Doing statistics of (grain size) distributions

(or why doing it easy if it can be infinitely complicated?)

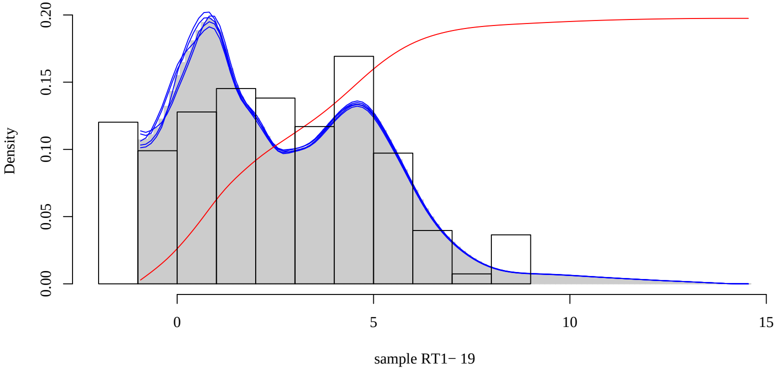
Raimon Tolosana-Delgado

Sedi-Seminar
6 November 2007

outline

1. grain-size distribution characterization
2. geometry and statistics of distributions
   - characteristics
   - CoDa analysis
   - Hilbert space
3. application
   - generalities
   - a glacial data set from the Aar-Gotthard massif (Alps)
   - the Darss sill data set
4. final comments

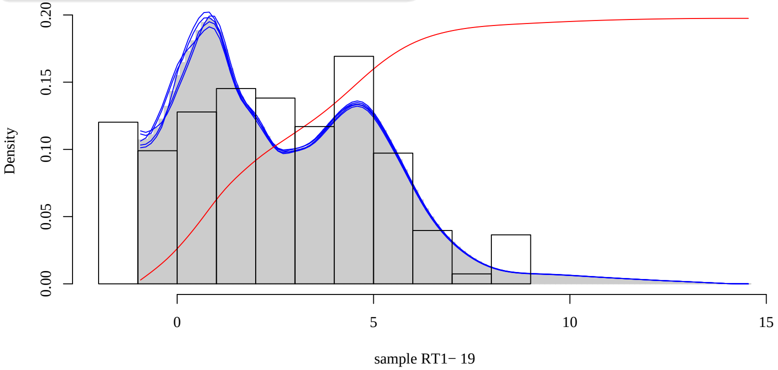
a grain-size distribution example



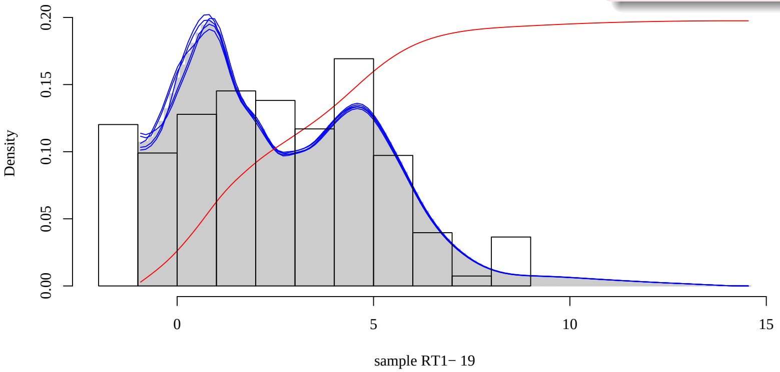
a grain-size distribution example

two measurement types

- quasi continuous LPS (x 5)
- binned sieve data



a grain-size distribution example



distribution!

- increasing
- bounded in [0,1]

traditional procedure (Folk, 1979; *Petrology of Sedimentary Rocks*)

1. compute the cummulative curve (distribution)
2. read some quantiles form the line, and derive distribution measures:

- central tendency
  - mode: most frequent ϕ
  - median: ϕ_{50}
  - $\frac{1}{2}(\phi_{16} + \phi_{84})$
  - combined $\frac{1}{3}(\phi_{16} + \phi_{84} + \phi_{50})$
- sorting
  - interquartile range: $\frac{1}{2}(\phi_{75} - \phi_{25})$ (bad)
  - $\frac{1}{2}(\phi_{84} - \phi_{16})$
  - combined $\frac{1}{4}(\phi_{84} - \phi_{16}) + \frac{1}{6.6}(\phi_{95} - \phi_{5})$
- skewness
- kurtosis

Argument: *statistics require many calculations* !

- computers are here to do some dirty work

what *must* have a curve to be a density?

- it must be positive
- its integral (area under the curve) must be one

what *must* have a curve to be a density?

- it must be positive
- its integral (area under the curve) must be one

what *can* have/be a distribution?

- several modes vs. one single maximum
- continuous vs. discontinuous
- smooth curve vs. fractal character (extreme irregularity)
- symmetry vs. skewness
- virtually any shape (not just a normal one)

what *must* have a curve to be a density?

- it must be positive
- its integral (area under the curve) must be one

what *can* have/be a distribution?

- several modes vs. one single maximum
- continuous vs. discontinuous
- smooth curve vs. fractal character (extreme irregularity)
- symmetry vs. skewness
- virtually any shape (not just a normal one)

compositional data: a brief summary (I)

composition: definitions

- a composition is . . . a vector $\mathbf{x} = [x_1, \ldots x_D]$ with
  - D positive components ($x_i \geq 0$)
  - of constant sum ($\sum x_i = 100$)

compositional data: a brief summary (I)

composition: definitions

- a composition is ... a vector $\mathbf{x} = [x_1, \ldots x_D]$ with
  - D positive components ($x_i \geq 0$)
  - of constant sum ($\sum x_i = 100$)
- its sample space is:
  - the set of *possible* measurements
  - the simplex ($\mathbb{S}^D$), *e.g.* if $D = 3$ the ternary diagram

compositional data: a brief summary (I)

composition: definitions

- a composition is ... a vector $\mathbf{x} = [x_1, \ldots x_D]$ with
  - D positive components ($x_i \geq 0$)
  - of constant sum ($\sum x_i = 100$)
- its sample space is:
  - the set of *possible* measurements
  - the simplex ($\mathbb{S}^D$), *e.g.* if $D = 3$ the ternary diagram
- its scale is:
  - a *subjective* assessment of the difference between two values
  - a *relative* one, an increment from 100 ppm to 200 ppm is more significant that from 1000 ppm to 1100 ppm

compositional data: a brief summary (I)

composition: definitions

- a composition is . . . a vector $\mathbf{x} = [x_1, \ldots x_D]$ with
  - D positive components ($x_i \geq 0$)
  - of constant sum ($\sum x_i = 100$)
- its sample space is:
  - the set of *possible* measurements
  - the simplex ($\mathbb{S}^D$), *e.g.* if $D = 3$ the ternary diagram
- its scale is:
  - a *subjective* assessment of the difference between two values
  - a *relative* one, an increment from 100 ppm to 200 ppm is more significant that from 1000 ppm to 1100 ppm

compositional data: a brief summary (II)

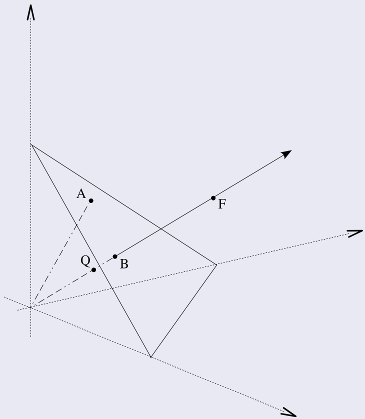
the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}\left[\cdot\right] : \mathbb{R}^D_+ \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$

compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}[\cdot] : \mathbb{R}_+^D \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$



compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}\left[\cdot\right] : \mathbb{R}_+^D \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\left[x_1 \cdot y_1, \ldots, x_D \cdot y_D\right]$

compositional data: a brief summary (II)

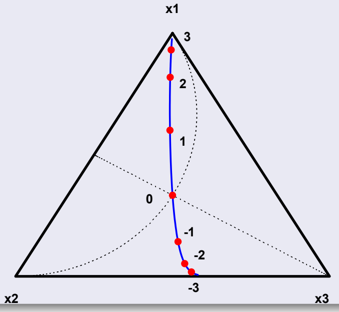
the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}\left[\cdot\right] : \mathbb{R}_+^D \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\left[x_1 \cdot y_1, \ldots, x_D \cdot y_D\right]$
  - "zero", neutral element $\mathbf{n} = [1/D, \ldots 1/D]$
  - complement operation: power transformation $\lambda \odot \mathbf{x} = \mathcal{C}\left[x_1^\lambda, \ldots, x_D^\lambda\right]$

compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}\left[\cdot\right] : \mathbb{R}_+^D \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\left[x_1 \cdot y_1, \ldots, x_D \cdot y_D\right]$
  - "zero", neutral element $\mathbf{n} = [1/D, \ldots 1/D]$
  - complement operation: power transformation $\lambda \odot \mathbf{x} = \mathcal{C}\left[x_1^\lambda, \ldots, x_D^\lambda\right]$



compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}\left[\cdot\right] : \mathbb{R}_{+}^{D} \longrightarrow \mathbb{S}^{D} : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\left[x_1 \cdot y_1, \ldots, x_D \cdot y_D\right]$
  - “zero”, neutral element $\mathbf{n} = [1/D, \ldots 1/D]$
  - complement operation: power transformation $\lambda \odot \mathbf{x} = \mathcal{C}\left[x_1^{\lambda}, \ldots, x_D^{\lambda}\right]$
- Aitchison distance: $d_A^2(\mathbf{x}, \mathbf{y}) = \sum_{i>j} \left(\ln \frac{x_i}{y_i} - \ln \frac{x_j}{y_j}\right)^2$

compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

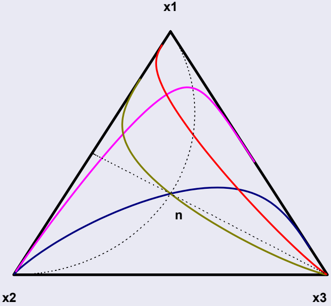
- the closure: $\mathcal{C}\left[\cdot\right] : \mathbb{R}_{+}^{D} \longrightarrow \mathbb{S}^{D} : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\left[x_1 \cdot y_1, \ldots, x_D \cdot y_D\right]$
  - "zero", neutral element $\mathbf{n} = [1/D, \ldots 1/D]$
  - complement operation: power transformation $\lambda \odot \mathbf{x} = \mathcal{C}\left[x_1^{\lambda}, \ldots, x_D^{\lambda}\right]$
- Aitchison distance: $d_A^2(\mathbf{x}, \mathbf{y}) = \sum_{i>j} \left(\ln \frac{x_i}{y_i} - \ln \frac{x_j}{y_j} \right)^2$
  - Aitchison scalar product $\langle \mathbf{x}, \mathbf{y} \rangle_A = \sum_{i>j} \ln \frac{x_i}{y_i} \cdot \ln \frac{x_j}{y_j}$
  - Aitchison norm $||\mathbf{x}||_A = d(\mathbf{n}, \mathbf{x})$

compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}[\cdot] : \mathbb{R}_+^D \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}[x_1 \cdot y_1, \ldots, x_D \cdot y_D]$
  - "zero
  - com on $\lambda \odot \mathbf{x} = \mathcal{C}[x_1^\lambda, \ldots, x_D^\lambda]$
- Aitchison $- \ln \frac{x_j}{y_j} \Big)^2$
  - Aitc $\ln \frac{x_i}{y_i} \cdot \ln \frac{x_j}{y_j}$
  - Aitc

parallel and orthogonal lines

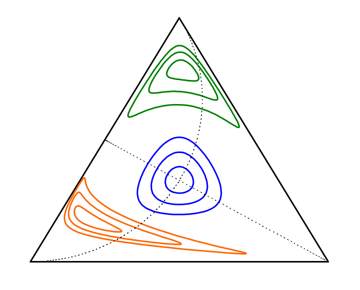


compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}[\cdot] : \mathbb{R}_+^D \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}[x_1 \cdot y_1, \ldots, x_D \cdot y_D]$
  - "zerc
  - comp on $\lambda \odot \mathbf{x} = \mathcal{C}[x_1^\lambda, \ldots, x_D^\lambda]$
- Aitchison $- \ln \frac{x_j}{y_j} \Big)^2$
  - Aitcl $\ln \frac{x_i}{y_i} \cdot \ln \frac{x_j}{y_j}$
  - Aitcl

ellipses and circles



compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

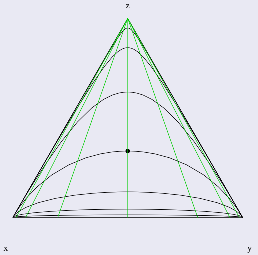
- the closure: $\mathcal{C}\left[\cdot\right] : \mathbb{R}_+^D \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\left[x_1 \cdot y_1, \ldots, x_D \cdot y_D\right]$
    - "zero", neutral element $\mathbf{n} = [1/D, \ldots 1/D]$
    - complement operation: power transformation $\lambda \odot \mathbf{x} = \mathcal{C}\left[x_1^\lambda, \ldots, x_D^\lambda\right]$
- Aitchison distance: $d_A^2(\mathbf{x}, \mathbf{y}) = \sum_{i>j} \left(\ln \frac{x_i}{y_i} - \ln \frac{x_j}{y_j}\right)^2$
    - Aitchison scalar product $\langle \mathbf{x}, \mathbf{y} \rangle_A = \sum_{i>j} \ln \frac{x_i}{y_i} \cdot \ln \frac{x_j}{y_j}$
    - Aitchison norm $||\mathbf{x}||_A = d(\mathbf{n}, \mathbf{x})$
- basis: a set of $D-1$ compositions (*orthonormal*) $\mathbf{e}_1, \ldots, \mathbf{e}_{D-1} \in \mathbb{S}^D$

compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}[\cdot] : \mathbb{R}^D_+ \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}[x_1 \cdot y_1, \ldots, x_D \cdot y_D]$
  - "zero
  - comp ... $, x_D^\lambda]$
- Aitchison
  - Aitch
  - Aitch
- basis: a s ... $\in \mathbb{S}^D$

2 systems of axis



compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}[\cdot] : \mathbb{R}^D_+ \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}[x_1 \cdot y_1, \ldots, x_D \cdot y_D]$
  - "zero", neutral element $\mathbf{n} = [1/D, \ldots 1/D]$
  - complement operation: power transformation $\lambda \odot \mathbf{x} = \mathcal{C}[x_1^\lambda, \ldots, x_D^\lambda]$
- Aitchison distance: $d_A^2(\mathbf{x}, \mathbf{y}) = \sum_{i>j} \left(\ln \frac{x_i}{y_i} - \ln \frac{x_j}{y_j} \right)^2$
  - Aitchison scalar product $\langle \mathbf{x}, \mathbf{y} \rangle_A = \sum_{i>j} \ln \frac{x_i}{y_i} \cdot \ln \frac{x_j}{y_j}$
  - Aitchison norm $||\mathbf{x}||_A = d(\mathbf{n}, \mathbf{x})$
- basis: a set of $D-1$ compositions (*orthonormal*) $\mathbf{e}_1, \ldots, \mathbf{e}_{D-1} \in \mathbb{S}^D$
- coordinates: log-ratios (relative scale) $\xi_1, \ldots \xi_{D-1} \in \mathbb{R}$ such that $\mathbf{x} = \xi_1 \odot \mathbf{e}_1 \oplus \xi_2 \odot \mathbf{e}_2 \oplus \cdots \oplus \xi_{D-1} \odot \mathbf{e}_{D-1}$

compositional data: a brief summary (II)

the Simplex: a meaningful *Euclidean* space structure

- the closure: $\mathcal{C}\left[\cdot\right] : \mathbb{R}_+^D \longrightarrow \mathbb{S}^D : \mathbf{x}' = \frac{\mathbf{x}}{\sum x_i}$
- change operation: perturbation $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}\left[x_1 \cdot y_1, \ldots, x_D \cdot y_D\right]$
  - "zero", neutral element $\mathbf{n} = [1/D, \ldots 1/D]$
  - complement operation: power transformation $\lambda \odot \mathbf{x} = \mathcal{C}\left[x_1^\lambda, \ldots, x_D^\lambda\right]$
- Aitchison distance: $d_A^2(\mathbf{x}, \mathbf{y}) = \sum_{i>j} \left(\ln \frac{x_i}{y_i} - \ln \frac{x_j}{y_j}\right)^2$
  - Aitchison scalar product $\langle \mathbf{x}, \mathbf{y} \rangle_A = \langle \mathrm{clr}(\mathbf{x}), \mathrm{clr}(\mathbf{y}) \rangle_{euc}$
  - Aitchison norm $||\mathbf{x}||_A = d(\mathbf{n}, \mathbf{x})$
- basis: a set of $D-1$ compositions (*orthonormal*) $\mathbf{e}_1, \ldots, \mathbf{e}_{D-1} \in \mathbb{S}^D$
- coordinates: log-ratios (relative scale) $\xi_1, \ldots \xi_{D-1} \in \mathbb{R}$ such that $\mathbf{x} = \xi_1 \odot \mathbf{e}_1 \oplus \xi_2 \odot \mathbf{e}_2 \oplus \cdots \oplus \xi_{D-1} \odot \mathbf{e}_{D-1}$
  - centered log-ratio transformation:
    $\mathrm{clr}(\mathbf{x}) = \ln \frac{\mathbf{x}}{\sqrt[D]{\prod_i x_i}} = \ln(\mathbf{x}) - \frac{1}{D} \sum_i \ln x_i$

from compositions to distributions

distributions: *compositions of infinite parts*

- a distribution is ... a curve $f(x)$ with
  - positive value $f(x) > 0$ for all x

from compositions to distributions

distributions: *compositions of infinite parts*

- a distribution is . . . a curve $f(x)$ with
  - positive value $f(x) > 0$ for all x
  - of sum equals 100% $\Longrightarrow$ integral equal to one $\int f(x)dx = 1$

from compositions to distributions

distributions: *compositions of infinite parts*

- a distribution is ... a curve $f(x)$ with
  - positive value $f(x) > 0$ for all x
  - of integral equal to one $\int f(x)dx = 1$
- its sample space is an infinite-dimensional "simplex", one of the Aitchison spaces of measures (Egozcue and Diaz-Barrero, 2003; van den Boogaart, 2005)

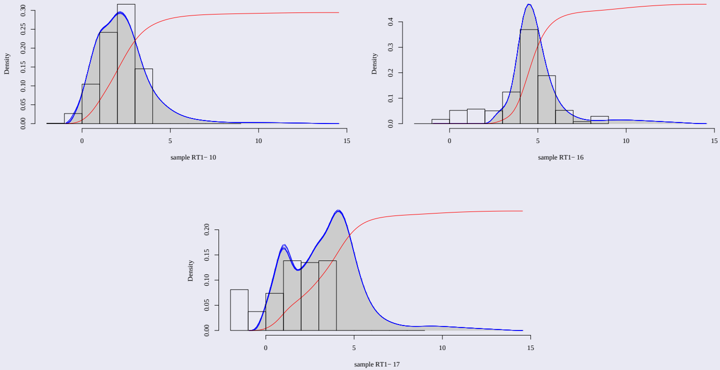
from compositions to distributions

distributions: *compositions of infinite parts*

- a distribution is ... a curve $f(x)$ with
  - positive value $f(x) > 0$ for all x
  - of integral equal to one $\int f(x)dx = 1$
- its sample space is an infinite-dimensional "simplex", one of the Aitchison spaces of measures (Egozcue and Diaz-Barrero, 2003; van den Boogaart, 2005)
- its scale is ?

from compositions to distributions

how do we compare grain-size curves?



from compositions to distributions

the Hilbert space of distributions: operations

- Hilbert space: "Euclidean space of infinite dimension" (but ...)
  - not unique!

from compositions to distributions

the Hilbert space of distributions: operations

- Hilbert space: "Euclidean space of infinite dimension" (but ...)
    - not unique!
- perturbation (translation, sum), powering (multiplication)
    - the neutral element must be chosen: $n(x)$ (normal, exponential, uniform)

from compositions to distributions

the Hilbert space of distributions: operations

- Hilbert space: "Euclidean space of infinite dimension" (but ...)
  - not unique!
- perturbation (translation, sum), powering (multiplication)
  - the neutral element must be chosen: $n(x)$ (normal, exponential, uniform)
- scalar product $\Rightarrow$ angles, distances, lengths

from compositions to distributions

the Hilbert space of distributions: operations

- Hilbert space: "Euclidean space of infinite dimension" (but ...)
  - not unique!
- perturbation (translation, sum), powering (multiplication)
  - the neutral element must be chosen: $n(x)$ (normal, exponential, uniform)
- scalar product ⇒ angles, distances, lengths

$$\langle f(x), g(x) \rangle = \int_{dom(n)} \mathrm{clr}\,(f(x)) \cdot \mathrm{clr}\,(g(x))\, n(x) dx$$

from compositions to distributions

the Hilbert space of distributions: operations

- Hilbert space: "Euclidean space of infinite dimension" (but ...)
  - not unique!
- perturbation (translation, sum), powering (multiplication)
  - the neutral element must be chosen: $n(x)$ (normal, exponential, uniform)
- scalar product ⇒ angles, distances, lengths

$$\langle f(x), g(x) \rangle = \int_{dom(n)} \operatorname{clr}(f(x)) \cdot \operatorname{clr}(g(x))\, n(x) dx$$

- clr transformation

$$\operatorname{clr}(f(x)) = \ln \frac{f(x)}{n(x)} - \int_{dom(n)} \ln \frac{f(x)}{n(x)} n(x) dx$$

from compositions to distributions

the Hilbert space of distributions: basis and coordinates

- infinite dimension $\Rightarrow$ infinite coordinates, infinite elements in a basis
- orthonormal:

$$\int_{d(n)} f^*(x) \cdot g^*(x) n(x) dx = 0 \quad \text{and} \quad \int_{d(n)} \left(f^*(x)\right)^2 \cdot n(x) dx = 1$$

from compositions to distributions

the Hilbert space of distributions: basis and coordinates

- infinite dimension $\Rightarrow$ infinite coordinates, infinite elements in a basis
- orthonormal:

$$\int_{d(n)} f^*(x) \cdot g^*(x) n(x) dx = 0 \quad \text{and} \quad \int_{d(n)} \left(f^*(x)\right)^2 \cdot n(x) dx = 1$$

- series of polynomials (standard physical-mathematical tool)
  - normal density: Hermite polynomials
  - exponential density: Laguerre polynomials
  - uniform density: Legendre polynomials (beta density: Jacobi polynomials)

from compositions to distributions

the Hilbert space of distributions: basis and coordinates

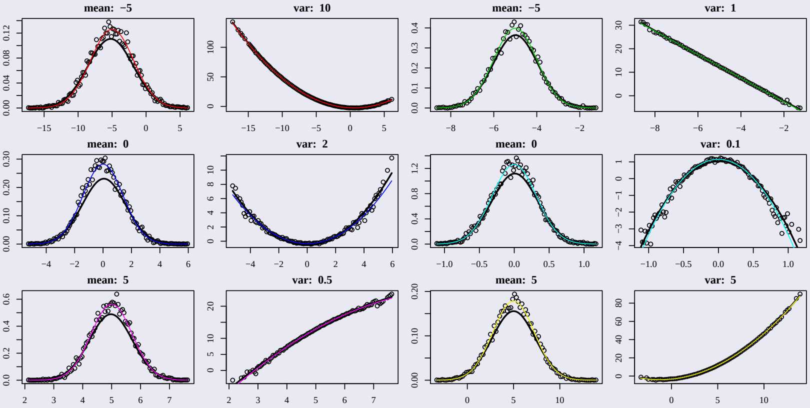
- infinite dimension ⇒ infinite coordinates, infinite elements in a basis
- orthonormal:

$$\int_{d(n)} f^*(x) \cdot g^*(x) n(x) dx = 0 \quad \text{and} \quad \int_{d(n)} \left(f^*(x)\right)^2 \cdot n(x) dx = 1$$

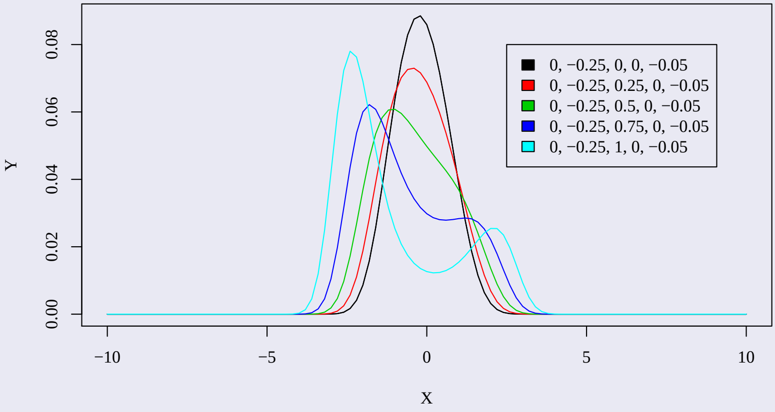
- series of polynomials (standard physical-mathematical tool)
  - normal density: Hermite polynomials
  - exponential density: Laguerre polynomials
  - uniform density: Legendre polynomials (beta density: Jacobi polynomials)
- coordinates
  - Hermite polynomials: $\frac{\mu}{\sigma^2}$, $\frac{\sigma^2-1}{\sqrt{2}\sigma^2}$, etc. (deviation from normality)
  - Laguerre polynomials: $\lambda - 1$

example of Hermite polynomials

- polynomials: $H_0 = 1,\ H_1 = x,\ H_2 = \frac{1}{\sqrt{2}}(1 - x^2)$
- decomposition: $\phi(x;\mu,\sigma) = \phi(x;0,1) \oplus \frac{\mu}{\sigma^2} \odot e^{H_1} \oplus \frac{\sigma^2 - 1}{\sqrt{2\sigma^2}} \odot e^{H_2}$



example of Hermite polynomials



algorithm

1. look at the data
2. choose a reasonable “zero” distribution, reference distribution

algorithm

1. look at the data
2. choose a reasonable "zero" distribution, reference distribution
    - diffusion ⇒ normal distribution
    - comminution ⇒ fractal, exponential distribution
    - bounded phenomena ⇒ uniform distribution

algorithm

1. look at the data
2. choose a reasonable "zero" distribution, reference distribution
   - diffusion $\Rightarrow$ normal distribution
   - comminution $\Rightarrow$ fractal, exponential distribution
   - bounded phenomena $\Rightarrow$ uniform distribution
3. compute the "standardized" density:
   ln(your histogram/zero density)$=Y$
4. fix (which, how many) polynomials to use, orthonormal with respect to the zero density ($\mathbf{X}$)

algorithm

1. look at the data
2. choose a reasonable "zero" distribution, reference distribution
   - diffusion $\Rightarrow$ normal distribution
   - comminution $\Rightarrow$ fractal, exponential distribution
   - bounded phenomena $\Rightarrow$ uniform distribution
3. compute the "standardized" density:
   ln(your histogram/zero density)=Y
4. fix (which, how many) polynomials to use, orthonormal with respect to the zero density ($\mathbf{X}$)
5. $Y = \mathbf{X} \cdot \beta \Longrightarrow$ regression $\Longrightarrow \hat{\beta}$ new "data" (meaningless β_0!)

algorithm

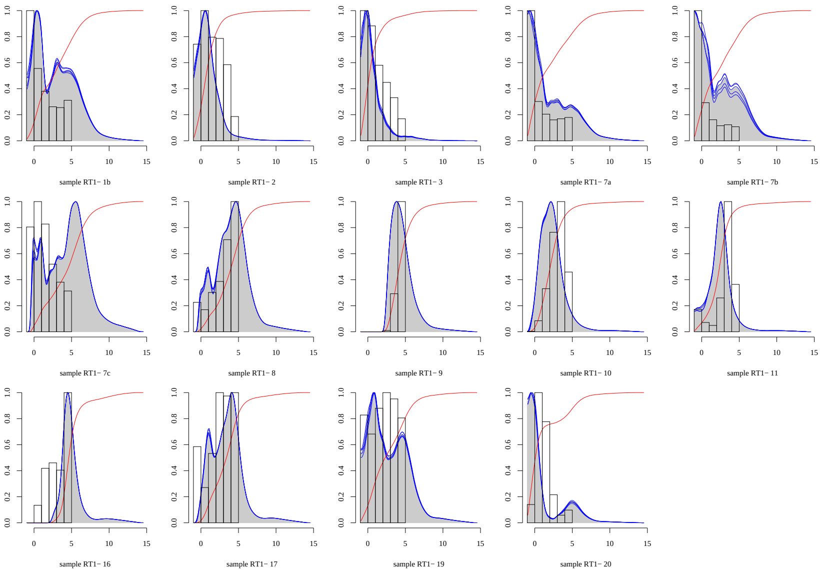
1. look at the data
2. choose a reasonable "zero" distribution, reference distribution
   - diffusion ⇒ normal distribution
   - comminution ⇒ fractal, exponential distribution
   - bounded phenomena ⇒ uniform distribution
3. compute the "standardized" density:
   ln(your histogram/zero density)=Y
4. fix (which, how many) polynomials to use, orthonormal with respect to the zero density ($\mathbf{X}$)
5. $Y = \mathbf{X} \cdot \beta \Longrightarrow$ regression $\Longrightarrow \hat{\beta}$ new "data" (meaningless β_0!)
6. apply the method you want to $\hat{\beta}$
   - principal component analysis, non-supervised classification
   - mapping

a glacial data set from the Aar-Gotthard massif (Alps)

goal and location

- characterize the effect of (weathering and) comminution on grain-size/geochemistry/mineralogy compositions
- focus: comminution $\sim$ glacial comminution (fluvial sorting)
- Aar-Gotthard granitic massif, central Alps (Switzerland); Rhone Gletscher, Damma Gletscher, Tiefer Gletscher
- fluvial sediments, recent central morraines, older lateral morraines
- grain size analyses:
  - laser particle sizer (LPS), 116 classes of $\phi \in [-0.865, 14.61]$
  - sieves, 11 classes from $\phi < -1$ to $\phi > 8$
- analyses of major oxides and trace element geochemistry (not now)

measured densities



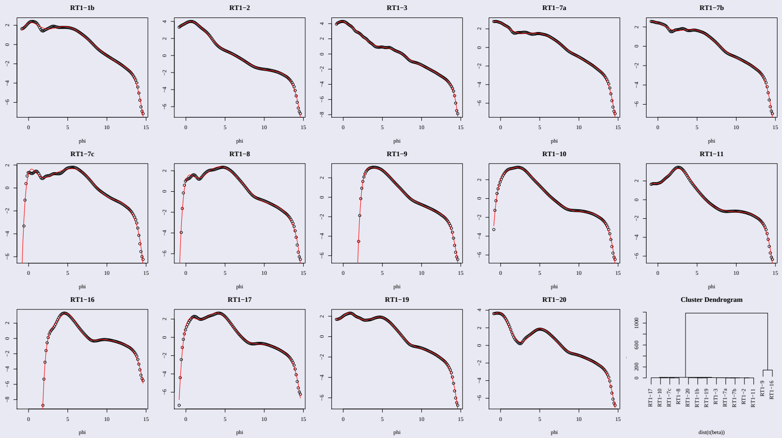
measured densities

summary of characteristics

sample	glacier	placement	coarsest	modes	finest
1b	Rhone	side	yes	2	yes
2	Rhone	side	yes	1	no
3	Rhone	side	yes	1	no
7a	Damma	side (older)	yes	3	yes
7b	Damma	side (older)	yes	3	no
7c	Damma	side (older)	yes	many	yes
8	Damma	side (older)	yes	2	yes
9	Damma	front	no	1	yes
10	Damma	front	yes	1	no
11	Damma	front	yes	1	no
16	Rhone	washed	no	1	yes
17	Rhone	washed	yes	2	yes
19	Tiefer	front	yes	2	yes
20	Tiefer	front	yes	2	no

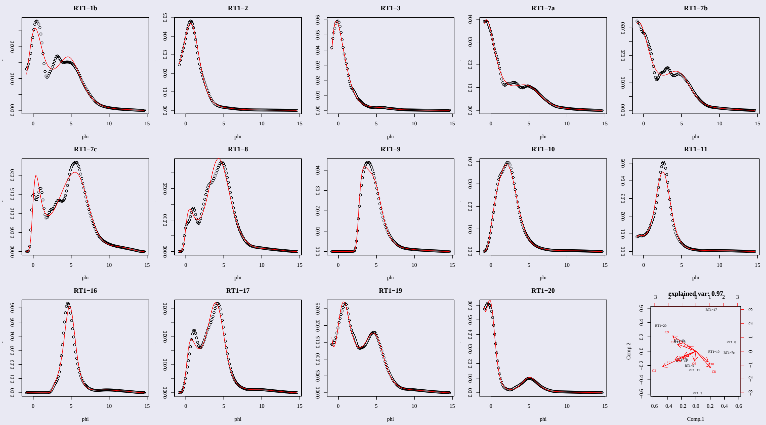
clr-transformed densities

fitted model: 11 Laguerre polynomials (exponential reference)



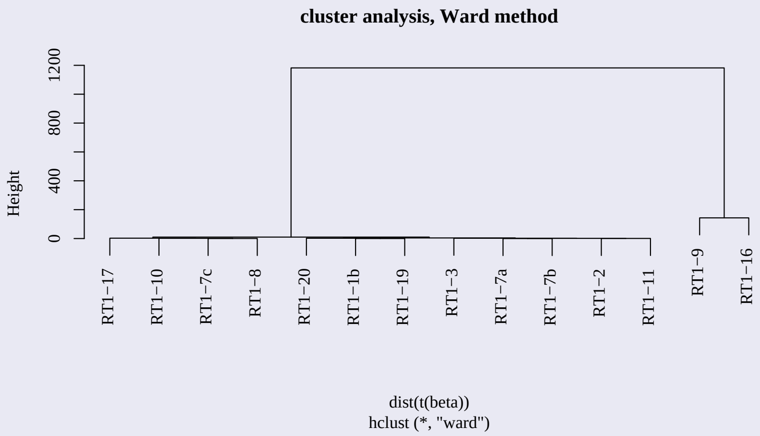
original densities

fitted model: 11 Laguerre polynomials (exponential reference)

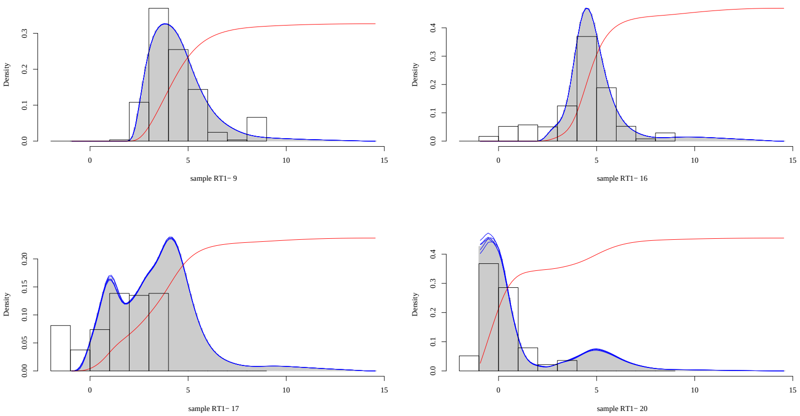


comparing densities: cluster analysis of $\hat{\beta}$'s

non-supervised classification: Ward cluster analysis



comparing densities: 9 and 16 against the other?

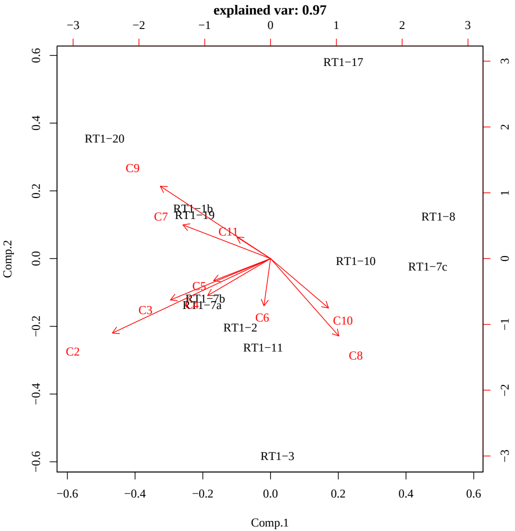


comparing densities: a "map" of differences

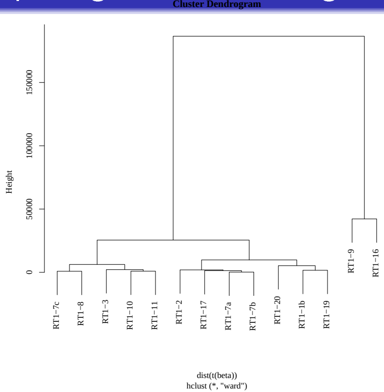
characteristics

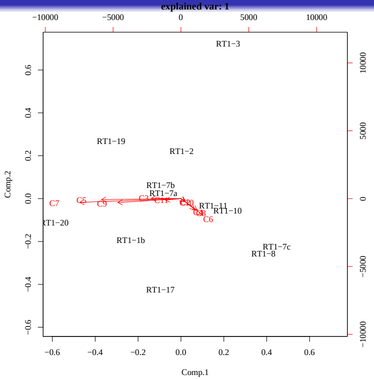
- morraines:
  - old side: 7(a,b,c), 8
  - recent side: 1b, 2, 3
  - front: (9) 10, 11, 19, 20
  - washed sediment: (16) 17
- unimodal: 2, 3, 10, 11
- tail to finest:
  - heavy: 7c, (16), 17, 19
  - yes: 1b, 7a, 8, (9)
  - no: 2, 3, 7b, 10, 11, 20

C2: pure exponential



comparing densities using the normal reference





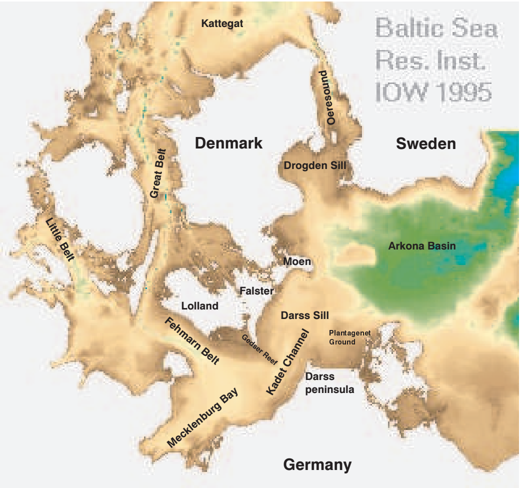
characteristics

old side morraines: 7(a,b,c), 8; recent side: 1b, 2, 3; front: (9) 10, 11, 19, 20; washed sediment: (16) 17

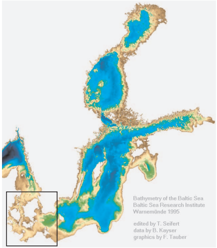
heavy tail to finest: 7c, (16), 17, 19; some fine: 1b, 7a, 8, (9); no fine: 2, 3, 7b, 10, 11, 20

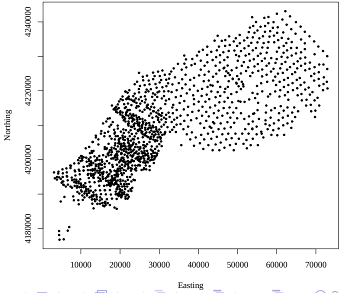
unimodal: 2, 3, 10, 11; C2: mean; C3: inverse variance

the Darss sill data set



Investigation area



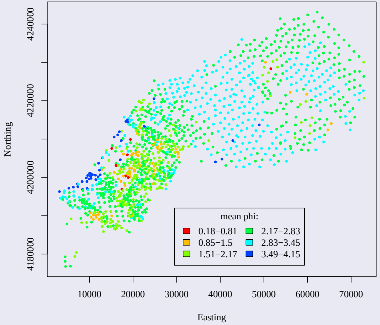


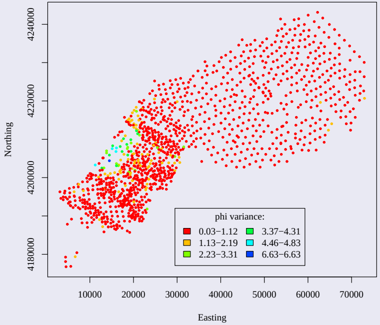
the Darss sill data set

the Darss sill data set

- Baltic Sea (412,560 km^2): world large brackish water bodies
- oceanographic conditions controlled by: (i) large river freshwater input, (ii) restricted ocean connections (Danish Straits)
- $\sim 73\%$ of water exchange is via the Darss Sill (minimum water depth: 18 m bsl)
    - prevailing outflow of brackish Baltic waters in the upper part of the water column usually along the Danish coast
    - inflow of more saline water in the southern part of the area and in the deeper part of the Kadet Channel
    - tidal currents are negligible; sediment-transporting bottom currents are intermittent
    - outcropping till genetically related to an ice marginal zone, short re-advance of the retreating Late Weichselian ice sheet ($\sim$ 13,500 years BP); thin cover of lag sediments (locally, stones and blocks $>$ 1 m)

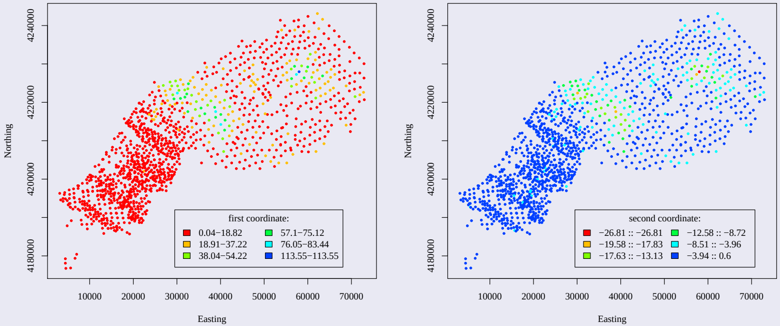
conventional analysis: average (μ) and sorting (σ^2)





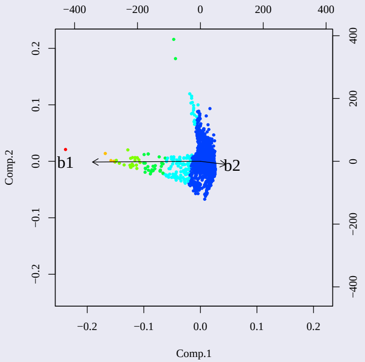
- Kadet Channel, Danish coast ($\uparrow \bar{\phi}; \uparrow \sigma^2_\phi;$)
- Kadet Channel, German coast ($\sim \bar{\phi}; \downarrow \sigma^2_\phi;$)
- Ground, Plantagenet basin? ($\uparrow \bar{\phi}; \downarrow \sigma^2_\phi;$)
- Bar, Darss Sill itself? ($\downarrow \bar{\phi}; \downarrow \sigma^2_\phi;$)

proposed analysis (I): average ($\beta_1 = \frac{\mu}{\sigma^2}$) and sorting ($\beta_2 = \frac{\sigma^2 - 1}{\sqrt{2}\sigma^2}$)



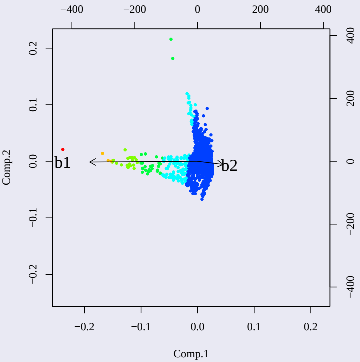
- two areas with extremely low variance ($\downarrow\downarrow \sigma^2_\phi$;)
- undiscriminated ground ($\sim \sigma^2_\phi$;)

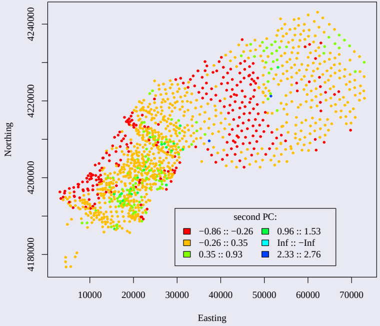
proposed analysis (II): a non-redundant characterization



- β_1 and β_2 almost proportional (inverse)

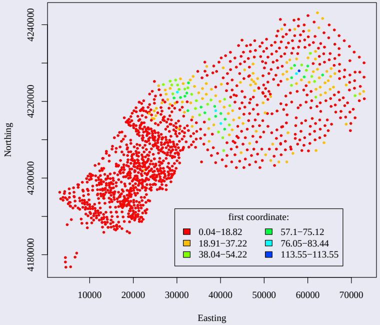
proposed analysis (II): a non-redundant characterization

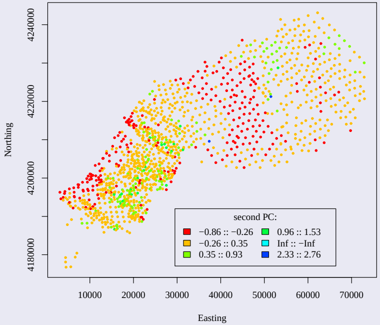




- β_1 and β_2 almost proportional (inverse)
- complementary direction to $\beta_1 || \beta_2$ (non-normality?)

proposed analysis (II): a non-redundant characterization





- Kadet Channel, Danish coast (low departure)
- Kadet Channel, German coast (average departure)
- two singular areas (extreme sorting, average-high departure)
- undiscriminated ground (low departure) $\sim$ Danish coast

final comments

1. grain size curves can be seen as (infinite-dimensional) compositions
    - high-dimensional
    - further structure (ordered bins, smoothness)
    - relative scale

final comments

1. grain size curves can be seen as (infinite-dimensional) compositions
   - high-dimensional
   - further structure (ordered bins, smoothness)
   - relative scale
2. grain size information can be summarized in a few parameters
   - depend on a reference curve
   - orthogonal $\Longrightarrow$ independent
   - increasingly complex

final comments

1. grain size curves can be seen as (infinite-dimensional) compositions
    - high-dimensional
    - further structure (ordered bins, smoothness)
    - relative scale
2. grain size information can be summarized in a few parameters
    - depend on a reference curve
    - orthogonal $\Longrightarrow$ independent
    - increasingly complex
3. these parameters may be statistically treated
    - to uncover groups (cluster)
    - to display patterns of variation (biplot)

final comments

1. grain size curves can be seen as (infinite-dimensional) compositions
   - high-dimensional
   - further structure (ordered bins, smoothness)
   - relative scale
2. grain size information can be summarized in a few parameters
   - depend on a reference curve
   - orthogonal $\Longrightarrow$ independent
   - increasingly complex
3. these parameters may be statistically treated
   - to uncover groups (cluster)
   - to display patterns of variation (biplot)
   - to explain them (regression, anova)
   - to make several measures compatible (weighted average)