LOCAL NONPARAMETRIC INFERENCE ON FUNCTIONAL DATA WITH MANIFOLD DOMAIN

Alessia Pini

Dept. of Statistical Sciences, Università Cattolica del Sacro Cuore, Milan, Italy

Statistical Methods for Data Analysis and Decision Sciences

Third Conference of the Statistics and Data Science Group of the Italian Statistical Society

Niels Lundtorp Olsen
Techical University of Denmark

Simone Vantini
Politecnico di Milano

PROBLEM FORMULATION

Functional data on $L^2(D) \cap \mathcal{C}^0(D)$, where $D \subset \mathbb{R}^p$.
Aim: test **locally** a functional hypothesis H_0 against H_1.

We assume that the domain D is a Riemaniann manifold of $\mathbb{R}^p$

alessia.pini@unicatt.it

PROBLEM FORMULATION

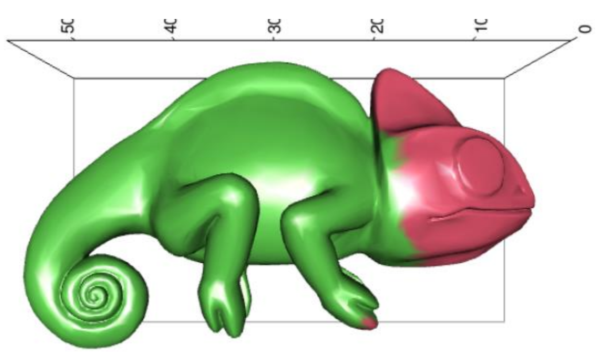
Functional data on $L^2(D) \cap \mathcal{C}^0(D)$, where $D \subset \mathbb{R}^p$.
Aim: test **locally** a functional hypothesis H_0 against H_1.

We assume that the domain D is a Riemaniann manifold of $\mathbb{R}^p$

Example. Testing differences between two groups of functional data defined on a complex chamaleon-shaped manifold

$$y_i(x) = \mu_j(x) + \varepsilon_{ij}(x) \quad j = 1, 2; \; i = 1, \ldots, n_j$$

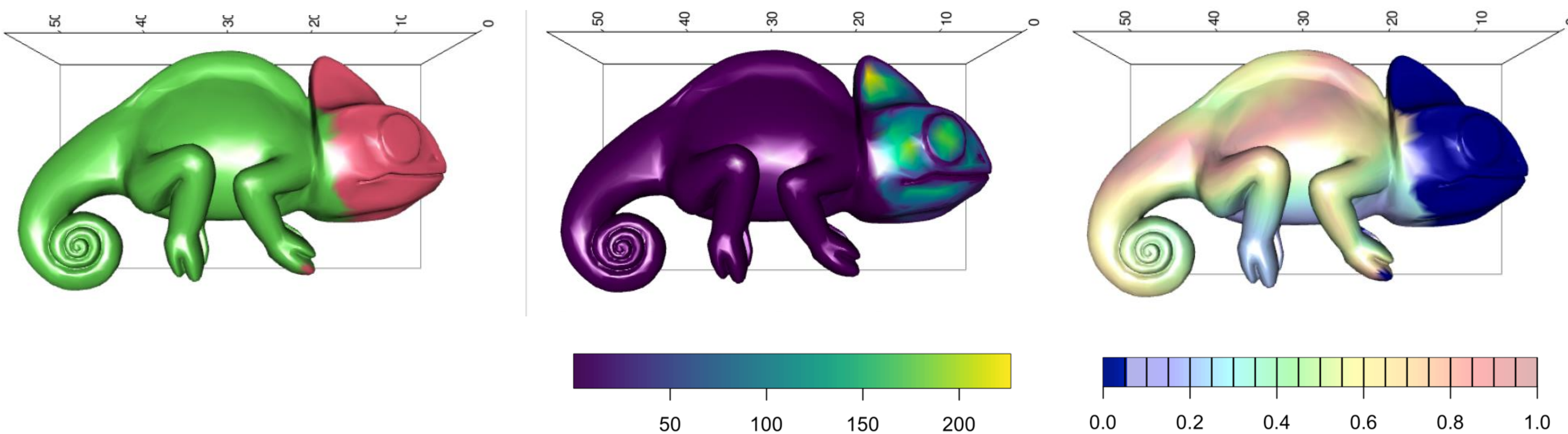
$H_0 : \mu_1(x) = \mu_2(x) \; \forall x \in D$; $\qquad H_1 : \mu_1(x) \neq \mu_2(x)$ for some $x \in D$



PROBLEM FORMULATION

Functional data on $L^2(D) \cap \mathcal{C}^0(D)$, where $D \subset \mathbb{R}^p$.
Aim: test **locally** a functional hypothesis H_0 against H_1.

True mean difference — Pointwise t-test statistic — Pointwise p-value



PROBLEM FORMULATION

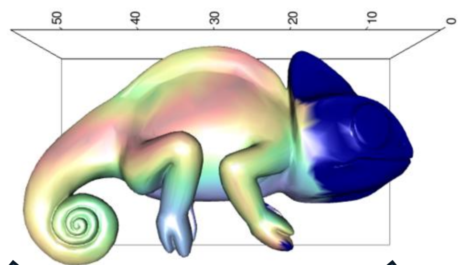
Multiplicity issue!
no control of the amount of type I errors over the domain.

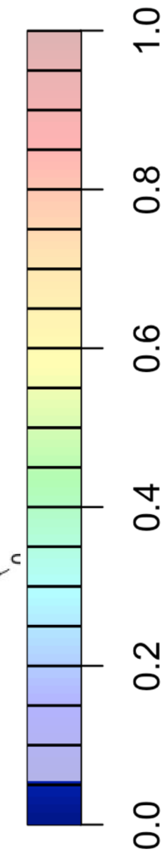
AIM

Computing an adjusted p-value function defined on D that controls an error rate defined on the whole domain.

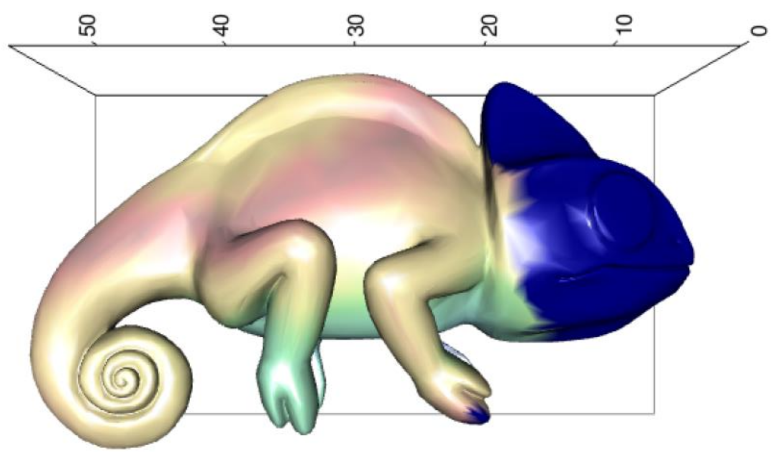
ADJUSTED P-VALUE FUNCTIONS

Pointwise p-value

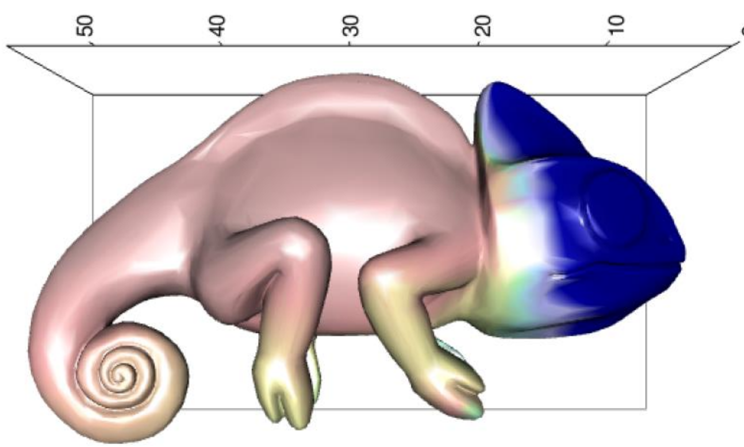




FDR adjusted p-value



Ball-wise adjusted p-value

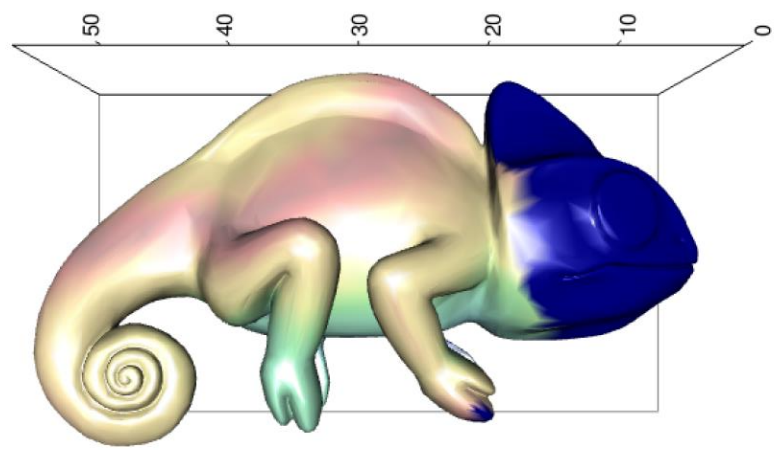


alessia.pini@unicatt.it

FDR-adjusted p-value function

Olsen et al 2021 TEST

$$\tilde{p}(t) = \min_{s \geq p(t)} \left\{ 1, \frac{\mu(D) \cdot s}{\mu(\{r : p(r) \leq s\})} \right\}$$



THE FUNCTIONAL FALSE DISCOVERY RATE

The FDR in multivariate statistics (Benjamini and Hochberg, 1995).

TABLE 1
Number of errors committed when testing m null hypotheses

	Declared non-significant	*Declared significant*	*Total*
True null hypotheses	**U**	**V**	m_0
Non-true null hypotheses	**T**	**S**	$m - m_0$
	$m - \mathbf{R}$	**R**	m

$$\mathrm{FDR} = \mathbb{E}\left[\frac{V}{R}\,1(R > 0)\right]$$

THE FUNCTIONAL FALSE DISCOVERY RATE

The FDR in functional data analysis.

	Declared non significant	**Declared significant**	
True null Hypotheses	U $\mu\{t \in D : H_{0t} \text{ true}, \tilde{p}(t) > \alpha\}$	V $\mu\{t \in D : H_{0t} \text{ true}, \tilde{p}(t) \leq \alpha\}$	m_0
False null Hypotheses	T $\mu\{t \in D : H_{0t} \text{ false}, \tilde{p}(t) > \alpha\}$	S $\mu\{t \in D : H_{0t} \text{ false}, \tilde{p}(t) \leq \alpha\}$	m_1
	$m - R$	R	m

$$\text{FDR} = \mathbb{E}\left[\frac{V}{R}\,1(R > 0)\right]$$

THE FUNCTIONAL BENJAMINI HOCHBERG PROCEDURE

1. Choose a suited functional test (either parametric or non-parametric).

$$H_0^{\mathcal{I}} : \mathcal{Y}_1^{\mathcal{I}} = \mathcal{Y}_2^{\mathcal{I}} \quad H_1^{\mathcal{I}} : \mathcal{Y}_1^{\mathcal{I}} \neq \mathcal{Y}_2^{\mathcal{I}}$$

2. Compute the unadjusted *p*-value function.

$$p(t) = \limsup_{\mathcal{I} \to t} p^{\mathcal{I}}$$

3. Perform functional FDR procedure:

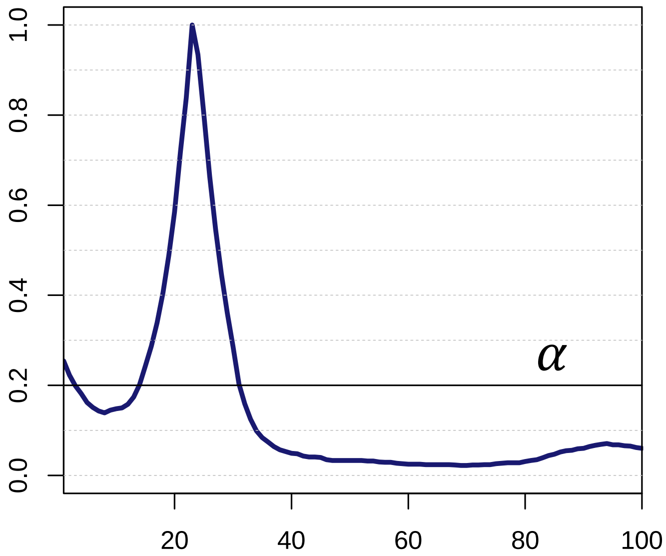
A. Adjust the threshold

$$p(t) < \alpha^*, \quad \text{with} \quad \alpha^* = \arg\max_{\bar{p}} \left\{ \frac{\alpha \mu\{r : p(r) \leq \bar{p}\}}{\mu(D)} \geq \bar{p} \right\}$$

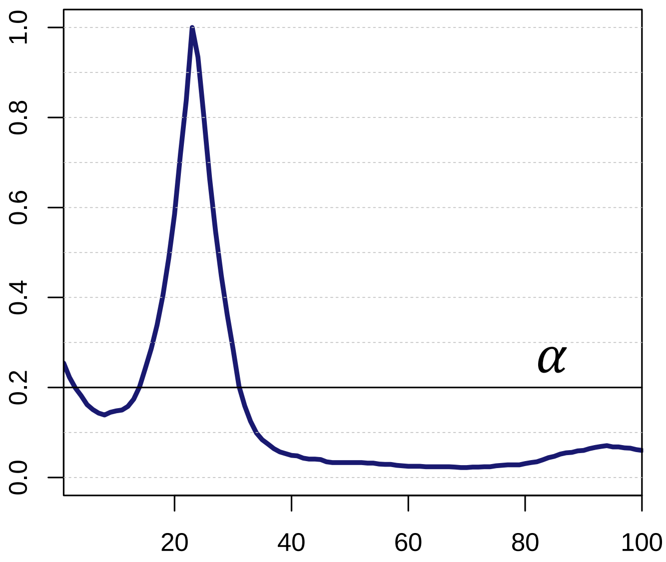
B. Adjust the *p*-value

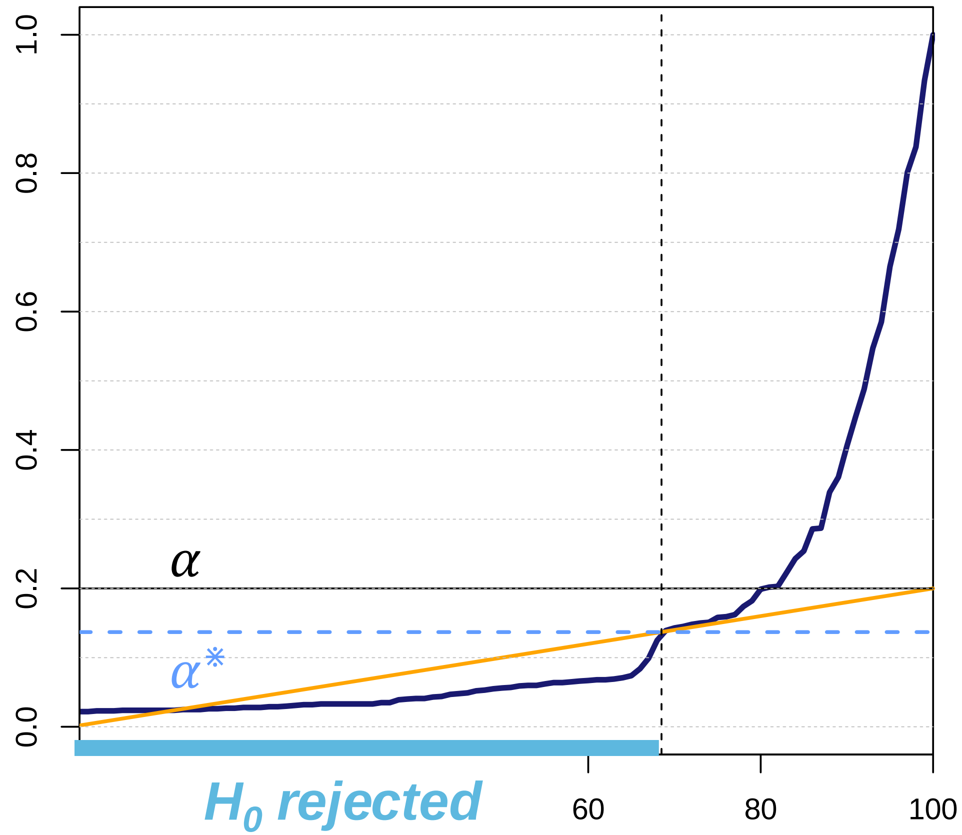
$$\tilde{p}(t) < \alpha, \quad \text{with} \quad \tilde{p}(t) = \min_{s \geq p(t)} \left\{ 1, \frac{\mu(D)s}{\mu\{r : p(r) \leq s\}} \right\}$$

ADJUSTING THE THRESHOLD



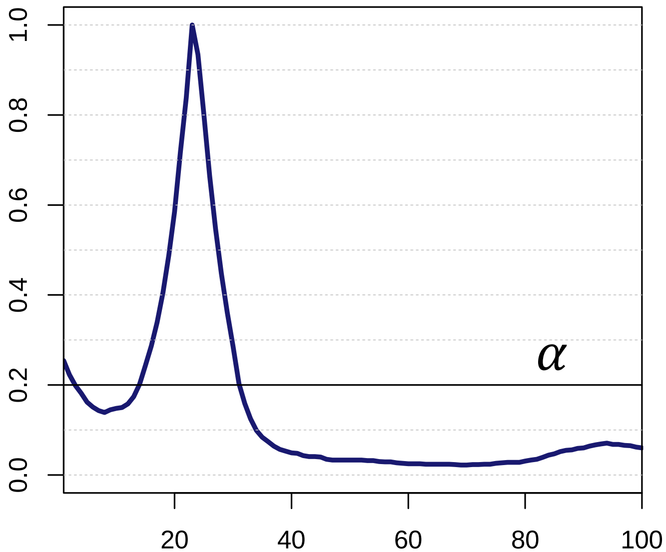
ADJUSTING THE THRESHOLD

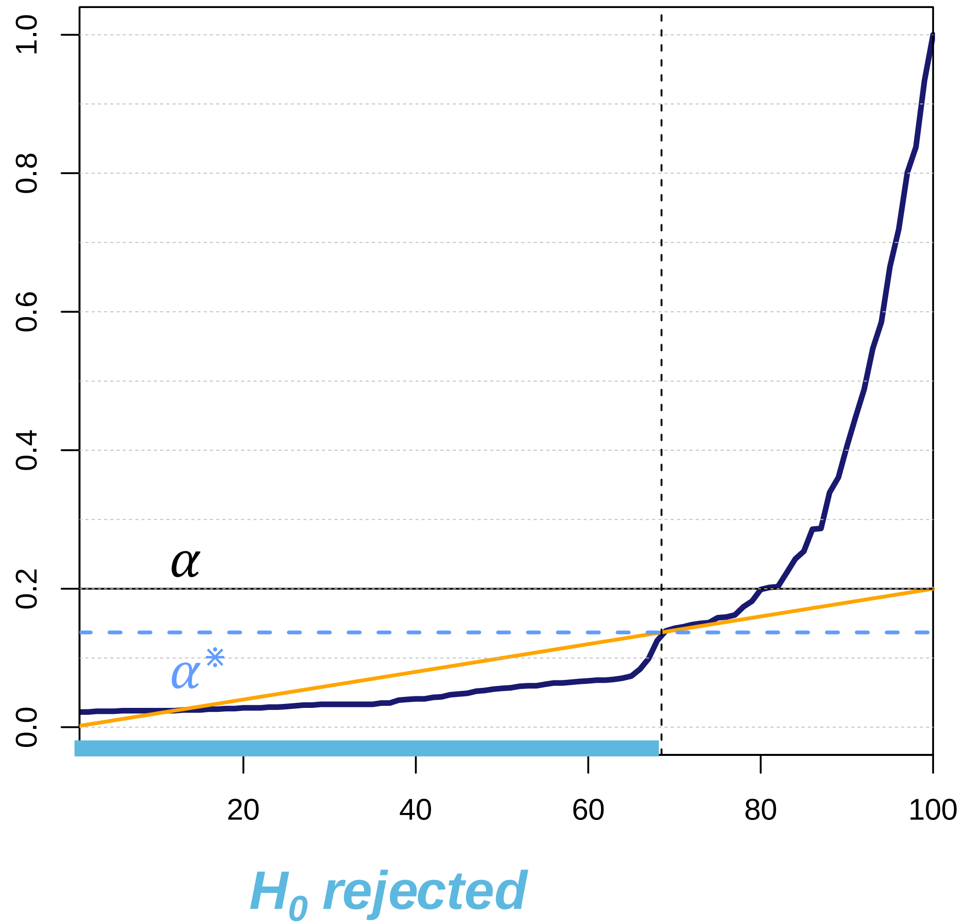


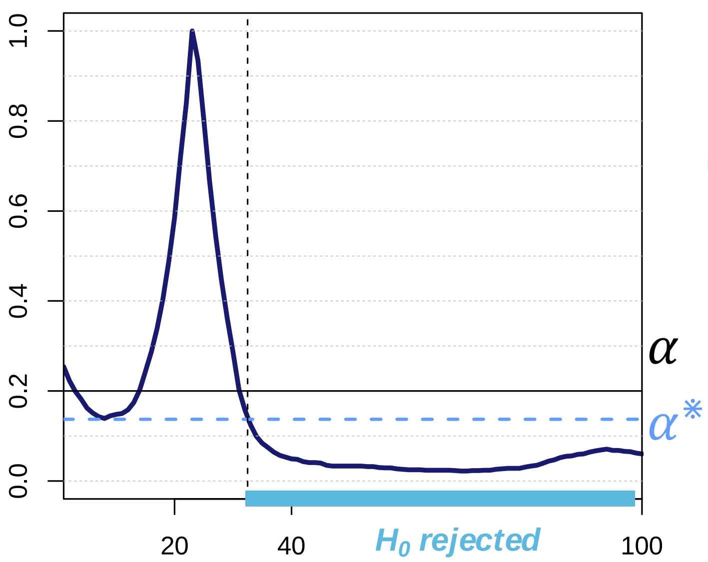


$$\alpha^* = \arg\max_{\bar{p}} \left\{ \frac{\alpha\mu\{r : p(r) \leq \bar{p}\}}{\mu(D)} \geq \bar{p} \right\}$$

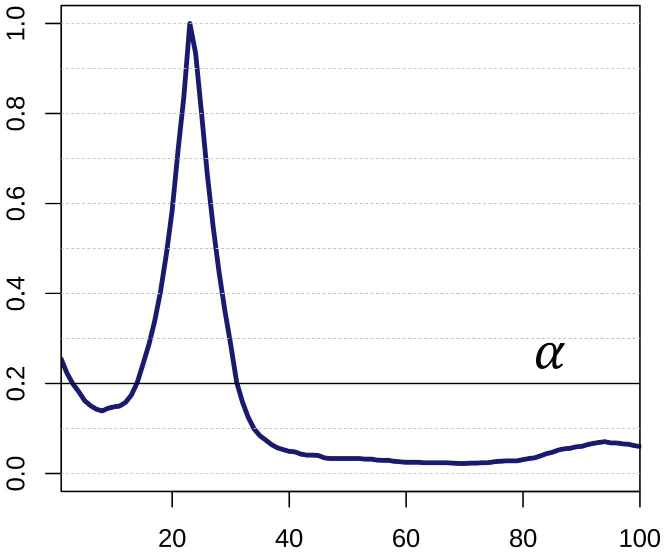
ADJUSTING THE THRESHOLD



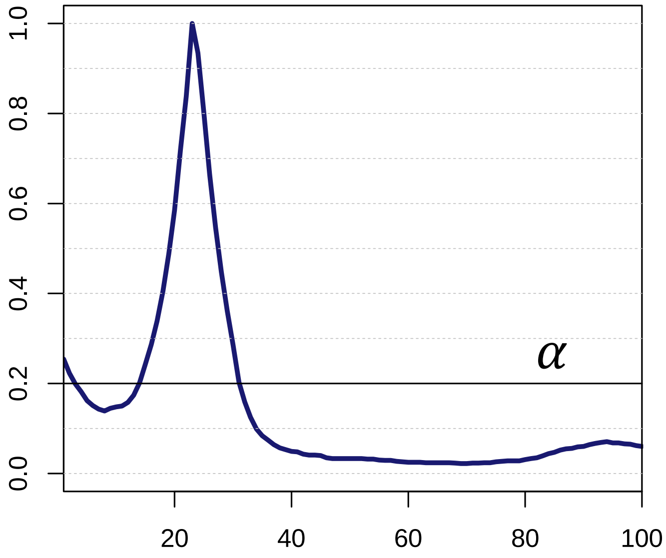


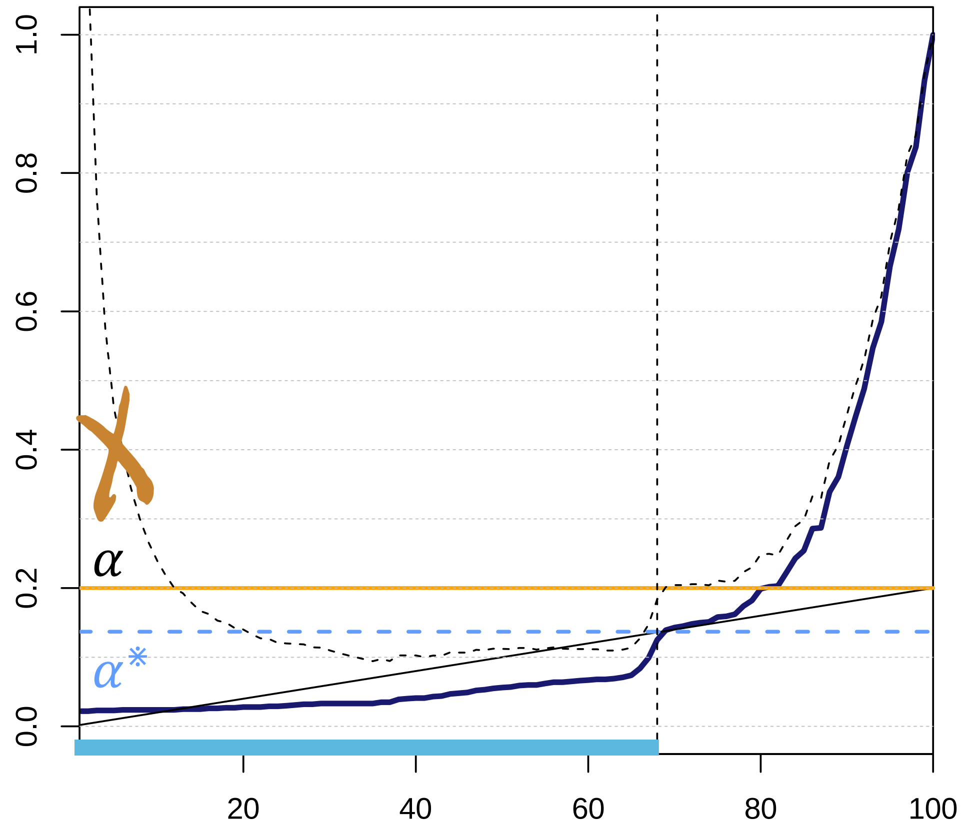


ADJUSTING THE P-VALUE FUNCTION



ADJUSTING THE P-VALUE FUNCTION



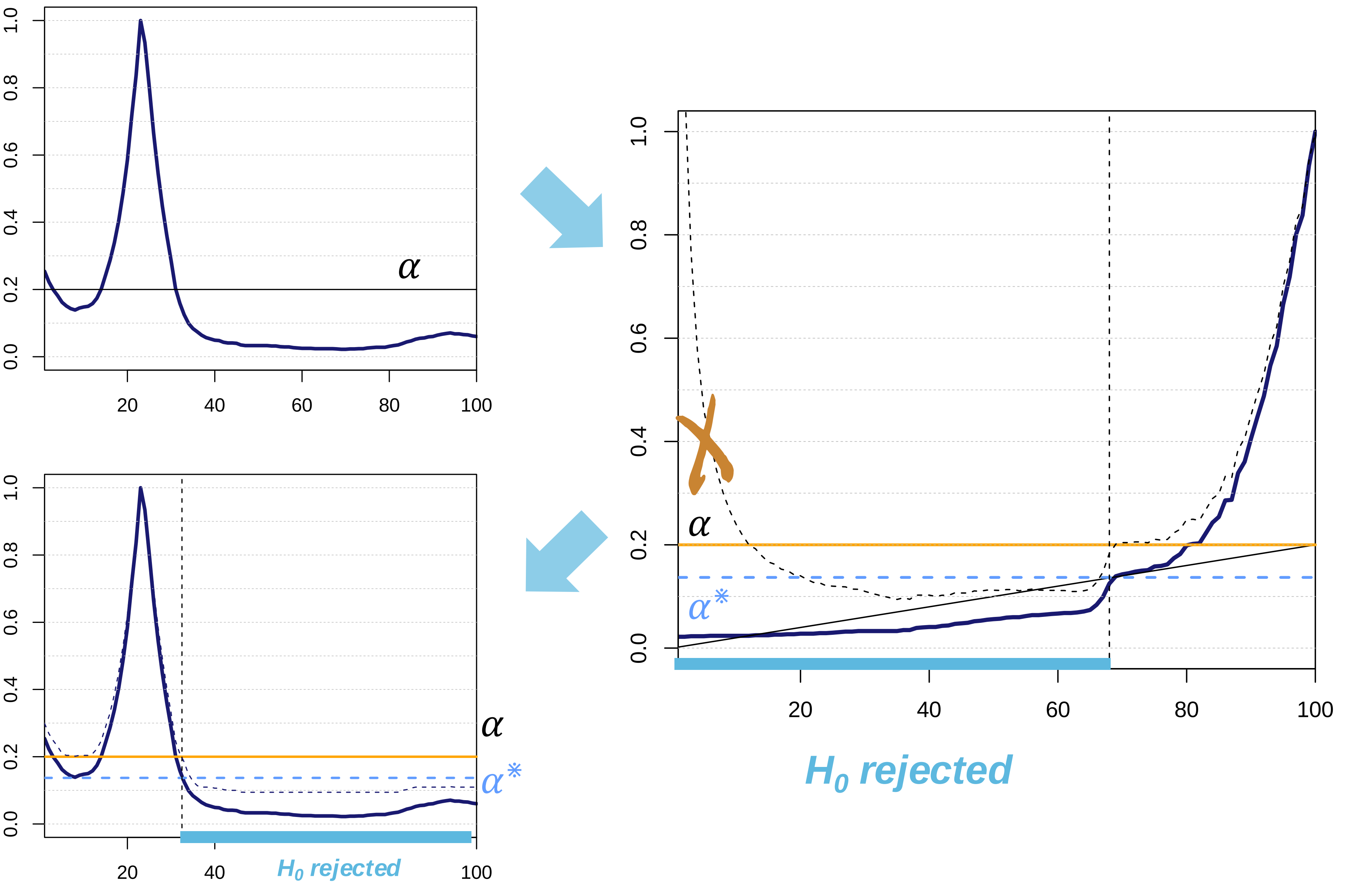


H_0 **rejected**

$$\tilde{p}(t) = \min_{s \geq p(t)} \left\{ 1, \frac{\mu(D)s}{\mu\{r : p(r) \leq s\}} \right\}$$

alessia.pini@unicatt.it

ADJUSTING THE P-VALUE FUNCTION



THE CONTROL OF THE FUNCTIONAL FALSE DISCOVERY RATE

$$\tilde{p}(t) < \alpha, \quad \text{with} \quad \tilde{p}(t) = \min_{s \geq p(t)} \left\{ 1, \frac{\mu(D)s}{\mu\{r : p(r) \leq s\}} \right\}$$

$$p(t) < \alpha^*, \quad \text{with} \quad \alpha^* = \arg\max_{\bar{p}} \left\{ \frac{\alpha\mu\{r : p(r) \leq \bar{p}\}}{\mu(D)} \geq \bar{p} \right\}$$

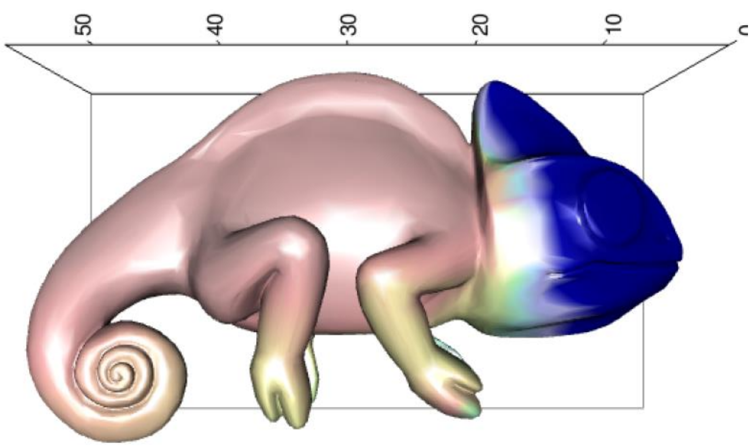
Control of the functional FDR:

$$FDR = \mathbb{E}\left[\frac{V}{R}1(R > 0)\right] \leq \alpha \frac{\mu(D_0)}{\mu(D)} \leq \alpha$$

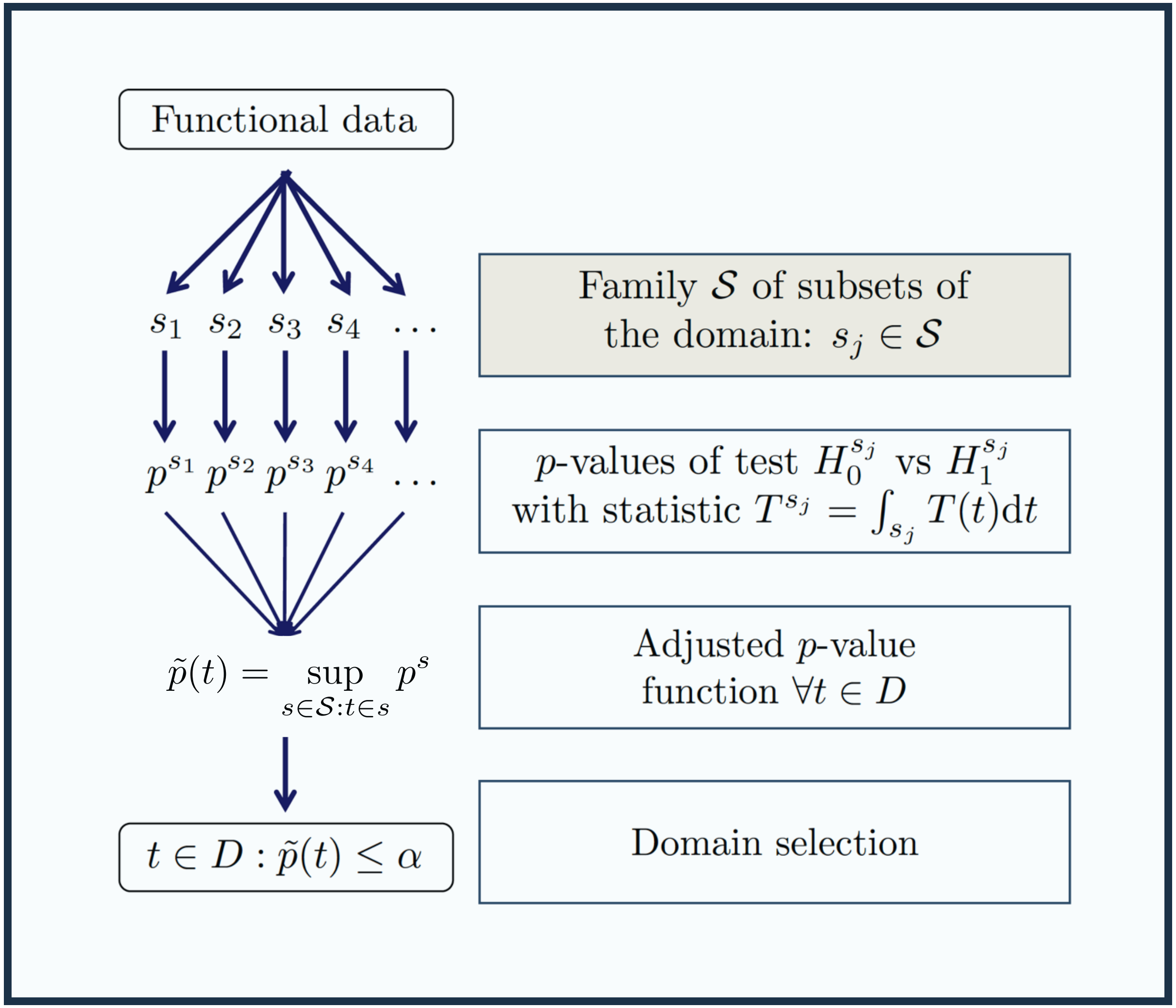
Ball-wise adjusted p-value function

Olsen et al 2025+

$$\tilde{p}(t) = \sup_{\mathcal{I} \ni t} p^{\mathcal{I}}$$

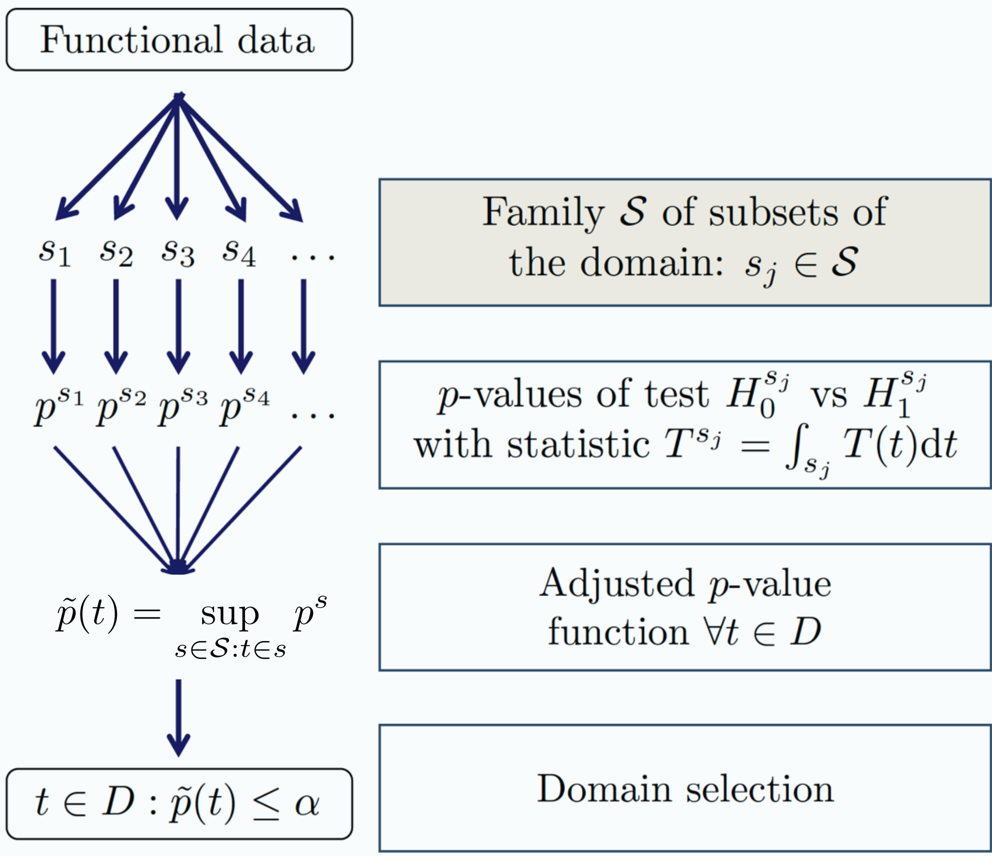


UNIFIED FRAMEWORK FOR LOCAL TESTING



alessia.pini@unicatt.it

UNIFIED FRAMEWORK FOR LOCAL TESTING



Local adjustment

adjustment subsets should take into account proximity

One dimensional domain: adjustment sets are intervals.

How to choose suitable adjustment subsets when the domain is a manifold?

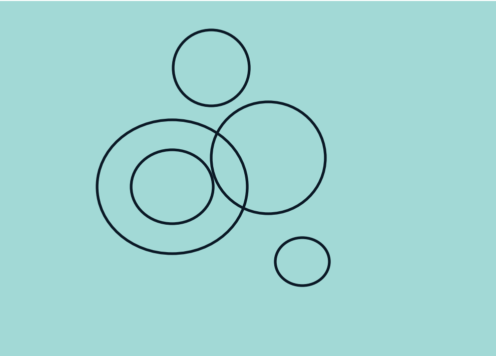
ADJUSTMENT SUBSETS

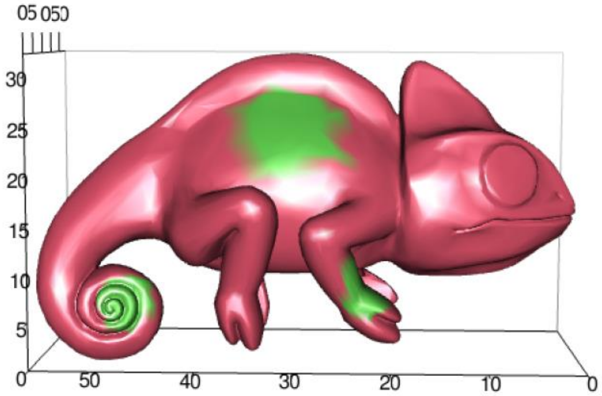
Idea: use balls as adjustment sets.
For a given manifold D with metric d, define $B(x;\epsilon)$ (the ball of center x and radius ϵ) as

$$B(x;\epsilon) = \{y \in D | d(x;y) < \epsilon\}\,;\; x \in D; \epsilon > 0$$

The adjustment family $\mathcal{S}$ consists of the collection of all balls of D of radius smaller than a constant r:

$$\mathcal{S} = \{B(x;\epsilon)\}_{x \in D, \epsilon \leq r}$$





ADJUSTMENT SUBSETS

Idea: use balls as adjustment sets.
For a given manifold D with metric d, define $B(x;\epsilon)$ (the ball of center x and radius ϵ) as

$$B(x;\epsilon) = \{y \in D | d(x;y) < \epsilon\}\,;\; x \in D; \epsilon > 0$$

The adjustment family $\mathcal{S}$ consists of the collection of all balls of D of radius smaller than a constant r:

$$\mathcal{S} = \{B(x;\epsilon)\}_{x \in D, \epsilon \leq r}$$

Theorem 1. *$\tilde{p}(x)$ controls the ball-wise error rate:*

$$\forall \alpha \in (0,1), \quad \forall B(x,\epsilon) \subseteq D : H_0^I \text{ is true}, \mathbb{P}(\exists s \in I : \tilde{p}(s) \leq \alpha) \leq \alpha$$

alessia.pini@unicatt.it

ICE COVER DATA ANALYSIS

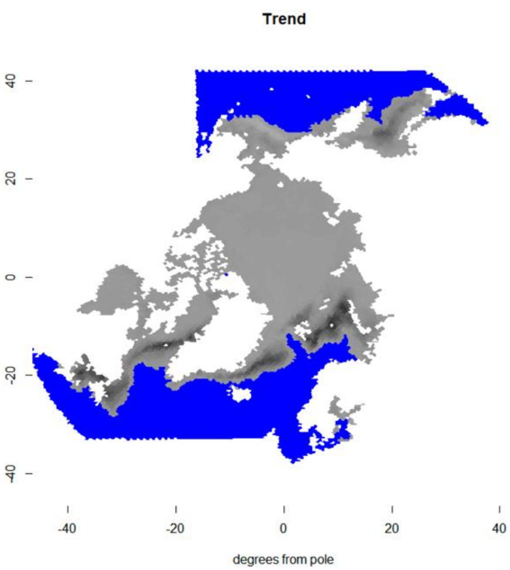
Example. Yearly measurements (1987-2015) of ice cover on the northern hemisphere, as measured by the satellites of Copernicus Programme. The aim is to test for significant change in ice cover during the period.

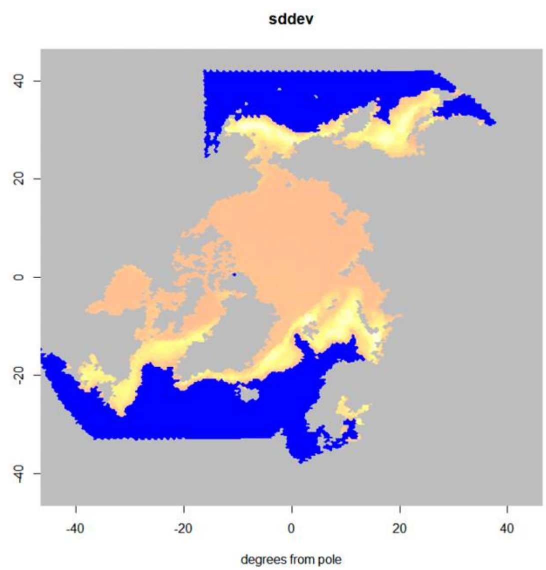
$$y_i(x) = a(x) + b(x)t_i + \varepsilon_i(x)$$

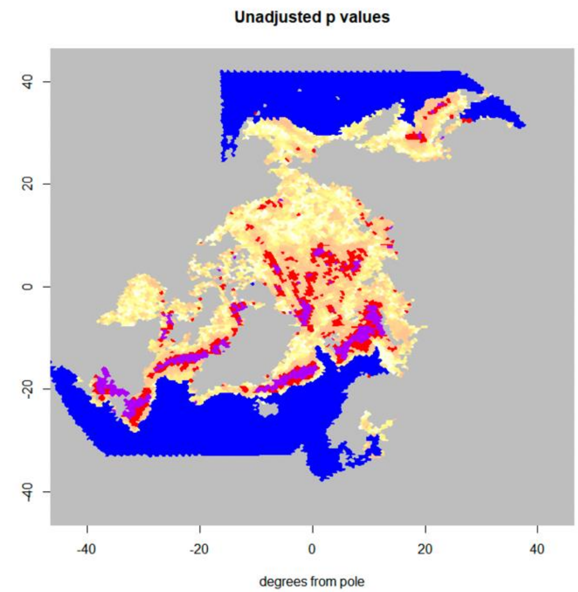
$$H_0 : b(x) = 0 \; \forall x \in D; \qquad H_1 : b(x) \neq 0 \text{ for some } x \in D$$

PROBLEM FORMULATION

Example. Yearly measurements (1987-2015) of ice cover on the northern hemisphere, as measured by the satellites of Copernicus Programme. The aim is to test for significant change in ice cover during the period.





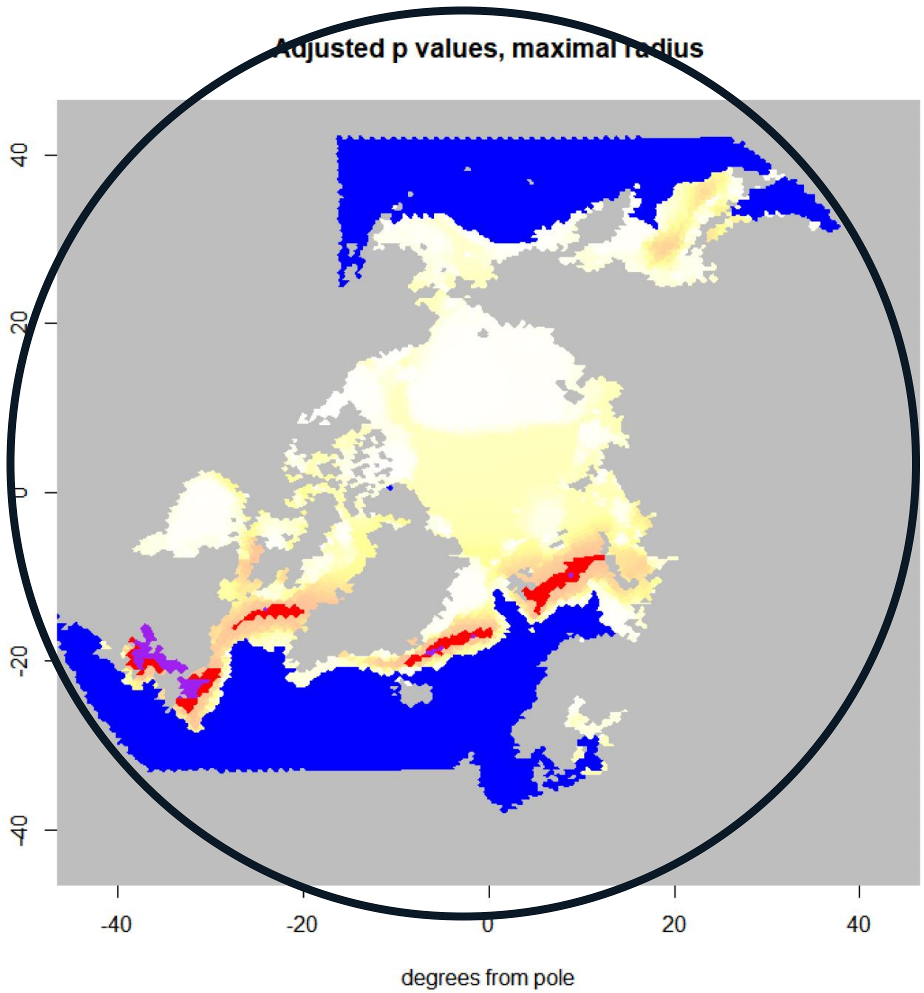


Constant zero ice cover
Negative trend
Zero trend

Constant zero ice cover
Small sd
Large sd

Constant zero ice cover
p-value smaller than 5%
p-value smaller than 1%

ICE COVER DATA ANALYSIS



Constant zero ice cover

Adjusted p-value smaller than 5%

Adjusted p-value smaller than 1%

DISCUSSION

- A method for local inference of functional data with manifold domains.
- A novel way to adjust for multiplicity when the domain is complex.
- The adjustment methods can be plugged-in with every available testing procedure.

Future works

- Are balls the only available type of adjustment sets for local adjustment?
- Is it possible to tune r with a data driven approach?
- Is it possible to develop a local adjustment for the false discovery rate?

Niels Lundtorp Olsen, Alessia Pini, and Simone Vantini. False discovery rate for functional data. *Test*, 2021.

Niels Lundtorp Olsen, Alessia Pini, and Simone Vantini. Local inference for functional data on manifold domains using permutation tests https://arxiv.org/abs/2306.07738

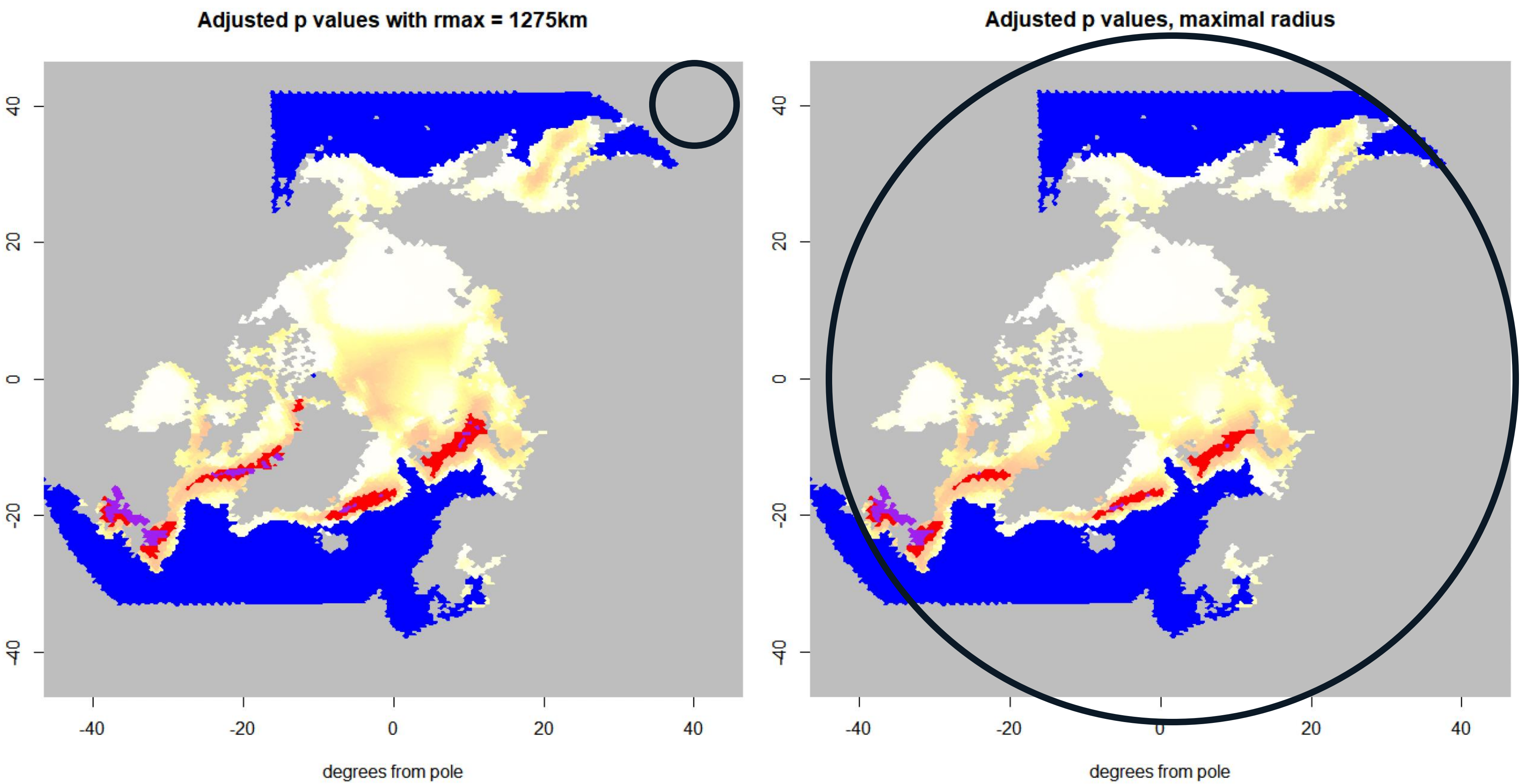
alessia.pini@unicatt.it

SELECTED REFERENCES

- Konrad Abramowicz, Alessia Pini, Lina Schelin, Sara Sj□ostedt de Luna, Aymeric Stamm, and Simone Vantini. Domain selection and familywise error rate for functional data: A unified framework. *Biometrics*, 2022.
- Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society Series B*, 1995.
- David Freedman and David Lane. A nonstochastic interpretation of reported significance levels. *Journal of Business & Economic Statistics*, 1(4) (1983).
- Sture Holm: A simple sequentially rejective multiple test procedure *Scandinavian Journal of Statistics*, 6(2) (1979)
- Dominik Liebl and Matthew Reimherr. Fast and fair simultaneous confidence bands for functional parameters. *Journal of the royal statistical society Series B*, 2023.
- Niels Lundtorp Olsen, Alessia Pini, and Simone Vantini. False discovery rate for functional data. *Test*, 2021.
- Ruth Marcus, Eric Peritz, and K. R. Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3) (1976).
- Alessia Pini and Simone Vantini. Interval-wise testing for functional data. *Journal of Nonparametric Statistics*, 29(2), 2017.
- Olga Vsevolozhskaya, Marc Greenwood, and Dmitri Holodov. Pairwise comparison of treatment levels in functional analysis of variance with application to erythrocyte hemolysis. *The Annals of Applied Statistics*, 8 (2014).

THANK YOU FOR YOUR ATTENTION!

ICE COVER DATA ANALYSIS



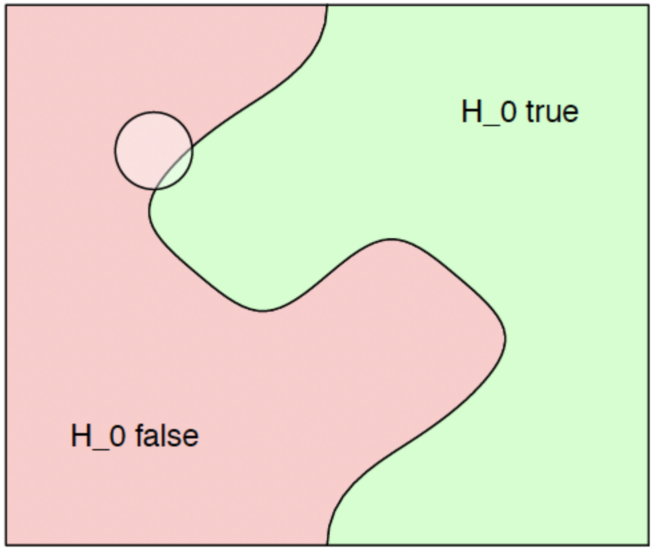
Constant zero ice cover
Adjusted p-value smaller than 5%
Adjusted p-value smaller than 1%

alessia.pini@unicatt.it

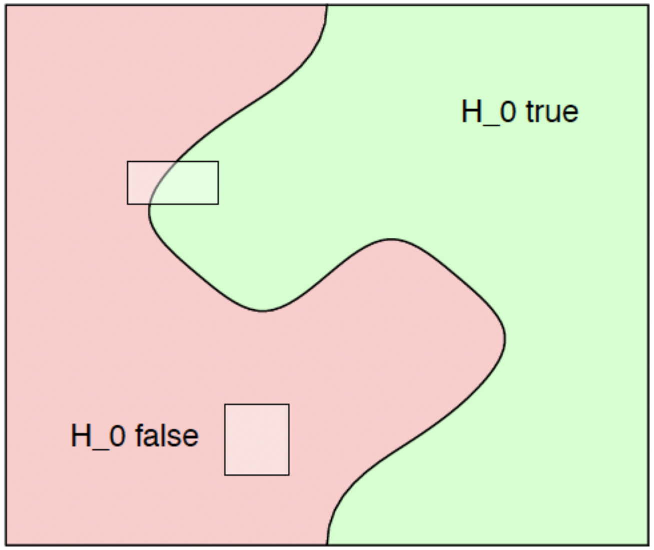
EFFECT OF THE METRIC

- The control that is obtained is strictly related to the type of adjustment sets that are used, which depends on the metric.
- By changing it, one can define different adjustment sets:

Spatial domain



Spatiotemporal domain



alessia.pini@unicatt.it

EFFECT OF THE RADIUS

- The radius r also influences the adjustment family and hence the control.
- Two extreme cases:

$$r \to 0$$

Adjustment subsets collapse to points. No adjustment is performed and:

$$\tilde{p}(x) = p(x).$$

The power is maximized, but the error control is minimal.

$$r \to \infty$$

Adjustment subsets include all possible balls. The procedure is very conservative since the adjustment family is very large. The power is minimized, but the error control is maximal.

alessia.pini@unicatt.it

CURRENT LITERATURE OF LOCAL INFERENCE

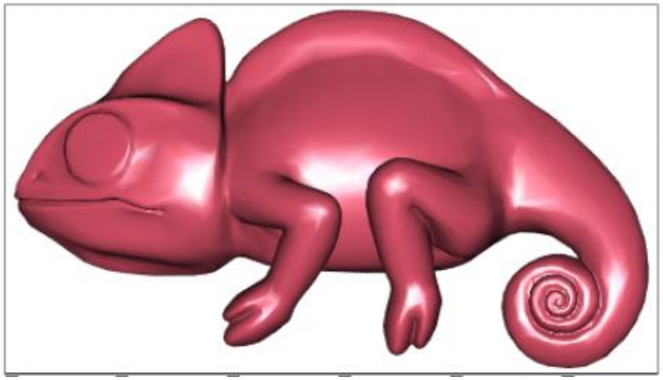
- One dimensional domains
- Manifold domains

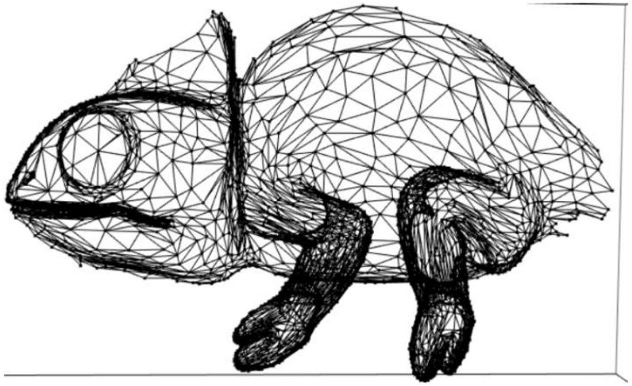
- Global adjustment: the adjustment is based on the pointwise p-values only
- Local adjustment: the adjustment is based on the topological structure of the domain

	One dimensional D	Manifold D
Global adjustment	Multivariate approaches (Holm, Bonferroni, Benjamini Hochberg) Functional FDR (Olsen etal 2021)	**Functional FDR (Olsen etal 2021)** Threshold Wise Testing (Abramowicz etal 2022)
Local adjustment	Interval Wise Testing (Pini Vantini 2017) Functional confidence bands (Liebl Rheimerr 2023) Partition closed testing (Vsevolozhskaya etal 2014)	**Ball Wise Testing**

NUMERICAL IMPLEMENTATION

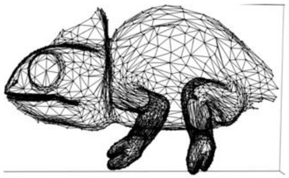
- Triangulation to approximate the points of the manifold.



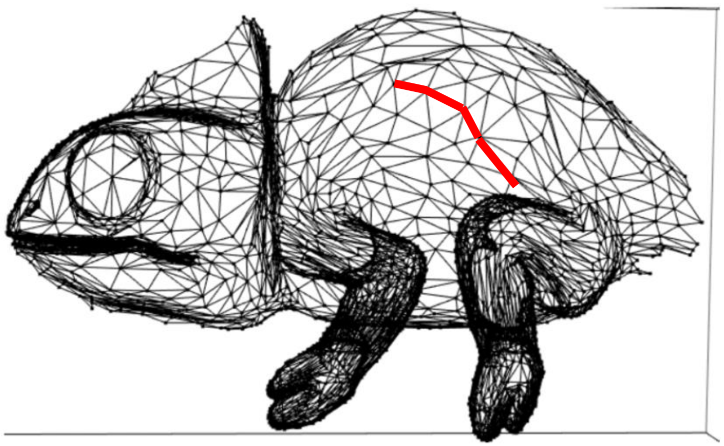


NUMERICAL IMPLEMENTATION

- Triangulation to approximate the points of the manifold.

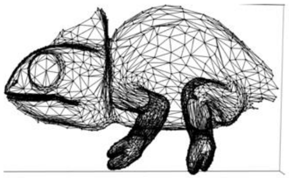


- Dijkstra's algorithm to compute geodesic distance on the manifold.

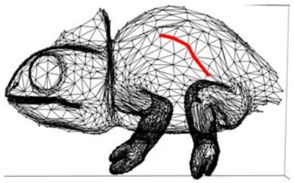


NUMERICAL IMPLEMENTATION

- Triangulation to approximate the points of the manifold.



- Dijkstra's algorithm to compute geodesic distance on the manifold.



- Integral approximation for computing the test statistic on balls.

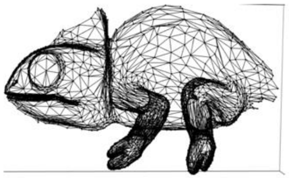
$$\int_{B(x,r)} f(u)\,\mathrm{d}u = \sum_{\{v \in E : d(x,e) < r\}} W(e) f(e)$$

$$W(e) = \frac{1}{3} \sum_{S : e \text{ is a vertex of } S} A(S), \quad e \in E$$

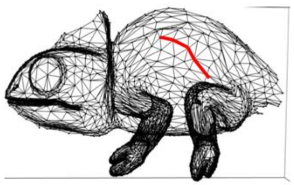
alessia.pini@unicatt.it

NUMERICAL IMPLEMENTATION

- Triangulation to approximate the points of the manifold.



- Dijkstra's algorithm to compute geodesic distance on the manifold.



- Integral approximation for computing the test statistic on balls.

$$\int_{B(x,r)} f(u)\,\mathrm{d}u = \sum_{\{v\in E:d(x,e)<r\}} W(e)f(e)$$

$$W(e) = \frac{1}{3}\sum_{S:e \text{ is a vertex of } S} A(S), \quad e \in E$$

- Permutation tests to evaluate the p-value of tests on balls.

$$y_i(x) = a(x) + b(x)t_i + \varepsilon_i(x)$$

$$\text{Under } H_0 : y_i(x) = a(x) + \varepsilon_i(x)$$

$$\boxed{\widehat{\varepsilon}_i(x)} = y_i(x) - \widehat{a}(x)$$

Freedman and Lane method: Permutation of the estimated residuals under the null hypothesis