Revisiting cokriging of **indicator functions** and compositions

Andrei Borisovitch Vistelius Award for Young Scientists
Annual Conference of the International Association for Mathematical Geology

Raimon Tolosana-Delgado raimon.tolosana@geo.uni-goettingen.de
Department of Sedimentology and Environmental Geology
Georg-August-Universität Göttingen

1. presentation
   - a case study: assessing water quality
   - Indicator Kriging: interpolating uncertain categories
   - sketch of solution

2. theory on compositional data
   - geometry
   - statistics
   - geostatistics

3. application
   - obtention and variography of the categorical variables
   - estimation of parameter vectors at sampled locations
   - computation of coordinates
   - conventional geostatistical inventory on the coordinates
   - extract probabilities for unsampled locations

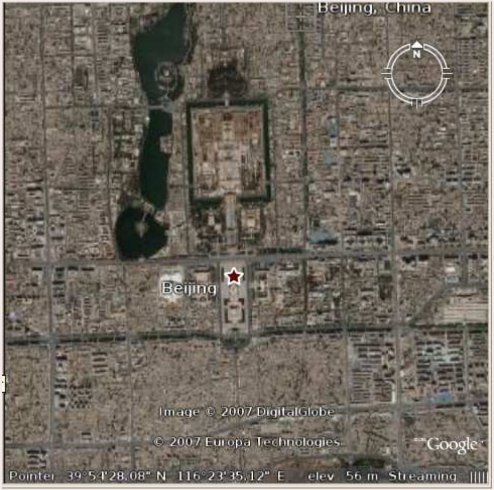
4. conclusions

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

measured variables

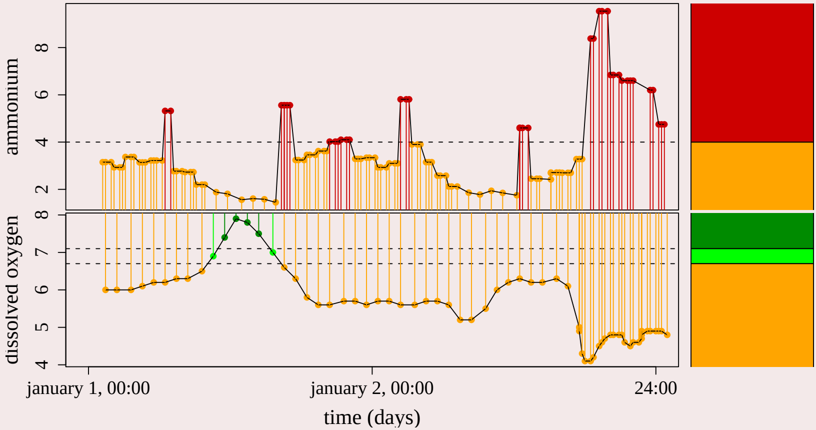
- conductivity, pH, ammonium concentration, (water temperature, dissolved $O_2, \ldots$)
- main interest: potential of ammonia production
- ammonia (NH_3): lethal (fishes, macroinvertebrates), but volatile
- ammonium (NH_4^+): much less dangerous on itself, but

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

$$\log \frac{[NH_3]}{\left[NH_4^+\right]} = f\left(T_w, pH\right)$$

water quality assessment: a particular case

uncertain category assessment



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

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

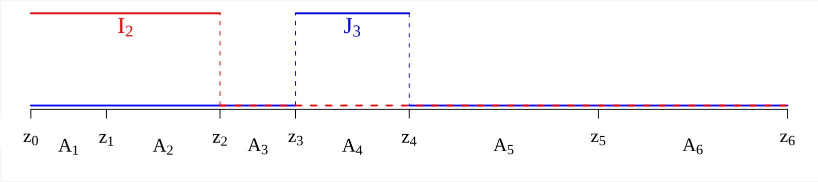
- often results are not valid probabilities:
  - $\hat{I}_i$'s are not ordered
  - $\hat{J}_i$'s are negative, or they do not sum up to one
-

geostatistics for categorical variables

treatment: Indicator Kriging (IK; Journel, 1983)

1. (re)define the categories as indicator functions

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



- often results are not valid probabilities:
  - $\hat{I}_i$'s are not ordered
  - $\hat{J}_i$'s are negative, or they do not sum up to one

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

- often results are not valid probabilities:
  - $\hat{I}_i$'s are not ordered
  - $\hat{J}_i$'s are negative, or they do not sum up to one
- the scale of I (or J) is the scale of $\Pr[Z(x_0) \in A]$

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

- often results are not valid probabilities:
  - $\hat{I}_i$'s are not ordered
  - $\hat{J}_i$'s are negative, or they do not sum up to one
- the scale of I (or J) is NOT the scale of $\mathrm{Pr}[Z(x_0) \in A_i]$

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:
$\hat{I}_i(x_0) \Rightarrow \mathrm{Pr}[Z(x_0) < z_i]$ or $\hat{J}_i(x_0) \Rightarrow \mathrm{Pr}[Z(x_0) \in A_i]$

problems

- often results are not valid probabilities:
  - $\hat{I}_i$'s are not ordered
  - $\hat{J}_i$'s are **negative, or they do not sum up to one**
- the scale of I (or J) is NOT the scale of $\mathrm{Pr}[Z(x_0) \in A_i]$

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

- often results are not valid probabilities:
  - $\hat{I}_i$'s are not ordered
  - $\hat{J}_i$'s are **negative, or they do not sum up to one**
- the scale of I (or J) is NOT the scale of $\mathtt{Pr}[Z(x_0) \in A_i]$

sketch of our solution

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)

algorithm: simplicial Indicator Kriging (sIK)

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! $\quad Pr[Z(x_0) = A_i] = p_i(x_0)$

sketch of our solution

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)

algorithm: simplicial Indicator Kriging (sIK)

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! $\Pr[Z(x_n) \in A_i] = \hat{p}_i(x_0)$

sketch of our solution

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)

algorithm: simplicial Indicator Kriging (sIK)

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 $\mathbf{p}(x_n) = [p_1, p_2, \ldots p_D]$ adequa
   log-ratio transformations)
4. compute variograms, fit models, interpol
5. extract desired probabilities from interpo
   multinomial! $\Pr[Z(x_n) \in A_i] = \hat{p}_i(x_0)$

a sharing matrix example

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

sketch of our solution

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)

algorithm: simplicial Indicator Kriging (sIK)

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 $\mathbf{p}(x_n) = [p_1, p_2, \ldots p_D]$ adequa
   log-ratio transformations)
4. compute variograms, fit models, interpol
5. extract desired probabilities from interpo
   multinomial! $\Pr[Z(x_n) \in A_i] = \hat{p}_i(x_0)$

a simpler sharing matrix

$$\mathbf{A} = \begin{pmatrix} 0.90 & 0.05 & 0.05 \\ 0.05 & 0.90 & 0.05 \\ 0.05 & 0.05 & 0.90 \end{pmatrix}$$

sketch of our solution

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)

algorithm: simplicial Indicator Kriging (sIK)

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! $\Pr[Z(x_n) \in A_i] = \hat{p}_i(x_0)$

sketch of our solution

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)

algorithm: simplicial Indicator Kriging (sIK)

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! $\quad \mathtt{Pr}[Z(x_n) \in A_i] = \hat{p}_i(x_0)$

geometry (I)

the *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
- the sample space of compositions, the D-part simplex ($\mathcal{S}^D$) is an Euclidean space (Billheimer et al.; Pawlowsky-Glahn and Egozcue, 2001)
- orthonormal basis and coordinates

$$\xi = \Psi \cdot \ln \mathbf{x} \iff \mathbf{x} = \mathcal{C}\left[\exp\left(\left(\Psi^t \cdot \xi\right)\right)\right]$$

geometry (I)

the *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
- the sample space of compositions, the D-part simplex ($\mathcal{S}^D$) is an Euclidean space (Billheimer et al.; Pawlowsky-Glahn and Egozcue, 2001)
- orthonormal basis and coordinates

$$\xi = \Psi \cdot \ln \mathbf{x} \quad \Longleftrightarrow \quad \mathbf{x} = \mathcal{C}\left[\exp\left(\left(\Psi^t \cdot \xi\right)\right)\right]$$

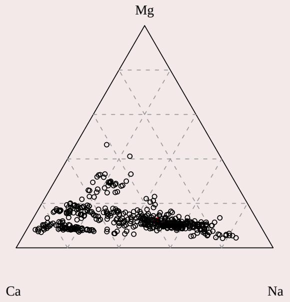
geometry (I)

the *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
- the sample space of compositions, the *L* Euclidean space (Billheimer et al.; Pawl Egozcue, 2001)
- orthonormal basis and coordinates

$$\xi = \Psi \cdot \ln \mathbf{x} \quad \Longleftrightarrow \quad \mathbf{x} = \mathcal{C}\left[\right.$$

$\mathcal{S}^3$



geometry (I)

the *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
- the sample space of compositions, the D-part simplex ($\mathcal{S}^D$) is an Euclidean space (Billheimer et al.; Pawlowsky-Glahn and Egozcue, 2001)
- orthonormal basis and coordinates

$$\xi = \mathbf{\Psi} \cdot \ln \mathbf{x} \quad \Longleftrightarrow \quad \mathbf{x} = \mathcal{C}\left[\exp\left(\left(\mathbf{\Psi}^t \cdot \xi\right)\right)\right]$$

geometry (I)

the *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 inf
- the sample space of compositions, tl Euclidean space (Billheimer et al.; P Egozcue, 2001)
- orthonormal basis and coordinates

ilr coordinate matrix

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

$$\xi = \mathbf{\Psi} \cdot \ln \mathbf{x} \iff \mathbf{x} = \mathcal{C}\left[\exp\left(\left(\mathbf{\Psi}^t \cdot \xi\right)\right)\right]$$

geometry (II)

relevance for a probability vector parameter of a Multinomial variable

- closure: likelihood vectors $\equiv$ probability vectors
- $\oplus$: discrete Bayes Theorem
- $||\cdot||_A$: information measure
- ξ are log-contrasts; alr are log-odds (logistic regression)

statistics

working on coordinates (Pawlowsky-Glahn, 2003)

1. choose an orthonormal basis, compute coordinates (ilr)
2. statistics with the coordinates: e.g. mean $\boldsymbol{\mu} : (D-1)$-real vector, variance $\boldsymbol{\Sigma} : (D-1, D-1)$-SPD matrix
3. apply results to the basis, if useful: e.g. mean becomes $\mathcal{C}\left[\exp\left(\boldsymbol{\Psi}^t \cdot \boldsymbol{\mu}\right)\right] = \mathbf{m} \in \mathcal{S}^D$
   - positive, summing up to one

- do results depend on the basis? NO
- Eaton (1983) reasons: "expectation is defined for real variables" + "orthonormal projection is real" $\Longrightarrow \mathrm{E}_{\mathcal{S}}[\mathbf{Z}] = \mathbf{m}$

statistics

working on coordinates (Pawlowsky-Glahn, 2003)

1. choose an orthonormal basis, compute coordinates (ilr)
2. statistics with the coordinates: e.g. mean $\boldsymbol{\mu}$: $(D-1)$-real vector, variance $\boldsymbol{\Sigma}$: $(D-1, D-1)$-SPD matrix
3. apply results to the basis, if useful: e.g. mean becomes $\mathcal{C}\left[\exp\left(\left(\boldsymbol{\Psi}^t \cdot \boldsymbol{\mu}\right)\right]\right. = \mathbf{m} \in \mathcal{S}^D$
   - positive, summing up to one

- do results depend on the basis? NO
- Eaton (1983) reasons: “expectation is defined for real variables” + “orthonormal projection is real”$\Longrightarrow \mathrm{E}_{\mathcal{S}}[\mathbf{Z}] = \mathbf{m}$

geostatistics

geostatistics for vectors

- alr ⇒ analyse ⇒ back-trasform (Pawlowsky-Glahn and Olea, 2004)
- compute coordinates ⇒ analyse ⇒ 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 $d_A(\hat{\mathbf{z}}_0, \mathbf{Z}_0)$
- results DO NOT depend on the basis

Proposition: kriging transformed 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 $\hat{\mathbf{z}}_0 = \mathbf{T} \cdot \hat{\mathbf{y}}_0$.
- logical, linear operators commute; Myers (1982-84, *Math. Geol.*)

geostatistics

geostatistics for vectors

- 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 $d_A(\hat{\mathbf{z}}_0, \mathbf{Z}_0)$
- results DO NOT depend on the basis

Proposition: kriging transformed 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 $\hat{\mathbf{z}}_0 = \mathbf{T} \cdot \hat{\mathbf{y}}_0$.
- logical, linear operators commute; Myers (1982-84, *Math. Geol.*)

geostatistics

geostatistics for vectors

- 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 $d_A(\hat{\mathbf{z}}_0, \mathbf{Z}_0)$
- results DO NOT depend on the basis

Proposition: kriging transformed 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 $\hat{\mathbf{z}}_0 = \mathbf{T} \cdot \hat{\mathbf{y}}_0$.
- logical, linear operators commute; Myers (1982-84, *Math. Geol.*)

data set

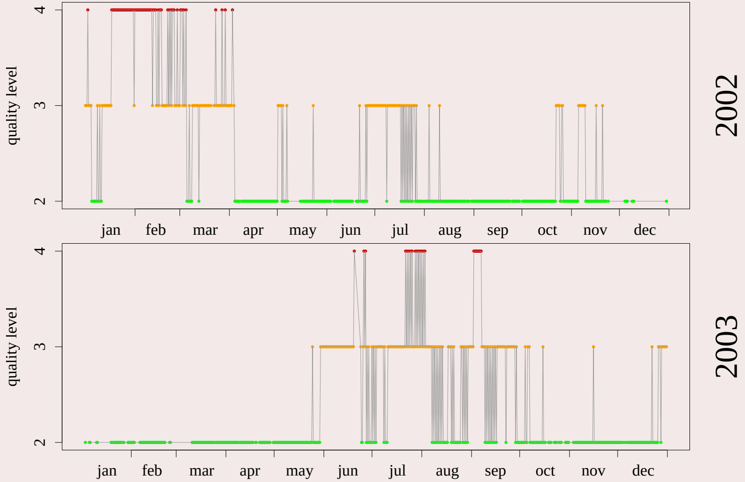
obtaining the data set of water quality categories

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

4. final quality category: the worse

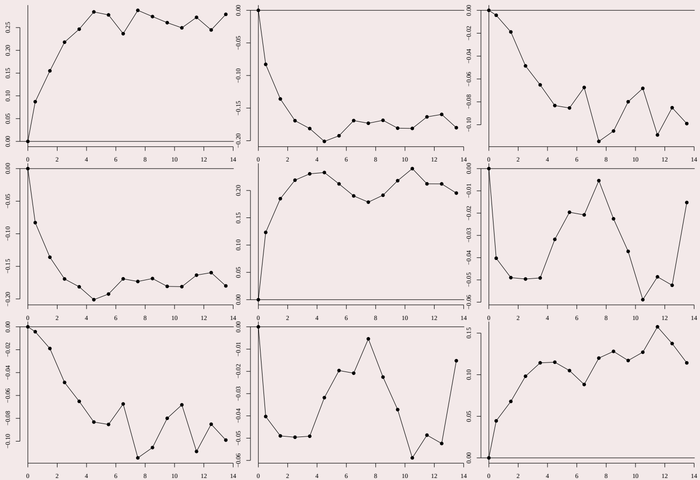
data set

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



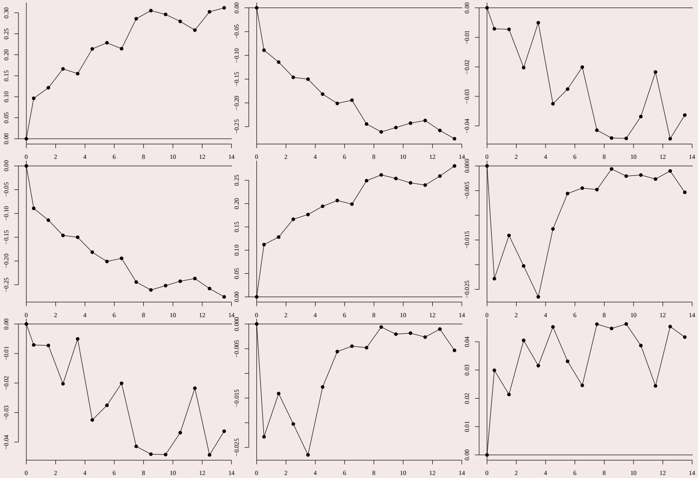
1. preliminary variography of disjunctive indicators

variogram system, year 2002



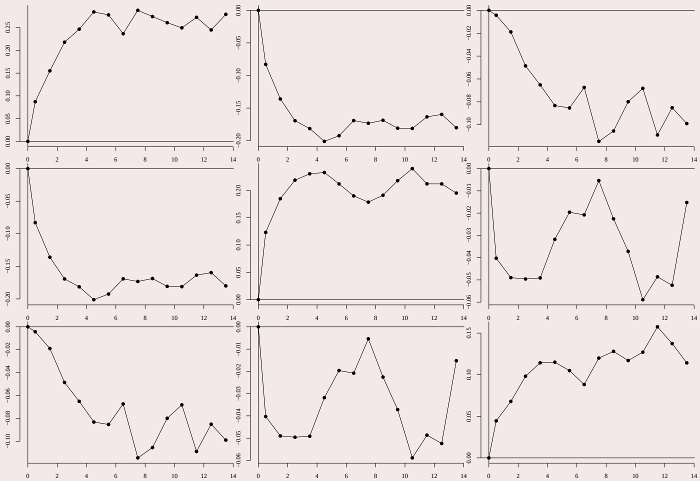
1. preliminary variography of disjunctive indicators

variogram system, year 2003



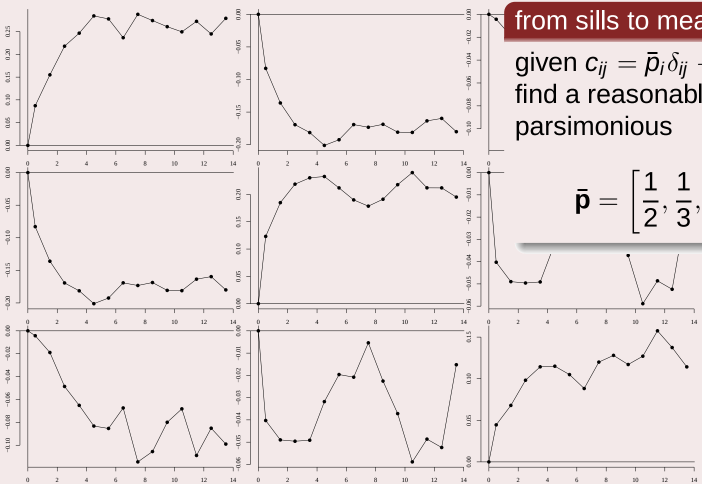
1. preliminary variography of disjunctive indicators

variogram system, year 2002



1. preliminary variography of disjunctive indicators

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

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

2. estimate parameter vectors of sampled locations

a simple estimation

- $\mathbf{J}_n$ are “observed”, but $\mathbf{p}_n$ must be estimated
- for instance:

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

- an alternative expression:

$$\hat{p}_i = \begin{cases} 1 - \alpha & J_i = 1, \\ \alpha/(D-1) & J_i = 0, \end{cases}$$

where $\alpha (= 0.05)$ probability of missclassification

- purely multinomial, no proximity effects between classes

3. computation of coordinates

relationships between coordinates and disjunctive indicators

- coordinates of a vector of probabilities

$$\boldsymbol{\pi} = \boldsymbol{\Psi} \cdot \ln \mathbf{p}, \qquad \boldsymbol{\Psi} = \begin{pmatrix} \frac{+2}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} \\ 0 & \frac{+1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$$

- coordinates of an estimated vector of probabilities (in general)

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

- coordinates of an estimated vector of probabilities (simplified)

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

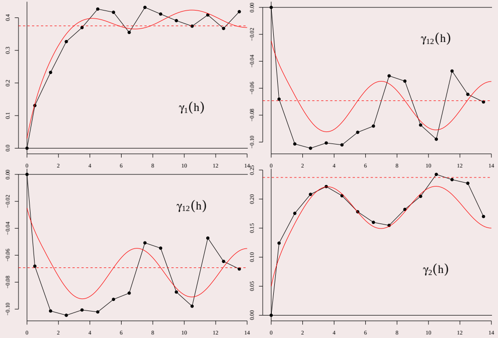
4. apply conventional geostatistics (I)

variography

- variograms for indicators and coordinates consistent: $\boldsymbol{\Gamma}^{\pi}(h) = \beta^2 \cdot \boldsymbol{\Psi} \cdot \boldsymbol{\Gamma}^{J}(h) \cdot \boldsymbol{\Psi}^t$
- 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)

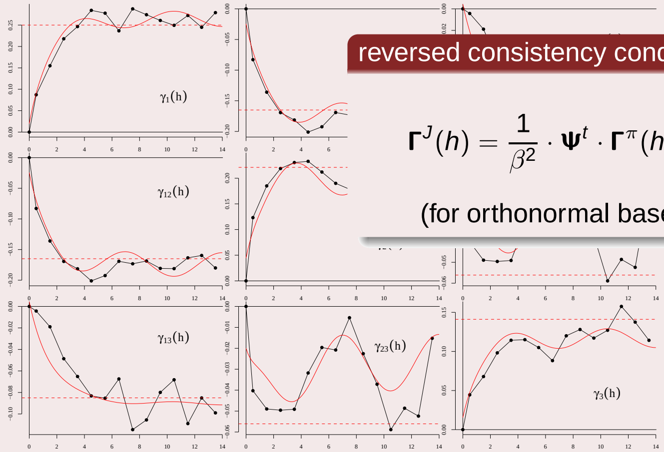
coordinate variograms (2002)



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

4. apply conventional geostatistics (II)

checking what happened with indicator variograms (2002)



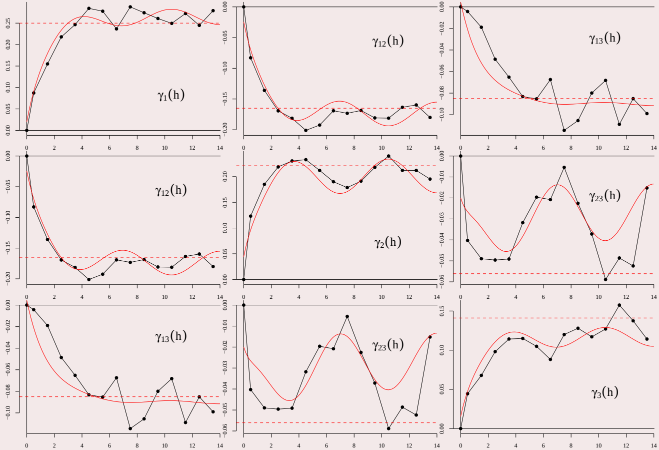
reversed consistency conditions

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

(for orthonormal bases!)

4. apply conventional geostatistics (II)

checking what happened with indicator variograms (2002)



4. apply conventional geostatistics (III)

kriging coordinates or kriging indicators?

- recall: $\hat{\pi}_0 = \beta \cdot \mathbf{\Psi} \cdot \mathbf{J}_0$, $\longrightarrow$ invertible! $\mathbf{J}_0 = \frac{1}{\beta} \cdot \mathbf{\Psi}^t \cdot \hat{\pi}_0 + \frac{1}{D}\mathbf{1}$
- proposition ensures that results for $\hat{\pi}_0$ are equivalent:
  - cokriging indicators ($\hat{\mathbf{j}}_0$) and transforming them
  - cokriging coordinates directly ($\hat{\pi}_0$)
- 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

4. apply conventional geostatistics (III)

kriging coordinates or kriging indicators?

- recall: $\hat{\pi}_0 = \beta \cdot \mathbf{\Psi} \cdot \mathbf{J}_0$, $\longrightarrow$ invertible! $\mathbf{J}_0 = \frac{1}{\beta} \cdot \mathbf{\Psi}^t \cdot \hat{\pi}_0 + \frac{1}{D}\mathbf{1}$
- proposition ensures that results for $\hat{\pi}_0$ are equivalent:
  - cokriging indicators ($\hat{\mathbf{j}}_0$) and transforming them
  - cokriging coordinates directly ($\hat{\pi}_0$)
- 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 $\mathbf{\Psi} \implies$ 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

4. apply conventional geostatistics (III)

kriging coordinates or kriging indicators?

- recall: $\hat{\pi}_0 = \beta \cdot \mathbf{\Psi} \cdot \mathbf{J}_0$, $\longrightarrow$ invertible! $\mathbf{J}_0 = \frac{1}{\beta} \cdot \mathbf{\Psi}^t \cdot \hat{\pi}_0 + \frac{1}{D}\mathbf{1}$
- proposition ensures that results for $\hat{\pi}_0$ are equivalent:
  - cokriging indicators ($\hat{\mathbf{j}}_0$) and transforming them
  - cokriging coordinates directly ($\hat{\pi}_0$)
- 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

4. apply conventional geostatistics (III)

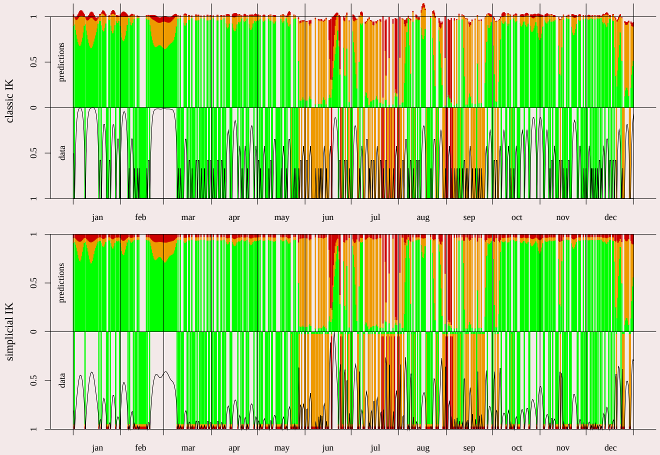
kriging coordinates or kriging indicators?

- recall: $\hat{\pi}_0 = \beta \cdot \boldsymbol{\Psi} \cdot \mathbf{J}_0, \longrightarrow$ invertible! $\mathbf{J}_0 = \frac{1}{\beta} \cdot \boldsymbol{\Psi}^t \cdot \hat{\pi}_0 + \frac{1}{D}\mathbf{1}$
- proposition ensures that results for $\hat{\pi}_0$ are equivalent:
  - cokriging indicators ($\hat{\mathbf{j}}_0$) and transforming them
  - cokriging coordinates directly ($\hat{\pi}_0$)
- if we apply kriged results to the basis used:
  $\hat{\mathbf{p}}_0 = \mathcal{C}\left(\exp\left(\boldsymbol{\Psi}^t \cdot \hat{\pi}_0\right)\right) = \mathcal{C}\left(\exp\left(\beta \cdot \hat{\mathbf{j}}_0\right)\right)$
  - always valid: positive, summing up to one
  - no $\boldsymbol{\Psi} \Longrightarrow$ 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

5. extract probabilities for unsampled locations

results 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]



conclusions

simplicial Indicator Kriging (sIK)

- 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)
- geostatistical procedure: not dependent on the preliminary **p** estimation (β, α, matrix **A**)
- final cokriging results: not dependent on the basis chosen

generalization: geostatistics for vector observations

- IF sample space has an Euclidean structure, compatible with data scale, THEN (geo)statistics can/should be applied to the coordinates
- results honour the space conditions (bounds: positive components, constant sum) and are BLUE with respect to the data scale (additive, multiplicative, etc.)

conclusions

simplicial Indicator Kriging (sIK)

- 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)
- geostatistical procedure: not dependent on the preliminary **p** estimation (β, α, matrix **A**)
- final cokriging results: not dependent on the basis chosen

generalization: geostatistics for vector observations

- IF sample space has an Euclidean structure, compatible with data scale, THEN (geo)statistics can/should be applied to the coordinates
- results honour the space conditions (bounds: positive components, constant sum) and are BLUE with respect to the data scale (additive, multiplicative, etc.)

more material

further reading

- Pawlowsky-Glahn, V., 2003. Statistical modelling on coordinates, in: *Compositional Data Analysis Workshop – CoDaWork'03, Proceedings*
- Tolosana-Delgado, R., 2006. *Geostatistics for constrained variables: positive data, compositions and probabilities. Application to environmental hazard monitoring*. Ph.D. thesis (U. Girona, Spain)
- Tolosana-Delgado, R., Pawlowsky-Glahn, V., Egozcue, J. J. Indicator kriging without order relation violations. *Mathematical Geology*
- 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 Compositional Data Analysis, Girona (Spain), May 27 to 30, 2008.

Revisiting cokriging of indicator functions and compositions

Andrei Borisovitch Vistelius Award

Annual Conference of the International Association for Mathematical Geology

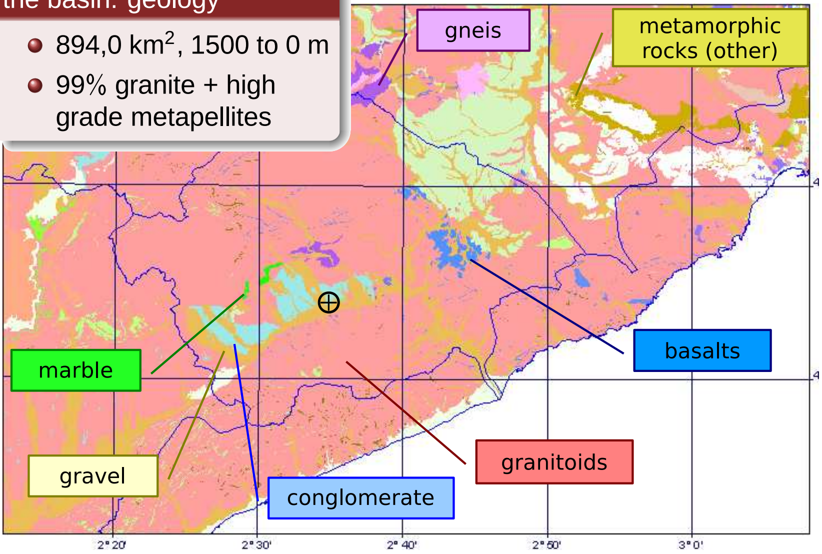
many thanks to:

- all of you, for your attention
- Vera Pawlowsky-Glahn, Juan José Egozcue, Gerald van den Boogaart, the whole Girona Group on CoDa analysis, and Hilmar von Eynatten, for their assistance, help, ideas
- the IAMG, for the award, and a 2004 student grant
- the organizers of the conference.

water quality assessment: a particular case

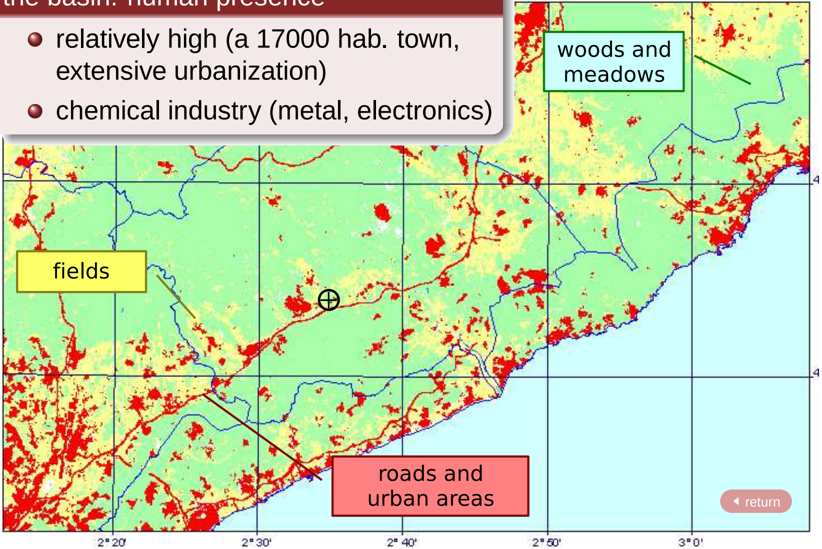
the basin: geology

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



the basin: human presence

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



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

entropy vs. Aitchison norm

- Aitchison norm

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

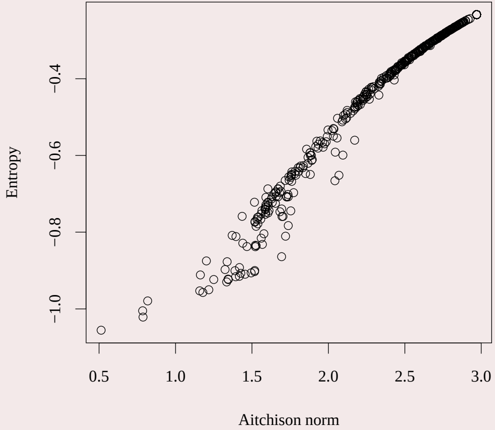
- Shannon entropy

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

return

measures of information in a probability vector

comparison



entrop

- A

- S

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

og p_3

return