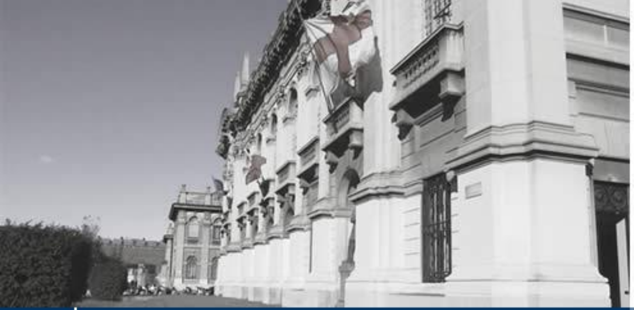


28th November 2014, Bozen-Bolzano

unibz
Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

POLITECNICO DI MILANO

Functional Data Analysis and Cluster Analysis: a Marriage with some Constraints

Simone VANTINI

joint with L.M. SANGALLI, P. SECCHI, V. VITELLI

MOX – Dept. of Mathematics, Politecnico di Milano

A marriage…

Cluster Analysis
(Identification of groups)

+

Functional Data Analysis
(Analysis of functions)

Functional Data Clustering
(Identification of groups of functions)

didn't

They lived happily ever after…

… with some constraints

There are indeed **some serious issues** going on in this marriage:

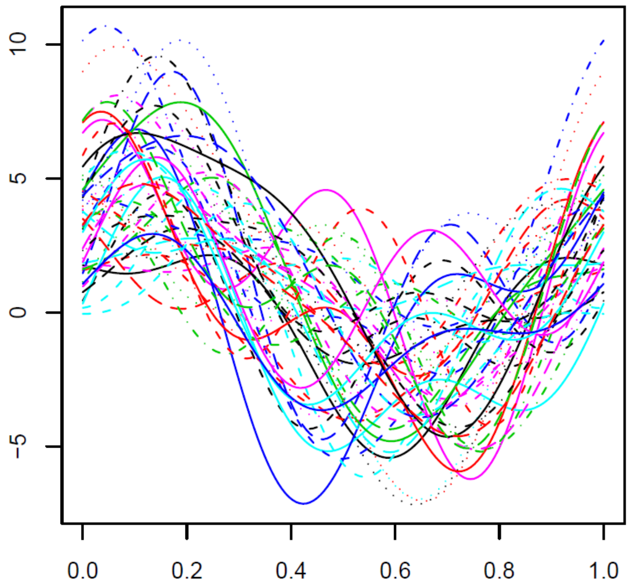
All Connected!

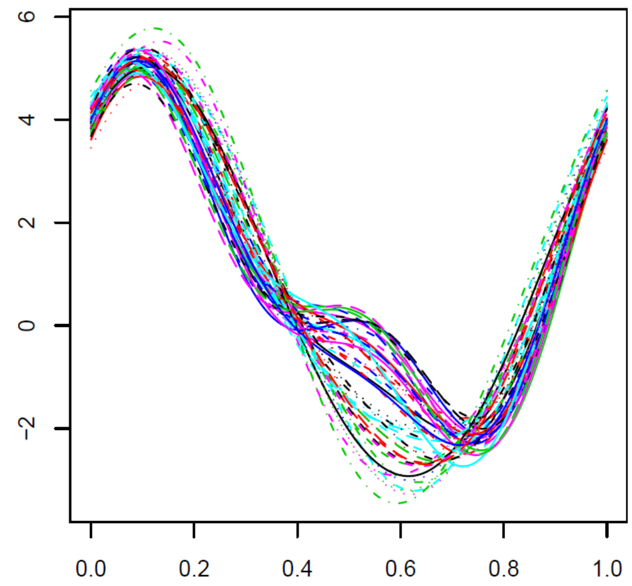
- **No model at hand**
  (probability density function does not exist,
  basically impossible to assess the validity of the model)
- **Choice of the Smoothing**
  (functions and their derivatives need to be estimated from point-wise noisy evaluations)
- **Choice of the Metric**
  (huge variety of distances wrt to the multivariate framework)
- **Choice of the Group of Warping Functions**
  (data should generally be horizontaly aligned)

An Example:

K-mean Clustering of Misaligned Data Using a Derivative-based Metric

Same row data (smoothed using B-splines with different number of knots)





Weak Smoothing

One Cluster

Strong Smoothing

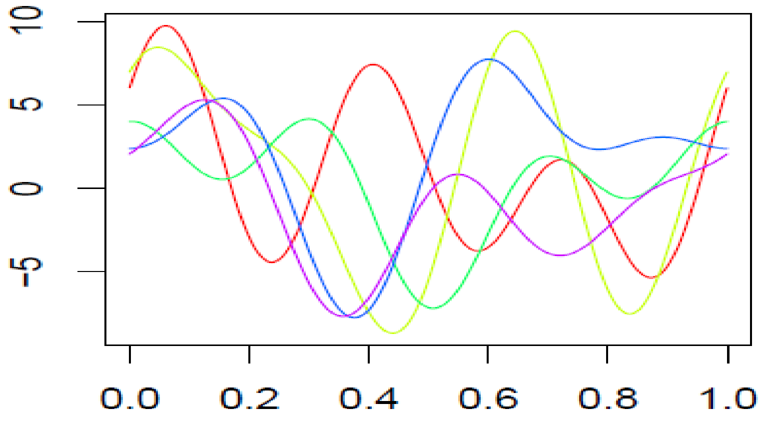
Two Clusters

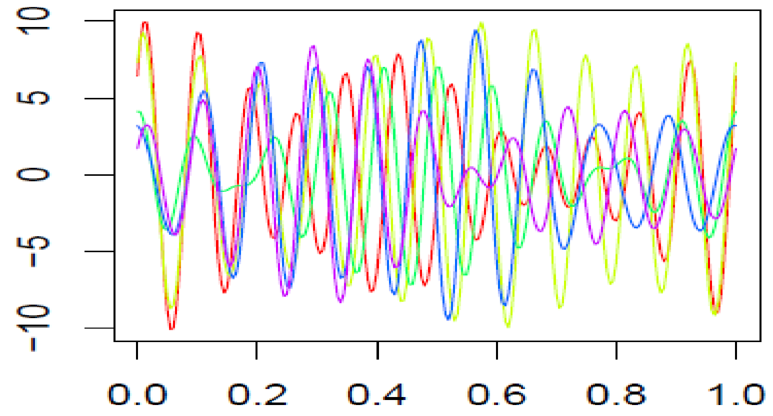
Figures are courtesy of Secchi, P., Vantini, S., Vitelli, V. (2013): " Bagging Voronoi classifiers for clustering spatial functional data", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 22, pp. 53-64

Choice of the Metric

- $\xi_1(\boldsymbol{s}), \ldots, \xi_7(\boldsymbol{s})$: 1D i.i.d. random variable
- $\{e_k, k \geq 1\}$: Fourier basis

$$\chi_{\boldsymbol{s}}^{(1)} = \sum_{k=1}^{7} \xi_k(\boldsymbol{s}) e_k \qquad \chi_{\boldsymbol{s}}^{(2)} = \sum_{k=19}^{25} \xi_{k-18}(\boldsymbol{s}) e_k$$





- From an ***L*²** perspective the two datasets shows the **same variability** across curves.
- From an ***H*¹** perspective the two datasets shows a **different variability** across curves (lower the former, larger the latter).

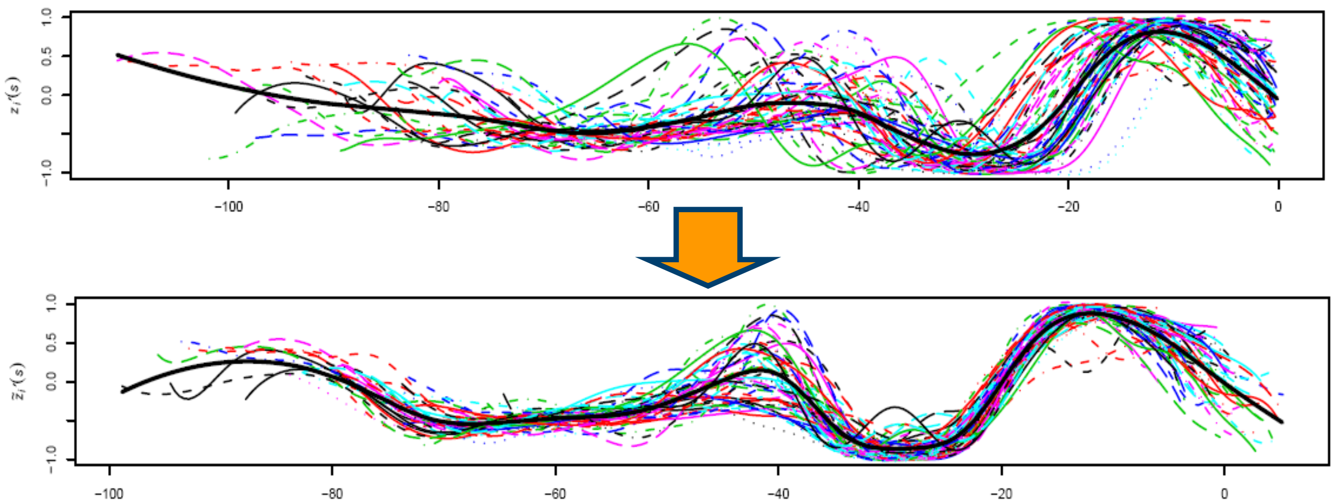
Figures are courtesy of A. Menafoglio; P. Secchi; M. Dalla Rosa (2013), "A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space". Electronic Journal of Statistics 7, 2209–2240

A Snapshot at Functional Data Registration/Alignment

Registration of a set of functions

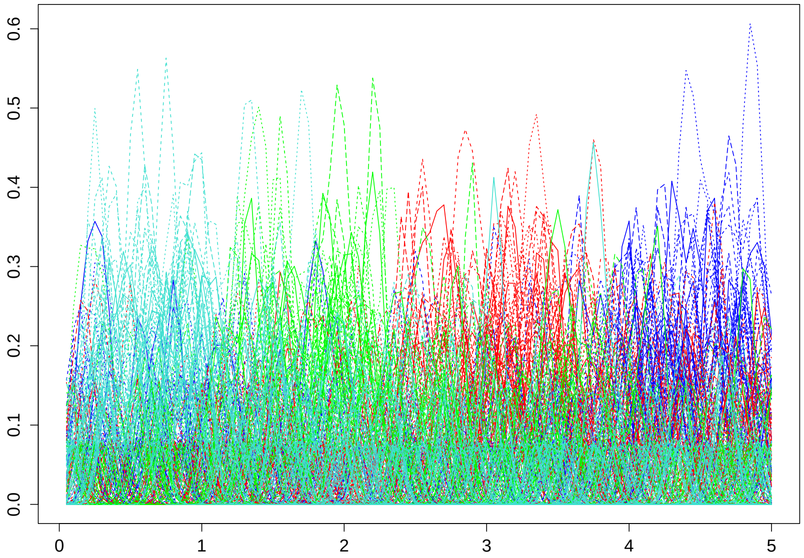
Find **suitable** warping functions $h_1, \ldots, h_n$ such that $c_1 \circ h_1, \ldots, c_n \circ h_n$ are the most **similar**.

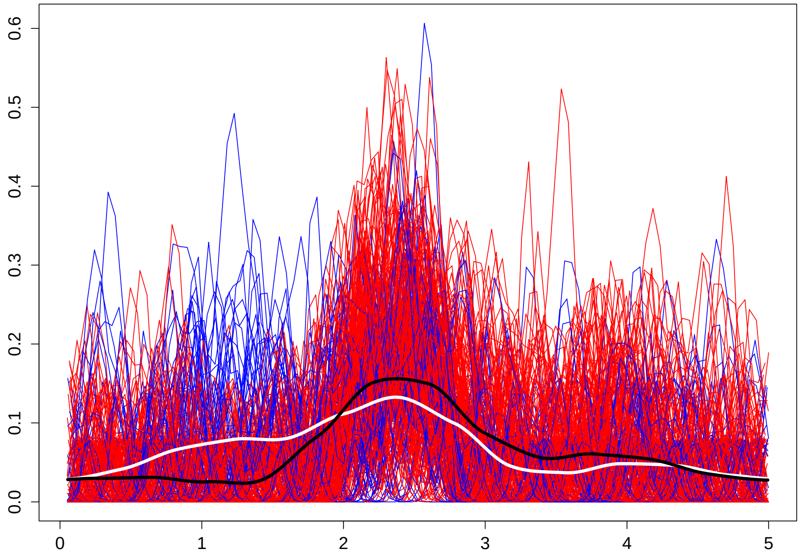
- **Landmark Approach** (similar means that functions are warped along the x-axis such that **each** (known) **landmark** occurs at the same point along the x-axis)
- **Continuous Approach** (similar means that functions are warped along the x-axis such that for **each point** along the x-axis functions present close values along the y-axis)



Choice of the Group of Warping Functions

200 periodic curves (spike-trains of neuronal activity)





Not Aligned Curves

Four Clusters

Aligned Curves

Two Clusters

Figures are courtesy of Patriarca, M., Sangalli, L.M., Secchi, P., Vantini, S.: "Analysis of Spike Train Data: an Application of K-mean Alignment", Electronic Journal of Statistics, 8, 1769-177, Special Section on Statistics of Time Warpings and Phase Variations

Clustering of Misaligned Functional Data

The Algorithm

↓

A Toy Example

↓

The Theory

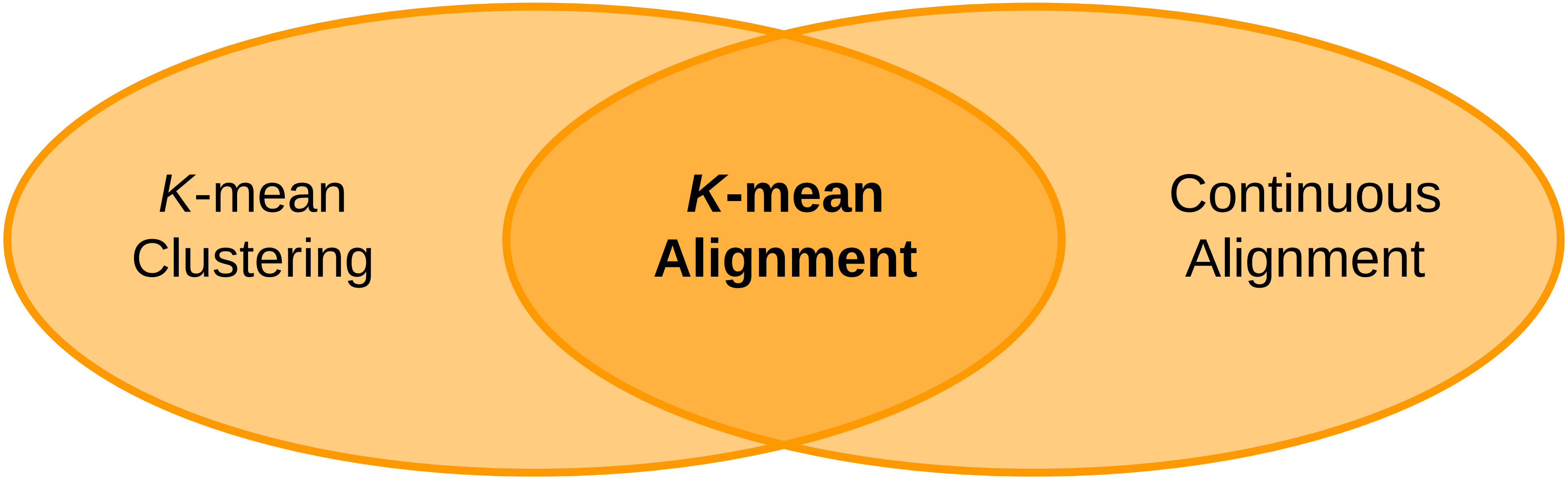
↓

The Case Study

The Algorithm

Framework

Functional Clustering **Alignment / Registration**



Goals

Goal of Continuous Alignment:
Decoupling Phase and Amplitude Variability

+

Goal of *K*-mean Clustering:
Decoupling Within and Between-cluster (Amplitude) Variability

=

Goal of *K*-mean Alignment:
Decoupling Phase Variability, Within-cluster Amplitude Variability, and Between-cluster Amplitude Variability

Continuous Alignment

n unregistered curves

Estimation

Estimate the template curve $\boldsymbol{c}_0$

Maximization

For each curve, find h_i that maximizes similarity between template curve $\boldsymbol{c}_0$ and patient warped curve $\boldsymbol{c}_i \circ h_i$

n registered curves and n warping functions

K-mean Clustering
(e.g. Tarpey and Kinateder 2003)

_K_-mean Clustering

n unregistered curves

Estimation

Estimate the *K* centroids curves $\{c_0^k\}_{k=1,2,..,K}$

Assignment

Assign c_i to the *k*-th cluster if the similarity between c_0^k and c_i is maximal over $k = 1, 2, \ldots, K$

K clusters

K-mean Clustering *vs* Continuous Alignment

Continuous Alignment	***K*-mean Clustering**
n unregistered curves	n unregistered curves
Estimation	***Estimation***
Estimate the template curve $\boldsymbol{c}_0$	Estimate the K centroids curves $\{\boldsymbol{c}_0^k\}_{k=1,2,..,K}$
Maximization	***Assignment***
For each curve, find h_i that maximizes similarity between template curve $\boldsymbol{c}_0$ and patient warped curve $\boldsymbol{c}_i \circ h_i$	Assign $\boldsymbol{c}_i$ to the k-th cluster if the similarity between $\boldsymbol{c}_0^k$ and $\boldsymbol{c}_i$ is maximal over $k = 1, 2, \dots, K$
n registered curves and n warping functions	K clusters

K-mean Alignment

n unregistered curves

Estimation

Estimate the K template/centroid curves $\{\boldsymbol{c}_0^k\}_{k=1,2,..,K}$

Maximization

For each curve, find h_i^k that maximizes similarity between each template curve $\boldsymbol{c}_0^k$ and the candidate warped curve $\boldsymbol{c}_i \circ h_i^k$

Assignment

Assign $\boldsymbol{c}_i$ to the k-th cluster if the similarity between $\boldsymbol{c}_0^k$ and $\boldsymbol{c}_i \circ h_i^k$ is maximal over $k = 1, 2, \ldots, K$ and then warp $\boldsymbol{c}_i$ along $h_i = h_i^k$

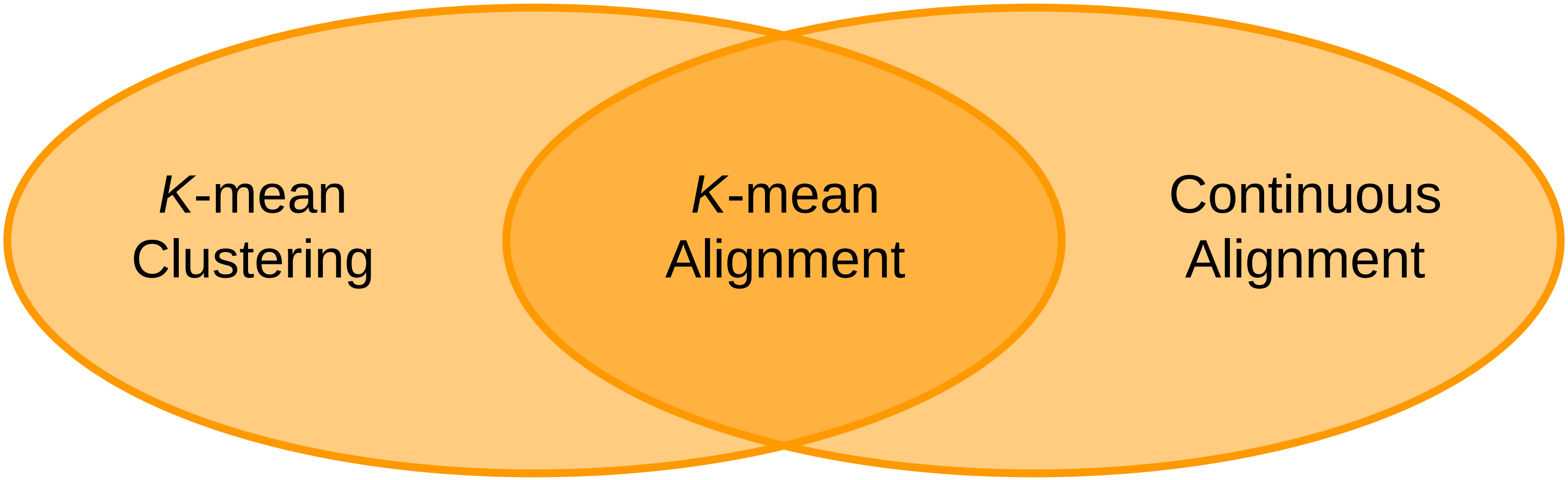
n registered curves, *n* warping functions and K clusters

$K = 1$ → Continuous Registration Algorithm (e.g. Sangalli *et al.* 2009)

$W = \{1\}$ → Functional K-mean Clustering (e.g. Tarpey and Kinateder 2003)

K-mean Alignment: A Double-Facet Method

Functional Clustering | **Alignment / Registration**



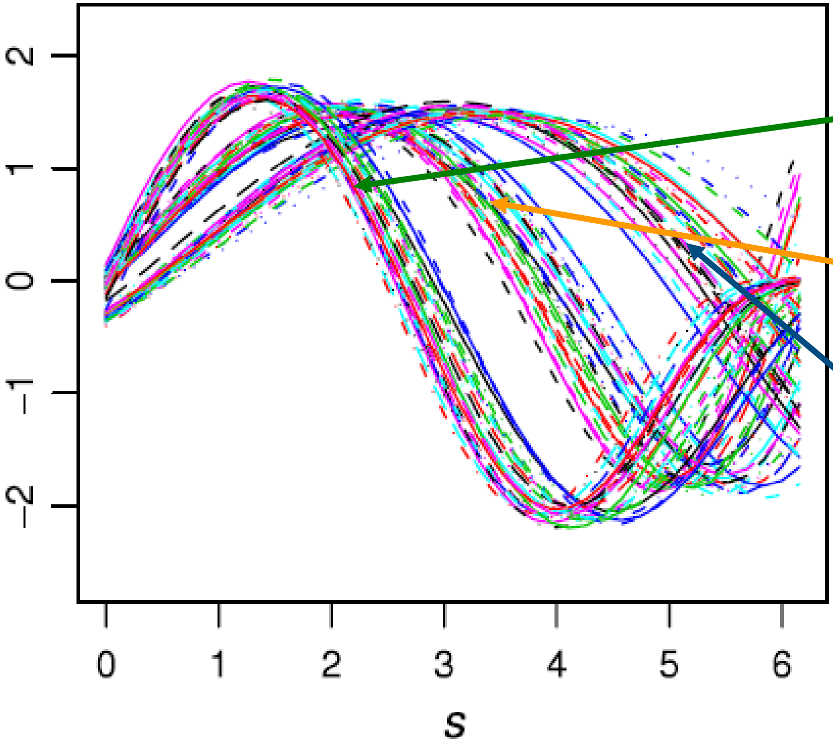
It is a *K*-mean Clustering Algorithm where warping is allowed

It is an Alignment Algorithm with *K* templates

A Toy Example

A Simulated Toy Example: Simulation Details

2 Amplitude Clusters (2 template curves)
with further clustering in the phase



$$2 * \sin(s) - 1 * \sin\left(\frac{s^2}{2\pi}\right)$$

$$1 * \sin(s) + 1 * \sin\left(\frac{s^2}{2\pi}\right)$$

$$1 * \sin\left(\frac{1}{3} + \frac{3}{4}s\right) + 1 * \sin\left(\frac{(\frac{1}{3} + \frac{3}{4}s)^2}{2\pi}\right)$$

Variability in both Amplitude and Phase

$$(1 + \varepsilon_{1i}) * \sin(\varepsilon_{3i} + (1 + \varepsilon_{4i})s) + (1 + \varepsilon_{2i}) * \sin\left(\frac{(\varepsilon_{3i} + (1 + \varepsilon_{4i})s)^2}{2\pi}\right)$$

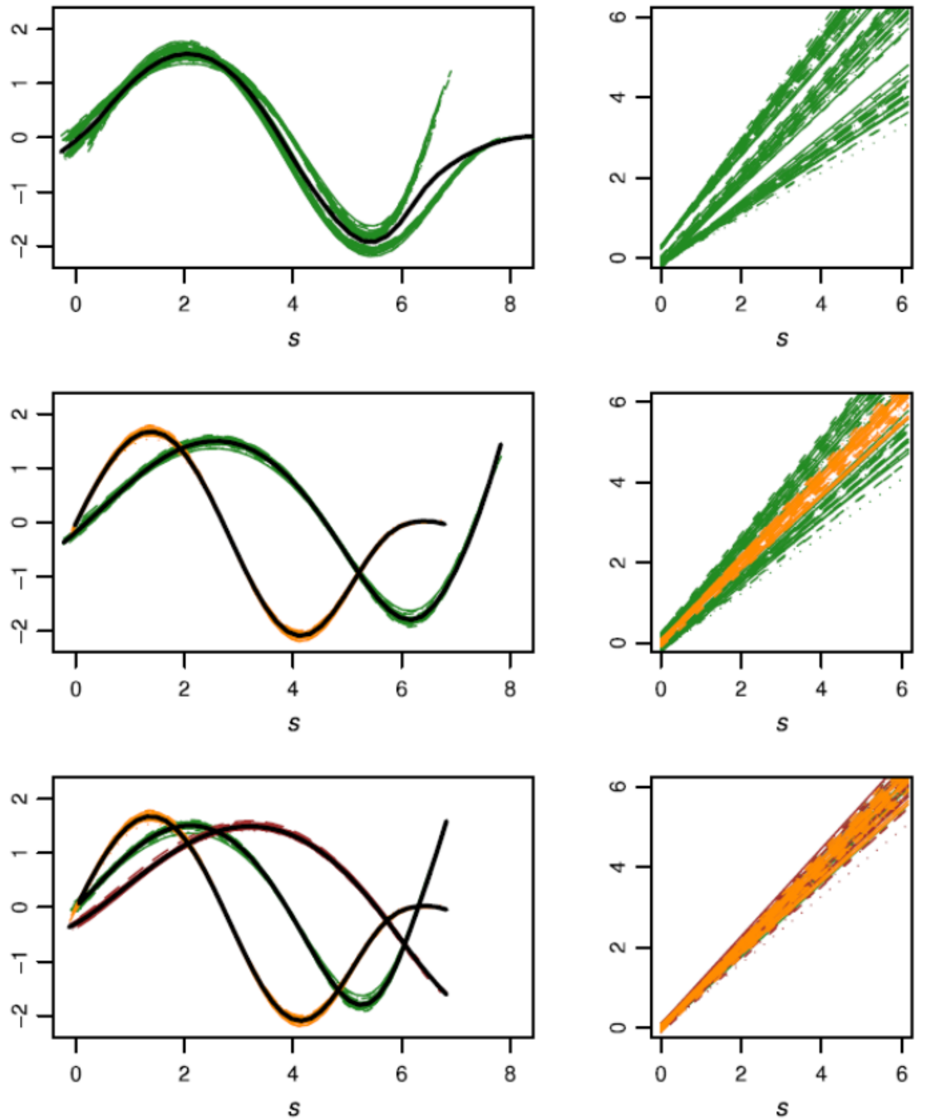
A Simulated Toy Example: Algorithm Results

Aligned and clustered curves | Warping functions

K = 1

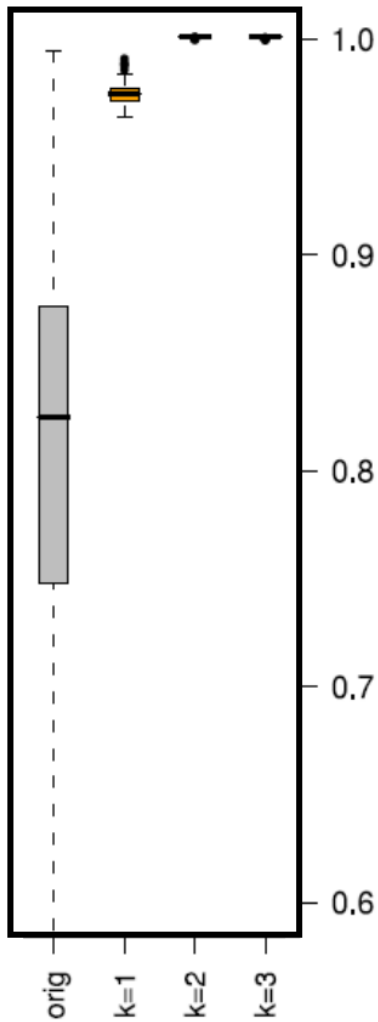
K = 2

K = 3

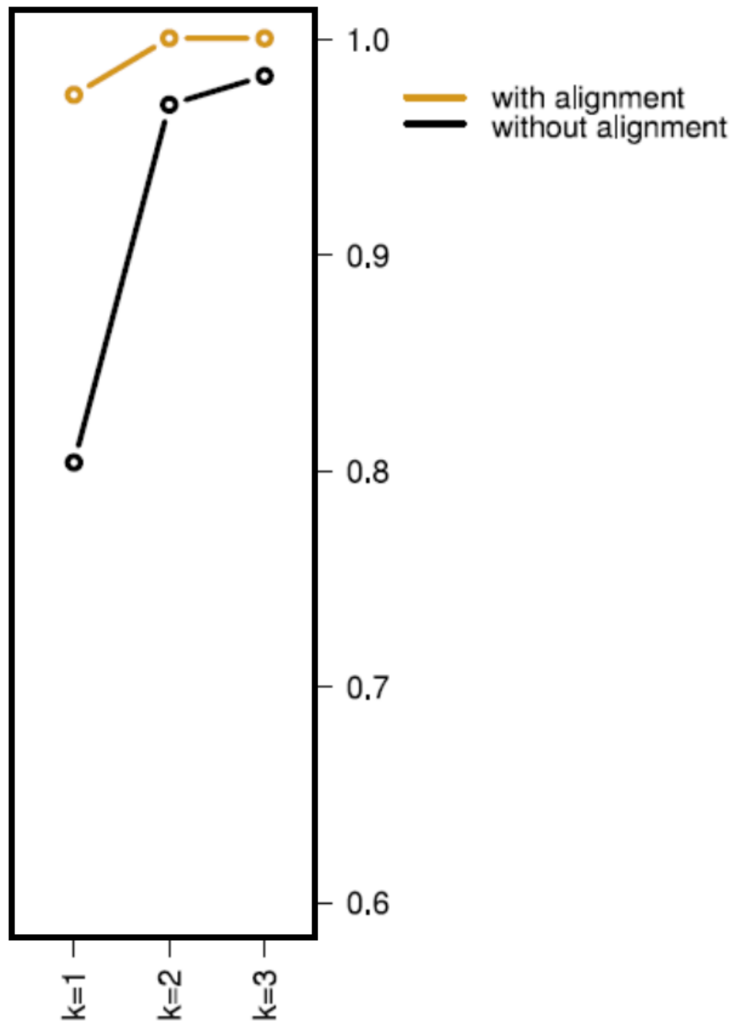


A Simulated Toy Example: Algorithm Results

Boxplots of the similarity indices between curves and templates

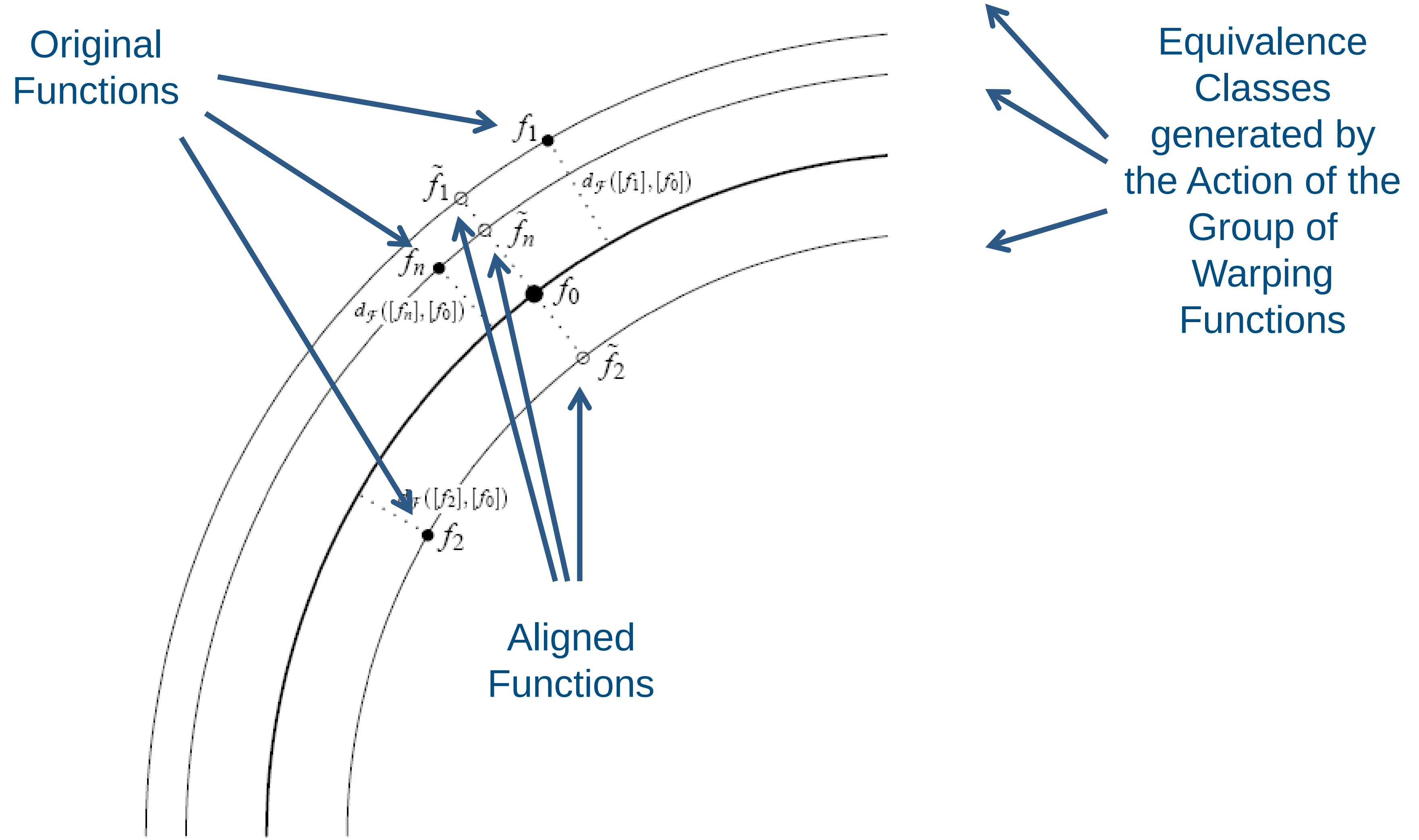


Means of the similarity indices between curves and templates

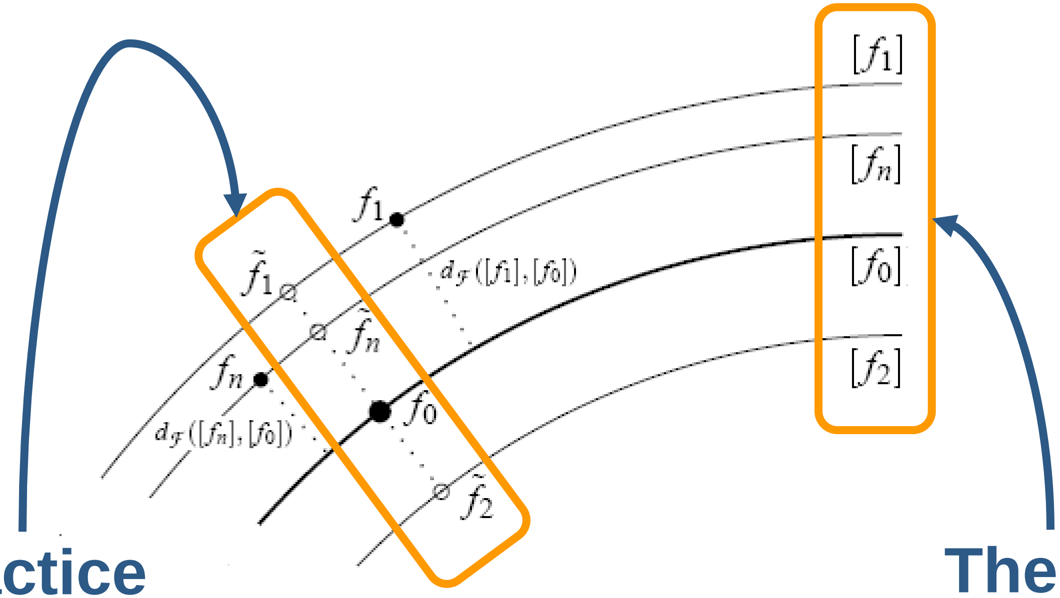


The Theory

Abstract Visualization of Alignment



Meta-Equivalence Result



Practice

Analysis of a registered functional data set with respect to the metric *d*

(e.g., K-mean Alignment in the Functional Space)

=

Theory

Analysis of a set of equivalence classes
(induced by the application of *W* to the original functions)
with respect to a new metric $d_{\mathcal{F}}$
(jointly defined by *d* and *W*).

(e.g., K-mean Clustering in the Quotient Space)

Meta-Equivalence Theorem

(a) $\boxed{F}$ is a metric space according to a distance $\boxed{d}$ $F \times F \longrightarrow \mathcal{R}_0^+$ whose elements are functions: $\Omega \subseteq \mathcal{R}^p \longrightarrow \Psi \subseteq \mathcal{R}^q$,

(b) $\boxed{W}$ is a compact (with respect to a metric d_G) subgroup (with respect to ordinary composition $\circ$) of the group G of the continuous automorphisms: $\Omega \subseteq \mathcal{R}^p \longrightarrow \Omega \subseteq \mathcal{R}^p$,

(c) $\forall f \in F$ the map $f\circ : h \in W \longmapsto (f\circ)(h) = (f \circ h) \in F$ is continuous;

(d) Given any couple of elements $f_1, f_2 \in F$ and an element $h \in W$, the distance between f_1 and f_2 is invariant under the composition of f_1 and f_2 with h, i.e.:

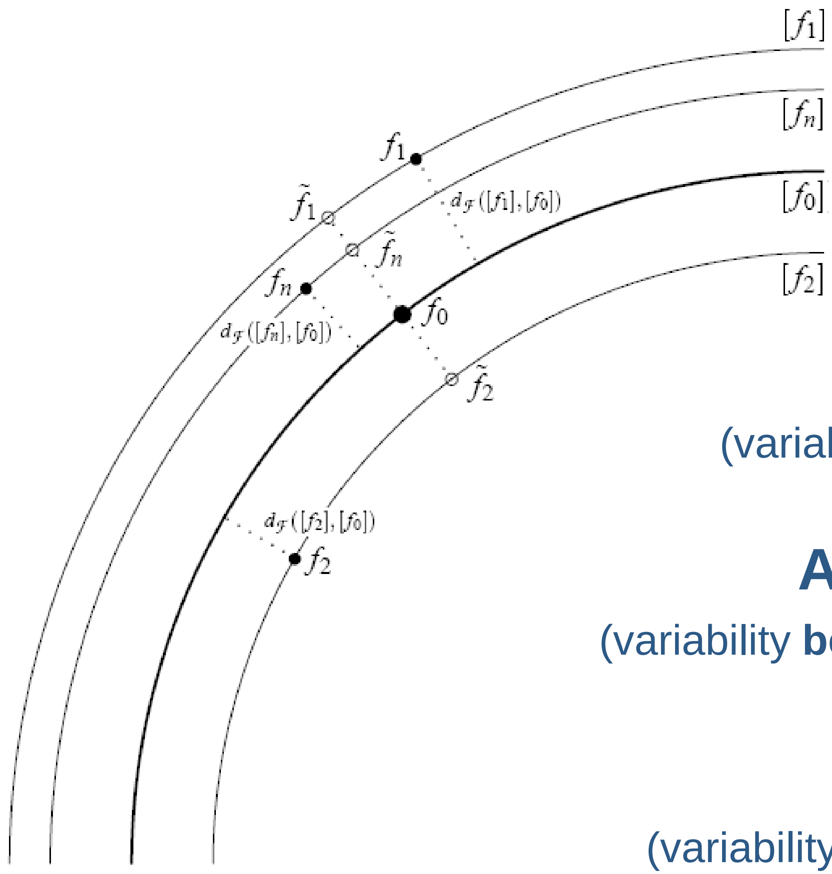
$$\boxed{d(f_1, f_2) = d(f_1 \circ h, f_2 \circ h)\ ;}$$

we will refer to this property as W-*invariance* of d.

$$\boxed{d_{\mathcal{F}}([f_1],[f_2]) = \min_{h_1,h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2) \textit{ is a metric on the quotient set } \mathcal{F}}$$

A Topological Characterization of Phase and Amplitude Variability

The introduction of
a metric/semi-metric *d* and of a group *W* of warping functions,
with respect to which the metric/semi-metric is invariant,
enables a not ambiguous definition of phase and amplitude variability.



Total Variability

(variability between elements of *F*)

Amplitude Variability

(variability **between** equivalence classes)

Phase Variability

(variability **within** equivalence classes)

Ancillary variability

In many situations, d is not a metric but a semi-metric, i.e.: $d(f_1, f_2) = 0 \nRightarrow f_1 = f_2$.

the presented theory still holds if F is replaced with $\bar{F} = F/\odot$

$$f_1 \odot f_2 \Leftrightarrow d(f_1, f_2) = 0$$

Hierarchy of Quotient Spaces

Functions belonging to F

are grouped in

equivalence classes belonging to $\bar{F}$

that are grouped in

equivalence classes belonging to $\mathcal{F}$

Some examples of *W*-invariant semi-metrics

Metric / Semi-metric	Maximal *W* (Phase Variability)	Ancillary Variability
$\|\|f_1 - f_2\|\|_{L^2}$	H-translations	Ø
$\|\|f_1' - f_2'\|\|_{L^2}$	H-translations	V-translations
$\|\|(f_1 - \bar{f}_1) - (f_2 - \bar{f}_2)\|\|_{L^2}$	H-translations	V-translations
$\|\|(f_1' - \bar{f}_1') - (f_2' - \bar{f}_2')\|\|_{L^2}$	H-translations	V-translations V-linear trends
$\left\|\left\|\frac{f_1}{\|\|f_1\|\|_{L^2}} - \frac{f_2}{\|\|f_2\|\|_{L^2}}\right\|\right\|_{L^2}$	H-translations H-dilations	V-dilations
$\left\|\left\|\frac{f_1'}{\|\|f_1'\|\|_{L^2}} - \frac{f_2'}{\|\|f_2'\|\|_{L^2}}\right\|\right\|_{L^2}$	H-translations H-dilations	V-translations V-dilations
…	…	…
$\left\|\left\|\mathrm{sign}(f_1')\sqrt{\|f_1'\|} - \mathrm{sign}(f_2')\sqrt{\|f_2'\|}\right\|\right\|_{L^2}$	H-diffeomorphisms	V-translations

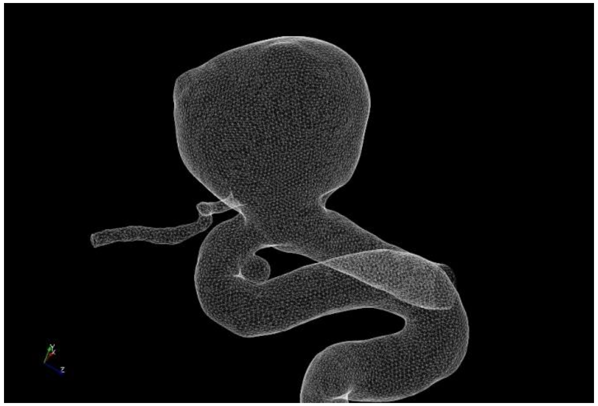
The Case Study

Cerebral Aneurysms

Cerebral aneurysms: malformations of cerebral arteries, in particular of arteries belonging to or connected to the Circle of Willis.

EPIDEMIOLOGICAL STATISTICS

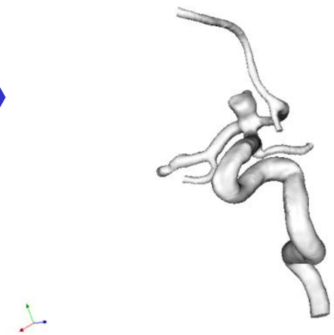
- Incidence rate of cerebral aneurysms:
  1/20 people
- Incidence rate of ruptured cerebral aneurysms per year:
  1/10000 people per year
- Mortality due to a ruptured aneurysm:
  > 50%: Out of 9 patients with a ruptured aneurysm:
  - 3 are expected to die before arriving at the hospital
  - 2 to die after having arrived at the hospital
  - 2 to survive with permanent cerebral damages
  - 2 to survive without permanent cerebral damages

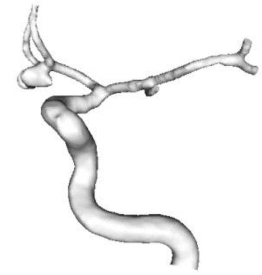


Upper

High Aneurysm

33

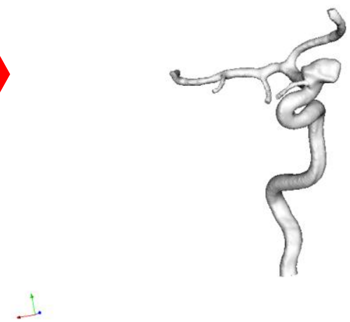


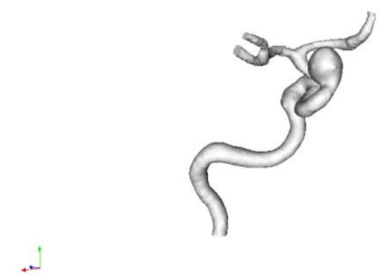


Lower

Low Aneurysm

25

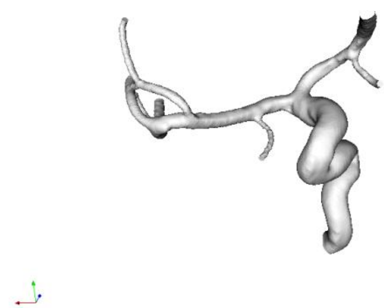




Healthy

No Aneurysm

7

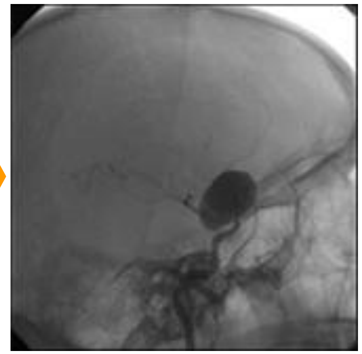


From X-rays to Centerlines

Contrast Fluid Injections

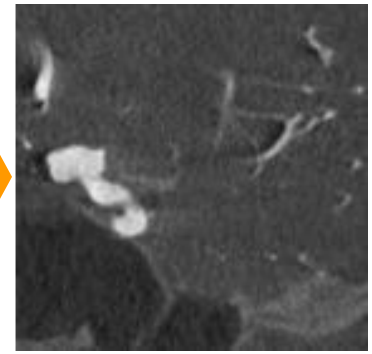
1

X-rays (one direction)



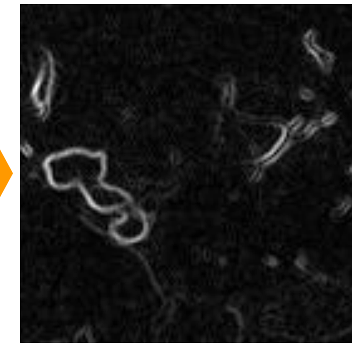
2

3d-array (one slice)



3

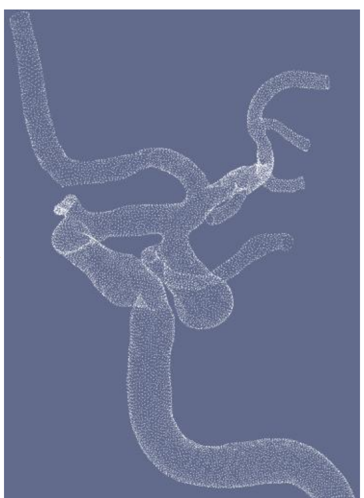
Gradient 3d-array (one slice)



4

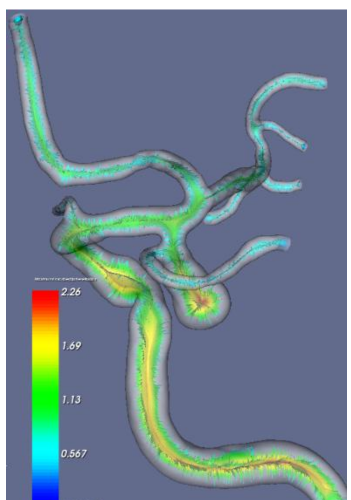
Surface Points

4



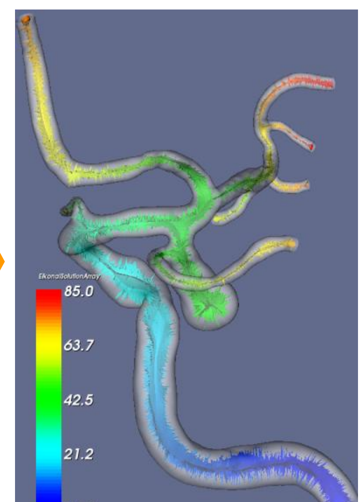
5

Voronoi Diagram



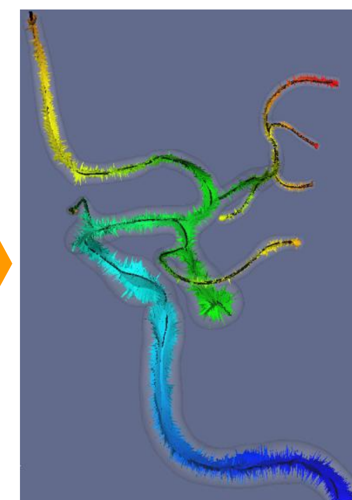
6

Eikoinal Equation

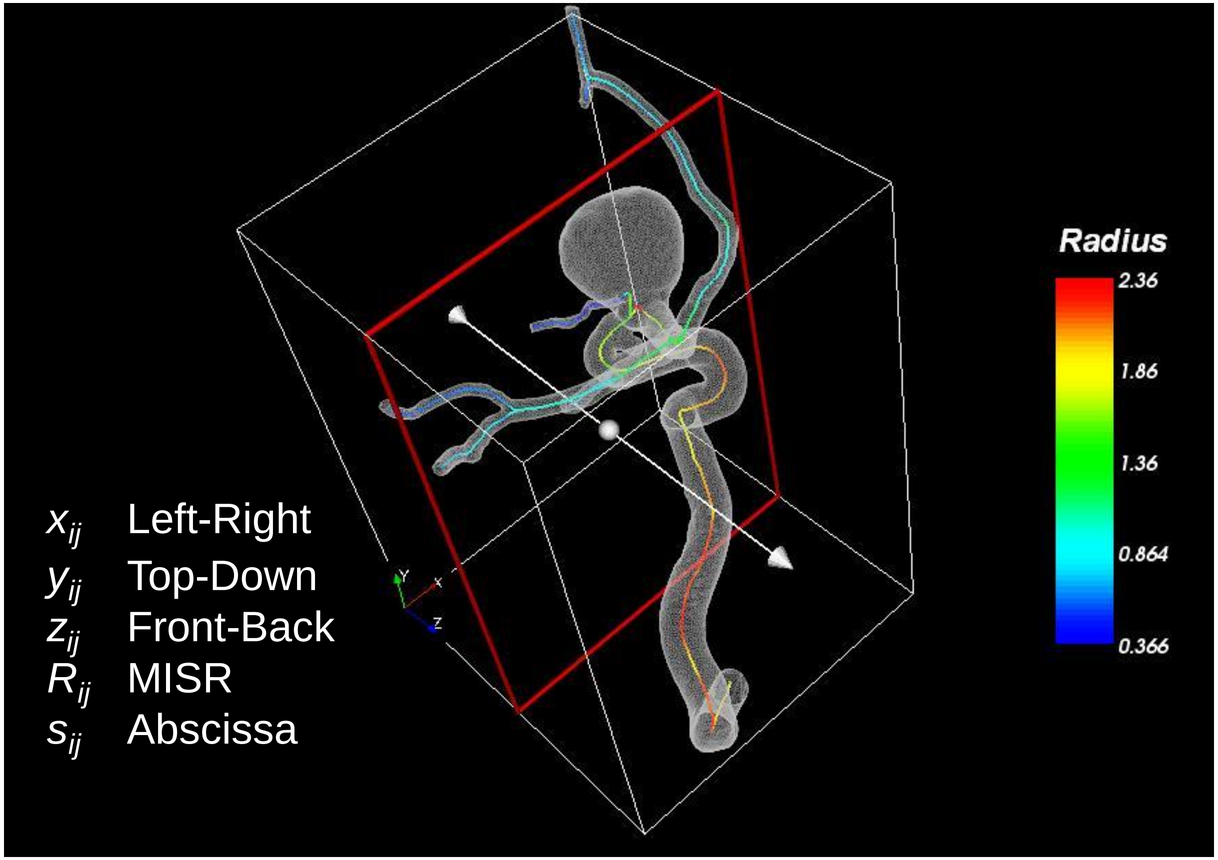


7

Centerline

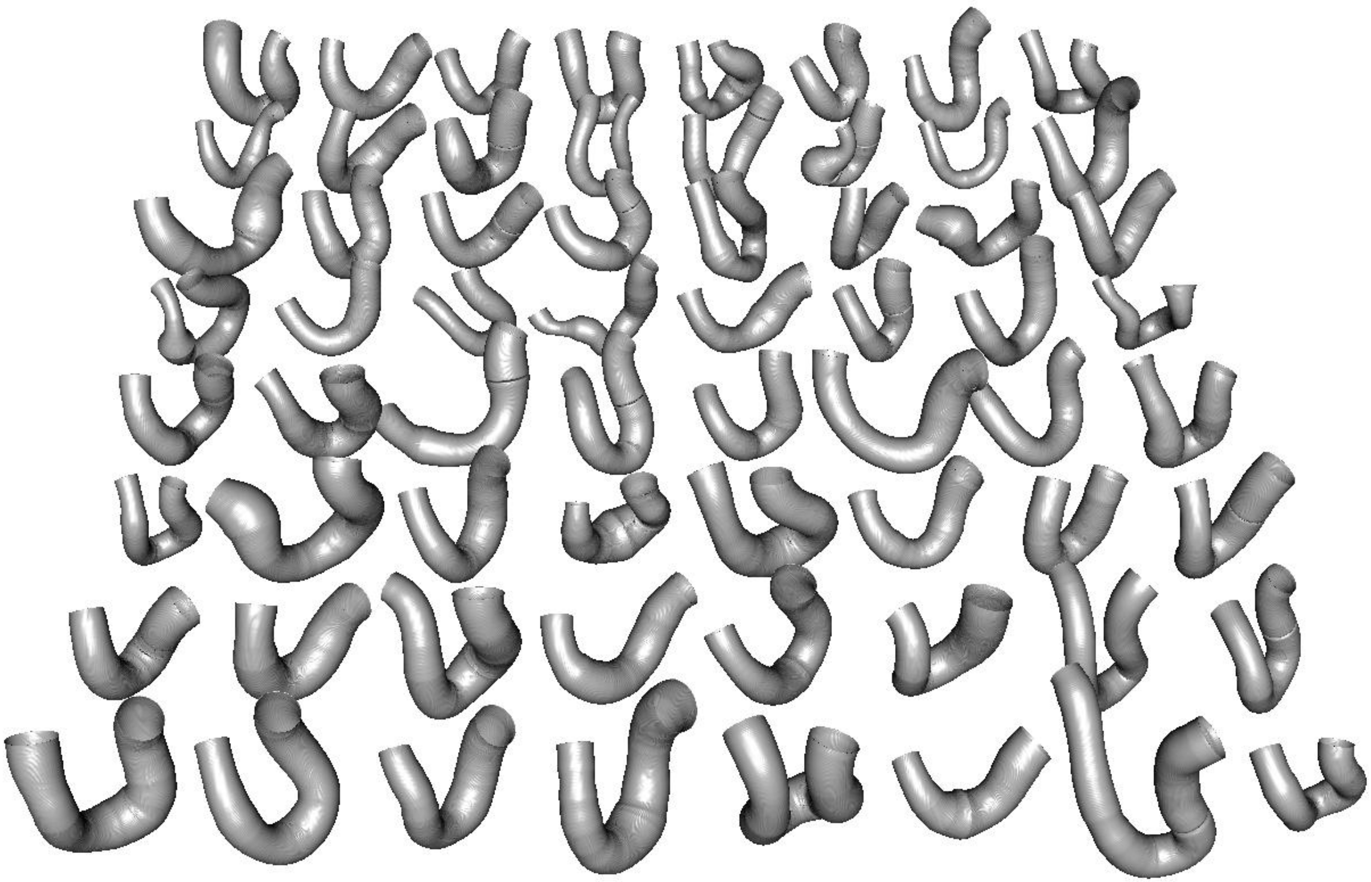


Discrete Data

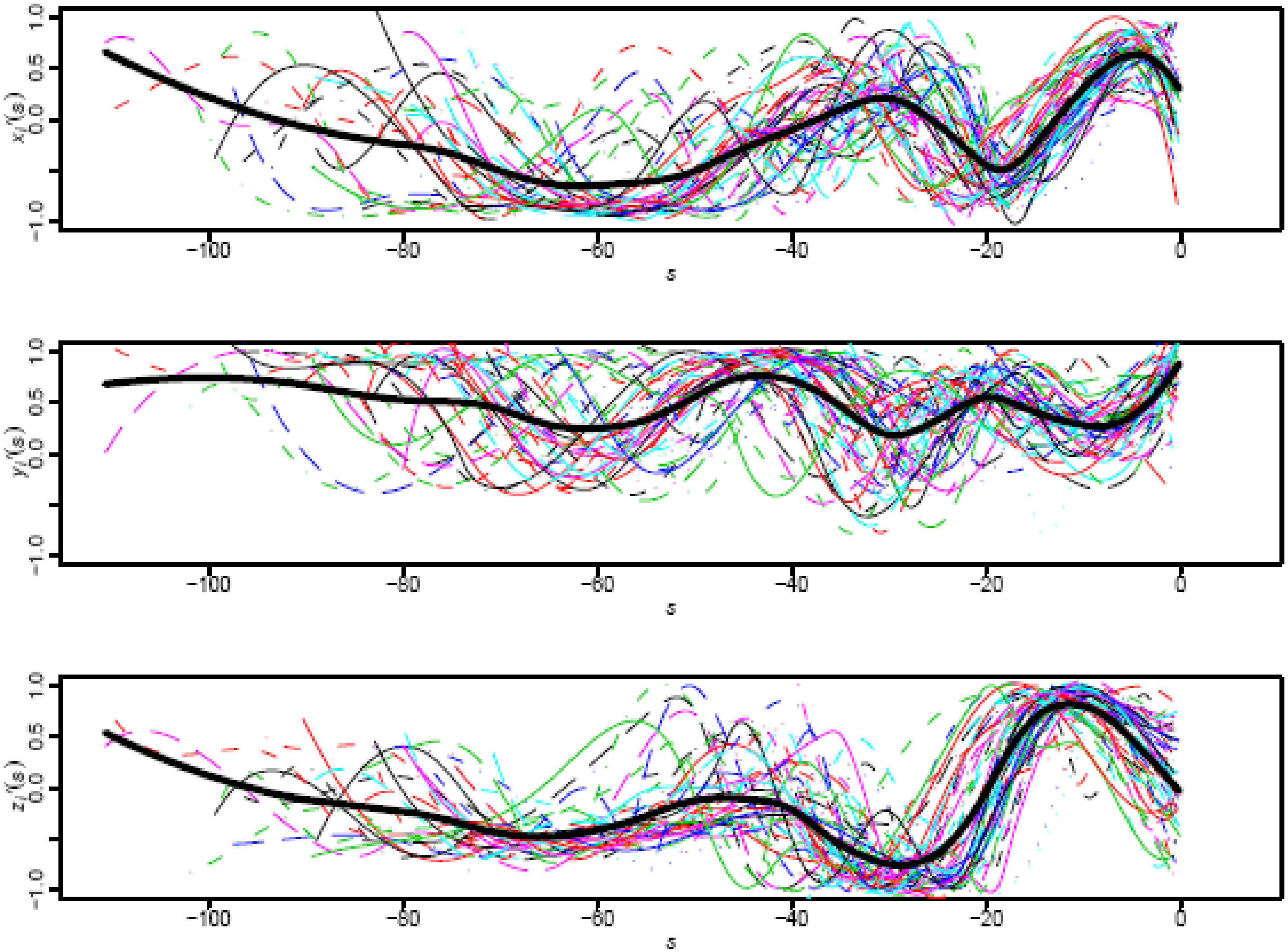


Observational Study conducted at Ospedale Ca' Granda Niguarda – Milano relative to 65 patients hospitalized from September 2002 to October 2005.

The sample of 65 ICA: each patient is represented by the centerline of their ICA



ICA centerline First Derivatives



simone.vantini@polimi.it

K-mean Alignment: Theoretical Choices

Similarity Index between Curves

$$\rho(\mathbf{c}_i, \mathbf{c}_j) = \frac{1}{3} \cdot [\rho(x_i, x_j) + \rho(y_i, y_j) + \rho(z_i, z_j)]$$

Group of Warping Functions

$$W = \{h : h(s) = ms + p \quad \text{with} \quad m \in \mathbb{R}^+, p \in \mathbb{R}\}$$

Properties

$$|\rho(\mathbf{c}_i, \mathbf{c}_j)| \leq 1$$

$$\rho(\mathbf{c}_i, \mathbf{c}_j) = 1 \Leftrightarrow \exists \mathbf{A} \in (\mathbb{R}^+)^3, \mathbf{B} \in \mathbb{R}^3 : \begin{cases} x_i = A_x x_j + B_x \\ y_i = A_y y_j + B_y \\ z_i = A_z z_j + B_z \end{cases}$$

Minimal Joint Properties

$$\rho(\mathbf{c}_i, \mathbf{c}_j) = \rho(\mathbf{c}_i \circ h, \mathbf{c}_j \circ h) \qquad \forall h \in W$$

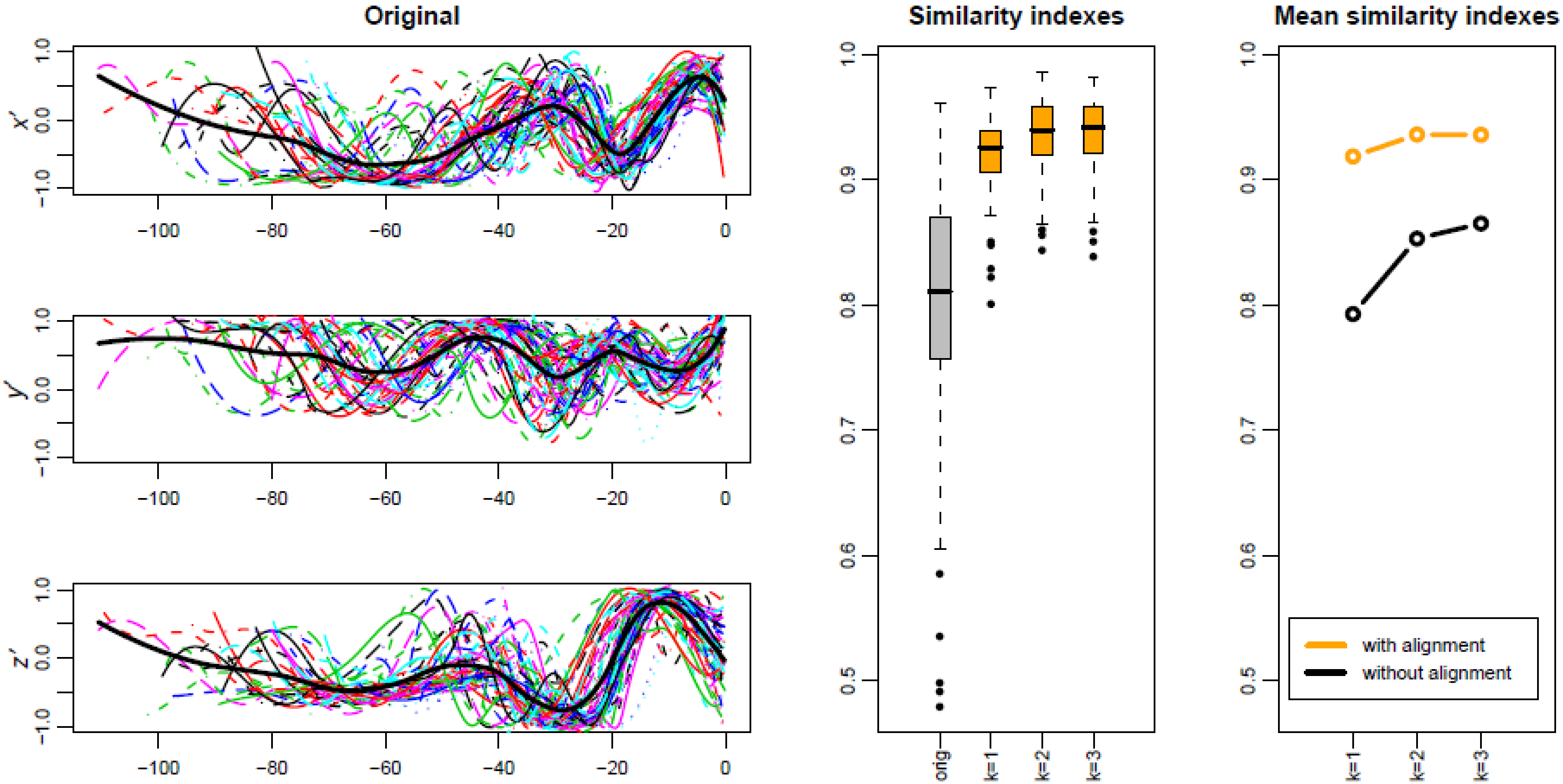
$$\rho(\mathbf{c}_i \circ h, \mathbf{c}_j) = \rho\left(\mathbf{c}_i, \mathbf{c}_j \circ h^{-1}\right) \qquad \forall h \in W$$

$$\sup_{h \in W} \rho(\mathbf{c}_i \circ h, \mathbf{c}_j) = \sup_{h \in W} \rho(\mathbf{c}_i, \mathbf{c}_j \circ h)$$

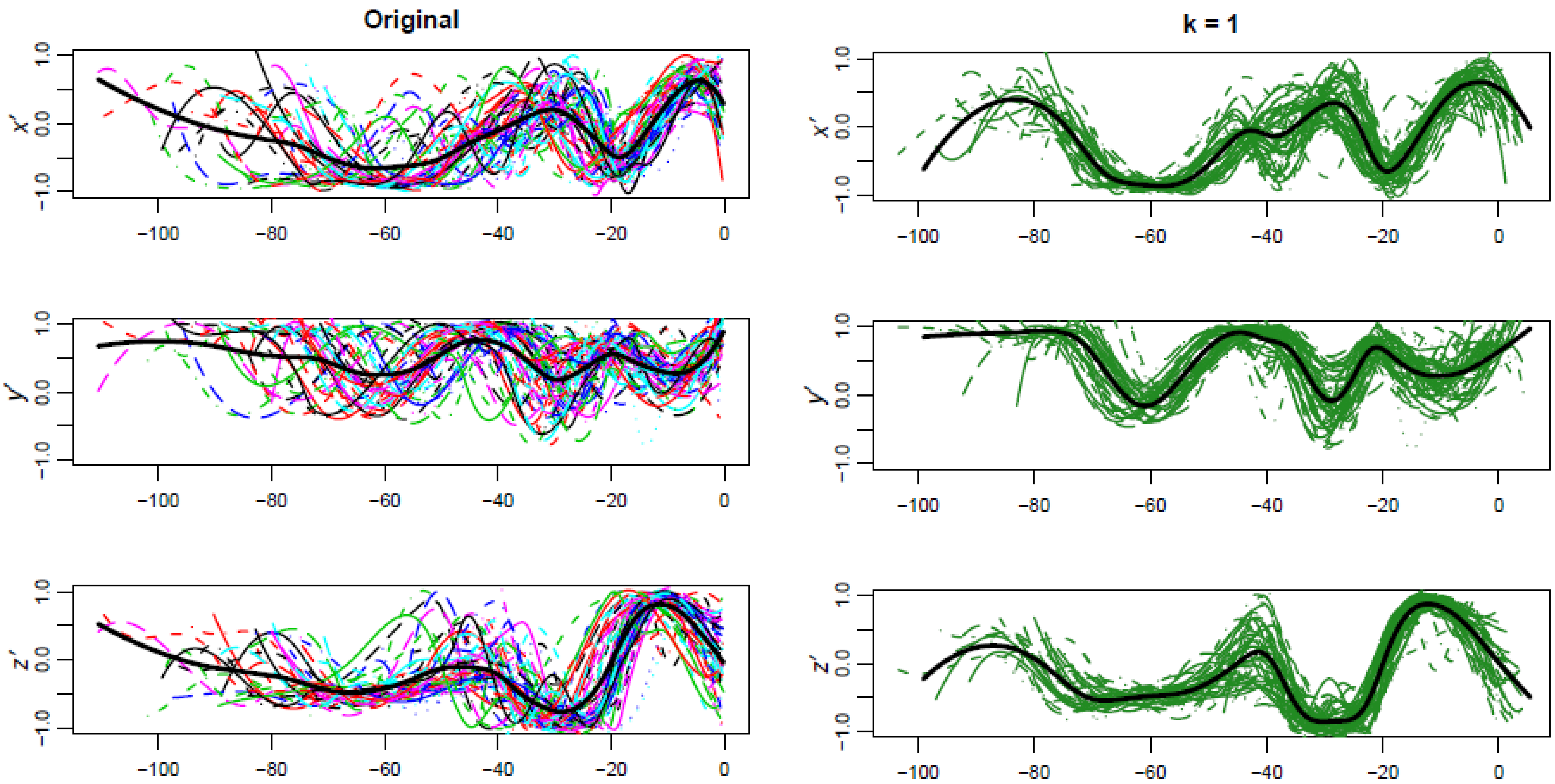
find $\underline{\varphi} = \{\varphi_1, \ldots, \varphi_k\} \subset \mathcal{C}$ and $\underline{h} = \{h_1, \ldots, h_N\} \subset W$ such that

$$\frac{1}{N} \sum_{i=1}^{N} \rho(\varphi_{\lambda(\underline{\varphi}, \mathbf{c}_i)}, \mathbf{c}_i \circ h_i) \geq \frac{1}{N} \sum_{i=1}^{N} \rho(\psi_{\lambda(\underline{\psi}, \mathbf{c}_i)}, \mathbf{c}_i \circ g_i)$$

K-mean Alignment Performances



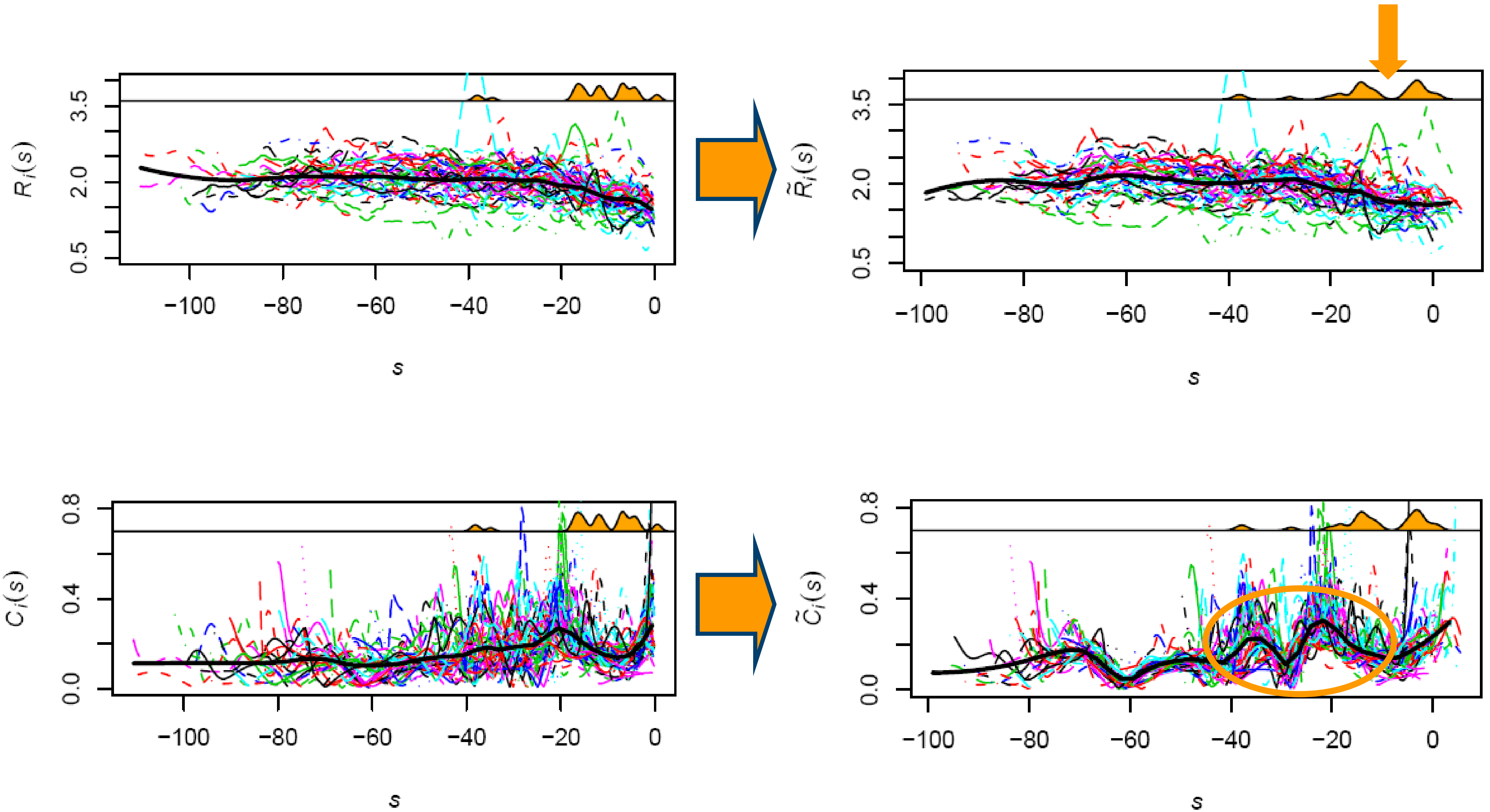
One-mean Alignment



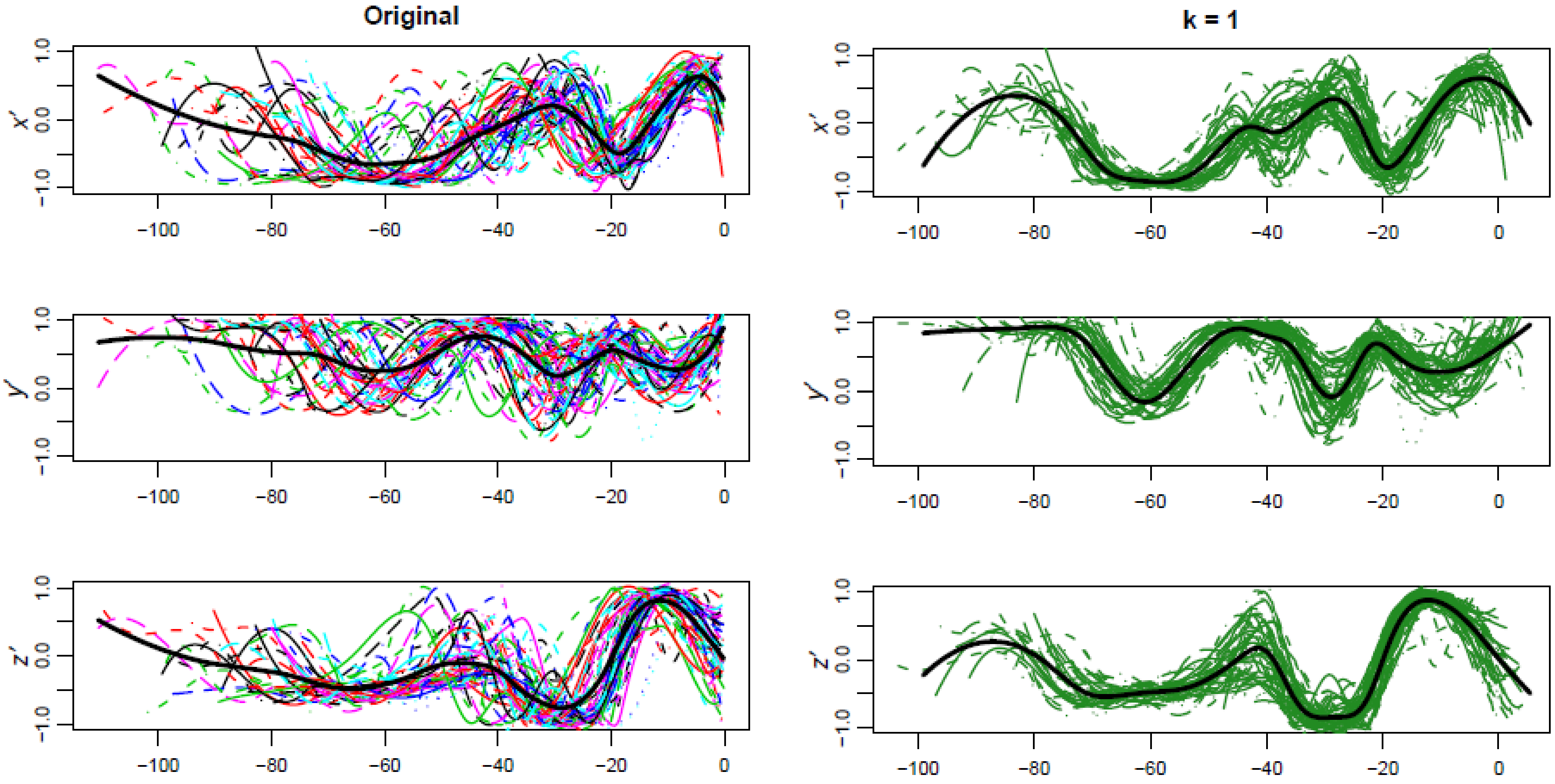
One-mean Alignment: aneurysm location on registered ICA radius and curvature profiles

Unregistered Radius and Curvature Profiles

Registered Radius and Curvature Profiles

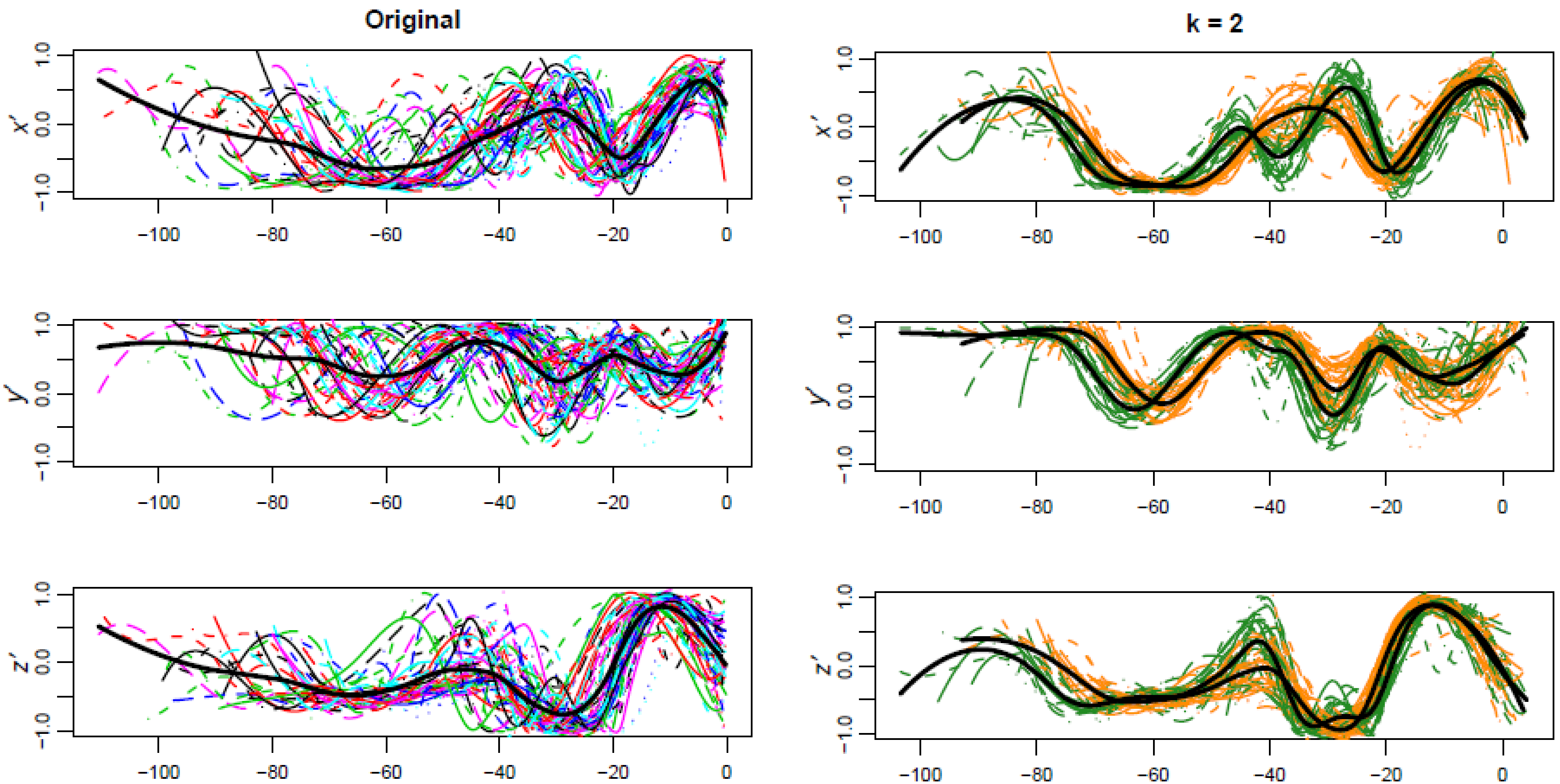


One-mean Alignment

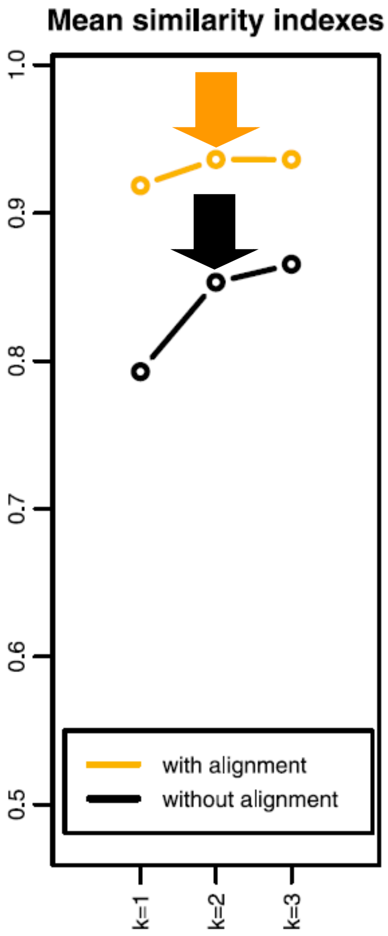


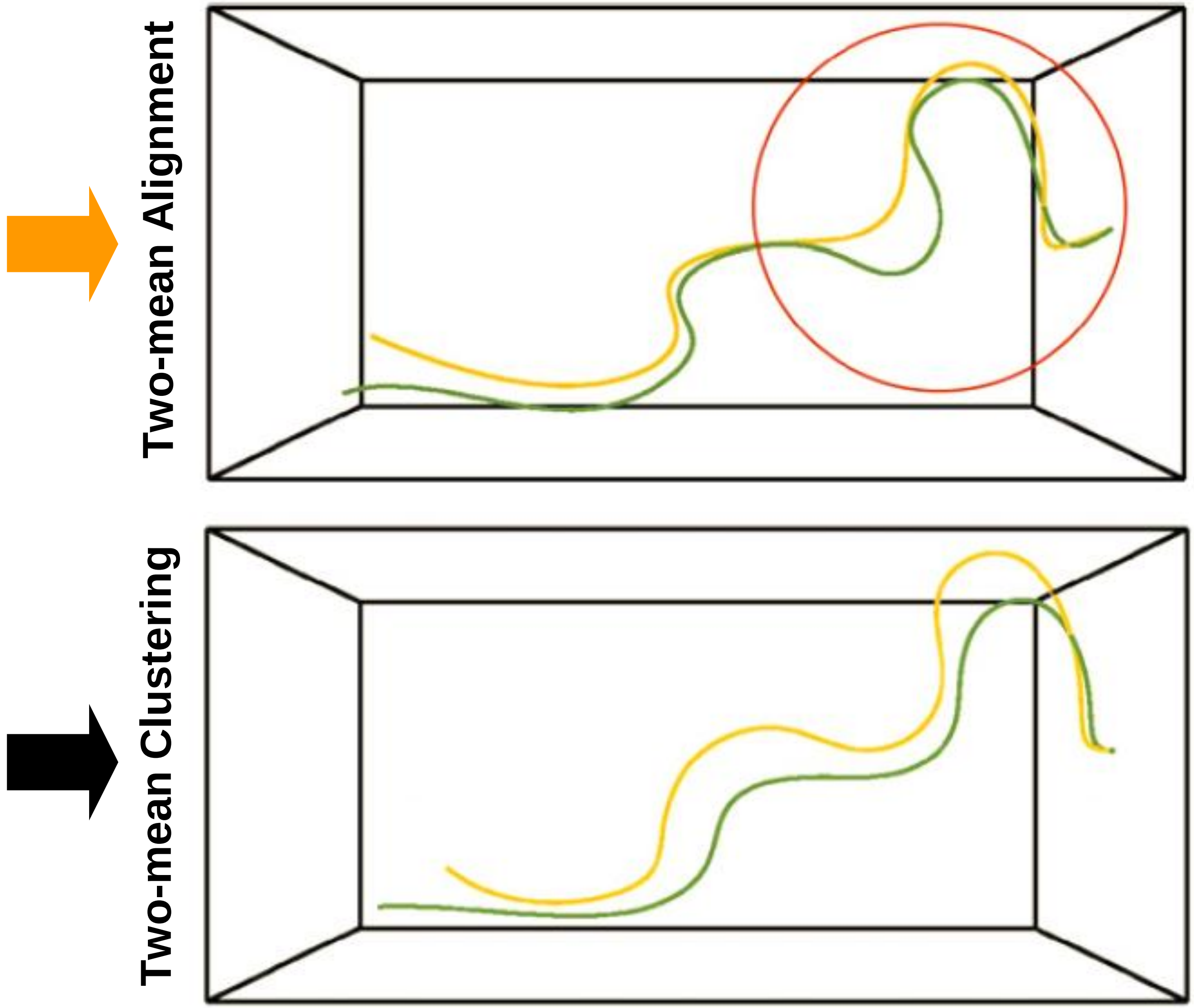
simone.vantini@polimi.it

Two-mean Alignment



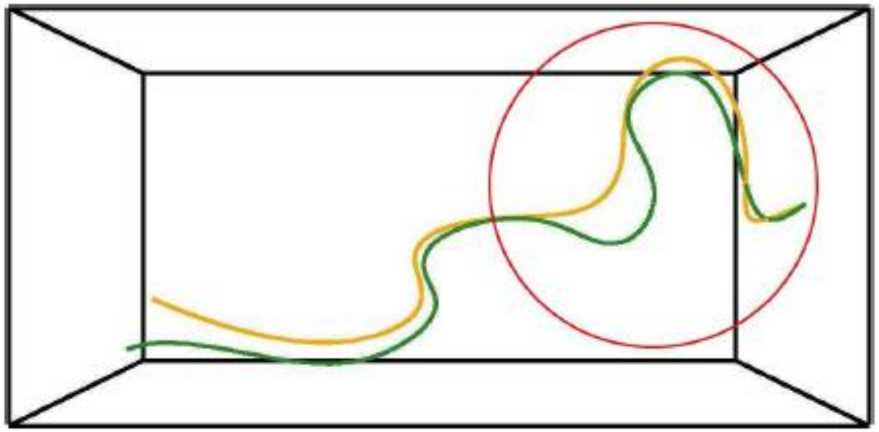
Two-mean Alignment vs Two-mean Clustering





Clustering

Two-mean Alignment



Clusters that are morphologically different

30 *S*-shaped ICAs
vs
35 *Ω*-shaped ICAs

Krayenbuehl et al. (1982)

	No Aneurysm	**Aneurysm along ICA**	**Aneurysm downstream ICA**
S-shaped ICAs	100%	52%	30%
Ω-shaped ICAs	0%	48%	70%

P-value of Pearson's Chi-squared test for independence equal to 0.0013

Fluid-dynamical interpretation of the onset of cerebral aneurysms

References

- Ramsay, J. O., Silverman, B. W. (2005):
  Functional Data Analysis, 2nd edition, Springer New York NY .
- Sangalli, L. M., Secchi, P., Vantini, S., Vitelli, V. (2010):
  "K-mean Alignment for Curve Clustering",
  Computational Statistics and Data Analysis, Vol. 54., pp. 1219-1233
- Vantini, S. (2013):
  "On the Definition of Phase and Amplitude Variability in Functional Data Analysis",
  Test, Vol. 21(4), pp. 676-696.
- Sangalli, L. M., Secchi, P., Vantini, S. (2014):
  "Analysis of AneuRisk65 data: k-mean alignment" [with discussion and rejoinder],
  Electronic Journal of Statistics, 8, 1891–1904, Special Section on Statistics of Time Warpings and Phase Variations.
- Krayenbuehl, H., Huber, P., Yasargil, M. G. (1982),
  Krayenbuhl/Yasargil Cerebral Angiography, Thieme Medical Publishers, 2nd ed.
- AneuRisk65 data are freely downloadable at
  `http://mox.polimi.it/it/progetti/aneurisk/`
  `http://ecm2.mathcs.emory.edu/aneuriskweb/a65`
- Parodi, A., Patriarca, M., Sangalli, L. M., Secchi, P., Vantini, S., Vitelli, V. (2014):
  `fdakma`: *Clustering and alignment of a given set of curves*,
  R package version 1.1.1.