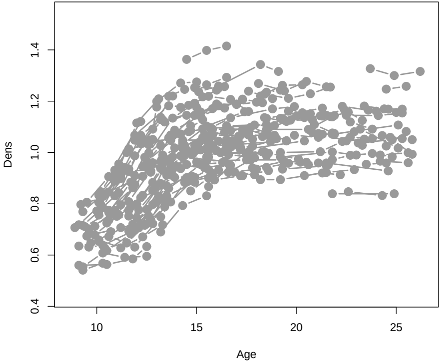
Componentes Principales Funcionales (FPCA)
Parte 2

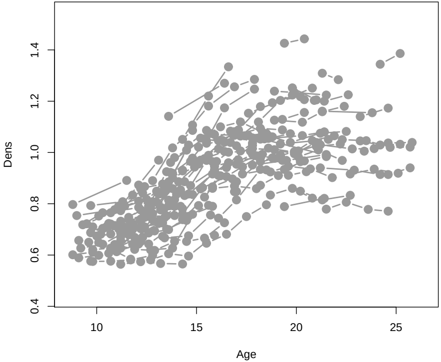
Graciela Boente

Densidad ósea

Mujeres ($n = 153$)



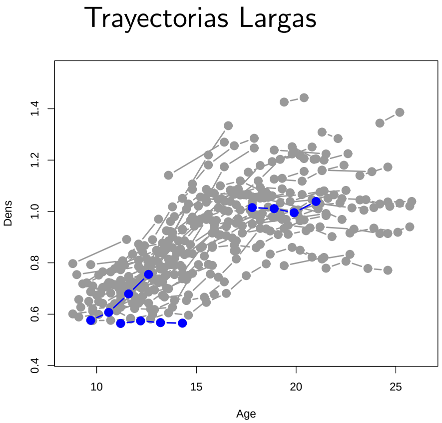
Hombres ($n = 127$)

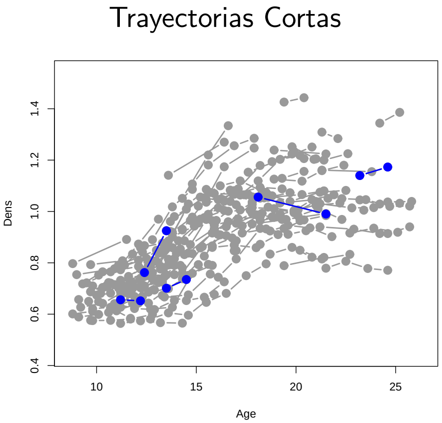


Densidad de la espina dorsal (g/cm^2) tomadas en distintas edades. En total hay 860 observaciones medidas sobre un período de dos décadas, pero solo hay 2-4 medidas para cada individuo.

Gareth James. James & Sugar (2003). Yao, Wu & Zou (2016)

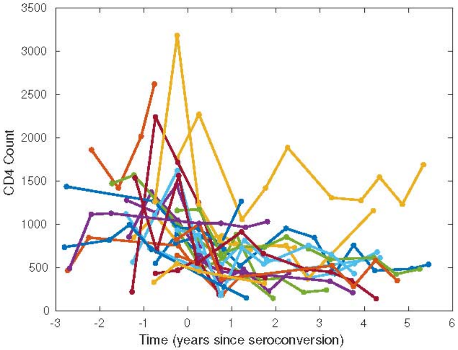
Densidad ósea: Hombres





CD4

- Conteos de CD4 en 369 pacientes de SIDA.
- El número de mediciones por individuo varía de 1 a 12, con número mediano de conteos igual a 6
- Ejemplo de datos funcionales esparsos medidos en tiempos diferentes e irregulares para cada individuo



Extraído y se agradece a Wang, J.L., Chiou, J., Müller, H.G. (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application*, **3**, 257-295

PACE (Yao, Müller & Wang, 2005)

PACE = Principal Analysis by Conditional Estimation

Supongamos que

$$Y_{ij} = X_i(t_{ij}) \qquad 1 \leq j \leq n_i \qquad 1 \leq i \leq N$$

También consideran el caso de errores de medición, es decir, se tiene un proceso $X(t)$ tal que $\mathbb{E}X(t) = \mu(t)$ y $\mathrm{Cov}(X(s), X(t)) = \gamma(s, t)$ y observamos

$$Y_{ij} = X_i(t_{ij}) + \epsilon_{ij} \qquad 1 \leq j \leq n_i \qquad 1 \leq i \leq N$$

- ϵ_{ij} son i.i.d. independientes de X_i y de los tiempos (que pueden ser aleatorios)

PACE (Yao, Müller & Wang, 2005)

- **Paso 1** Empezaremos estimando μ mediante un polinomio local, o sea, para cada $t_0 \in \mathcal{I}$, sean

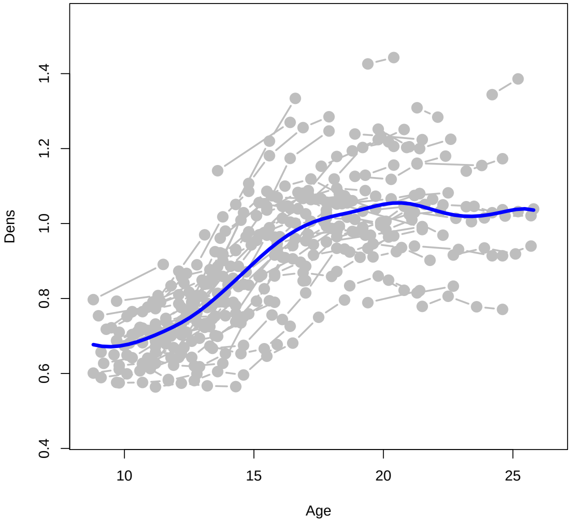
$$\left(\widehat{\beta}_0(t_0), \widehat{\beta}_1(t_0)\right)^{\mathrm{T}} = \underset{\beta_0,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \mathcal{K}\left(\frac{t_{ij} - t_0}{h}\right) \left(Y_{ij} - \beta_0 - \beta_1(t_0 - t_{ij})\right)^2 ,$$

  - $\mathcal{K} : \mathbb{R} \to \mathbb{R}$ es un núcleo, o sea, $\mathcal{K} \geq 0$, $\int \mathcal{K}(u)\, du < \infty$, $\mathcal{K}$ es continua
  - $h = h_n$ es la ventana o parámetro de suavizado.

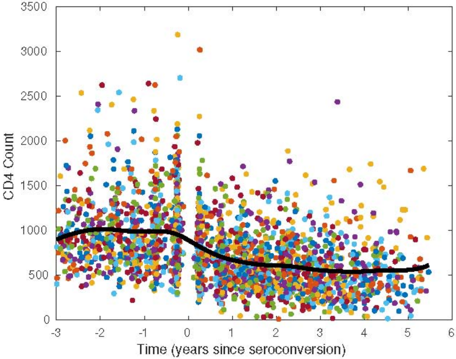
- El estimador de $\mu(t_0)$ es igual a

$$\widehat{\mu}(t_0) = \widehat{\beta}_0(t_0) .$$

Densidad Osea: $\widehat{\mu}$, $h = 0.85$



CD4: $\widehat{\mu}$, $h = 0.3$

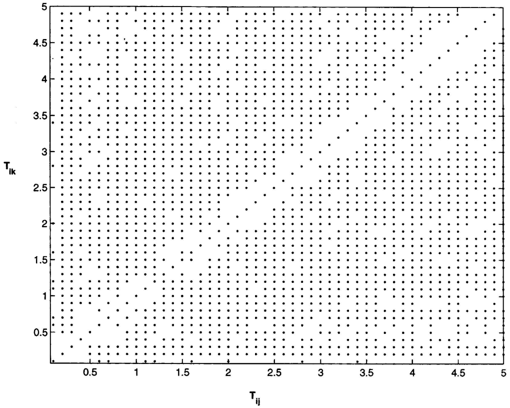


El estimador de la media $\widehat{\mu}$ muestra que

- los conteos de CD4 son estables hasta 6 meses antes de la seroconversión (tiempo 0),
- baja seis meses antes y después de la seroconversión y
- se estabiliza nuevamente, un año despues de la seroconversión.

Extraído y se agradece a Wang, J.L., Chiou, J., Müller, H.G. (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application*, **3**, 257-295

CD4



- Aunque los datos por individuo son ralos,
- Los datos combinados llenan el dominio de la superficie $\gamma(s, t)$ bastante densamente

Extraído y se agradece a Yao, F., Müller, H.G., Wang, J.L. (2005). Functional data analysis for sparse longitudinal data. *J. American Statistical Association*, **100**, 577-590.

PACE (Yao, Müller & Wang, 2005)

Supongamos primero que no hay errores de medición, o sea, para $1 \leq i \leq N$

$$Y_{ij} = X_i(t_{ij}) \qquad 1 \leq j \leq n_i$$

PACE (Yao, Müller & Wang, 2005)

- **Paso 2** Como

$$\gamma(s,t) = \mathrm{Cov}\left(X(s), X(t)\right)$$
$$= \mathbb{E}\Big[\left(X(s)-\mu(s)\right)\left(X(t)-\mu(t)\right)\Big]$$

uno puede estimar $\gamma(s,t)$ usando un smoother bivariado de los productos cruzados (*covarianzas crudas*)

$$\widehat{\gamma}_{ij\ell} = \left(X_i(t_{ij})-\widehat{\mu}(t_{ij})\right)\left(X_i(t_{i\ell})-\widehat{\mu}(t_{i\ell})\right) = \left(Y_{ij}-\widehat{\mu}(t_{ij})\right)\left(Y_{i\ell}-\widehat{\mu}(t_{i\ell})\right)$$

$$\longrightarrow \quad \widehat{\gamma}(s,t) \quad \longrightarrow \quad \widehat{\lambda}_k\,, \widehat{\phi}_k$$

PACE (Yao, Müller & Wang, 2005)

- **Paso 3** Los escores $\xi_{ij} = \langle X_i - \mu, \phi_j \rangle$ no se pueden aproximar por
$$\sum_{\ell} (X_i(t_{i\ell}) - \widehat{\mu}(t_{i\ell}))\ \widehat{\phi}_j((t_{i\ell})\ ((t_{i\ell}) - (t_{i(\ell-1)})$$
pues hay pocos puntos!!

- $\mathbf{t} = (t_1, \ldots, t_m)^{\mathrm{T}}$, $\mathbf{X} = (X(t_1), \ldots, X(t_m))^{\mathrm{T}}$.

- Si las observaciones son Gaussianas
$$\mathbb{E}\left(\xi_k \,\middle|\mathbf{X}\right) = \lambda_k\, \phi_k^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

  - $\boldsymbol{\phi}_k = (\phi_k(t_1), \ldots, \phi_k(t_m))^{\mathrm{T}}$,
  - $\boldsymbol{\mu} = (\mu(t_1), \ldots, \mu(t_m))^{\mathrm{T}}$
  - el elemento (ℓ, j) de $\boldsymbol{\Sigma}$ es $\gamma(t_\ell\,,\, t_j)$.

PACE (Yao, Müller & Wang, 2005)

- **Paso 3:** $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})^{\mathrm{T}}$.

$$\widehat{\xi}_{ik} = \widehat{\lambda}_k \, \widehat{\boldsymbol{\phi}}_k^{\mathrm{T}} \, \widehat{\boldsymbol{\Sigma}}^{-1} \, (\mathbf{X}_i - \widehat{\boldsymbol{\mu}}_i)$$

  - $\mathbf{X}_i = (X_i(t_{i1}), \ldots X_i(t_{in_i}))^{\mathrm{T}}$
  - $\widehat{\boldsymbol{\phi}}_k = \left(\widehat{\phi}_k(t_{i1}), \ldots, \widehat{\phi}_k(t_{in_i})\right)^{\mathrm{T}}$,
  - $\widehat{\boldsymbol{\mu}}_i = (\widehat{\mu}(t_{i1}), \ldots, \widehat{\mu}(t_{in_i}))^{\mathrm{T}}$
  - el elemento (ℓ, j) de $\widehat{\boldsymbol{\Sigma}}$ es $\widehat{\gamma}(t_{i\ell}\,,\, t_{ij})$.

- $$\widehat{X}_i(t) = \widehat{\mu}(t) + \sum_{k=1}^{q} \widehat{\xi}_{ik} \widehat{\phi}_k(t)$$

$$\mathbb{E}\left(\xi_k \middle| \mathbf{X}\right) = \lambda_k\, \phi_k^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

- Para cada $s \in \mathcal{I}$, $(X(s) - \mu(s), X(t_1), X(t_2), \cdots, X(t_m))^{\mathrm{T}}$ tiene distribución normal multivariada $N((0, \boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}}, \boldsymbol{\Sigma}_1)$ donde

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} \gamma(s,s) & (\gamma(s,t_1), \ldots, \gamma(s,t_m)) \\ (\gamma(s,t_1), \ldots, \gamma(s,t_m))^{\mathrm{T}} & \boldsymbol{\Sigma} \end{pmatrix}.$$

- Si llamamos $\boldsymbol{\gamma}(s) = (\gamma(s,t_1), \ldots, \gamma(s,t_m))^{\mathrm{T}}$, tenemos que

$$\begin{aligned}
\mathbb{E}\left(\xi_k \middle| \mathbf{X}\right) &= \mathbb{E}\left(\langle X - \mu, \phi_k\rangle \middle| \mathbf{X}\right) = \langle \mathbb{E}\left(X - \mu \middle| \mathbf{X}\right), \phi_k\rangle \\
&= \int_{\mathcal{I}} \phi_k(s)\, \mathbb{E}\left(X(s) - \mu(s) \middle| \mathbf{X}\right)\, ds \\
&= \int_{\mathcal{I}} \phi_k(s) \boldsymbol{\gamma}(s)^{\mathrm{T}}\, ds\, \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \\
&= \lambda_k \left(\phi_k(t_1), \ldots, \phi_k(t_m)\right)^{\mathrm{T}}\, \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}).
\end{aligned}$$

PACE (Yao, Müller & Wang, 2005)

Supongamos ahora que hay errores de medición

$$Y_{ij} = X_i(t_{ij}) + \epsilon_{ij} \qquad 1 \leq j \leq n_i$$

PACE (Yao, Müller & Wang, 2005)

- **Paso 2** Recordemos que

$$\begin{aligned}\gamma\,(s,t) &= \textsc{Cov}\,(X(s),X(t)) \\ &= \mathbb{E}\Big[(X(s)-\mu(s))\;(X(t)-\mu(t))\Big]\end{aligned}$$

Luego,

$$\begin{aligned}\textsc{Cov}\,(Y_{ij},Y_{i\ell}) &= \textsc{Cov}\,(X_i(t_{ij}),X_i(t_{i\ell}))+\textsc{Cov}\,(\epsilon_{ij},\epsilon_{i\ell}) \\ &= \gamma(t_{ij},t_{i\ell})+\sigma^2\,\delta_{j\ell}\end{aligned}$$

PACE (Yao, Müller & Wang, 2005)

- **Paso 2:** Los elementos diagonales de las covarianzas crudas

$$G_i(t_{ij}, t_{i\ell}) = (Y_{ij} - \widehat{\mu}(t_{ij}))\,(Y_{i\ell} - \widehat{\mu}(t_{i\ell}))$$

deben ser eliminados al suavizar.

- O sea suavizamos, $G_{i,j\ell}$ para $j \neq \ell$. Para ello definamos

$$H(\beta_0, \beta_1, \beta_2) = \sum_{i=1}^{N} \sum_{1 \leq j \neq \ell \leq n_i} \mathcal{K}\left(\frac{t_{ij} - s}{h_G}\right) \mathcal{K}\left(\frac{t_{i\ell} - t}{h_G}\right) \times \left(G_i(t_{ij}, t_{i\ell}) - \beta_0 - \beta_1(s - t_{ij}) - \beta_2(t - t_{i\ell})\right)^2$$

- Sea $(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2) = \operatorname{argmin} H(\beta_0, \beta_1, \beta_2)$, definimos

$$\widehat{\gamma}(s, t) = \widehat{\beta}_0 = \widehat{\beta}_0(s, t).$$

PACE (Yao, Müller & Wang, 2005)

- **Paso 2:** Para obtener el estimador de $\gamma(t,t)$
  - Primero se rotan los ejes de los tiempos 45 grados obteniendo

$$\begin{pmatrix} t^\star_{ij} \\ t^\star_{i\ell} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} t_{ij} \\ t_{i\ell} \end{pmatrix}$$

  - Definimos $\widetilde{\gamma}(s,t)$ minimizando en (ν_0, ν_1, ν_2) la cantidad

$$H_1(\nu_0,\nu_1,\nu_2) = \sum_{i=1}^{N} \sum_{1\le j\ne \ell\le n_i} \mathcal{K}\left(\frac{t^\star_{ij}-s}{h_G}\right)\mathcal{K}\left(\frac{t^\star_{i\ell}-t}{h_G}\right) \times \left(G_i(t^\star_{ij}, t^\star_{i\ell}) - \nu_0 - \nu_1\left(s - t^\star_{ij}\right) - \nu_2\left(t - t^\star_{i\ell}\right)^2\right)^2$$

  obteniendo $\widetilde{\gamma}(s,t) = \widehat{\nu}_0 = \widehat{\nu}_0(s,t)$

- Debido a la rotación el estimador de $\gamma(t,t)$ es

$$\widehat{\gamma}(t,t) = \widetilde{\gamma}\left(\sqrt{2}\,t, 0\right)$$

PACE (Yao, Müller & Wang, 2005)

- **Paso 2: continuación**

- Sea $\widehat{V}(t)$ un smoother lineal de los elementos diagonales, es decir,

$$\widehat{V}(t) = \widehat{\alpha}_0$$

donde

$$(\widehat{\alpha}_0, \widehat{\alpha}_1) = \underset{(\alpha_0,\alpha_1)}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{1 \leq j \leq n_i} \mathcal{K}\left\{\frac{t_{ij} - t}{h_G}\right) \{G_i(t_{ij}, t_{ij}) - \alpha_0 - \alpha_1 (t - t_{ij})\}^2$$

- Estimamos σ^2 por

$$\widehat{\sigma}^2 = \frac{1}{|\mathcal{I}|} \int_{\mathcal{I}} \left\{\widehat{V}(t) - \widehat{\gamma}(t,t)\right\} dt$$

si $\widehat{\sigma}^2 > 0$ y $\widehat{\sigma}^2 = 0$ otro caso.

PACE (Yao, Müller & Wang, 2005)

- **Paso 3:** Los escores $\xi_{ij} = \langle X_i - \mu, \phi_j \rangle$ no se pueden aproximar por

$$\sum_{\ell} (X_i(t_{i\ell}) - \widehat{\mu}(t_{i\ell}))\ \widehat{\phi}_j((t_{i\ell})\ ((t_{i\ell}) - (t_{i(\ell-1)})$$

pues hay pocos puntos!!

- $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})^{\mathrm{T}}$.

$$\widehat{\xi}_{ik} = \widehat{\lambda}_k\, \widehat{\phi}_k^{\mathrm{T}}\, \widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}^{-1} (\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i) \qquad \mathbf{Y}_i = (Y_{i1}, \ldots Y_{in_i})^{\mathrm{T}}$$

- El elemento (j, ℓ) de $\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}$ es

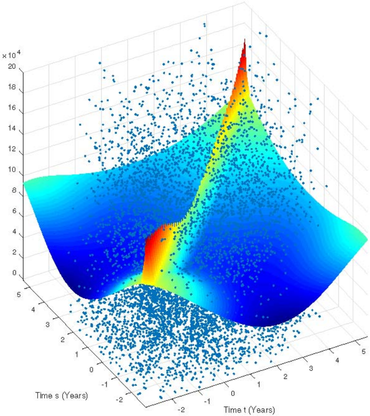
$$\widehat{\gamma}(t_{ij}, t_{i\ell}) + \widehat{\sigma}^2\ \delta_{j\ell}$$

- $\widehat{X}_i(t) = \widehat{\mu}(t) + \sum_{k=1}^{q} \widehat{\xi}_{ik} \widehat{\phi}_k(t)$

FPCA esparso

- Selección de las ventanas:
  - Convalidación cruzada
- Selección del número de direcciones principales q
  - Proporción de la variabilidad explicada

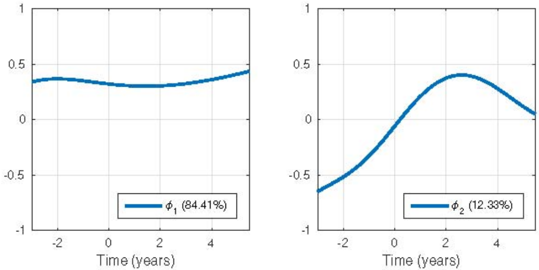
CD4: $h = 1.4$



- Los puntos son las covarianzas *crudas*
- Superficie $\widehat{\gamma}$ obtenida omitiendo los datos en la diagonal
- la diagonal forma una cresta debido a los errores de medición en los datos.

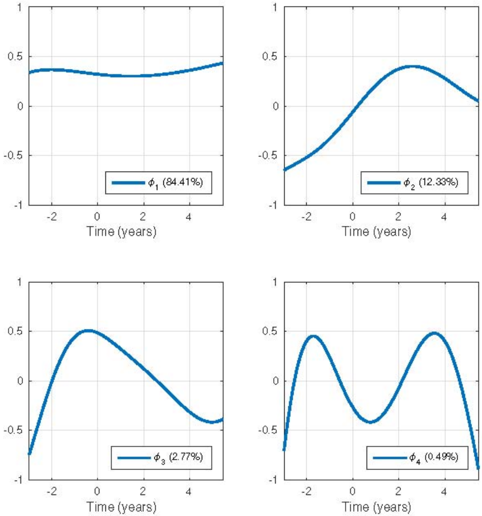
Extraído y se agradece a Wang, J.L., Chiou, J., Müller, H.G. (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application*, **3**, 257-295

CD4: $\widehat{\phi}_j$



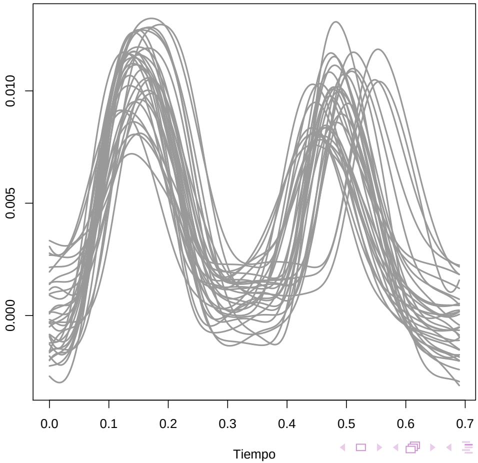
- La primer dirección es casi constante. Luego, la primer componente es en la magnitud media específica de los conteos de CD4.
- La segunda dirección muestra una variación alrededor de una tendencia temporal lineal por partes con un punto de corte cercano a 2.5 años después de la seroconversión, lo que refleja que la segunda componente es una diferencia de escala entre sujetos a lo largo de la dirección de esta función lineal por partes.

CD4: $\widehat{\phi}_j$

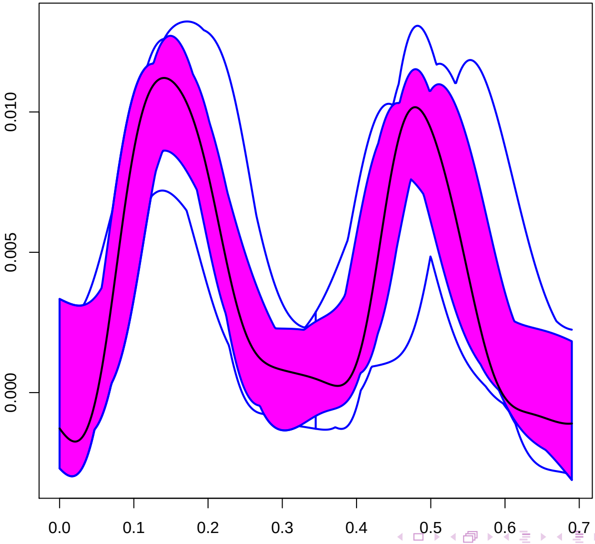


Lip Data

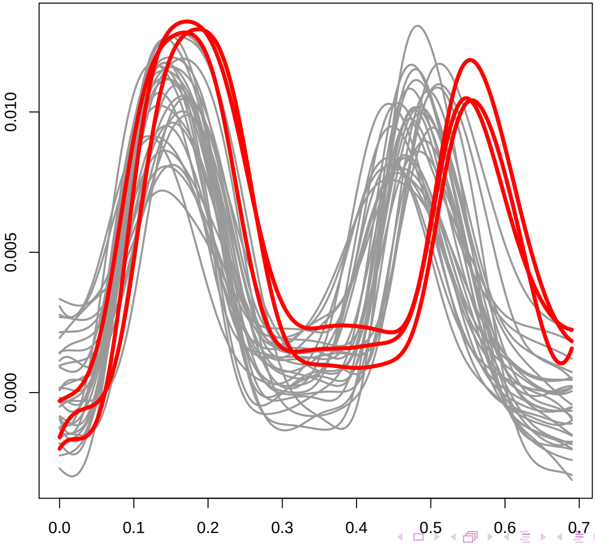
Trayectorias (32) del centro del labio inferior de un individuo que pronuncia repetidamente la palabra *BOB* (Malfait & Ramsay, 2003 and Gervini, 2008)



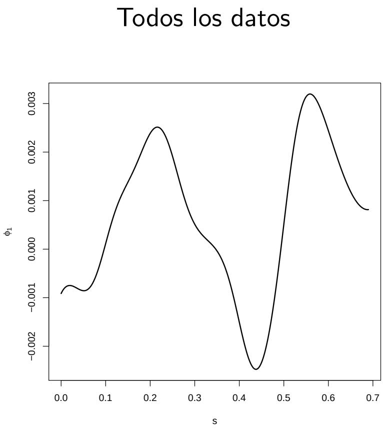
Lip Data - Functional boxplot

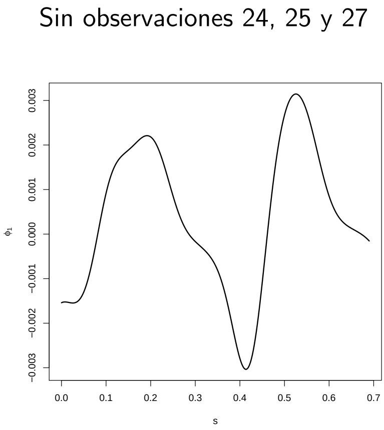


Lip Data



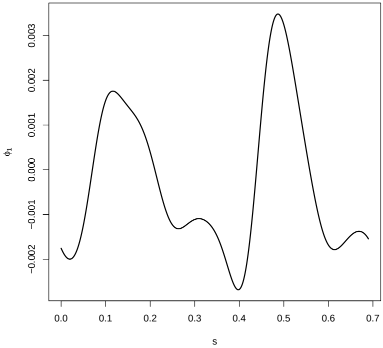
Lip Data: $\widehat{\phi}_1$



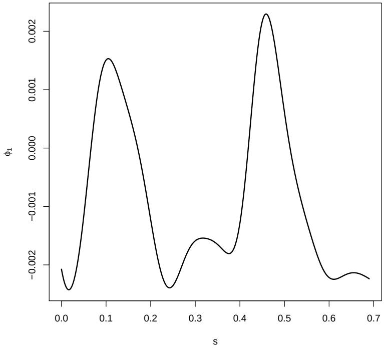


Lip Data: $\widehat{\phi}_2$

Todos los datos

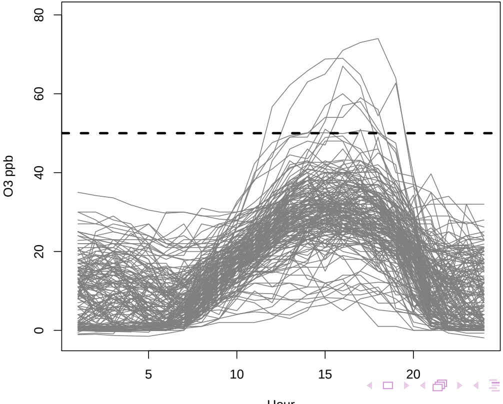


Sin observaciones 24, 25 y 27

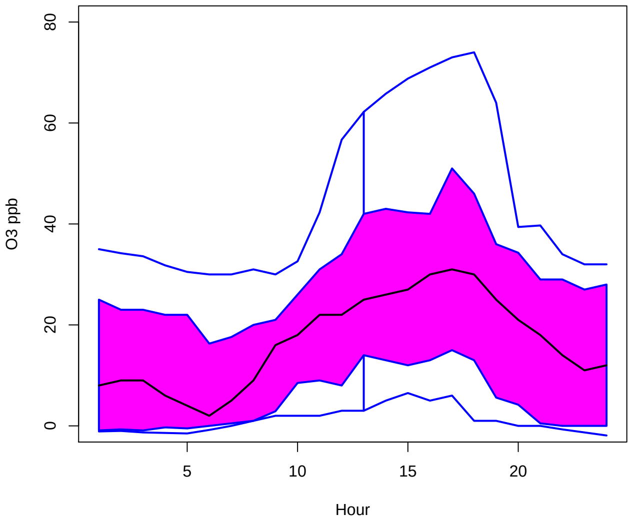


Ozone data

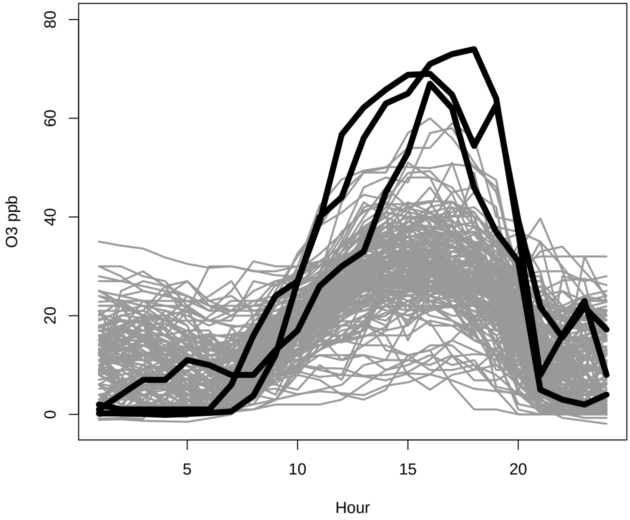
- Mediciones horarios del nivel de concentración de Ozono ($O3$ ppb)
- Agosto, 2004 a 2012, 176 días con mediciones horarias promedio
- Máximo nivel deseado establecido por el Canadian National Ambient Air Quality Objectives: 50 ppb



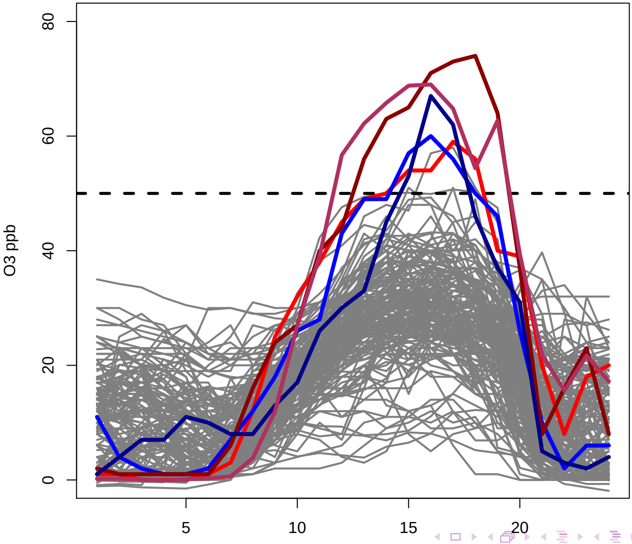
Richmond Ozone - Functional boxplot



Richmond Ozone

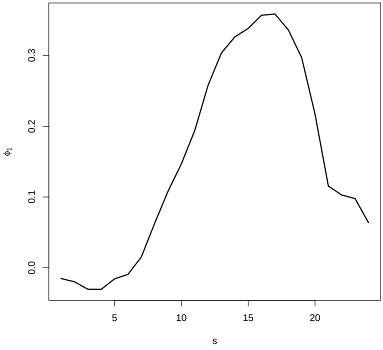


Richmond Ozone

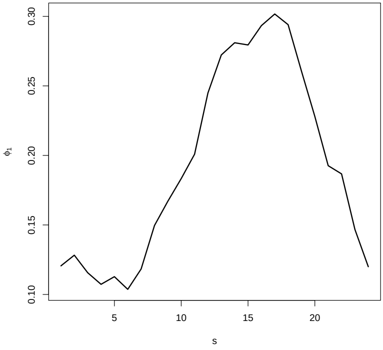


Richmond Ozone: $\widehat{\phi}_1$

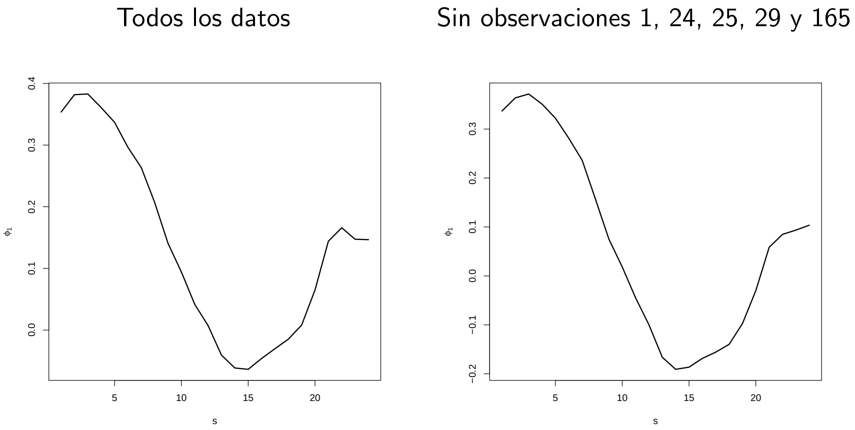
Todos los datos



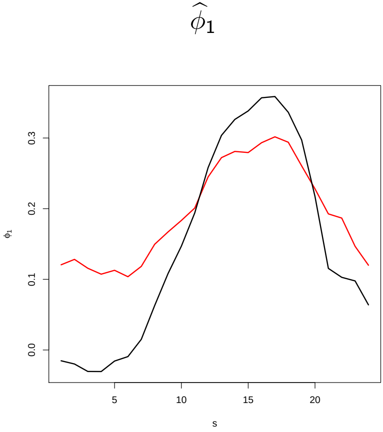
Sin observaciones 1, 24, 25, 29 y 165

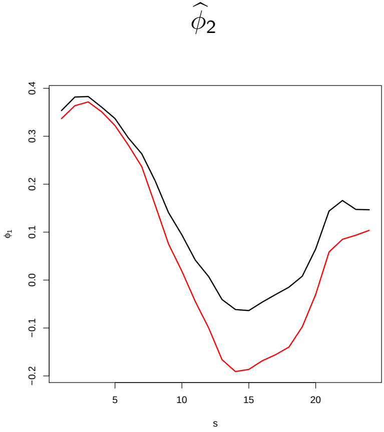


Richmond Ozone: $\widehat{\phi}_2$



Richmond Ozone: $\widehat{\phi}_j$



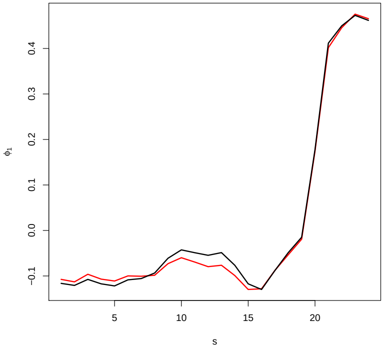


En negro: las estimaciones con todos los datos

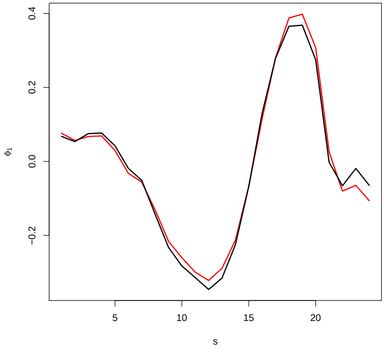
En rojo: las estimaciones sin las observaciones 1, 24, 25, 29 y 165

Richmond Ozone: $\widehat{\phi}_j$

$\widehat{\phi}_3$



$\widehat{\phi}_4$



En negro: las estimaciones con todos los datos

En rojo: las estimaciones sin las observaciones 1, 24, 25, 29 y 165

Propiedades FPCA

- **Propiedad 1.**
  La primer dirección principal corresponde a la autofunción de Γ asociada al mayor autovalor y $\mathrm{VAR}(\xi_1) = \lambda_1$.

- **Propiedad 2.**
  La primer dirección principal maximiza $\mathrm{VAR}\left(\langle\alpha, X\rangle\right)$ sobre $\mathcal{S}_1 = \{\alpha : \|\alpha\| = 1\}$.

- **Propiedad 3.**
  Sea $Y = X - \mu$. Si $\mathcal{L}_0$ es el subespacio generado por $\phi_1, \ldots, \phi_p$, con $\lambda_p > \lambda_{p+1}$, dado $\mathcal{L}$ de dimensión p,

$$\mathbb{E}(\|Y - \pi(Y, \mathcal{L}_0)\|^2) \leq \mathbb{E}(\|Y - \pi(Y, \mathcal{L})\|^2)$$

$\pi(X, \mathcal{L})$ es la proyección ortogonal de X sobre $\mathcal{L}$.

Propiedad 1: autofunción de operador de covarianza

- $\widehat{\mu} = \text{argmin}_{\theta \in \mathcal{H}} \sum_{i=1}^{n} \left(\|X_i - \theta\| - \|X_i\| \right) / n$: mediana geométrica
- $Y_i = \dfrac{X_i - \widehat{\mu}}{\|X_i - \widehat{\mu}\|}$
- Operador de covarianza espacial

$$\widehat{\Gamma}^{\mathrm{S}} = \frac{1}{n} \sum_{i=1}^{n} Y_i \otimes Y_i$$

- Las componentes principales esféricas son las autofunciones ϕ_j^{S} de $\widehat{\Gamma}^{\mathrm{S}}$
  - Locantore *et al.* (1999), Gervini (2008)
  - Boente, Salibian-Barrera & Tyler (2014), Boente, Rodriguez & Sued (2019)

Componentes principales esféricas

Por qué funcionan las componentes principales esféricas?

- $\widehat{\Gamma}^{\mathrm{S}} \xrightarrow{a.s.} \Gamma^{\mathrm{S}} = \mathbb{E}\left(\dfrac{(X-\mu)\otimes(X-\mu)}{\|X-\mu\|^2}\right)$
- Γ^{S} tiene las mismas autofunciones que Γ si
  - $X \sim \mathcal{E}(\mu, \Gamma, \varphi)$
    Boente, Salibian-Barrera & Tyler (2014)
  - $X = \mu + \sum_{j=1}^{p} Z_j \lambda_j^{1/2} \phi_j$ y $\mathbf{Z} = (Z_1, \ldots, Z_p)$ tiene marginales simétricas e intercambiables.
    Gervini (2008)
- $\boldsymbol{\Gamma}^{\mathrm{S}} = \mathbb{E}\left(\dfrac{(\mathbf{X}-\boldsymbol{\mu})\otimes(\mathbf{X}-\boldsymbol{\mu})}{\|\mathbf{X}-\boldsymbol{\mu}\|^2}\right) = \sum_{\ell\geq 1} \boldsymbol{\lambda}_\ell^{\mathrm{S}}\, \boldsymbol{\phi}_\ell \otimes \boldsymbol{\phi}_\ell$

$$\lambda_\ell^{\mathrm{S}} = \lambda_\ell \mathbb{E}\left(\frac{Z_\ell^2}{\sum_{s\geq 1} \lambda_s Z_s^2}\right).$$

FPCA robusto

- Propiedad 1: autofunción de operador de covarianza
  - Locantore *et al.* (1999), Gervini (2008), Kraus & Panaretos (2012), Sawant, Billor & Shin (2012)
- Propiedad 2: "Direcciones" de máxima variabilidad (Projection-pursuit)
  - Hyndman & Ullah (2007); Bali, B., Tyler & Wang (2011)

Se basa en el hecho que *la primer dirección maximiza* $\mathrm{VAR}\left(\langle\alpha, X\rangle\right)$ *sobre* $\mathcal{S}_1 = \{\alpha : \|\alpha\| = 1\}$.

Propiedad 2: "Direcciones" de máxima variabilidad

- $s_n(\alpha)$ una escala robusta de la muestra $\{\langle\alpha, X_i\rangle\}_{i=1}^n$
  - MAD
  - $M-$escala, es decir, si $Y_i = \langle\alpha, X_i - \widehat{\mu}\rangle$, el $M-$estimador de escala es la solución $s_n(\alpha)$ de

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i}{s_n(\alpha)}\right) = \delta\ .$$

- El estimador de la primer dirección principal se define como

$$\widehat{\phi}_1 = \underset{\|\alpha\|=1}{\operatorname{argmax}}\, s_n^2(\alpha)$$

- Los siguientes como

$$\widehat{\phi}_m = \underset{\alpha\in\widehat{\mathcal{B}}_m}{\operatorname{argmax}}\, s_n^2(\alpha) \quad 2 \le m,$$

$$\widehat{\mathcal{B}}_m = \{\alpha \in \mathcal{H} :\ \|\alpha\| = 1, \langle\alpha, \widehat{\phi}_j\rangle = 0\ ,\ \forall\ 1 \le j \le m-1\}.$$

- $\widehat{\lambda}_m = s_n^2(\widehat{\phi}_m)$

Un enfoque PP suavizado

- A veces nos interesan direcciones principales suaves.
  - $D : \mathcal{H}_\mathrm{S} \to \mathcal{H}$ un operador lineal ("diferenciador", e.g., $D\alpha = \alpha''$ when $\mathcal{H} = L^2(\mathcal{I})$).
  - Permite definir un producto interno de penalización $\lceil \alpha, \beta \rceil = \langle D\alpha, D\beta \rangle$.
  - "Operador de penalización" $\Psi : \mathcal{H}_\mathrm{S} \to \mathbb{R}$, $\Psi(\alpha) = \lceil \alpha, \alpha \rceil$
  - Producto interno penalizado $\langle \alpha, \beta \rangle_\tau = \langle \alpha, \beta \rangle + \tau \lceil \alpha, \beta \rceil$.
  - Norma penalizada $\|\alpha\|_\tau^2 = \|\alpha\|^2 + \tau\Psi(\alpha)$

Un fnfoque unificado robusto via PP

- Algunas veces en lugar de las direcciones principales empíricas, interesa obtener direcciones principales suaves.

- $s_n(\alpha) = \sigma_{\mathrm{R}}(P_n(\alpha])$

$$\begin{cases} \widehat{\phi}_1 & = \underset{\|\alpha\|_\tau=1}{\operatorname{argmax}} \left\{ s_n^2(\alpha) - \rho\Psi(\alpha) \right\} \\ \widehat{\phi}_m & = \underset{\alpha\in\widehat{\mathcal{B}}_{m,\tau}}{\operatorname{argmax}} \left\{ s_n^2(\alpha) - \rho\Psi(\alpha) \right\} \quad 2 \leq m, \end{cases}$$

$$\widehat{\mathcal{B}}_{m,\tau} = \{\alpha \in \mathcal{H} : \ \|\alpha\|_\tau = 1, \langle \alpha, \widehat{\phi}_j \rangle_\tau = 0 \ , \ \forall\, 1 \leq j \leq m-1\}.$$

- $\widehat{\phi}_m \xrightarrow{a.s.} \phi_{\mathrm{R},m}$

FPCA robusto

- Propiedad 1: autofunción de operador de covarianza
  - Locantore *et al.* (1999), Gervini (2008), Kraus & Panaretos (2012), Sawant, Billor & Shin (2012)
- Propiedad 2: "Direcciones" de máxima variabilidad (Projection-pursuit)
  - Hyndman & Ullah (2007); Bali, B., Tyler & Wang (2011)
- Propiedad 3: "Mejor" aproximación finito-dimensional (BLDA)
  - Lee, Shin, Billor, (2013); B., Salibián–Barrera (2015); Ceballos, Salibian–Barrera & Van Aelst (2016)

BLDA se basa en la propiedad

Sea $X \sim \mathcal{E}(\mu, \Gamma, \varphi)$.
Dado cualquier subespacio lineal de dimensión q, $\mathcal{L}$.
Llamemos $\widetilde{X} = X - \mu$ y $\mathcal{L}_q = \text{span}\,(\phi_1, \ldots, \phi_q)$

$$\Rightarrow \qquad \|\widetilde{X} - \pi(\widetilde{X}, \mathcal{L}_q)\|^2 \leq_s \|\widetilde{X} - \pi(\widetilde{X}, \mathcal{L})\|^2$$

lo que permite definir estimadores robustos usando

- $S-$estimadores: B. & Salibian–Barrera (2015)
- Least trimmed estimadores: Ceballos, Salibian–Barrera & Van Aelst (2016).

$S-$ y $LTS-$estimadores son Fisher–consistentes si $X \sim \mathcal{E}\,(\mu, \Gamma, \varphi)$ y $\lambda_q > \lambda_{q+1}$.

BLDA: $X_i \in \mathcal{H}$ $1 \leq i \leq n$

- $\{\delta_i\}_{i\geq 1}$: base de $\mathcal{H}$. $p_n \uparrow \infty$

$$x_{ij} = \langle X_i, \delta_j \rangle \qquad \mathbf{x}_i = (x_{i\,1}, \ldots, x_{i\,p_n})^{\mathrm{T}}$$

$$\|X_i - \widehat{X}_i\|^2_{\mathcal{H}} \simeq \|\mathbf{x}_i - \widehat{\mathbf{x}}_i\|^2_{\mathbb{R}^p}$$

- Aplique un ajuste robusto a $\mathbf{x}_i = (x_{i\,1}, \ldots, x_{i\,p_n})^{\mathrm{T}}$, para obtener

$$\widehat{\mathbf{x}}_i = \widehat{\boldsymbol{\mu}} + \sum_{\ell=1}^{q} \widehat{a}_{i\ell} \widehat{\mathbf{b}}^{(\ell)}$$

  - $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_{p_n})^{\mathrm{T}}$
  - $\widehat{\mathbf{b}}^{(\ell)} = (\widehat{b}_{\ell\,1}, \ldots, \widehat{b}_{\ell\,p_n})^{\mathrm{T}} \in \mathbb{R}^{p_n}$, $\|\widehat{\mathbf{b}}^{(\ell)}\| = 1$

$S-$estimadores

Modificamos

$$\min_{\boldsymbol{\mu},\mathbf{A},\mathbf{B}} \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i - \widehat{\mathbf{x}}_i\|^2 = \min_{\boldsymbol{\mu},\mathbf{A},\mathbf{B}} \sum_{j=1}^{p} \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \widehat{x}_{ij})^2$$

to

$$\min_{\boldsymbol{\mu},\mathbf{A},\mathbf{B}} \sum_{j=1}^{p} \widehat{\sigma}_j^2$$

donde $\widehat{\sigma}_1, \ldots, \widehat{\sigma}_p$ estimadores robustos de escala, e.g.:

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{x_{ij} - \widehat{x}_{ij}}{\widehat{\sigma}_j}\right) = b$$

$\rho : \mathbb{R} \to \mathbb{R}_+$ es una función par, no–decreciente en $|t|$, creciente para $t > 0$ cuando $\rho(t) < \sup_{t\in\mathbb{R}} \rho(t) = 1$, $\rho(0) = 0$.

BLDA: Aproximación en el espacio original $\mathcal{H}$

- $\widehat{\mu} = \sum_{j=1}^{p_n} \widehat{\mu}_j \, \delta_j$

- Base asociada de autofunciones estimadas en $\mathcal{H}$ es

$$\widehat{\phi}_\ell = \sum_{j=1}^{p_n} \widehat{b}_{\ell\, j} \, \delta_j \quad 1 \leq \ell \leq q$$

- Los valores ajusteados en $\mathcal{H}$ son

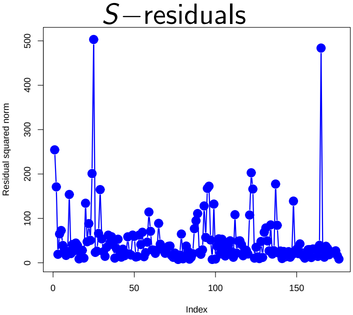
$$\widehat{X}_i = \widehat{\mu} + \sum_{\ell=1}^{q} \widehat{a}_{i\ell} \, \widehat{\phi}_\ell$$

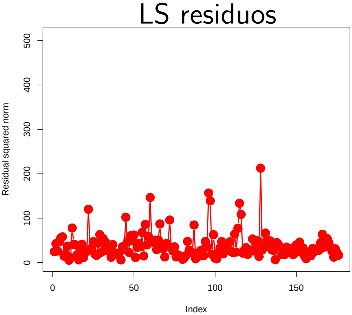
Datos de Ozono

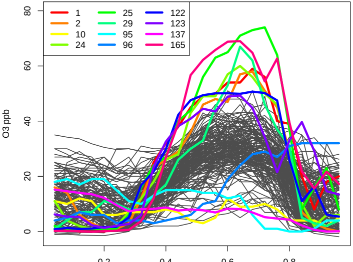
- Mediciones horarios del nivel de concentración de Ozono ($O3$ ppb)
- Agosto, 2004 a 2012, 176 días con mediciones horarias promedio

- Usamos un $B-$spline cúbico con 10 nodos, $q = 1$

Datos de Ozono - Residuos



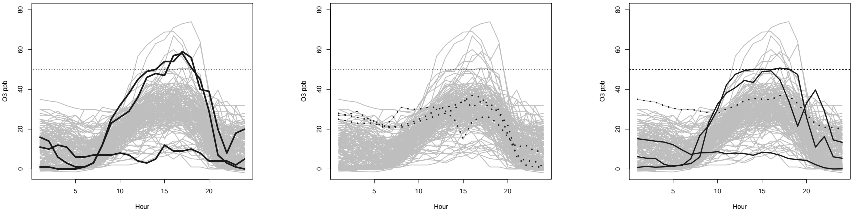




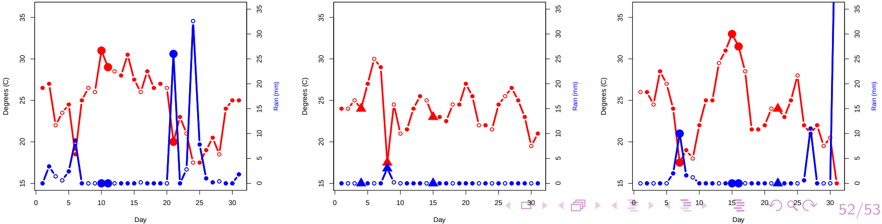
Datos atípicos

Nivel Ozono: $S-$est. (Líneas continuas) y LS (Líneas punteadas)

2004 2006 2010



Temperatura and Lluvia: $S-$est. ($\bullet$) y LS ($\blacktriangle$)



Datos atípicos

Outliers identificados por el ajuste robusto corresponden a d
ıas con picos relativamente altos de concentración de O3, pero algunos días tienen un perfil “llano”.

- El ajuste robusto identifica como atípicos los días con temperaturas o muy altas o muy bajas.
- Los días atípicos con perfil O3 “plano” corresponden a máxima temperatura baja.

  Los días con un pico agudo de O3 peak corresponden a días calurosos.
- El método robusto identifica los outliers potenciales que corresponden a valores extremos de variables de una variable meteorológica claramente asociada.