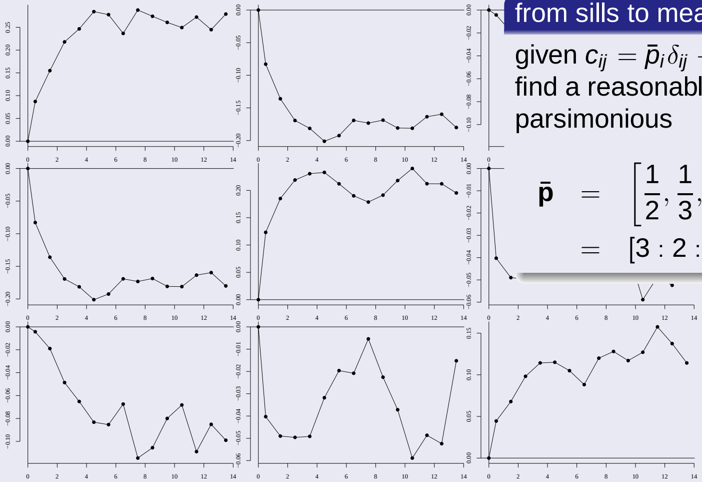
- $\mathbf{J} = [J_1, \ldots J_D]$: multinomial variable; interest in its parameter $\mathbf{p}$
- respect the scale of the interpolated object (compositional scale)

five-step algorithm

1. first look at **J** structure (variogram: nugget, sill, range)
2. estimate $p_i(x_n)$ at sampled locations: $\hat{\mathbf{p}}(x_n) = \mathbf{A} \cdot \mathbf{J}(x_n)$
3. represent $\mathbf{p}(x_n) = [p_1, p_2, \ldots p_D]$ adequately in its scale (apply log-ratio transformations)
4. compute variograms, fit models, interpolate, in transformed scale
5. extract desired probabilities from interpolations
   multinomial! $\quad \mathtt{Pr}[Z(x_n) \in A_i] = \hat{p}_i(x_0)$

variography of disjunctive indicators (step 1)

variogram system, year 2002



from sills to means of **p**

given $c_{ij} = \bar{p}_i \delta_{ij} - \bar{p}_i \bar{p}_j$, find a reasonable, parsimonious

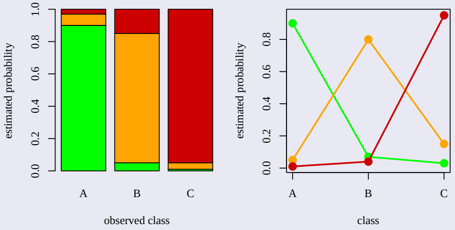
$$\bar{\mathbf{p}} = \left[\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right] = $$
$$= [3 : 2 : 1]$$

estimation of **p** at sampled locations (step 2)

a little bit more on the sharing matrices

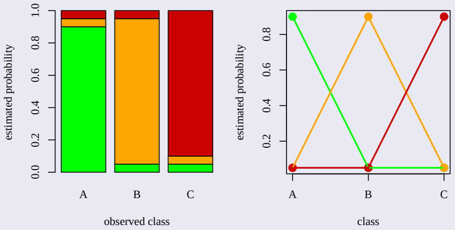
- the complex one

$$\begin{pmatrix} 0.90 & 0.05 & 0.01 \\ 0.07 & 0.80 & 0.04 \\ 0.03 & 0.15 & 0.95 \end{pmatrix}$$



- the simple one

$$\begin{pmatrix} 0.90 & 0.05 & 0.05 \\ 0.05 & 0.90 & 0.05 \\ 0.05 & 0.05 & 0.90 \end{pmatrix}$$

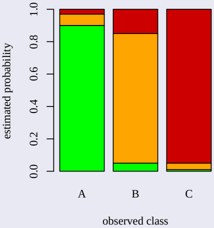


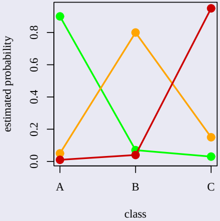
estimation of **p** at sampled locations (step 2)

a little bit more on the sharing matrices

- the complex one

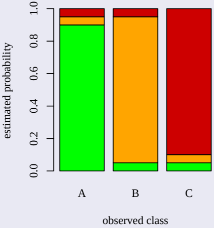
$$\begin{pmatrix} 0.90 & 0.05 & \color{red}{0.01} \\ 0.07 & 0.80 & \color{red}{0.04} \\ 0.03 & 0.15 & \color{red}{0.95} \end{pmatrix}$$

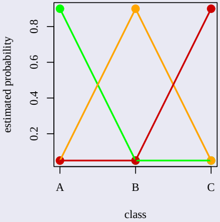




- the simple one

$$\begin{pmatrix} 0.90 & 0.05 & 0.05 \\ 0.05 & 0.90 & 0.05 \\ 0.05 & 0.05 & 0.90 \end{pmatrix}$$



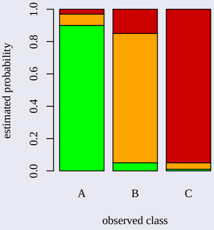


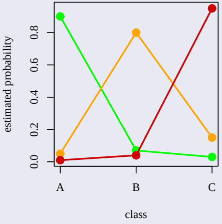
estimation of **p** at sampled locations (step 2)

a little bit more on the sharing matrices

- the complex one

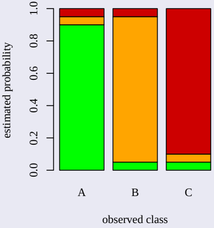
$$\begin{pmatrix} \color{red}{0.90} & 0.05 & 0.01 \\ 0.07 & \color{red}{0.80} & 0.04 \\ 0.03 & 0.15 & \color{red}{0.95} \end{pmatrix}$$

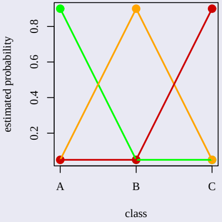




- the simple one

$$\begin{pmatrix} 0.90 & 0.05 & 0.05 \\ 0.05 & 0.90 & 0.05 \\ 0.05 & 0.05 & 0.90 \end{pmatrix}$$





computing coordinates (step 3)

reviewing *scale* and *sample space* of compositional data

- compositions can be freely closed: $\mathbf{x} \equiv \mathcal{C}[\mathbf{x}] = \mathbf{x}/sum(\mathbf{x})$
- compositions convey only relative information
- sample space, the D-part simplex ($\mathcal{S}^D$), Euclidean space
- orthonormal basis and coordinates
  $\xi = \boldsymbol{\Psi} \cdot \ln \mathbf{x} \quad \Longleftrightarrow \quad \mathbf{x} = \mathcal{C}\left[\exp\left(\left(\boldsymbol{\Psi}^t \cdot \xi\right)\right)\right]$

relevance for p

- $\mathcal{C}$: likelihood vectors $\equiv$ probability vectors

$$\mathcal{C}[3,2,1] = \frac{1}{3+2+1}[3,2,1] = \left[\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right] \equiv [3:2:1]$$

- $\oplus$, discrete Bayes Theorem; $||\cdot||_a$: information measure
- ξ are log-contrasts (logistic regression)

computing coordinates (step 3)

reviewing *scale* and *sample space* of compositional data

- compositions can be freely closed: $\mathbf{x} \equiv \mathcal{C}[\mathbf{x}] = \mathbf{x}/sum(\mathbf{x})$
- compositions convey only relative information
- sample space, the D-part simplex ($\mathcal{S}^D$), Euclidean space
- orthonormal basis and coordinates
  $\xi = \boldsymbol{\Psi} \cdot \ln \mathbf{x} \iff \mathbf{x} = \mathcal{C}\left[\exp\left(\left(\boldsymbol{\Psi}^t \cdot \xi\right)\right)\right]$

relevance for p

- $\mathcal{C}$: likelihood vectors $\equiv$ probability vectors

$$\mathcal{C}[3,2,1] = \frac{1}{3+2+1}[3,2,1] = \left[\frac{1}{2},\frac{1}{3},\frac{1}{6}\right] \equiv [3:2:1]$$

- $\oplus$, discrete Bayes Theorem; $||\cdot||_a$: information measure
- ξ are log-contrasts (logistic regression)

computing coordinates (step 3)

reviewing *scale* and *sample space* of compositional data

- compositions can be freely closed: $\mathbf{x} \equiv \mathcal{C}[\mathbf{x}] = \mathbf{x}/sum(\mathbf{x})$
- compositions convey only relative information
- sample space, the D-part simplex ($\mathcal{S}^D$), Euclidean space
- orthonormal basis and coordinates
  $\boldsymbol{\xi} = \boldsymbol{\Psi} \cdot \ln \mathbf{x} \quad \Longleftrightarrow \quad \mathbf{x} = \mathcal{C}\left[\exp\left(\left(\boldsymbol{\Psi}^t \cdot \boldsymbol{\xi}\right)\right)\right]$

relevance for p

- $\mathcal{C}$: likelihood vectors $\equiv$ probability vectors

$$\mathcal{C}[3,2,1] = \frac{1}{3+2+1}[3,2,1] = \left[\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right] \equiv [3:2:1]$$

- $\oplus$, discrete Bayes Theorem; $||\cdot||_a$: information measure
- ξ are log-contrasts (logistic regression)

computing coordinates (step 3)

reviewing *scale* and *sample space* of compositional data

- compositions can be freely closed: $\mathbf{x} \equiv \mathcal{C}\left[\mathbf{x}\right] = \mathbf{x}/sum(\mathbf{x})$
- compositions convey only relative information
- sample space, the D-part simplex ($\mathcal{S}^D$), Euclidean space
- orthonormal basis and coordinates
  $\xi = \boldsymbol{\Psi} \cdot \ln \mathbf{x} \quad \Longleftrightarrow \quad \mathbf{x} = \mathcal{C}\left[\exp\left(\left(\boldsymbol{\Psi}^t \cdot \xi\right)\right)\right]$

relevance for p

- $\mathcal{C}$: likelihood vectors $\equiv$ probability ve

$$\mathcal{C}[3,2,1] = \frac{1}{3+2+1}[3,2,1] =$$

- $\oplus$, discrete Bayes Theorem; $||\cdot||_a$: in
- ξ are log-contrasts (logistic regression)

ilr coordinate matrix

$$\boldsymbol{\Psi} = \begin{pmatrix} \frac{+2}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} \\ 0 & \frac{+1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$$

computing coordinates (step 3)

reviewing *scale* and *sample space* of compositional data

- compositions can be freely closed: $\mathbf{x} \equiv \mathcal{C}[\mathbf{x}] = \mathbf{x}/sum(\mathbf{x})$
- compositions convey only relative information
- sample space, the D-part simplex ($\mathcal{S}^D$), Euclidean space
- orthonormal basis and coordinates
  $\xi = \mathbf{\Psi} \cdot \ln \mathbf{x} \quad \Longleftrightarrow \quad \mathbf{x} = \mathcal{C}\left[\exp\left(\left(\mathbf{\Psi}^t \cdot \xi\right)\right)\right]$

relevance for **p**

- $\mathcal{C}$: likelihood vectors $\equiv$ probability vectors

$$\mathcal{C}[3,2,1] = \frac{1}{3+2+1}[3,2,1] = \left[\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right] \equiv [3:2:1]$$

- $\oplus$, discrete Bayes Theorem; $||\cdot||_a$: information measure
- ξ are log-contrasts (logistic regression)

geostatistics on coordinates (step 4)

review of geostatistics for compositions

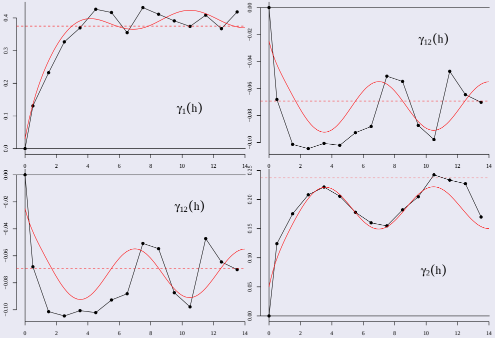
- alr $\Rightarrow$ analyse $\Rightarrow$ back-trasform (Pawlowsky-Glahn and Olea, 2004)
- compute coordinates $\Rightarrow$ analyse $\Rightarrow$ apply to the basis
  - unbiased, $\mathrm{E}_{\mathcal{S}}[\hat{\mathbf{z}}_0] = \mathrm{E}_{\mathcal{S}}[\mathbf{Z}_0]$
  - minimal error variance, or minimal expected distance $d_A(\hat{\mathbf{z}}_0, \mathbf{Z}_0)$
- proposition $\Longrightarrow$ results DO NOT depend on the basis
  - any change of basis is a full-rank linear transformation

variography of coordinates (step 4)

coordinate variography

- easier to model: less components
- positive definiteness
- no further conditions

coordinate variograms (2002)



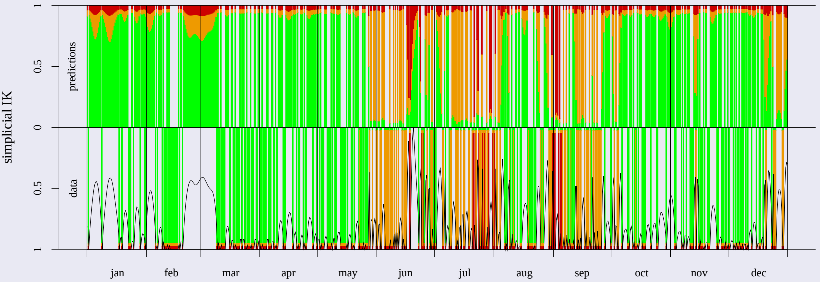
$\frac{\Gamma^{\pi}(h)}{100\beta^2}$	$\gamma_1(h)$	$\gamma_2(h)$	$\gamma_{12}(h)$
nugget	3	5	-2.5
exponential$(r=1.5)$	3	5	-2
exponential$(r=5)$	30	4	-2
hole$(T=7)$	3	4	-2

extraction of the sought probabilities (step 5)

1. interpolated $\hat{\mathbf{p}}_0 \Rightarrow$ apply to the basis: $\hat{\mathbf{p}}_0 = \mathcal{C}\left[\exp\left(\boldsymbol{\Psi} \cdot \hat{\mathbf{p}}_0\right)\right]$
2. sought probability: $\Pr\left[Z_0 \in A_i\right] = \left(\hat{\mathbf{p}}_0\right)_i$

extraction of the sought probabilities (step 5)

1. interpolated $\hat{\mathbf{p}}_0 \Rightarrow$ apply to the basis: $\hat{\mathbf{p}}_0 = \mathcal{C}\left[\exp\left(\mathbf{\Psi} \cdot \hat{\mathbf{p}}_0\right)\right]$
2. sought probability: $\Pr\left[Z_0 \in A_i\right] = \left(\hat{\mathbf{p}}_0\right)_i$

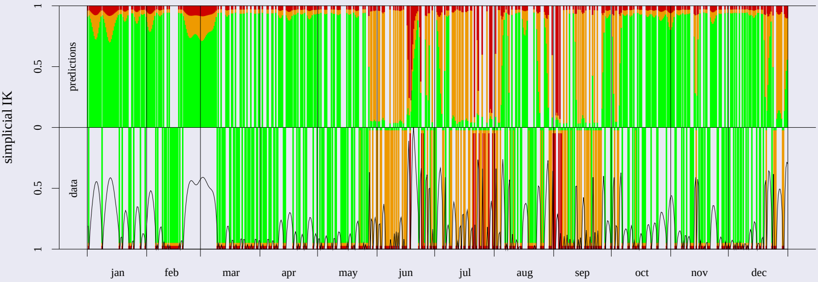


information measure: Aitchison norm $\|\hat{\mathbf{p}}_0\|_a$ scaled in [0.5, 3]

$\|\hat{\mathbf{p}}_0\|_a \longrightarrow 0 \Leftrightarrow \hat{\mathbf{p}}_0 \longrightarrow n \Leftrightarrow Z_0$ less certain

extraction of the sought probabilities (step 5)

1. interpolated $\hat{\mathbf{p}}_0 \Rightarrow$ apply to the basis: $\hat{\mathbf{p}}_0 = \mathcal{C}\left[\exp\left(\mathbf{\Psi} \cdot \hat{\mathbf{p}}_0\right)\right]$
2. sought probability: $\Pr\left[Z_0 \in A_i\right] = \left(\hat{\mathbf{p}}_0\right)_i$

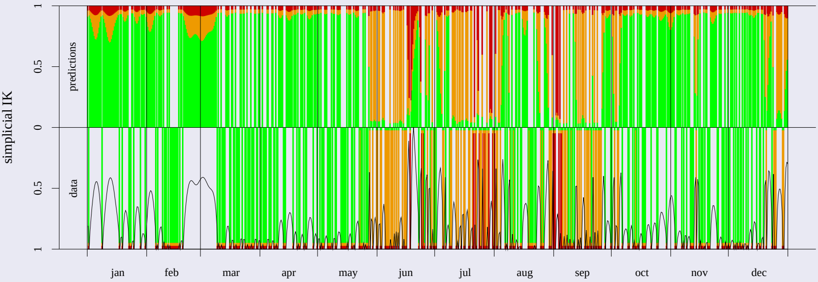


information measure: Aitchison norm $||\hat{\mathbf{p}}_0||_a$ scaled in [0.5, 3]

- $||\hat{\mathbf{p}}_0||_a \longrightarrow 0 \Leftrightarrow \hat{\mathbf{p}}_0 \longrightarrow \mathbf{n} \Leftrightarrow Z_0$ less certain
- $||\hat{\mathbf{p}}_0||_a \longrightarrow +\infty \Leftrightarrow (\hat{\mathbf{p}}_0)_j \longrightarrow 0 \Leftrightarrow Z_0$ more certain (A_j impossible)

extraction of the sought probabilities (step 5)

1. interpolated $\hat{\mathbf{p}}_0 \Rightarrow$ apply to the basis: $\hat{\mathbf{p}}_0 = \mathcal{C}\left[\exp\left(\mathbf{\Psi}\cdot\hat{\mathbf{p}}_0\right)\right]$
2. sought probability: $\Pr\left[Z_0 \in A_i\right] = \left(\hat{\mathbf{p}}_0\right)_i$

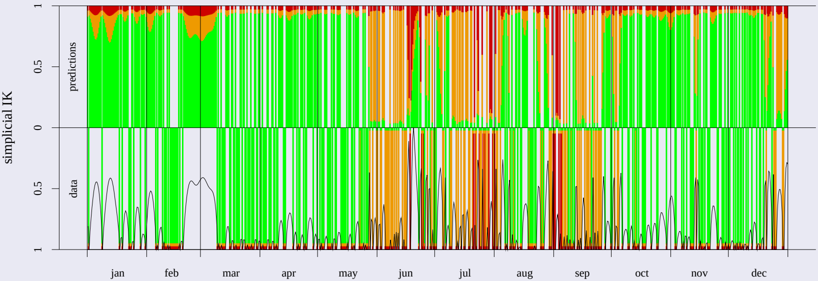


information measure: Aitchison norm $||\hat{\mathbf{p}}_0||_a$ scaled in [0.5, 3]

- $||\hat{\mathbf{p}}_0||_a \longrightarrow 0 \Leftrightarrow \hat{\mathbf{p}}_0 \longrightarrow \mathbf{n} \Leftrightarrow Z_0$ less certain
- $||\hat{\mathbf{p}}_0||_a \longrightarrow +\infty \Leftrightarrow (\hat{\mathbf{p}}_0)_j \longrightarrow 0 \Leftrightarrow Z_0$ more certain (A_j impossible)

extraction of the sought probabilities (step 5)

1. interpolated $\hat{\mathbf{p}}_0 \Rightarrow$ apply to the basis: $\hat{\mathbf{p}}_0 = \mathcal{C}\left[\exp\left(\mathbf{\Psi} \cdot \hat{\mathbf{p}}_0\right)\right]$
2. sought probability: $\Pr\left[Z_0 \in A_i\right] = \left(\hat{\mathbf{p}}_0\right)_i$



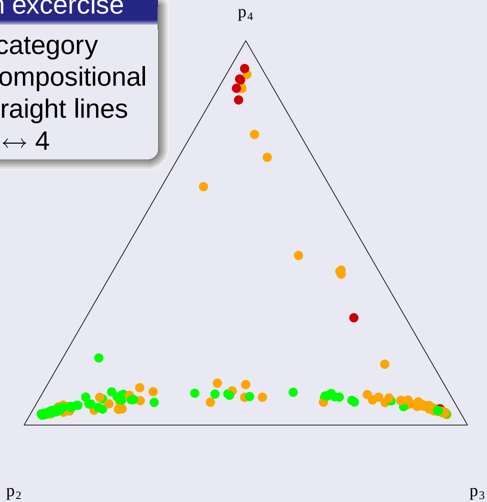
information measure: Aitchison norm $||\hat{\mathbf{p}}_0||_a$ scaled in [0.5, 3]

- $||\hat{\mathbf{p}}_0||_a \longrightarrow 0 \Leftrightarrow \hat{\mathbf{p}}_0 \longrightarrow \mathbf{n} \Leftrightarrow Z_0$ less certain
- $||\hat{\mathbf{p}}_0||_a \longrightarrow +\infty \Leftrightarrow (\hat{\mathbf{p}}_0)_j \longrightarrow 0 \Leftrightarrow Z_0$ more certain (A_j impossible)

extraction of the sought probabilities (step 5)

cross-validation excercise

colours ∼ true category
predictions ∼ compositional straight lines
no confusion 2 ↔ 4



summary

- simplicial indicator kriging algorithm
  1. variography of disjunctive indicators
  2. local estimation of **p** through sharing matrix
  3. representation of **p** in log-ratio coordinates
  4. geostatistics of the coordinates
  5. obtention of probabilities: application of interpolated coordinates to the basis
- simplicial indicator kriging advantadges
  1. old software is useful, estimation of the average of **p**
  2. opportunity to include assessment of reliability (instrumental error vs. unclear classification, local vs. global $\Longrightarrow$ GIS potential)
  3. interpretable coordinates: Bayesian addition of information
  4. easier modeling of variograms in coordinates; invertible cokriging systems
  5. final **p** estimates always valid; no correction needed

summary

- simplicial indicator kriging algorithm
  1. variography of disjunctive indicators
  2. local estimation of **p** through sharing matrix
  3. representation of **p** in log-ratio coordinates
  4. geostatistics of the coordinates
  5. obtention of probabilities: application of interpolated coordinates to the basis
- simplicial indicator kriging advantadges
  1. old software is useful, estimation of the average of **p**
  2. opportunity to include assessment of reliability (instrumental error vs. unclear classification, local vs. global $\Longrightarrow$ GIS potential)
  3. interpretable coordinates: Bayesian addition of information
  4. easier modeling of variograms in coordinates; invertible cokriging systems
  5. final **p** estimates always valid; no correction needed

properties of simplicial indicator kriging

summary of properties of sIK already seen

- estimator is BLU Estimator:
  - Best: minimal (metric) variance,
  - Linear transformation of observed data
  - Unbiased: expected estimation = expected true value

  . . . in a compositional sense
- results are always valid probability vectors
- independent of the working basis

the simple sharing matrix case: coordinates vs. indicators

1. observed $\mathbf{J}_n$ at location x_n
2. sharing matrix:

$$\hat{\mathbf{p}}_n = \mathbf{A} \cdot \mathbf{J}_n; \qquad \mathbf{A} = \begin{pmatrix} 0.950 & 0.025 & 0.025 \\ 0.025 & 0.950 & 0.025 \\ 0.025 & 0.025 & 0.950 \end{pmatrix}$$

$$\hat{p}_i = \begin{cases} 1-\alpha & J_i = 1, \\ \alpha/(D-1) & J_i = 0, \end{cases} \quad \alpha (= 0.05) \text{ prob. missclassification}$$

3. coordinates:
   - of a generic vector of probabilities

$$\hat{\boldsymbol{\pi}}_n = \boldsymbol{\Psi} \cdot \ln(\mathbf{A} \cdot \mathbf{J}_n) = \boldsymbol{\Psi} \cdot \mathbf{B} \cdot \mathbf{J}_n, \qquad \mathbf{B} = (\ln \mathbf{A})$$

   - of the simple sharing matrix case

$$\hat{\boldsymbol{\pi}}_n = \beta \cdot \boldsymbol{\Psi} \cdot \mathbf{J}_n, \qquad \beta = \ln \frac{(1-\alpha)(D-1)}{\alpha}$$

the simple sharing matrix case: geostatistics

$$\hat{\pi}_n = \beta \cdot \mathbf{\Psi} \cdot \mathbf{J}_n, \qquad \beta = \ln \frac{(1-\alpha)(D-1)}{\alpha}$$

linear, invertible relationship ⇒ relations in geostatistics:

- consistency of covariance models:

$$\mathbf{\Gamma}^{\pi}(h) = \beta^2 \cdot \mathbf{\Psi} \cdot \mathbf{\Gamma}^{J}(h) \cdot \mathbf{\Psi}^t$$

- relation between predictions:

$$\hat{\pi}_0 = \beta \cdot \mathbf{\Psi} \cdot \hat{\mathbf{J}}_0 \longleftrightarrow \hat{\mathbf{J}}_0 = \frac{1}{\beta} \cdot \mathbf{\Psi}^t \cdot \hat{\pi}_0 + \frac{1}{D}\mathbf{1}$$

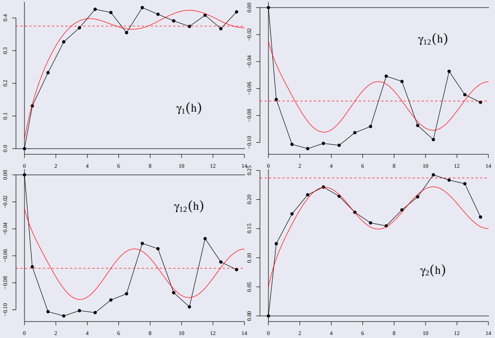
the simple sharing matrix case

variography

- variograms for indicators and coordinates consistent: $\mathbf{\Gamma}^{\pi}(h) = \beta^2 \cdot \mathbf{\Psi} \cdot \mathbf{\Gamma}^{J}(h) \cdot \mathbf{\Psi}^t$
- no need to recompute them!
- easier to model in coordinates: less components, NOT bound to:
  - sum to 0 by rows
  - sum to 0 by columns
  - sill condition, $c_{ij} = \bar{p}_i \delta_{ij} - \bar{p}_i \bar{p}_j$

  (as **J** does)

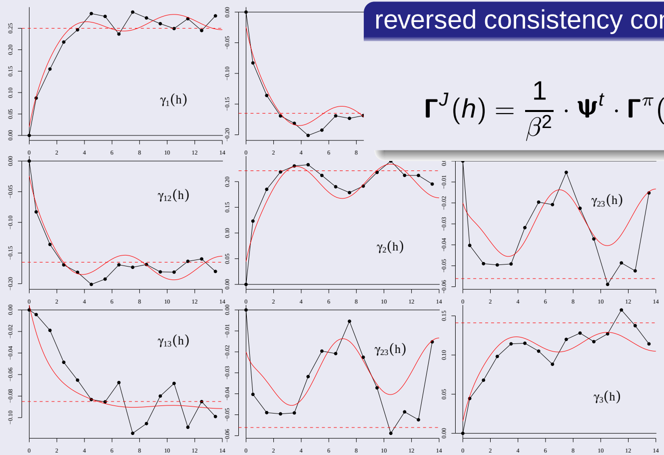
coordinate variograms (2002)



$\frac{\mathbf{\Gamma}^{\pi}(h)}{100\beta^2}$	$\gamma_1(h)$	$\gamma_2(h)$	$\gamma_{12}(h)$
nugget	3	5	-2.5
exponential$(r = 1.5)$	3	5	-2
exponential$(r = 5)$	30	4	-2
hole$(T = 7)$	3	4	-2

the simple sharing matrix case

checking what happened with indicator variograms (2002)



reversed consistency conditions

$$\mathbf{\Gamma}^J(h) = \frac{1}{\beta^2} \cdot \mathbf{\Psi}^t \cdot \mathbf{\Gamma}^\pi(h) \cdot \mathbf{\Psi}$$

the simple sharing matrix case

relations between cokriging predictions

- proposition $\Rightarrow$ results for $\hat{\pi}_0$ are equivalent:
  - cokriging $D-1$ coordinates directly ($\hat{\pi}_0$)
  - cokriging D indicators ($\hat{\mathbf{j}}_0$) and transforming them through

$$\hat{\pi}_0 = \beta \cdot \mathbf{\Psi} \cdot \mathbf{J}_0$$

- if we apply kriged results to the basis used:
  $\hat{\mathbf{p}}_0 = \mathcal{C}\left(\exp\left(\mathbf{\Psi}^t \cdot \hat{\pi}_0\right)\right) = \mathcal{C}\left(\exp\left(\beta \cdot \hat{\mathbf{j}}_0\right)\right)$
  - always valid: positive, summing up to one
  - no $\mathbf{\Psi} \Longrightarrow$ choice of basis modifies nothing
  - wait to fix β (or $\alpha = 0.05$) until the end
- only for cokriging!
- if cokriging is too complex?
  1. kriging j_i individually
  2. combine them with β

the simple sharing matrix case

relations between cokriging predictions

- proposition $\Rightarrow$ results for $\hat{\pi}_0$ are equivalent:
  - cokriging $D-1$ coordinates directly ($\hat{\pi}_0$)
  - cokriging D indicators ($\hat{\mathbf{j}}_0$) and transforming them through

$$\hat{\pi}_0 = \beta \cdot \mathbf{\Psi} \cdot \mathbf{J}_0$$

- if we apply kriged results to the basis used:

$$\hat{\mathbf{p}}_0 = \mathcal{C}\left(\exp\left(\mathbf{\Psi}^t \cdot \hat{\pi}_0\right)\right) = \mathcal{C}\left(\exp\left(\beta \cdot \hat{\mathbf{j}}_0\right)\right)$$

  - always valid: positive, summing up to one
  - no $\mathbf{\Psi} \Longrightarrow$ choice of basis modifies nothing
  - wait to fix β (or $\alpha = 0.05$) until the end
- only for cokriging!
- if cokriging is too complex?
  1. kriging j_i individually
  2. combine them with β

the simple sharing matrix case

relations between cokriging predictions

- proposition $\Rightarrow$ results for $\hat{\pi}_0$ are equivalent:
  - cokriging $D-1$ coordinates directly ($\hat{\pi}_0$)
  - cokriging D indicators ($\hat{\mathbf{j}}_0$) and transforming them through

$$\hat{\pi}_0 = \beta \cdot \mathbf{\Psi} \cdot \mathbf{J}_0$$

- if we apply kriged results to the basis used:

$$\hat{\mathbf{p}}_0 = \mathcal{C}\left(\exp\left(\mathbf{\Psi}^t \cdot \hat{\pi}_0\right)\right) = \mathcal{C}\left(\exp\left(\beta \cdot \hat{\mathbf{j}}_0\right)\right)$$

  - always valid: positive, summing up to one
  - no $\mathbf{\Psi} \Longrightarrow$ choice of basis modifies nothing
  - wait to fix β (or $\alpha = 0.05$) until the end
- only for cokriging!
- if cokriging is too complex?
  1. kriging j_i individually
  2. combine them with β

the simple sharing matrix case

relations between cokriging predictions

- proposition $\Rightarrow$ results for $\hat{\pi}_0$ are equivalent:
  - cokriging $D-1$ coordinates directly ($\hat{\pi}_0$)
  - cokriging D indicators ($\hat{\mathbf{j}}_0$) and transforming them through

$$\hat{\pi}_0 = \beta \cdot \mathbf{\Psi} \cdot \mathbf{J}_0$$

- if we apply kriged results to the basis used:

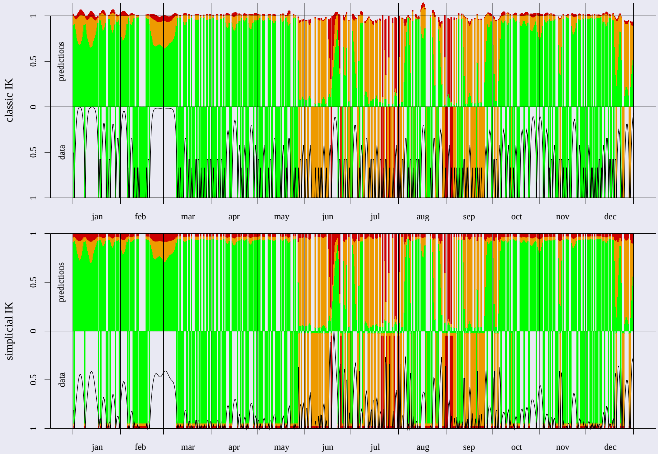
$$\hat{\mathbf{p}}_0 = \mathcal{C}\left(\exp\left(\mathbf{\Psi}^t \cdot \hat{\pi}_0\right)\right) = \mathcal{C}\left(\exp\left(\beta \cdot \hat{\mathbf{j}}_0\right)\right)$$

  - always valid: positive, summing up to one
  - no $\mathbf{\Psi} \Longrightarrow$ choice of basis modifies nothing
  - wait to fix β (or $\alpha = 0.05$) until the end
- only for cokriging!
- if cokriging is too complex?
  1. kriging j_i individually
  2. combine them with β

the simple sharing matrix case

comparing classic and simplicial IK, with $\alpha = 0.05$

- information measure: kriging variance of J_2 scaled in [0, 0.25]
- information measure: Aitchison norm $||\hat{\mathbf{p}}_0||_A$ scaled in [0.5, 3]



comparison

classic co-IK

$$\hat{\ }_0 = \sum_{n=1}^{N} \mathbf{\Lambda}_n \cdot \mathbf{J}_n$$

- many data to estimate variograms, strong conditions on the valid models
- negative components
- needed corrections

simplicial co-IK

$$\hat{\pi}_0 = \sum_{n=1}^{N} \mathbf{\Lambda}_n \cdot \pi_n$$

- many data to estimate variograms

classic IK

$$(\hat{\mathbf{j}}_0)_i = \sum_{n=1}^{N} \lambda_n \cdot (\mathbf{J}_n)_i$$

- suboptimal
- negative components
- sum $\neq 1$
- needed corrections
- ignores the variogram problem (does not solve it!)

simplicial IK

$$\hat{\mathbf{p}}_0 = \mathcal{C}\left[\exp\left(\beta \cdot \hat{\mathbf{j}}_0\right)\right]$$

- suboptimal

comparison

classic co-IK

$$\hat{}_0 = \sum_{n=1}^{N} \mathbf{\Lambda}_n \cdot \mathbf{J}_n$$

- many data to estimate variograms, strong conditions on the valid models
- negative components
- needed corrections

simplicial co-IK

$$\hat{\pi}_0 = \sum_{n=1}^{N} \mathbf{\Lambda}_n \cdot \pi_n$$

- many data to estimate variograms

classic IK

$$(\hat{\mathbf{j}}_0)_i = \sum_{n=1}^{N} \lambda_n \cdot (\mathbf{J}_n)_i$$

- suboptimal
- negative components
- sum $\neq 1$
- needed corrections
- ignores the variogram problem (does not solve it!)

simplicial IK

$$\hat{\mathbf{p}}_0 = \mathcal{C}\left[\exp\left(\beta \cdot \hat{\mathbf{j}}_0\right)\right]$$

- suboptimal

comparison

classic co-IK

$$\hat{}_0 = \sum_{n=1}^{N} \mathbf{\Lambda}_n \cdot \mathbf{J}_n$$

- many data to estimate variograms, strong conditions on the valid models
- negative components
- needed corrections

simplicial co-IK

$$\hat{\boldsymbol{\pi}}_0 = \sum_{n=1}^{N} \mathbf{\Lambda}_n \cdot \boldsymbol{\pi}_n$$

- many data to estimate variograms

classic IK

$$(\hat{\mathbf{j}}_0)_i = \sum_{n=1}^{N} \lambda_n \cdot (\mathbf{J}_n)_i$$

- suboptimal
- negative components
- sum $\neq 1$
- needed corrections
- ignores the variogram problem (does not solve it!)

simplicial IK

$$\hat{\mathbf{p}}_0 = \mathcal{C}\left[\exp\left(\beta \cdot \hat{\mathbf{j}}_0\right)\right]$$

- suboptimal

comparison

classic co-IK

$$\hat{}_0 = \sum_{n=1}^{N} \mathbf{\Lambda}_n \cdot \mathbf{J}_n$$

- many data to estimate variograms, strong conditions on the valid models
- negative components
- needed corrections

simplicial co-IK

$$\hat{\boldsymbol{\pi}}_0 = \sum_{n=1}^{N} \mathbf{\Lambda}_n \cdot \boldsymbol{\pi}_n$$

- many data to estimate variograms

classic IK

$$(\hat{\mathbf{j}}_0)_i = \sum_{n=1}^{N} \lambda_n \cdot (\mathbf{J}_n)_i$$

- suboptimal
- negative components
- sum $\neq 1$
- needed corrections
- ignores the variogram problem (does not solve it!)

simplicial IK

$$\hat{\mathbf{p}}_0 = \mathcal{C}\left[\exp\left(\beta \cdot \hat{\mathbf{j}}_0\right)\right]$$

- suboptimal

conclusions

- distinguish **J** (multinomial) from **p** (its parameter)
- geostatistics on the coordinates of **p** (as a composition)
  - easier modeling of covariance/variogram structures
  - yield always valid results (also individual kriging)
  - BLUE with respect to a compositional scale
  - interpretable in a Bayesian framework
- geostatistical procedure: not dependent on the preliminary **p** estimation (β, α, matrix **A**)
- final cokriging results: not dependent on the basis chosen

more material

further reading

- *all the stuff*:
  Tolosana-Delgado, R., 2006. *Geostatistics for constrained variables: positive data, compositions and probabilities. Application to environmental hazard monitoring*. Ph.D. thesis (U. Girona, Spain)
- *about simplicial indicator kriging*:
  Tolosana-Delgado, R., Pawlowsky-Glahn, V., Egozcue, J. J. Indicator kriging without order relation violations. *Mathematical Geology*
- *using the same technique with positive variables*:
  Tolosana-Delgado, R., Pawlowsky-Glahn, V., 2007. Kriging regionalized positive variables revisited: sample space and scale considerations. *Mathematical Geology*, in press

CoDaWork'08: 3rd International Workshop on CoDa

Girona (Spain), May 27 to 30, 2008.

simplicial indicator kriging

- **Thanks for your attention**

statistics for random vectors

the object way: use vectors + linear applications (Eaton, 1983)

- $\mathrm{E}[Z]$: expectation already defined if Z a real random variable
- projections have real *values*, $P_{\mathbf{u}}(\mathbf{z}) = (\mathbf{z}, \mathbf{u})_A$, with $\mathbf{u}$ a direction
- $\mathrm{E}_{\mathcal{S}}[\mathbf{Z}] = \mathbf{m}$ a vector capturing all projections, $\mathrm{E}\left[P_{\mathbf{u}}(\mathbf{Z})\right] = P_{\mathbf{u}}(\mathbf{m})$
- $\mathrm{Var}_{\mathcal{S}}[\mathbf{Z}] = \Sigma$ an endomorphism capturing all pairs of projections, $\mathrm{E}\left[P_{\mathbf{u}}(\mathbf{Z} \ominus \mathbf{m}) \cdot P_{\mathbf{v}}(\mathbf{Z} \ominus \mathbf{m})\right] = P_{\mathbf{u}}(\Sigma\mathbf{v})$

return

measures of information in a probability vector

entropy vs. Aitchison norm

- Aitchison norm

$$||\mathbf{p}||_A = \sqrt{\frac{1}{3}\left(\log^2\frac{p_1}{p_2} + \log^2\frac{p_2}{p_3} + \log^2\frac{p_1}{p_3}\right)}$$

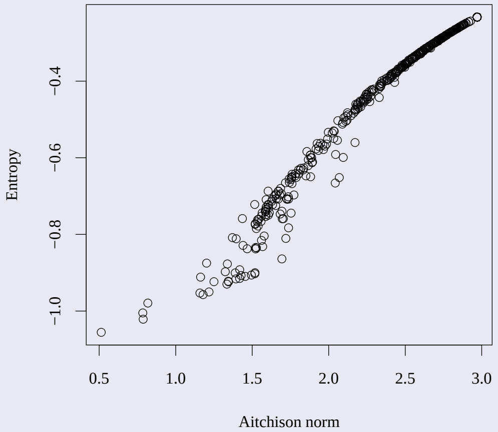
- Shannon entropy

$$H = p_1 \log p_1 + p_2 \log p_2 + p_3 \log p_3$$

return

measures of information in a probability vector

comparison



entrop

- A

$$\ldots og^2 \frac{p_1}{p_3} \Big)$$

- S

$$\ldots g\, p_3$$

return