

# Estimación robusta basada en interdistancias para correlación canónica funcional.

21 de Septiembre de 2016

Kudraszow Nadia L. (IMAS-UBA-CONICET/UNLP)

Graciela Boente (IMAS-UBA-CONICET)

LXV Reunión anual de comunicaciones científicas de la UMA  
Bahía Blanca



Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico

Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación

Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

Análisis de Correlación Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta basada en interdistancias

Análisis robusto de correlación canónica funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de suavizado  
Resultados

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico

Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

2

Sean  $\mathbf{x}$  e  $\mathbf{y}$  vectores aleatorios con dimensiones  $p$  y  $q$ , con matrices de covarianzas  $\Gamma_{11}$  and  $\Gamma_{22}$  respectivamente, y  $\Gamma_{12}$  la matriz que representa la covarianza entre  $\mathbf{x}$  e  $\mathbf{y}$ .

**OBJETIVO:** encontrar  $\{(\mathbf{t}_k, \mathbf{v}_k)\}_{k=1}^K$ , con  $K = \text{rango}(\text{Cov}(\mathbf{y}, \mathbf{x}))$ , tales que

$$(\mathbf{t}_k, \mathbf{v}_k) = \arg \max_{(\mathbf{t}, \mathbf{v})} \{\text{Corr}(\mathbf{t}'\mathbf{y}, \mathbf{v}'\mathbf{x})\}$$

donde  $(\mathbf{t}, \mathbf{v})$  satisfacen

$$\text{Var}(\mathbf{t}'\mathbf{y}) = \text{Var}(\mathbf{v}'\mathbf{x}) = 1$$

y

$$\text{Corr}(\mathbf{t}'_j\mathbf{y}, \mathbf{t}'_i\mathbf{y}) = \text{Corr}(\mathbf{v}'_j\mathbf{x}, \mathbf{v}'_i\mathbf{x}) = 0,$$

para  $k \geq 2$  y  $j = 1, \dots, k - 1$ .

25

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

3

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

25

Los vectores  $\{\mathbf{t}_k\}_1^K$  y  $\{\mathbf{v}_k\}_1^K$  son llamados **variables canónicas** de  $\mathbf{y}$  y  $\mathbf{x}$  respectivamente y

$$c_k = \text{Corr}(\mathbf{t}'_k \mathbf{y}, \mathbf{v}'_k \mathbf{x}) = \frac{\mathbf{t}'_k \Gamma_{12} \mathbf{v}_k}{\sqrt{\mathbf{t}'_k \Gamma_{11} \mathbf{t}_k} \sqrt{\mathbf{v}'_k \Gamma_{22} \mathbf{v}_k}}$$

es llamada la  **$k$ -ésima correlación canónica**.

**Versión Muestral:**

- ▶ Se reemplazan las matrices de covarianzas poblacionales por sus respectivas versiones muestrales.
- ▶  $n > p + q$  asegura que la matriz de covarianzas de  $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$  sea no singular.
- ▶ El método clásico es sensible a observaciones atípicas debido a que está basado en matrices de covarianzas muestrales.

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico

Multivariado

Funcional

Suavizado Funcional

4

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación

Muestras Simuladas

Contaminación

Estimadores

Parámetro de  
suavizado

Resultados

25

## Consideremos

- ▶  $\mathcal{H}$  un espacio de Hilbert separable,
- ▶  $\{(X_i(t), Y_i(t)), i = 1, \dots, n\}$  realizaciones independientes de un proceso estocástico bivariado  $(X(t), Y(t)) \in \mathcal{H} \times \mathcal{H}$ ,
- ▶  $t$  en un intervalo finito  $\mathcal{J}$ .

## Asumimos

- ▶  $EX = EY = 0$   
 $\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{21} \\ \Gamma_{12} & \Gamma_{22} \end{pmatrix}$  es el operador de covarianza donde
- ▶  $\Gamma_{11}$  y  $\Gamma_{22}$  son los operadores de covarianza de  $X$  e  $Y$ , respectivamente,
- ▶  $\Gamma_{12}$  y  $\Gamma_{21}$  son los operadores de covarianza cruzados.

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

5

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

25

Una aproximación no suave al problema de correlación canónica funcional consiste en buscar  $u, v \in \mathcal{H}$ , no nulos, que maximicen

$$\text{corr}(\langle u, X \rangle, \langle v, Y \rangle) = \frac{\langle u, \mathbf{\Gamma}_{12} v \rangle^2}{\langle u, \mathbf{\Gamma}_{11} u \rangle \langle v, \mathbf{\Gamma}_{22} v \rangle}.$$

Los estimadores de  $u, v$  se obtienen al maximizar la respectiva correlación al cuadrado muestral

$$\frac{\langle u, \hat{\mathbf{\Gamma}}_{12} v \rangle^2}{\langle u, \hat{\mathbf{\Gamma}}_{11} u \rangle \langle v, \hat{\mathbf{\Gamma}}_{22} v \rangle}. \quad (1)$$

**Problema:** Maximizar (1) no provee información relevante sobre los datos o el modelo (Leurgans, Moyeed y Silverman (1993)).

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

6

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

25

## Suavizado

- ▶  $D : \mathcal{H}_S \subset \mathcal{H} \rightarrow \mathcal{H}$  es un operador lineal al que llamaremos **diferenciación**, donde  $\alpha \in \mathcal{H}_S$  si  $\|D\alpha\| < \infty$ .
- ▶  $\Psi : \mathcal{H}_S \rightarrow \mathbb{R}$  con

$$\Psi(\alpha) = \langle D\alpha, D\alpha \rangle$$

será el **operador de penalización**.

Dados  $u, v \in \mathcal{H}_S$ , la **correlación canónica poblacional suavizada al cuadrado** está definida por

$$\frac{\langle u, \Gamma_{12}v \rangle^2}{(\langle u, \Gamma_{11}u \rangle + \tau\Psi(u)) (\langle v, \Gamma_{22}v \rangle + \tau\Psi(v))}.$$

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico

Multivariado  
Funcional  
Suavizado Funcional

7

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

25

## Análisis de Correlación Canónica Suavizado

Consiste en encontrar  $\{(u_k, v_k)\}_{k=1}^K \subset \mathcal{H}_S \times \mathcal{H}_S$ , tales que

$$(u_k, v_k) = \arg \max_{(u, v)} \frac{\langle u, \mathbf{\Gamma}_{12} v \rangle^2}{(\langle u, \mathbf{\Gamma}_{11} u \rangle + \tau \Psi(u)) (\langle v, \mathbf{\Gamma}_{22} v \rangle + \tau \Psi(v))}$$

donde  $(u, v)$  satisfacen

$$\langle u, \mathbf{\Gamma}_{11} u \rangle = \langle v, \mathbf{\Gamma}_{22} v \rangle = 1$$

y

$$\langle u_j, \mathbf{\Gamma}_{11} u \rangle = \langle v_j, \mathbf{\Gamma}_{22} v \rangle = 0,$$

para  $k \geq 2$  y  $j = 1, \dots, k-1$ .

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico

Multivariado  
Funcional  
Suavizado Funcional

7

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

25

## Análisis de Correlación Canónica Suavizado

Consiste en encontrar  $\{(u_k, v_k)\}_{k=1}^K \subset \mathcal{H}_S \times \mathcal{H}_S$ , tales que

$$(u_k, v_k) = \arg \max_{(u, v)} \frac{\langle u, \Gamma_{12} v \rangle^2}{(\langle u, \Gamma_{11} u \rangle + \tau \Psi(u)) (\langle v, \Gamma_{22} v \rangle + \tau \Psi(v))}$$

donde  $(u, v)$  satisfacen

$$\langle u, \Gamma_{11} u \rangle = \langle v, \Gamma_{22} v \rangle = 1$$

y

$$\langle u_j, \Gamma_{11} u \rangle = \langle v_j, \Gamma_{22} v \rangle = 0,$$

para  $k \geq 2$  y  $j = 1, \dots, k-1$ .

**Estimación:** Se reemplaza  $\Gamma_{ij}$  por su versión empírica  $\widehat{\Gamma}_{ij}$  obtenida del operador de covarianza muestral.

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

8

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

25

**Estimación:** Para calcular  $\{(\hat{u}_k, \hat{v}_k)\}_{k=1}^K$  se reemplaza  $\Gamma_{ij}$  por su versión empírica  $\hat{\Gamma}_{ij}$  obtenida del operador de covarianza muestral.

Por otro lado  $\{(\hat{u}_k, \hat{v}_k)\}_{k=1}^K$  pueden ser caracterizados encontrando los  $K$  autovalores  $\rho$  más grandes del sistema de ecuaciones

$$\begin{pmatrix} \mathbf{0} & \hat{\Gamma}_{12} \\ \hat{\Gamma}_{21} & \mathbf{0} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \rho \begin{pmatrix} \hat{\Gamma}_{11} + \tau D^2 & \mathbf{0} \\ \mathbf{0} & \hat{\Gamma}_{22} + \tau D^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

Es bien sabido que estos estimadores no son robustos: una pequeña proporción de datos atípicos podría causar un gran efecto sobre sus valores.

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

9

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

Para obtener estimadores robustos se propondrá reemplazar los operadores de covarianza muestrales por operadores de covarianza robustos, como los propuestos por Gervini (2010).

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

9

25

Para obtener estimadores robustos se propondrá reemplazar los operadores de covarianza muestrales por operadores de covarianza robustos, como los propuestos por Gervini (2010).

## Estimadores robustos de posición y escala funcionales:

- ▶ Sea  $X_1, \dots, X_n$  una muestra en un espacio de Hilbert  $\mathcal{H}$ .
- ▶ Una observación  $X_i$  es un valor atípico si está lejos de la mayoría de las otras observaciones.
- ▶ Para detectar las observaciones atípicas se utilizan las **interdistancias**  $d_{ij} = \|X_i - X_j\|$ .

# Estimación robusta basada en interdistancias

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

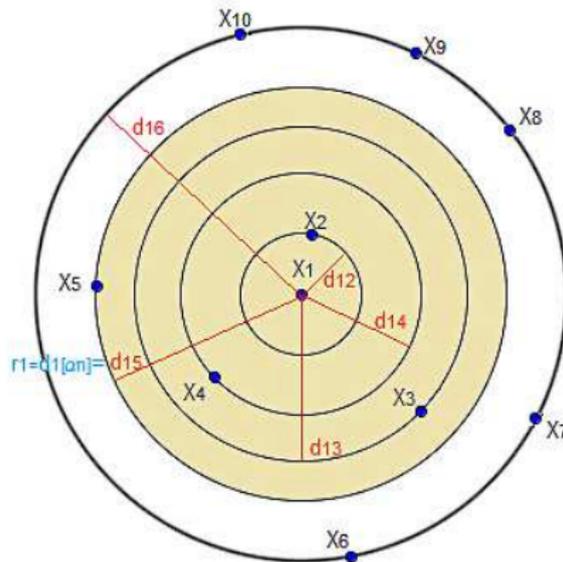
Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

10

25

- ▶ Dado  $0 \leq \alpha \leq 1/2$ .
- ▶ Llamamos  $\alpha$ -radio  $r_i$  a la distancia al  $[\alpha n]$ -ésimo punto más cercano.

Ejemplo: Si  $\alpha = 0.5$  y  $n = 10$  tenemos  $r_1 = d_{1[\alpha n]} = d_{15}$ .



Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico

Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

11

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

25

- ▶ El siguiente paso es eliminar o darle menos peso a las observaciones con radios más grandes.
- ▶ Dada **proporción de recorte**

$$\beta \in [0, 0.5].$$

- ▶ Podemos utilizar dos pesos distintos
  1. **Escalonado**: para una poda **“fuerte”** se utiliza

$$w(X_i) = \mathbb{I}\{r_i < r_{(\lceil(1-\beta)n\rceil)}\},$$

donde  $\mathbb{I}\{\cdot\}$  es la función indicador.

2. **Suave**: para una poda **“leve”** se utiliza

$$g(r) = \begin{cases} 1 & 0 \leq r \leq a, \\ (r-b) \left[ \frac{1}{a-b} + \frac{(r-a)\{2r-(a+b)\}}{(b-a)^3} \right] & a \leq r \leq b, \\ 0, & r \geq b \end{cases}$$

donde  $a = 1 - \beta_1$  con  $\beta_1 > \beta$  y  $b = 1 - \beta$ .

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

12

25

## Estimador robusto para la media

$$\hat{\mu}_R = \frac{1}{\sum_{i=1}^n w(X_i)} \sum_{i=1}^n w(X_i) X_i .$$

## Estimador robusto para el operador de covarianza

$$\hat{\mathbf{C}} = \frac{1}{\sum_{i=1}^n w(X_i)} \sum_{i=1}^n w(X_i) (X_i - \hat{\mu}_R) \otimes (X_i - \hat{\mu}_R) .$$

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico

Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

Las **direcciones canónicas** robustas se estiman encontrando las autofunciones del sistema

$$\begin{pmatrix} \mathbf{0} & \hat{\mathbf{C}}_{12} \\ \hat{\mathbf{C}}_{21} & \mathbf{0} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \rho \begin{pmatrix} \hat{\mathbf{C}}_{11} + \tau D^2 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{C}}_{22} + \tau D^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix},$$

donde  $\hat{\mathbf{C}}_{ij}$  es la estimación robusta de  $\mathbf{\Gamma}_{ij}$ .

**Problema:** El orden de los autovalores no necesariamente coincide con el orden de las correlaciones canónicas al cuadrado.

Sea  $\rho_R$  un funcional de correlación robusto, las direcciones canónicas robustas son ordenadas según los valores de

$$\rho_R^2(P_n[\langle u_k, X \rangle, \langle v_k, Y \rangle]),$$

que a su vez estiman las correlaciones canónicas robustas al cuadrado.

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico

Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

14

25

Basados en He, Müller y Wang (2004) se simularon muestras de tamaño 100 de los procesos

$$X(t) = \sum_{i=1}^k \xi_i \theta_i(t) \quad y \quad Y(t) = \sum_{i=1}^k \zeta_i \theta_i(t),$$

donde  $k = 21$  y  $\{\theta_n\}_{n \in \mathbb{N}}$  es la base de Fourier en  $[0, 50]$ .

- ▶ Los  $\{(\xi, \zeta)\}_{i=1}^{21}$  fueron generados de una  $\mathcal{N}(\mathbf{0}, \mathbf{A})$  con

$$\mathbf{A} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21}^T & \mathbf{R}_{22} \end{pmatrix}$$

donde

- ▶  $\mathbf{R}_{11} = \text{diag}\{10 \cdot 0.75^i\}$ ,  $i = 0, 0, 0, 1, 2, \dots, 18$ .
- ▶  $\mathbf{R}_{22} = \mathbf{R}_{11}$
- ▶  $\mathbf{R}_{12} = \text{diag}(7, 3, 1, 0, \dots, 0)$ .
- ▶ Se realizaron 100 replicas.

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico

Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación

Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

15

25

## Correlaciones canónicas:

- ▶  $\rho_1 = 0.7,$
- ▶  $\rho_2 = 0.3,$
- ▶  $\rho_3 = 0.1,$
- ▶  $\rho_i = 0$  si  $i > 3.$

## Direcciones canónicas:

- ▶  $\phi_1(t) = \psi_1(t) = \sqrt{0.1}\theta_1(t) = \sqrt{0.1/T},$
- ▶  $\phi_2(t) = \psi_2(t) = \sqrt{0.1}\theta_2(t) = \sqrt{0.2/T} \sin((t - T/2)2\pi/T),$
- ▶  $\phi_3(t) = \psi_3(t) = \sqrt{0.1}\theta_3(t) = \sqrt{0.2/T} \cos((t - T/2)2\pi/T).$

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
**Contaminación**  
Estimadores  
Parámetro de  
suavizado  
Resultados

16

También se simularon muestras con un 10% de los datos contaminados de la siguiente manera:

$$(X_c, Y_c) = (1 - S) \cdot (X, Y) + S \cdot (Z\theta_2, Z\theta_2)$$

donde

- ▶  $(X, Y)$  son los procesos antes definidos,
- ▶  $S \sim Be(0.1)$
- ▶  $Z \sim \mathcal{N}(25, 1)$ .

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

16

25

También se simularon muestras con un 10% de los datos contaminados de la siguiente manera:

$$(X_c, Y_c) = (1 - S) \cdot (X, Y) + S \cdot (Z\theta_2, Z\theta_2)$$

donde

- ▶  $(X, Y)$  son los procesos antes definidos,
- ▶  $S \sim Be(0.1)$
- ▶  $Z \sim \mathcal{N}(25, 1)$ .

Las correlaciones canónicas de  $(X_c, Y_c)$  son:

$$\rho_{c,1} \approx 0.9, \quad \rho_{c,2} = 0.7, \quad \rho_{c,3} = 0.1 \text{ y } \rho_{c,i} = 0 \text{ si } i > 3$$

y las direcciones canónicas son:

$$\begin{aligned} \phi_{c,1} &= \psi_{c,1} = \sqrt{0.1}\theta_2 \\ \phi_{c,2} &= \psi_{c,2} = \sqrt{0.1}\theta_1 \\ \phi_{c,3} &= \psi_{c,3} = \sqrt{0.1}\theta_3. \end{aligned}$$

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

17

25

- ▶ **Clásico:** Estimador clásico del análisis de correlación canónica funcional utilizando como operador de diferenciación una discretización del operador segunda derivada.
- ▶ **Robustos:** Estimadores para los operadores de covarianza
  - ▶ **con podado “fuerte”:** Con peso “escalonado” utilizando  $\alpha = 0.5$  y  $\beta = 0.1$ .
  - ▶ **con podado “leve”:** Con peso “suave” utilizando  $\alpha = \beta_1 = 0.5$  y  $\beta = 0.1$ .

Se utilizó el mismo operador de diferenciación que para el estimador robusto.

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

## Estimadores para las correlaciones canónicas

- ▶ **CL:** Estimador de Pearson.
- ▶ **OGKmad:** Utiliza el estimador de la matriz de covarianzas bivariada de Gnanadesikan-Kettenring para calcular el funcional de correlación, usando como escala la mad.
- ▶ **OGKtau:** Idem Rob–OGKmad pero usando un tau-estimador de escala.
- ▶ **M:** Utiliza el M-estimador de la matriz de covarianzas bivariada.
- ▶ **Sp:** Funcional de correlación de Spearman.

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico

Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación

Muestras Simuladas  
Contaminación  
Estimadores

Parámetro de  
suavizado  
Resultados

19

25

El parámetro  $\tau$  fue elegido mediante validación cruzada de la siguiente manera:

- ▶  $(\hat{u}_{\alpha,1}^{(i)}, \hat{v}_{\alpha,1}^{(i)})$  son las primeras direcciones canónicas estimadas usando el parámetro  $\alpha$  y calculadas con toda la muestra salvo el dato  $(X_i, Y_i)$ .
- ▶ Se calculan las proyecciones

$$\{(\hat{u}_{\alpha,1}^{(i)}, X_i)\}_{i=1}^n \text{ y } \{(\hat{v}_{\alpha,1}^{(i)}, Y_i)\}_{i=1}^n.$$

- ▶ Usamos como parámetro  $\alpha$  al valor que maximiza la correlación (robusta o clásica según el funcional que se utilice para calcular las correlaciones canónicas) de las proyecciones calculadas antes.

Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación

Muestras Simuladas

Contaminación

Estimadores

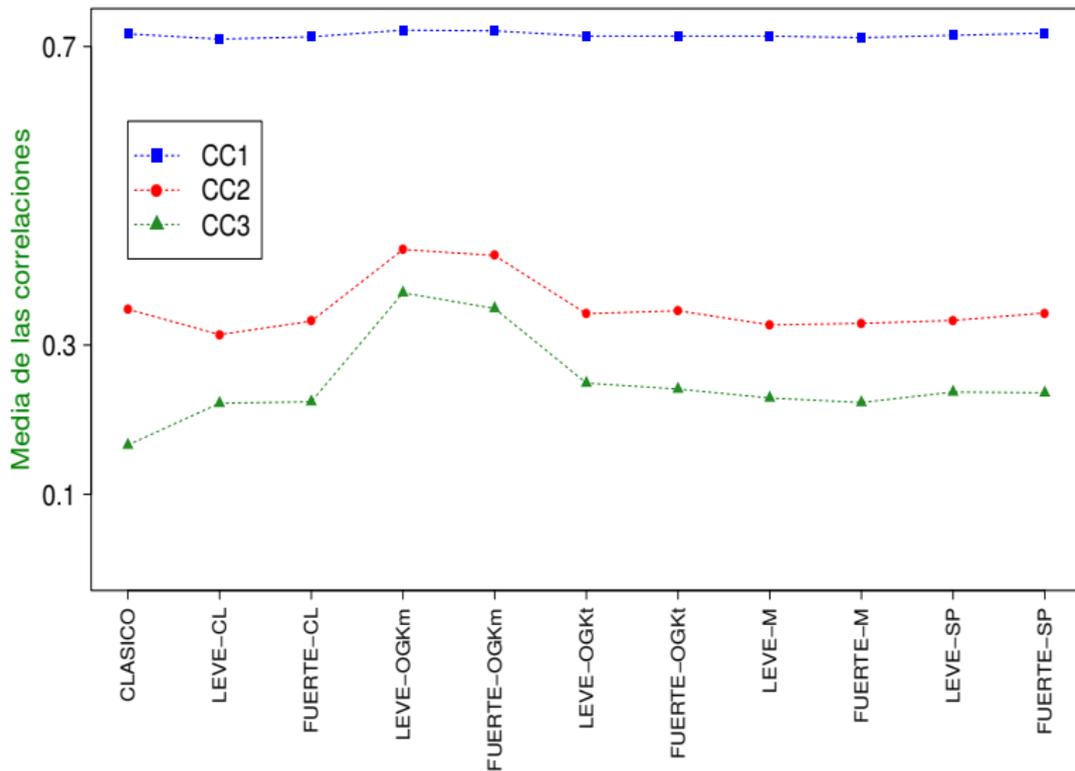
Parámetro de  
suavizado

Resultados

20

25

### Medias de la 1ra, 2da y 3er correlación canónica estimada sin contaminación



Estimación robusta  
basada en  
interdistancias para  
CCF.

Kudraszow Nadia L.  
(IMAS-UBA-  
CONICET/UNLP)

Graciela Boente  
(IMAS-UBA-  
CONICET)

Análisis de Correlación  
Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta  
basada en  
interdistancias

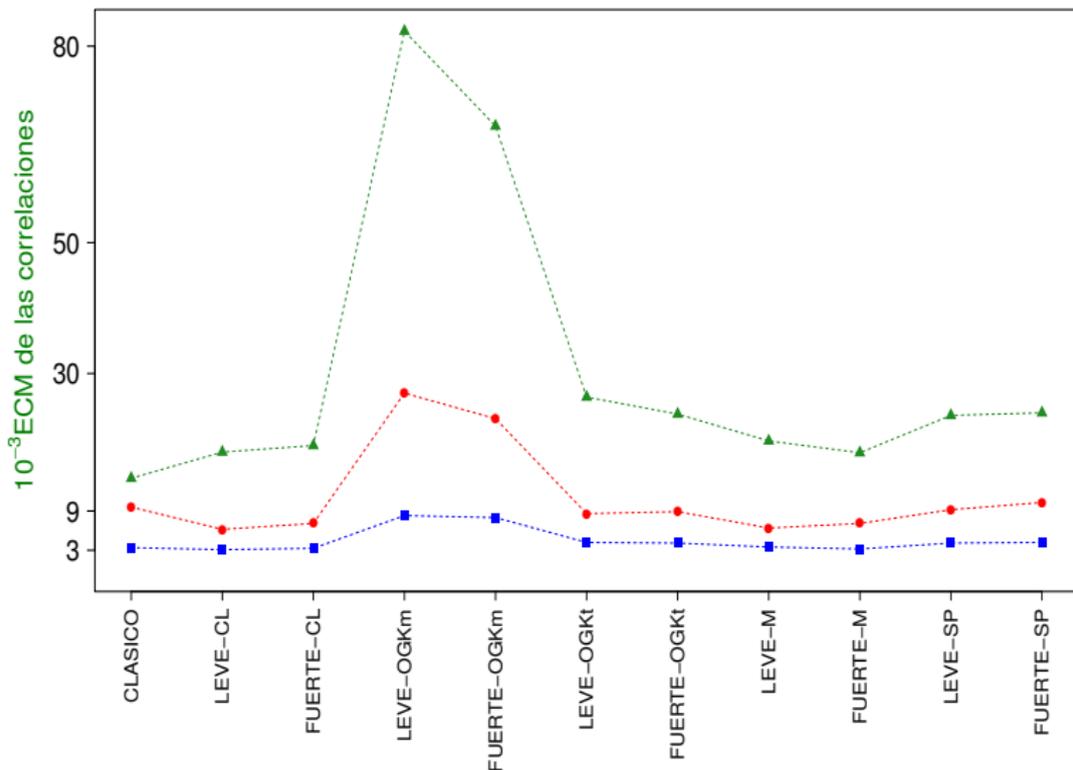
Análisis robusto de  
correlación canónica  
funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de  
suavizado  
Resultados

21

25

### ECM de la 1ra, 2da y 3er correlación canónica estimada sin contaminación



Estimación robusta basada en interdistancias para CCF.

Kudraszow Nadia L. (IMAS-UBA-CONICET/UNLP)

Graciela Boente (IMAS-UBA-CONICET)

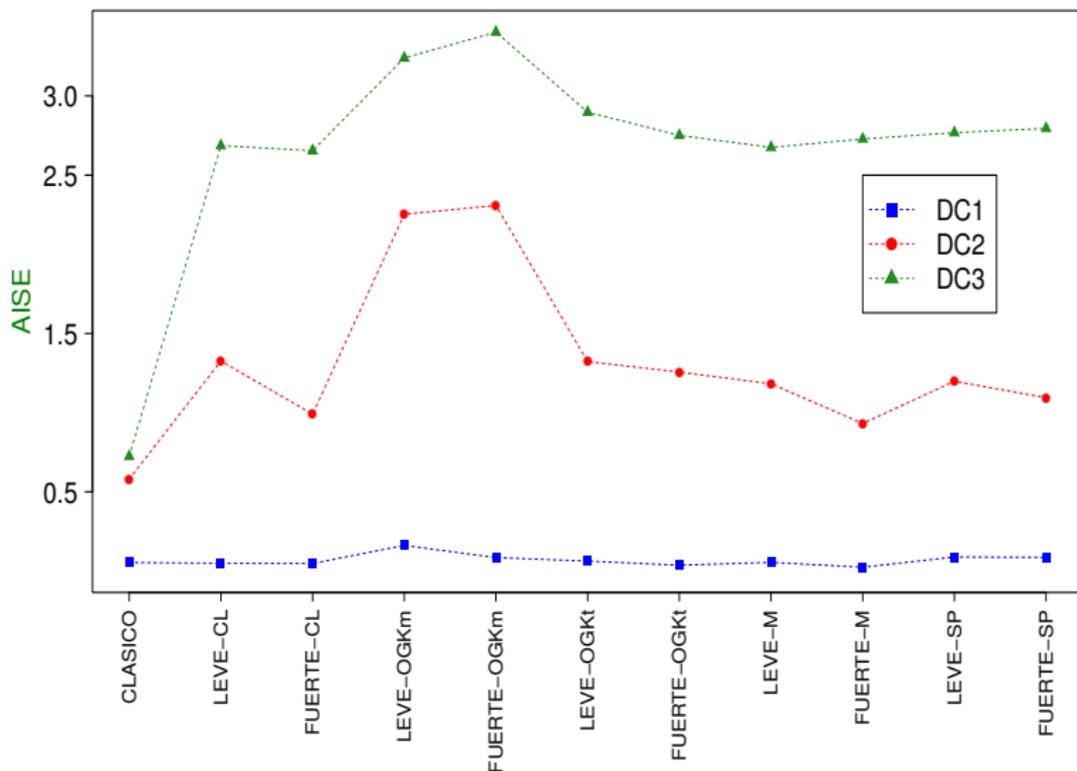
Análisis de Correlación Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta basada en interdistancias

Análisis robusto de correlación canónica funcional

Estudio de Simulación  
Muestras Simuladas  
Contaminación  
Estimadores  
Parámetro de suavizado  
Resultados

### AISE de las 1ras, 2das y 3ras direcciones canónicas estimadas sin contaminación



Estimación robusta basada en interdistancias para CCF.

Kudraszow Nadia L. (IMAS-UBA-CONICET/UNLP)

Graciela Boente (IMAS-UBA-CONICET)

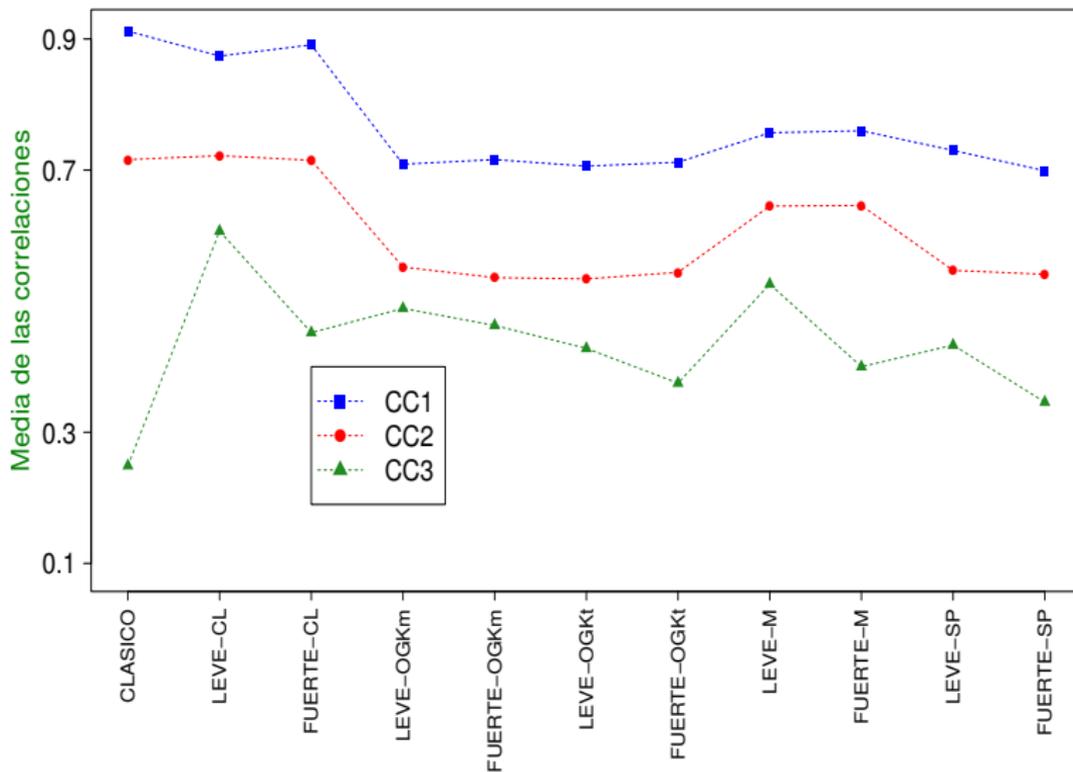
Análisis de Correlación Canónica Clásico Multivariado Funcional Suavizado Funcional

Estimación robusta basada en interdistancias

Análisis robusto de correlación canónica funcional

Estudio de Simulación Muestras Simuladas Contaminación Estimadores Parámetro de suavizado Resultados

**Medias de la 1ra, 2da y 3er correlación canónica estimada con contaminación**



Estimación robusta basada en interdistancias para CCF.

Kudraszow Nadia L. (IMAS-UBA-CONICET/UNLP)

Graciela Boente (IMAS-UBA-CONICET)

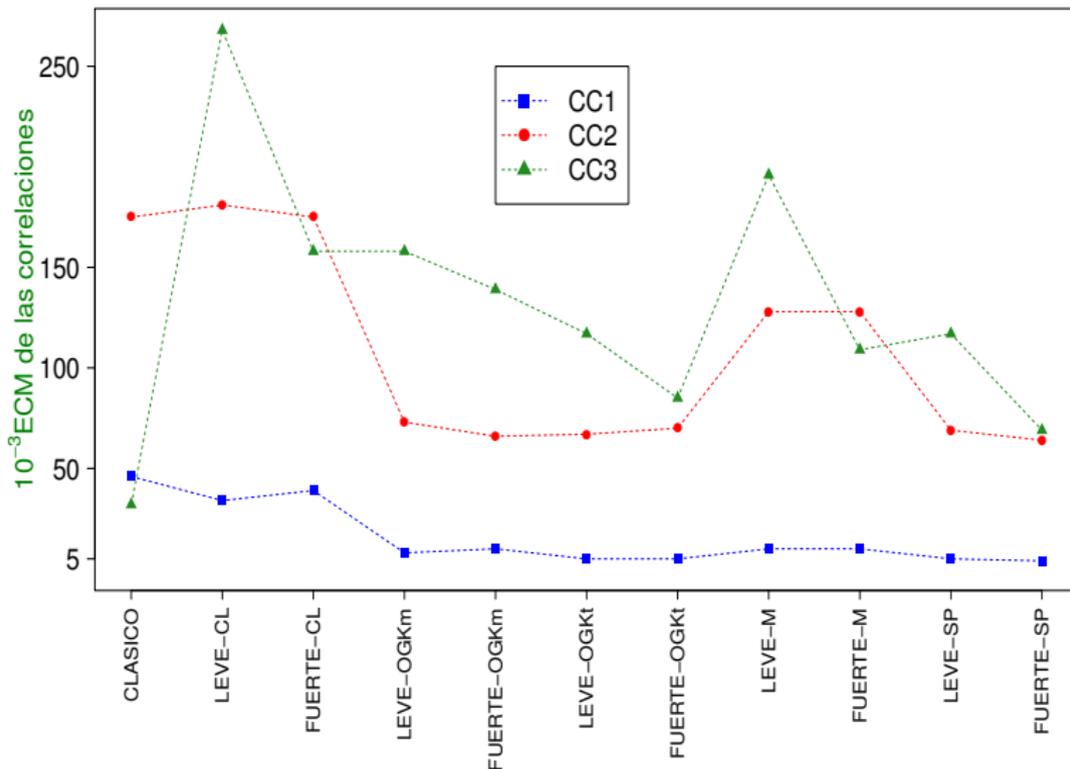
Análisis de Correlación Canónica Clásico Multivariado Funcional Suavizado Funcional

Estimación robusta basada en interdistancias

Análisis robusto de correlación canónica funcional

Estudio de Simulación Muestras Simuladas Contaminación Estimadores Parámetro de suavizado Resultados

ECM de la 1ra, 2da y 3er correlación canónica estimada con contaminación



Estimación robusta basada en interdistancias para CCF.

Kudraszow Nadia L. (IMAS-UBA-CONICET/UNLP)

Graciela Boente (IMAS-UBA-CONICET)

Análisis de Correlación Canónica Clásico  
Multivariado  
Funcional  
Suavizado Funcional

Estimación robusta basada en interdistancias

Análisis robusto de correlación canónica funcional

Estudio de Simulación

Muestras Simuladas

Contaminación

Estimadores

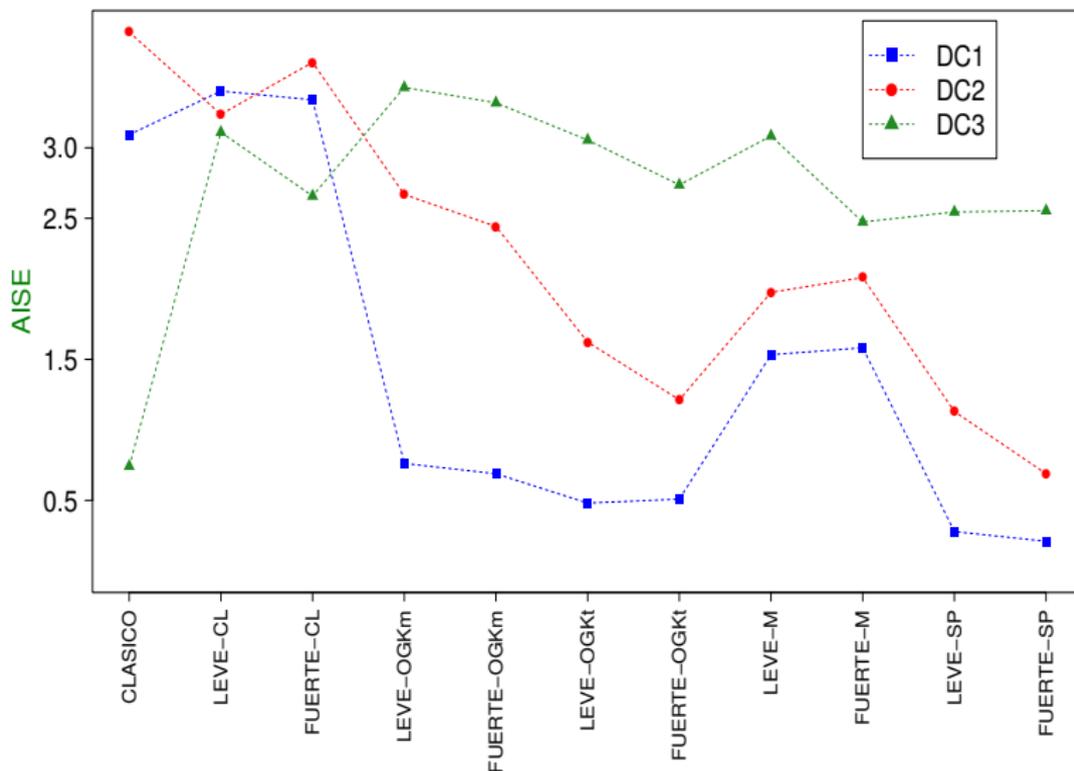
Parámetro de suavizado

Resultados

25

25

### AISE de las 1ras, 2das y 3ras direcciones canónicas estimadas con contaminación



Gracias!

