

**UNIVERSIDAD AUTÓNOMA DE QUERÉTARO**



**ELEMENTOS BÁSICOS DEL ANÁLISIS FUNCIONAL DE DATOS**

**TESIS**

**QUE COMO PARTE DE LOS REQUISITOS PARA  
OBTENER EL TÍTULO DE LICENCIADO EN**

**MATEMÁTICAS APLICADAS**

**PRESENTA**

**HUGO EDUARDO CALVA DÍAZ**

**DIRIGIDA POR**

**DR. EDUARDO CASTAÑO TOSTADO**

**CENTRO UNIVERSITARIO  
QUERÉTARO. QRO.- MÉXICO**

**2002**

**BIBLIOTECA CENTRAL UAQ**

**"ROBERTO RUIZ OBREGÓN"**

No. Adq. H67450

No. Título \_\_\_\_\_

Clas 510.24

C167e

\_\_\_\_\_

\_\_\_\_\_



ACUERDO 478 / 2002  
23 de septiembre del 2002

**C. Hugo Eduardo Calva Díaz**  
**Pasante de la Lic. en Matemáticas Aplicadas**  
**Presente.**

Con relación a la elección de opción de titulación, elaboración de tesis individual, le informo que el H. Consejo Académico en su sesión ordinaria del 23 de septiembre del año en curso, aprueba la opción de titulación y deberá trabajar en el tema **“Elementos Básicos del Análisis Funcional de Datos”** bajo la dirección del Dr. Eduardo Castaño Tostado, catedrático de esta Facultad, quién tuvo a bien aceptar el nombramiento.

## **CONTENIDO**

### **CAPÍTULO I**

#### **INTRODUCCIÓN**

##### **1.1 Datos Funcionales**

##### **1.2 Ejemplos**

###### **1.2.1 “Lip Data” (Datos de Labio)**

###### **1.2.2 “Growth Data” (Datos de Crecimiento)**

###### **1.2.3 Temperaturas Canadienses**

###### **1.2.4 “Pinch Force” (Fuerza de Presión del pulgar-índice)**

##### **1.3 Estrategia General del Análisis Funcional de Datos**

### **CAPÍTULO II**

#### **REPRESENTACIÓN DE DATOS FUNCIONALES COMO FUNCIONES SUAVES**

##### **2.1 Introducción**

##### **2.2 Modelo a seguir**

##### **2.3 Suavizamiento**

##### **2.4 Funciones Bases**

###### **2.4.1 Series de Fourier**

###### **2.4.2 Bases Polinomiales**

###### **2.4.3 Bases Spline**



## 2.5 Ponderaciones Local

2.5.1 Suavizamiento por Kernel

2.5.2 Estimadores por funciones base localizadas

2.5.3 Representación polinomial local

## 2.6 Representación de Datos Penalizando de Rugosidad

## 2.7 Propiedades de Splines

## 2.8 Propiedades Estadísticas de Suavizadores

### **CAPÍTULO III**

### **ANÁLISIS DE COMPONENTES PRINCIPALES FUNCIONALES**

#### 3.1 Introducción

#### 3.2 Análisis de Componentes Principales Funcionales

3.2.1 Bases empíricas ortonormales

3.2.2 Eigenanálisis

3.2.3 Interpretación de ACP

#### 3.3 Análisis de componentes Principales Regularizados

3.3.1 Encontrar ACP regularizados en la práctica

#### 3.4 Detalles Computacionales

### **CAPÍTULO IV**

### **REGISTRO DE DATOS FUNCIONALES**

#### 4.1 Introducción

4.1.1 Formulación del problema de registro

#### 4.2 Tipos de registro

4.2.1 Alineación landmark

4.2.2 Funciones "warping"

4.2.3 Estimación de la función "warping" por suavizamiento

4.2.4 Registro continuo

#### 4.3 Funciones Warping

#### 4.4 Detalles Computacionales



## **CAPÍTULO V**

### **APLICACIONES**

- 5.1 Perfiles sin Registro
- 5.2 Perfiles con Registro Landmark
- 5.3 Perfiles con Registro Continuo
- 5.4 Estadísticas Descriptivas
- 5.5 Análisis de Componentes Principales

### **BIBLIOGRAFÍA**

Falkner, F. (1960) Child Development: An international method of study. Baser: Kager.

Friedman, J. H. and Silverman, B. W. (1989) Flexible nonparametric smoothing and additive modeling (with discussion and response ). Journal of the American Statistical Association, 90, 1179-1188.

Green, P.J. and Silverman, B.W. (1994). Nonparametric Regression and Generalized Linear Models, A roughness penalty approach. CHAPMAN&HALL/CRC.

Gasser, Th. Y Müller, H. G. (1979). Kernel estimation of regression functions. Smoothing techniques for Curve Estimation (Th. Gasser and M. Rosenblatt, eds.), Springer Lecture Notes in Mathematics No. 757, Springer-Verlag, Berlin, pp. 23-68.

Hart, Jeffrey D. (2000). Nonparametric Smoothing and Lack-of-Fit Tests. Springer-Verlag, New York.

Kneip, A. y Gasser, T. (1992). Statistical tools to analyze data representing a sample of curves. The annals of Statistics, 20, 3, 1266-1305.

Mauricio Sánchez , R. A. (2001). Caracterización Físicoquímica, Térmica y Eléctrica de Razas Mexicanas de Maíz y evaluación de sus Posibles Usos en la Industria alimentaria. Tesis de Maestría en ciencia y Tecnología de los Alimentos, Universidad Autónoma de Querétaro, Facultad de Química. Programa de Posgrado en Alimentos del Centro de la República.



Nadaraya, E. A. (1964). On estimating regression Theory Probab. Appl., 9, 141-142.

Priestley, M.B. y Chao, M.T. (1972). Non-parametric function fitting. J. Roy. Statist. Soc. Ser. B 34, 385-392.

Ramsay, J.O., Bock, R.D. y Gasser, T. (1995) Comparision of height acceleration curves in the Fels, Zurichn and Berkeley growth data. Ann. Hum. Biol., 22, 413-426.

Ramsay, J.O. y Silverman, B. W. (1997). Functional Data Análisis. Springer Verlag, New York.

Ramsay, J.O. y Dalzell, C.J. (1991). Some tools for Functional Data Analysis. Journal of the Royal Statistical Society, B, 3, 539-572.

Ramsay, J.O. (1995). Some tools for the multivariate analysis of functional data. In Recent Advances in Descriptive Multivariate Analysis. Edited by Wojtek J. Krzanoswski. Claredon Press, Oxford.

Ramsay, J.O. (1998). Curve Registration. Journal of the Royal Statistical Society, B, 60, 365-375.

Ramsay, J.O. (1998). Estimating smooth functions. Journal of the Royal Statistical Society, B, 60, part 2, 365-375.

Ramsay, J.O. (2000). A Guide to Curve Registration. FTP ego.psych.amsay.

Ramsay, J.O. y Silverman, B.W. (1998). S-PLUS Functions for FDA.FTP ego.peych.megill.camsay.

Ramsay, J.O. and Li Xiaochun, (1998). Curve Registration. Journal of the Royal Statistical Society, B, 60, part 2, 351-363.



Ramsay, J.O. Wang, X. y Flanagan, R. (1995). A functional data analysis of the pinch force of human fingers. Appl. Statist., 44, 17-30.

Reinsch, C. (1967). Smoothing dy spline functions. Numer. Math., 10, 177-83

Rice, J.A. y silverman, B.W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society, B, 53, 233-243.

Silverman, B.W. (1984). Spline smoothing: the equivalent variable Kelnel method. Ann Statist, 12, 898-916.

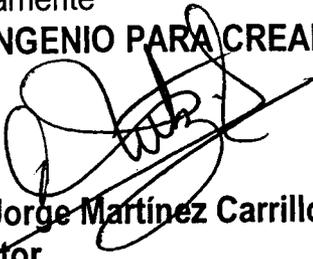
Watson, G.S. (1964). Smooth regression analysis. Sankhya Ser. A, 26, 359-372.

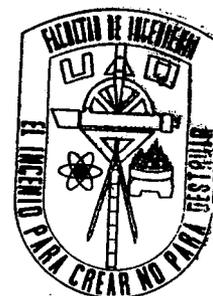
También hago de su conocimiento las disposiciones de nuestra Facultad, en el sentido de que antes de su Examen Profesional deberá cumplir con los requisitos de nuestra Legislación y deberá reimprimir el presente oficio en todos los ejemplares de su tesis.

Sin otro particular, quedo de usted.

Atentamente

**"EL INGENIO PARA CREAR, NO PARA DESTRUIR"**

  
Ing. Jorge Martínez Carrillo  
Director



DIRECCIÓN



C.U., 23 de septiembre del 2002.

**Ing. Jorge Martínez Carrillo**  
**Director de la Facultad de**  
**Ingeniería de la UAQ**  
**PRESENTE:**

Por este conducto me permito comunicar a Usted, que una vez revisada la tesis individual titulada **“Elementos Básicos del Análisis Funcional de Datos”** del Pasante de la licenciatura en Matemáticas Aplicadas, Hugo Eduardo Calva Díaz, de acuerdo al artículo 20 inciso h) del Reglamento de Titulación vigente.

Emito mi voto aprobatorio.

Atentamente,

**“EL INGENIO PARA CREAR, NO PARA DESTRUIR”**

**Dr. Eduardo Castaño Tostado**  
Director de Tesis

c.c Archivo  
\*sar



CU., 23 de Septiembre del 2002

**H. CONSEJO ACADÉMICO  
DE LA FACULTAD DE INGENIERÍA  
P R E S E N T E :**

Me permito comunicar a este Órgano Colegiado que, una vez revisada la tesis individual titulada "**Elementos Básicos del Análisis Funcional de Datos**", del pasante de la licenciatura en Matemáticas Aplicadas, Hugo Eduardo Calva Díaz, emito mi Voto Aprobatorio.

Atentamente,  
"EL INGENIO PARA CREAR, NO PARA DESTRUIR"

**Dr. Juan de Dios Figueroa Cárdenas  
Sinodal**

c.c Archivo  
\*sar



CU., 23 de Septiembre del 2002

**H. CONSEJO ACADÉMICO  
DE LA FACULTAD DE INGENIERÍA  
P R E S E N T E :**

Me permito comunicar a este Órgano Colegiado que, una vez revisada la tesis individual titulada "**Elementos Básicos del Análisis Funcional de Datos**", del pasante de la licenciatura en Matemáticas Aplicadas, Hugo Eduardo Calva Díaz, emito mi Voto Aprobatorio.

Atentamente,  
"EL INGENIO PARA CREAR, NO PARA DESTRUIR"

**M. en C. Héctor Mendieta Martínez**  
**Sinodal**

c.c Archivo  
\*sar



CU., 23 de Septiembre del 2002

**H. CONSEJO ACADÉMICO  
DE LA FACULTAD DE INGENIERÍA  
P R E S E N T E :**

Me permito comunicar a este Órgano Colegiado que, una vez revisada la tesis individual titulada **“Elementos Básicos del Análisis Funcional de Datos”**, del pasante de la licenciatura en Matemáticas Aplicadas, Hugo Eduardo Calva Díaz, emito mi Voto Aprobatorio.

Atentamente,  
“EL INGENIO PARA CREAR, NO PARA DESTRUIR”

**Mat. Rocío Hidalgo Flores  
Sinodal**

c.c Archivo  
\*sar

# **AGREDECIMIENTOS**

A Dios por ponerme en este camino, el cual me ha hecho crecer como persona y como profesional.

Al Dr. Eduardo Castaño Tostado por dirigir este trabajo con profesionalismo y cariño, además le agradezco lo que me enseñó como persona.

Al Dr. Juan de Dios Figueroa Cárdenas (CINVESTAV Unidad Querétaro) por todas las facilidades prestadas para la realización de este trabajo.

A todos mis sinodales por el tiempo invertido en este trabajo, a pesar de sus múltiples ocupaciones.

A todos y cada uno de mis maestros por enseñarme el amor por las matemáticas.

A todos mis amigos por hacer este trabajo mas divertido al igual que mi vida.

A mis padrinos y primos, Galito, Meche, Toño, Mechis y Marce, por todo el amor y el apoyo que me han dado durante 24 años.

A mis abuelitos, Nila, Chano, Tere y Noé, por darme esos cuatro ejemplos de vida.

A Ana por su amistad.

A mi familia, Oscar, Ofelia, Oscar Jr. y Mariana, por aguantar todos los enojos y desesperaciones, por enseñarme todo lo que soy ahora, por ser parte de mi vida y llevarme siempre con ustedes en donde quiera que estén.

**A Oscar, Ofelia, Oscar Jr., Mariana y Ana.**

# Índice General

<b>1</b>	<b>Introducción</b>	<b>6</b>
1.1	Datos Funcionales . . . . .	6
1.2	Ejemplos . . . . .	7
1.2.1	“Lip Data” (Datos de Labios) . . . . .	7
1.2.2	“Growth Data” (Datos de Crecimiento) . . . . .	8
1.2.3	Temperaturas Canadienses . . . . .	10
1.2.4	“Pinch Force” (Fuerza de Presión del pulgar-índice) . . . . .	11
1.3	Estrategia General del Análisis Funcional de Datos . . . . .	12
<b>2</b>	<b>Representación de Datos Funcionales como</b>	
	<b>Funciones Suaves</b>	<b>15</b>
2.1	Introducción . . . . .	15
2.2	Modelo a seguir . . . . .	15
2.3	Suavizamiento . . . . .	16
2.4	Funciones Base . . . . .	18
2.4.1	Series de Fourier . . . . .	21
2.4.2	Bases Polinomiales . . . . .	23
2.4.3	Bases Splines . . . . .	23
2.5	Ponderación Local . . . . .	24
2.5.1	Suavizamiento por kernel . . . . .	25
2.5.2	Estimadores por funciones base localizadas . . . . .	28
2.5.3	Representación polinomial local . . . . .	30
2.6	Representación de Datos Penalizando la Rugosidad . . . . .	30
2.7	Propiedades de Splines . . . . .	33
2.8	Propiedades Estadísticas de Suavizadores . . . . .	36
<b>3</b>	<b>Análisis de Componentes Principales</b>	<b>47</b>
3.1	Introducción . . . . .	47
3.2	Análisis de Componentes Principales Funcionales . . . . .	47
3.2.1	Bases empíricas ortonormales . . . . .	48
3.2.2	Eigenanálisis . . . . .	49

3.2.3	Interpretación de ACP . . . . .	51
3.3	Análisis de Componentes Principales Funcionales Regularizados . . . . .	53
3.3.1	Encontrar ACP Regularizados en la Práctica . . . . .	55
3.4	Detalles Computacionales . . . . .	58
<b>4</b>	<b>Registro de Datos Funcionales</b>	<b>60</b>
4.1	Introducción . . . . .	60
4.1.1	Formulación del problema de registro . . . . .	62
4.2	Tipos de Registro . . . . .	64
4.2.1	Alineación Landmark . . . . .	64
4.2.2	Funciones “Warping” . . . . .	66
4.2.3	Estimación de la función “warping” por suavizamiento . . . . .	71
4.2.4	Registro continuo . . . . .	72
4.3	Detalles Computacionales . . . . .	73
<b>5</b>	<b>Aplicaciones</b>	<b>77</b>
5.1	Perfiles sin Registrar . . . . .	78
5.2	Perfiles con Registro Landmark . . . . .	81
5.3	Perfiles con Registro Continuo . . . . .	82
5.4	Estadísticas Descriptivas . . . . .	83
5.5	Análisis de Componentes Principales . . . . .	86
<b>6</b>	<b>Conclusiones</b>	<b>96</b>
<b>7</b>	<b>Bibliografía</b>	<b>98</b>

# Índice de Figuras

<b>Figura 1.1</b>	Curvas suavizadas para los datos de labio. . . . .	8
<b>Figura 1.2</b>	Curvas del crecimiento humano. . . . .	9
<b>Figura 1.3</b>	Segunda derivada de las funciones del crecimiento humano. . . . .	10
<b>Figura 1.4</b>	Curvas suavizadas de temperaturas en las estaciones de Canadá. . . . .	11
<b>Figura 1.5</b>	Curvas de fuerza para las temperaturas de Canadá. . . . .	12
<b>Figura 1.6</b>	Curvas suavizadas para los datos de "Pinch force". . . . .	13
<b>Figura 1.7</b>	Función de correlación para las temperaturas canadienses . . .	14
<b>Figura 2.1</b>	Ventanas para datos que serán suavizados. . . . .	25
<b>Figura 3.1</b>	Primeras cuatro funciones de componentes principales estimadas de la expresión de los datos de clima canadiense en 12 términos de la serie de Fourier. . . . .	48
<b>Figura 3.2</b>	Cuatro componentes principales del clima canadiense con su porcentaje de variación explicada . . . . .	51
<b>Figura 3.3</b>	Escores de las estaciones del clima sobre los primeros CP. . . . .	52
<b>Figura 3.4</b>	Tres funciones CP de datos de labio. . . . .	59
<b>Figura 4.1</b>	Curvas de aceleración para el crecimiento humano. . . . .	60
<b>Figura 4.2</b>	Curvas suavizadas de los datos de pinch force, con registro y sus funciones warping. . . . .	61

<b>Figura 4.3</b>	Curvas no registradas (derecha) y curvas registradas (izquierda) de lip data.....	66
<b>Figura 4.4</b>	Funciones warping para el registro landmark para los datos de labio.....	71
<b>Figura 4.5</b>	Cuatro estaciones canadienses registradas.....	73
<b>Figura 4.6</b>	Resultado del ejercicio descrito. En negro curva por registrar, en rojo curva registrada y en azul función objetivo. Panel derecho función warping.....	74
<b>Figura 4.7</b>	Funciones de deformación para los datos de labios.....	76
<b>Figura 5.1</b>	Los 28 perfiles representados en forma discreta.....	80
<b>Figura 5.2</b>	Perfiles suavizados sin registro, con registro landmark y con registro continuo.....	81
<b>Figura 5.3</b>	Perfil derivada del maíz Con-M.....	83
<b>Figura 5.4</b>	Perfil derivada del maíz chalq9.....	84
<b>Figura 5.5</b>	Perfiles promedio con su respectivo conjunto de curvas sin registro, con registro landmark y con registro continuo. . .	85
<b>Figura 5.6</b>	Desviaciones estándar de los perfiles sin registrar (azul), con registro landmark (negro) y con registro continuo (rojo).....	85
<b>Figura 5.7</b>	Correlaciones de los perfiles sin registro (izquierda), con registro landmark (centro) y con registro continuo (derecha).....	86

<b>Figura 5.8</b>	Componentes principales de los perfiles correspondientes a un tipo de registro y sin registro. . . . .	88
<b>Figura 5.9</b>	Componentes principales rotados de los perfiles correspondientes a un tipo de registro y sin registro. . . . .	89
<b>Figura 5.10</b>	Scores correspondientes a los perfiles sin registrar, con registro landmark y con registro continuo . . . . .	90
<b>Figura 5.11</b>	Scores rotados correspondientes a los perfiles sin registrar, con registro landmark y con registro continuo. . . . .	90
<b>Figura 5.12</b>	Los 28 perfiles suavizados con sus nombres y su promedio. . . . .	92
<b>Figura 5.13</b>	Los 28 perfiles suavizados con registro landmark y su promedio. . . . .	93
<b>Figura 5.14</b>	Los 28 perfiles suavizados con registro continuo y su promedio. . . . .	94

# 1 Introducción

## 1.1 Datos Funcionales

En la actualidad hay procesos en los cuales se miden características de unidades experimentales o de muestreo de manera casi continua; en muchos casos esto motiva a que estas mediciones se pueden pensar como manifestaciones discretas de una función subyacente, un dato funcional.

Esto sugiere el analizar este tipo de procesos considerando a los datos agrupados como funciones.

Las metas del análisis funcional de datos son esencialmente las mismas que las de cualquier otra rama de la Estadística. Ellas incluyen las siguientes metas:

- Representar los datos de manera que favorezca más aún el análisis.
- Mostrar las características más importante de los datos.
- Estudiar importantes fuentes de variación o patrones entre los datos.
- Explicar la variación en un resultado o variable dependiente usando información de variables independientes.
- Comparar dos o más conjuntos de datos con respecto a ciertos tipos de variación, donde dos conjuntos de datos pueden contener diferentes conjuntos de repeticiones de las mismas funciones, o diferentes funciones por un conjunto de repeticiones.

Cada uno de estos puntos puede ser conducidos con técnicas apropiadas para lograr una meta.

Considere que un dato funcional para una réplica  $i$  se manifiesta como un conjunto de mediciones discretas,  $y_{i1}, \dots, y_{in}$ , la primera tarea es convertir estos valores a una función  $x_i$  con valores  $x_i(t)$  computables para cualquier valor  $t$ , donde  $t$  representa genéricamente, el tiempo, aunque puede representar otra escala de interés. Si suponemos que los valores

discretos son libres de error, entonces el proceso es de *interpolación*, pero si ellos están sujetos a un error observacional que necesita ser removido, entonces el proceso de conversión de datos discretos a una función involucra *suavizamiento*. El capítulo 1 ofrece un panorama general de estos procedimientos. El capítulo 2 generaliza un método estadístico multivariado para encontrar las fuentes de variación de un fenómeno o proceso más importantes, este método es llamado *análisis de componentes principales funcionales* y tiene el mismo objetivo que en el caso multivariado. En el capítulo 3 se presenta material de lo que se llama *registro de datos funcionales* que es otra técnica muy útil para esta clase de datos, ya que se encarga de alinear características de las curvas estimadas para después poder hacer comparaciones más válidas entre ellas. Finalmente, en el capítulo 4 se termina con una aplicación sobre el maíz, la cual es de mucha ayuda para caracterizar propiedades importantes del *proceso de gelatinización del maíz*. Proceso de mucha importancia en la industria de alimentos.

## 1.2 Ejemplos

A través de este trabajo estaremos utilizando algunos ejemplos desarrollados por Ramsay y Silverman (1997).

### 1.2.1 “Lip Data” (Datos de Labios)

En la Figura 1.1 se muestran 20 réplicas del movimiento del centro del labio inferior, cuando una persona dice la sílaba “bob”. El intervalo  $[0,1]$  es arbitrario. Los datos así mostrados ya han sido preprocesados por herramientas mencionadas en este trabajo. El objetivo era identificar la dirección en la que había más movimiento.

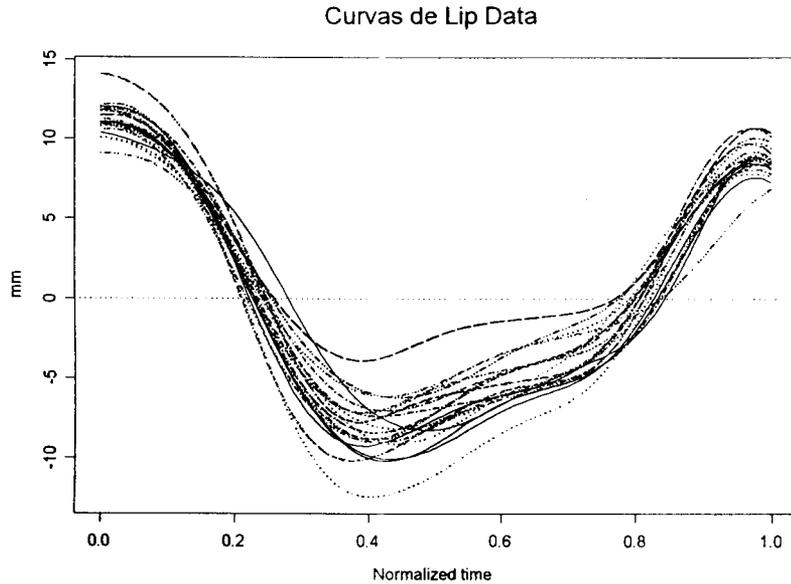


Figura 1.1. Curvas suavizadas para los datos de labio

### 1.2.2 “Growth Data” (Datos de Crecimiento)

La Figura 1.2 provee de otro prototipo para el tipo de datos que consideraremos. Muestra las estaturas de 39 niños suizos medidos a lo largo de 18 años en el Estudio Longitudinal del Crecimiento en Zurich (Falkner, 1960). Las mediciones no están equiespaciadas; existen mediciones anuales de 2 a diez años, seguidas por mediciones bianuales. Aunque se tuvo un gran cuidado en el proceso de medición, existe ruido en los valores de la estatura, con una desviación estándar de 5 mm. Aunque cada registro involucra únicamente 29 valores discretos, estos valores reflejan una variación suave en el crecimiento que podría ser tasado, en principio, tantas veces como se desee, y es entonces una función de la estatura. De este modo, los datos consisten de 39 observaciones funcionales  $Height_i(t)$ . La Figura 1.3 muestra las curvas de aceleración  $D^2Height_i(t)$  estimadas de estos datos por Ramsay, Bock y Gasser (1995). El estudio de las derivadas puede ser de mucha ayuda, además que nosotros podemos pensar en ellas ya que estamos hablando de que tenemos funciones continuamente derivables. En este ejemplo las derivadas son usadas para explotar la variación que se tiene entre curva y curva, además de que se encuentra más informativo este aspecto, como se verá más adelante.

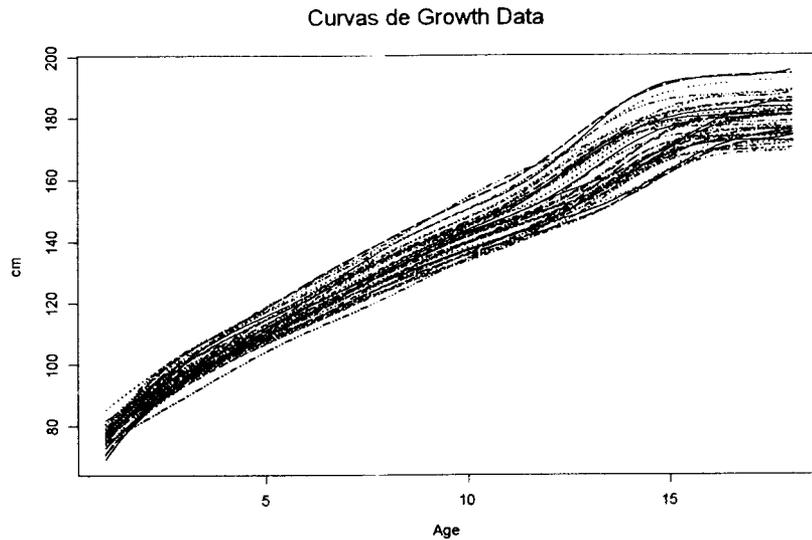


Figura 1.2. Curvas del crecimiento humano

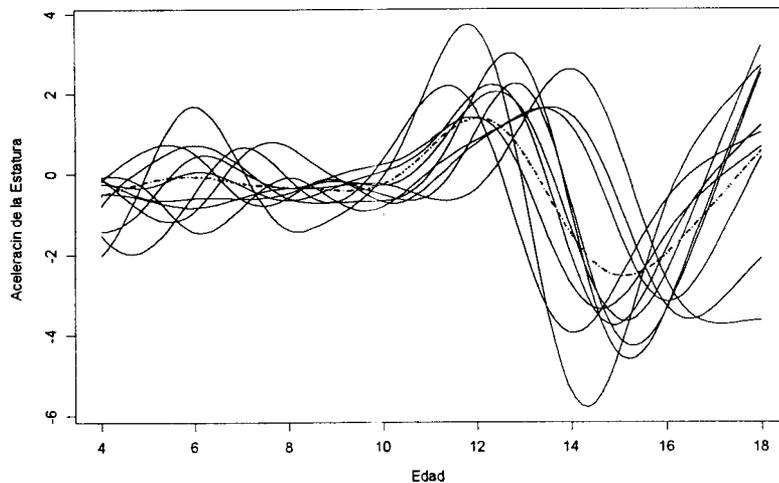


Figura 1.3. Segunda derivada de las funciones de crecimiento humano

La repetición de estas curvas de crecimiento invita a la exploración de cómo las curvas varían. Esto es potencialmente complejo. Por ejemplo, el rápido crecimiento durante la pubertad es visible en todas las curvas, pero el tiempo y la intensidad del crecimiento en la edad puberta, difieren de niño a niño. Algún tipo de análisis de componentes principales sería útil sin duda alguna, pero tenemos que adaptar un

procedimiento que tome en cuenta el espacio desigual entre edades y la suavidad de las funciones de crecimiento subyacentes.

### 1.2.3 Temperaturas Canadienses

Los datos en muchos campos pueden ser representados por procesos naturales y de aquí por funciones. Cambiando completamente de contexto del ejemplo anterior, consideremos la Figura 1.4, donde la temperatura media mensual de 35 estaciones del clima en Canadá son graficadas. Arvida, Beaverlodge, Calgary, Charlottetown, Churchill, Dowson, Edmonton, Fredericton, Halifax, Inuvik, Iqualuit, Kamloops, Kapuskasing, London, Montreal, Ottawa, Prince Albert, Prince George, Prince Rupert, Ciudad de Quebec, Regina, Resolute, Schefferville, Sherbrooke, St. John's, Sydney, The Pas, Thunder Bay, Toronto, Vancouver, Victoria, Whitehorse, Winnipeg, Yarmouth y Yellowknife. Uno esperaría que la temperatura se comportara primariamente en forma senoidal en ciclos periódicos anuales. Existe algún desfazamiento entre las estaciones debido a que los días más fríos en algunas de éstas ocurren más tarde. Consecuentemente, un modelo de la forma, para la  $i$ -ésima estación,

$$Temp_i(t) \approx c_{i1} + c_{i2} \text{sen}(\pi t/6) + c_{i3} \text{cos}(\pi t/6) \quad (1.1)$$

debería ser apropiado para los datos. El hecho aquí, es que existen claras desviaciones de una conducta senoidal o armónica. Una manera de ver esto es computando la función

$$LTemp = (\pi/6)^2 DTemp + D^3Temp$$

donde la notación  $D^m$  significa que se está tomando la  $m$ -ésima derivada., el resultado de la función anterior es algunas veces llamada *función de fuerza* que es un operador diferencial. Ahora si una función de temperatura es verdaderamente senoidal, entonces  $LTemp$  debería ser exactamente cero, como sería para cualquier función de la forma (1.1). Esto es, debe satisfacer la ecuación diferencial homogénea  $LTemp = 0$ . El resultado de aplicarle el operador diferencial antes mencionado se observa en la Figura 1.5. El punto aquí es que puede ser de interés quitar algunos efectos de un carácter simple, aplicando un operador diferencial, más que simplemente restarlos, como en el caso de modelos lineales clásicos. La experiencia sugiere que el uso de operadores diferenciales puede tener mayor impacto que el uso de modelos lineales multivariados.

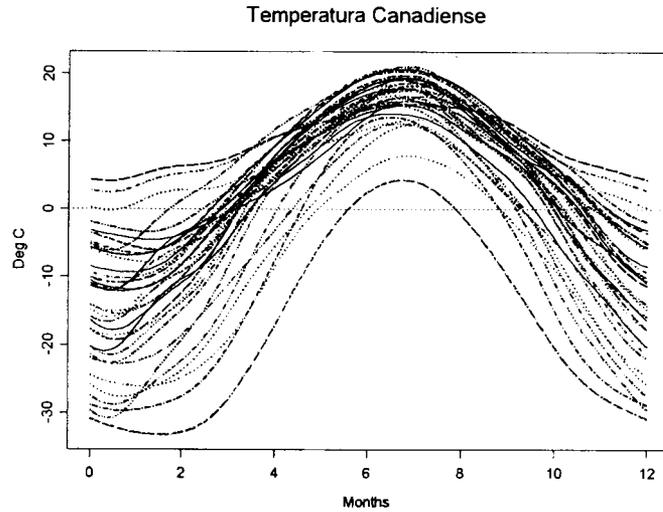


Figura 1.4. Curvas suavizadas de temperatura en las estaciones de Canadá

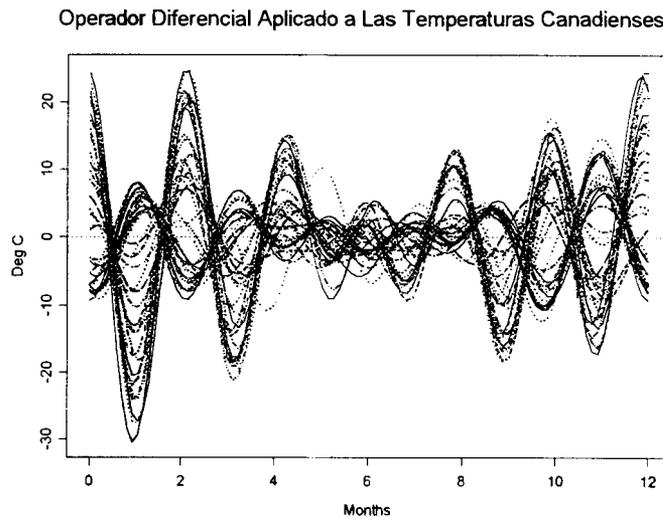


Figura 1.5. Curvas de fuerza (operador diferencial) para las temperaturas de Canadá

### 1.2.4 “Pinch Force” (Fuerza de Presión del pulgar-índice)

La Figura 1.6 muestra datos ya funcionales biomecánicos. Las curvas en la Figura son 20 registros de la fuerza ejercida sobre un medidor durante un breve “apretón”, por el dedo pulgar y el dedo índice. El propósito del experimento fue estudiar la neurofisiología del

grupo de músculos de los dedos pulgar-índice. Los datos fueron recolectados en la Unidad de Fisiología Aplicada de Cambridge. (ver Ramsay, Wang y Flanagan, 1995).

Estos datos ilustran un problema común en análisis funcional de datos. El comienzo del apretón está localizado en un tiempo arbitrario, y un primer paso es alinear los registros por algún corrimiento en el eje del tiempo. En el capítulo 3 presentamos las formas de cómo estimar este corrimiento, y cómo profundizar si es necesario estimar transformaciones del dominio, incluso transformaciones no lineales.

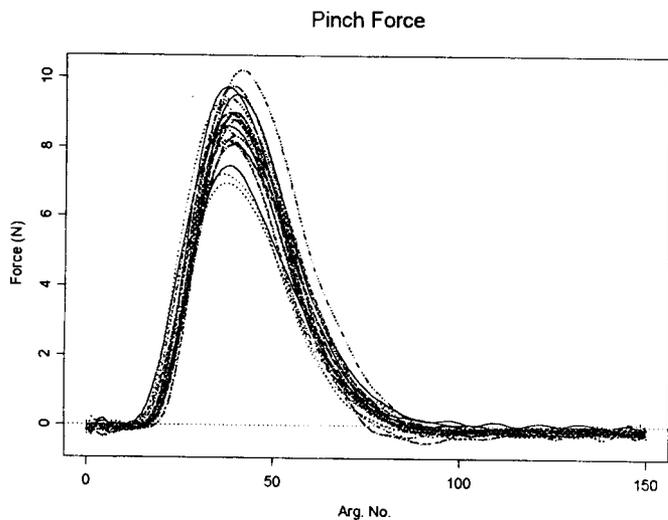


Figura 1.6. Curvas suavizadas para los datos de “pinch force”

### 1.3 Estrategia General del Análisis Funcional de Datos

A continuación enunciaremos una estrategia general del análisis funcional de datos, esto es para tener un panorama general de cómo se procederá en el análisis de este tipo de datos.

Partiendo de datos  $y_1, y_2, \dots, y_n$  obtenidos de medir en tiempos  $t_1, t_2, \dots, t_n$ , podemos suponer un modelo de la siguiente forma

$$y_j = x(t_j) + \epsilon_j, \quad \epsilon_j \sim (0, \sigma^2).$$

Una tarea de vital importancia después de plantear el modelo es *representar a  $x$  eficientemente*, para poder extraer la mejor información posible. Existen algunas técnicas para hacer esta representación entre ellas podemos mencionar dos: *usando funciones base y penalizando la rugosidad*, las cuales serán desarrolladas y analizadas en el capítulo 1

Habiendo representado a  $x$  eficientemente podemos pasar al análisis de datos, entre los que se encuentra *el registro, análisis de estadísticas descriptivas y análisis de componentes principales funcionales*. Las técnicas de registro o alineamiento serán presentadas en el capítulo 4, mientras que en el capítulo 3 se presenta el análisis de componentes principales funcional. En cuanto a estadísticas descriptivas funcionales, éstas se definen a continuación. La media, la varianza, desviación estándar, covarianza y correlación son simples generalizaciones de la estadística multivariada. De este modo la función media se define como

$$\bar{x}(t) = N^{-1} \sum_{i=1}^N x_i(t), \quad (1.2)$$

que es el promedio puntual de las funciones a través de las réplicas. La función varianza se define como

$$var_X(t) = (N - 1)^{-1} \sum_{i=1}^N [x_i(t) - \bar{x}(t)]^2, \quad (1.3)$$

y la función desviación estándar es la raíz cuadrada de la función varianza.

La función covarianza resume la dependencia de los registros a través de diferentes valores de  $t$ , y es computada para todo  $t_1$  y  $t_2$  por

$$cov_X(t_1, t_2) = N^{-1} \sum_{i=1}^N [x_i(t_1) - \bar{x}(t_1)][x_i(t_2) - \bar{x}(t_2)]. \quad (1.4)$$

La función de correlación asociada es

$$corr_X(t_1, t_2) = \frac{cov_X(t_1, t_2)}{\sqrt{var_X(t_1)var_X(t_2)}}. \quad (1.5)$$

Para poder entender las curvas de nivel de la función de correlación, con base en la Figura 1.7, observemos que en el primer panel de la Figura 1.7 está graficada una recta de unos, “1”, la cual indica la correlación entre tiempos idénticos. Segmentos perpendiculares a esta recta indican qué tan rápido cae la correlación cuando los valores se empiezan a separar. Por ejemplo, se podría localizar una posición fija sobre un valor  $t$  y fijarnos en la recta de 1's sobre ese valor, y después movernos perpendicularmente y con esto podemos visualizar qué pasa con la correlación entre los valores en el par  $(t - \delta, t + \delta)$ , cuando la distancia perpendicular  $\delta$  crece.

Funcion de Correlacion de la Temperatura

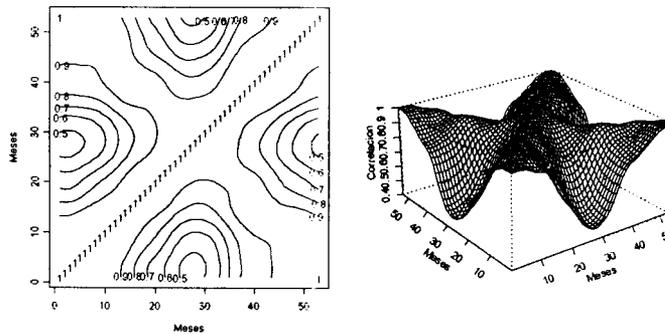


Figura 1.7. Función de correlación para las temperaturas canadienses

## 2 Representación de Datos Funcionales como Funciones Suaves

### 2.1 Introducción

La filosofía básica del análisis funcional de datos es: un conjunto de datos que representa mediciones repetidas de una unidad bajo estudio, se considera como producto de la manifestación discreta de una función subyacente  $x$ . El término funcional se refiere a la estructura intrínseca de los datos más que su forma explícita. Pensando en lo anterior, en este capítulo consideraremos algunas técnicas para convertir los datos funcionales en su verdadera forma como función.

### 2.2 Modelo a seguir

Un registro de una observación funcional de  $x$  consiste de  $n$  pares  $(t_j, y_j)$ , donde  $y_j = x(t_j)$ . Siempre se supone que el rango de los valores de interés para el argumento  $t$  es un intervalo acotado  $\Upsilon$ , además de que  $x$  satisface algunas condiciones de continuidad o suavidad sobre  $\Upsilon$ .

Si tenemos un vector  $y = (y_1, \dots, y_n)$  entonces podemos escribir el modelo como sigue:

$$y_j = x(t_j) + \epsilon_j, \quad \epsilon_j \sim (0, \sigma^2), \quad (2.1)$$

donde  $\epsilon_j$  representa un error, el cual contribuye a la rugosidad del vector de datos. Este error o ruido se puede dar por aspectos del proceso de medición. Una de las tareas en representar los datos como funciones es intentar filtrar el ruido eficientemente, o suavizar los resultados del análisis.

El modelo estadístico estándar para  $\epsilon_j$  supone que son independientemente distribuidos con media cero y varianza finita, es conveniente suponer que una varianza  $\sigma^2$  es común a estas distribuciones. Pero en casos especiales debemos tomar en cuenta la no homogeneidad de las varianzas y covarianzas para vecindades de las  $\epsilon_j$ 's. Se debe tener en mente la posibilidad de que los errores tengan propiedades multiplicativas más que de adición, para estos casos es conveniente trabajar con el logaritmo de los datos.

El tamaño muestral es una clave determinante de lo que es posible en análisis funcional de datos (AFD). El aspecto más importante en análisis funcional de datos, es una propiedad esencialmente local de estos: la densidad de los valores  $t_j$  relativa a la cantidad de curvatura en los datos más que simplemente el número  $n$ . La curvatura en un punto  $t$  se mide en general por el valor absoluto de la segunda derivada en este punto ( $|D^2x(t)|$ ). Cuando la curvatura es alta, es primordial tener suficientes puntos para estimar la función efectivamente. La suficiencia depende del tamaño de  $\epsilon_j$ ; cuando el nivel de error es pequeño y la curvatura es moderada, como es el caso de los datos de la temperatura en Canadá, Figura 1.4, se puede tener éxito con un tamaño de muestra pequeño.

Se analizarán varias maneras en las cuales las observaciones discretas  $y_j$  pueden ser representadas apropiadamente por una función suave  $x$ .

Aunque las técnicas explicadas en este capítulo son diseñadas para filtrar el ruido en las funciones mismas usando el modelo (2.1), no es el único modelo posible para la variación de una observación funcional a otra.

Contando con estos elementos el paso a seguir es la representación de  $x$  como una función suave, para de ahí poder llegar al análisis de datos. La mayoría de los métodos que consideramos en este capítulo son suavizadores lineales; de este modo procederemos primero a hacer una explicación de esta clase general de suavizadores.

## 2.3 Suavizamiento

Un suavizador lineal estima el valor de la función  $x(t)$  mediante una combinación lineal de observaciones discretas

$$\hat{x}(t) = \sum_{j=1}^n S_j(t)y_j. \quad (2.2)$$

La conducta del suavizador en  $t$  está determinada por los pesos  $S_j(t)$ . Estos pesos no dependen únicamente del valor de  $t$  en el cual la función esta siendo estimada y del subíndice

$j$ . En general existe alguna clase de dependencia sobre el patrón completo de los puntos de observación  $(t_1, \dots, t_n)$ , pero esto no está expresado explícitamente en la notación.

Los suavizadores lineales pueden ser representados en forma matricial. Supongamos que tenemos una sucesión  $s_1 < s_2 < \dots < s_m$  de puntos de evaluación en un intervalo  $\Upsilon$  en el cual la función está siendo estimada. Notemos que los puntos de evaluación no necesariamente son los mismos que los puntos de observación  $t_j$ . Sea  $\hat{x}$  el vector de longitud  $m$  con valores  $x(s_i)$  y  $y$  el vector de datos observados  $y_j$ . Luego podemos escribir

$$\hat{x} = \mathbf{S}y \tag{2.3}$$

donde  $\mathbf{S}_{ij} = S_j(s_i)$ .

Una amplia gama de suavizadores usados son lineales. La linealidad de un suavizador es una característica deseable por muchas razones: La propiedad de linealidad

$$\mathbf{S}(ay + bz) = a\mathbf{S}y + b\mathbf{S}z$$

es importante para poder trabajar con varias propiedades de la representación suave, además de que la simplicidad de los suavizadores lineales implican cálculos relativamente rápidos. Por otro lado, algunos suavizadores no lineales se pueden adaptar fácilmente a diferentes conductas en diversas partes del rango de observación, y pueden ser robustos ante observaciones atípicas.

Rapidez de cómputo puede ser crítica; un suavizador útil para pocos cientos de puntos de evaluación o de observación, puede ser completamente no práctico para miles de estos puntos. Suavizadores que requieren del orden de  $n$ , denotado por  $O(n)$ , operaciones para computar  $n$  valores suavizados  $\hat{x}(s_i)$  son virtualmente esenciales para  $n$  grande, pero suavizadores lineales que son menos eficientes computacionalmente pueden ser muy atractivos en otras estimaciones. Si  $\mathbf{S}$  tiene estructura de banda, significa que sólo un número pequeño  $K$  de valores sobre ese lado de su diagonal en cualquier renglón son no cero, entonces el cómputo de  $O(n)$  es seguro. Sin embargo, no únicamente para  $\mathbf{S}$  con

estructura de banda es posible lograr cálculos de  $O(n)$ ; ver, por ejemplo, la discusión de *spline smoothing* en Green y Silverman (1994).

## 2.4 Funciones Base

Uno de los métodos o procedimientos más familiares para suavizar es representar la función  $x(t)$  por una combinación lineal de  $K$  funciones base conocidas como  $\phi_k$ ,

$$x(t) = \sum_{k=1}^K c_k \phi_k(t). \quad (2.4)$$

El grado al cual los datos  $y_j$  están suavizados está determinado por  $K$  funciones base. El suavizador lineal más simple es obtenido si determinamos los coeficientes  $c_k$  de la expansión minimizando el criterio de mínimos cuadrados

$$SMSSE(y | c) = \sum_{j=1}^n \left[ y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2, \quad (2.5)$$

o en términos matriciales,

$$SMSSE(y | c) = (y - \Phi c)' (y - \Phi c) = \|y - \Phi c\|^2,$$

donde el vector  $c$  de magnitud  $K$  contiene los coeficientes  $c_k$  y  $\Phi$  es una matriz de  $n \times K$  donde  $\Phi_{j,k} = \{\phi_k(t_j)\}$ , es decir son las funciones base evaluadas en los puntos de observación. Este criterio es minimizado por la solución  $c = (\Phi' \Phi)^{-1} \Phi' y$ .

Suponiendo que los puntos de evaluación son idénticos a los puntos de observación, la matriz de suavidad  $\mathbf{S}$  está dada por

$$\mathbf{S} = \Phi (\Phi' \Phi)^{-1} \Phi'. \quad (2.6)$$

La matriz de suavidad  $\mathbf{S}$  en este caso es una matriz de proyección ortogonal, debido a que es simétrica y satisface la relación de idempotencia  $\mathbf{S}^2 = \mathbf{S}$ . De este

modo, un suavizador de funciones expandidas en una base y usando mínimos cuadrados es simplemente una proyección ortogonal de un registro sobre el espacio expandido por las columnas de la matriz  $\Phi$ .

En general, si los puntos de evaluación,  $s_i$ 's son distintos que los puntos de observación,  $t_j$ 's se define  $\tilde{\Phi}_{m \times n} = \{\phi_j(s_i)\}$ . Entonces la matriz de suavidad está dada por:

$$S = \tilde{\Phi}(\Phi' \tilde{\Phi})^{-1} \Phi'.$$

Así, podemos extender el criterio mínimos cuadrados a la forma

$$SMSSSE(y|c) = (y - \Phi c)' \mathbf{W} (y - \Phi c) = \|y - \Phi c\|_{\mathbf{W}}^2, \quad (2.7)$$

donde  $\mathbf{W}$  es una matriz conocida, simétrica y definida positiva que permite dar pesos distintos de cuadrados y productos de residuales. Esta extensión puede ser importante si conocemos que la varianza de los errores  $\epsilon_j$  no es constante, o que los errores no sean independientemente distribuidos en alguna forma conocida. Los estimados de los coeficientes  $c$  entonces están dados por  $(\Phi' \mathbf{W} \Phi)^{-1} \Phi' \mathbf{W} y$ . En el caso donde los puntos de evaluación y los puntos de observación sean iguales, la correspondiente matriz de suavidad es

$$\mathbf{S} = \Phi (\Phi' \mathbf{W} \Phi)^{-1} \Phi' \mathbf{W},$$

que continúa siendo un operador de proyección ortogonal, debido a que  $\mathbf{W} \mathbf{S}$  es simétrica. El mapeo definido por  $\mathbf{S}$  se dice algunas veces que es una *proyección en la métrica  $\mathbf{W}$* .

Una pregunta de vital importancia es ¿cómo escoger el orden de la expansión  $K$ ? Debido a que la expansión en funciones base por mínimos cuadrados es esencialmente un problema de regresión múltiple, existe literatura, la cual es una guía sobre la selección de buenos subconjuntos de variables, escogidos dentro de un conjunto muy grande. Sin embargo, el hecho de que  $K$  pueda tomar únicamente valores enteros positivos, indica que el control de

la suavidad puede ser comparativamente cruda. En la siguiente sección se hablará de cómo se puede obtener un control más fino de la suavidad.

Cuando  $n$  es grande, la eficiencia en computación es crítica, hay 3 tareas primordiales en calcular los estimados para puntos de evaluación en general:

- 1) Computar productos internos, de los cuales hay  $K$  en  $\Phi'$  y  $K(K + 1)/2$  en  $\Phi'\Phi$ .
- 2) Resolver el sistema lineal  $\Phi'\Phi c = \Phi'y$ .
- 3) Computar los  $m$  productos  $\tilde{\Phi}c$ , donde la matriz  $\tilde{\Phi}$  contiene las funciones evaluadas en las  $s_i$ 's.

Algoritmos eficientes y estables por mínimos cuadrados pueden hacer estos cálculos o sus equivalentes en  $O[(n + m)K^2]$  operaciones, y esto es aceptable dado que  $K$  es pequeña y fija relativa a  $n$  y  $m$ . Pero para  $K$  grande es extremadamente útil que la matriz  $\Phi'\Phi$  tenga una estructura de banda con pocos ceros en un número fijo de posiciones sobre ese lado de la diagonal, y así poder tener estabilidad numérica y cálculos económicos. Una base ortogonal con respecto a la suma sobre  $j$  es el caso extremo donde la matriz  $\Phi'\Phi$  es diagonal; los coeficientes del ajuste por mínimos cuadrados son computados multiplicando el vector de datos  $y$  por una matriz explícita de  $K \times n$ , sin la necesidad de resolver un sistema de ecuaciones lineales.

En el peor de los casos, sin una estructura de banda y con  $K$  y  $m$  del tamaño de  $O(n)$ , los cálculos son hechos en  $O(n^3)$ , lo cual no es aceptable cuando  $n$  se puede contar por varios cientos.

Haciendo un resumen del proceso de estimación para representar a  $x$  en forma matricial tenemos que

$$\begin{aligned}
y &= x + \epsilon \text{ (modelo)} \\
y &= \Phi c + \epsilon \text{ (sustitución de } x \text{ por funciones base)} \\
\hat{c} &= \min_c (y - \Phi c)' (y - \Phi c) \\
\hat{c} &= (\Phi' \Phi)^{-1} \Phi' y \text{ (solución sobre } c) \\
\hat{x} &= \Phi (\Phi' \Phi)^{-1} \Phi' y \text{ (función estimada)} \\
\mathbf{S} &= \Phi (\Phi' \Phi)^{-1} \Phi' \text{ (matriz de suavidad)} \\
\hat{x} &= \mathbf{S} y \text{ (} x \text{ estimada por un suavizador lineal).}
\end{aligned}$$

Nuestro objetivo ahora es, si ya sabemos cómo representar a la función deseada por funciones base, seleccionar una base que represente de una manera fiel a los datos.

Una característica deseable de las funciones base es que concuerden con aquellos aspectos conocidos de las funciones a ser estimadas. Una base debería ser escogida de tal manera que se logre una excelente aproximación usando valores pequeños de  $K$ . La selección de bases es particularmente importante para estimar derivadas,

$$D\hat{x}(t) = \sum_{k=1}^K c_k D\phi_k(t). \quad (2.8)$$

Bases que trabajan bien para estimar la función pueden dar una muy pobre estimación de las derivadas. Esto es debido a que una representación más exacta de las observaciones pueden forzar a  $\hat{x}$  a tener pequeñas oscilaciones, pero con una frecuencia muy alta con terribles consecuencias para sus derivadas. Uno de los aspectos para escoger una buena base puede ser analizar cómo la expansión aproxima a un cierto grado de derivadas y viendo si el comportamiento de éstas es razonable.

### 2.4.1 Series de Fourier

Una base muy conocida está dada por series de Fourier:

$$\hat{x}(t) = c_0 + c_1 \text{sen} \omega t + c_2 \cos \omega t + c_3 \text{sen} 2\omega t + c_4 \cos 2\omega t + \dots \quad (2.9)$$

Definida por  $\phi_0(t) = 1$ ,  $\phi_{2r-1}(t) = \text{sen}(r\omega t)$ ,  $\phi_{2r}(t) = \text{cos } r\omega t$ , esta base es periódica y  $\omega$  determina el período  $2\pi/\omega$ , el cual es igual a la longitud del intervalo  $\Upsilon$ , sobre el cual se está trabajando. Si los valores de  $t_j$  están igualmente espaciados sobre  $\Upsilon$ , entonces la base es ortogonal en el sentido de que el producto cruz de la matriz  $\Phi'\Phi$  es diagonal, y puede ser transformado en la identidad, dividiendo las funciones base por constantes adecuadas,  $\sqrt{n}$  para  $j = 0$  y  $\sqrt{\frac{n}{2}}$  para  $j \neq 0$ .

La transformada rápida de Fourier (TRF) hace posible encontrar los coeficientes eficientemente; cuando  $n$  es una potencia de 2 y los argumentos están equiespaciados, entonces podemos encontrar los coeficientes  $c_k$  y los  $n$  valores suaves en  $x(t_j)$  en el orden de  $n \log n$  operaciones [ $O(n \log n)$ ]. Esta es una de las características que ha hecho a la serie de Fourier la base tradicional para series de tiempo muy largas.

La estimación de las derivadas es simple ya que

$$D \text{sen}(r\omega t) = r\omega \text{cos } r\omega t$$

$$D \text{cos}(r\omega t) = -r\omega \text{sen}(r\omega t).$$

Esto implica que los coeficientes de la expansión de  $Dx$  son:

$$(0, -\omega c_2, \omega c_1, -2\omega c_4, 2\omega c_3, \dots)$$

y de  $D^2x$

$$(0, -\omega^2 c_1, -\omega^2 c_2, -4\omega^2 c_3, -4\omega^2 c_4, \dots)$$

Una serie de Fourier es especialmente útil para funciones extremadamente estables, sin comportamientos locales fuertes y curvatura uniforme. Esta base produce expansiones que son uniformemente suaves, pero son inapropiadas para datos con algún grado de discontinuidad en la función misma o en algunas de sus derivadas de orden menor. Una serie de Fourier es como la margarina: “es barata y se puede untar donde sea, pero no esperes que el resultado de comerla sea excelente”.

## 2.4.2 Bases Polinomiales

La base monomial  $\phi_k(t) = (t - \varpi)^k$ ,  $k = 0, \dots, K$  es también clásica. Desafortunadamente, esta base puede generar singularidades en el producto cruz  $\Phi' \Phi$ , y el parámetro de corrimiento  $\varpi$  debe ser cuidadosamente escogido. Sin embargo, si los valores  $t_j$ 's están equiespaciados o pueden ser escogidos de tal manera que exhiban un patrón estándar, expansiones en polinomios ortogonales pueden ser obtenidas, implicando  $O[(n + m)K]$  operaciones para todos los valores suaves. En cualquier otro caso tendremos  $O[(n + m)K^2]$  operaciones.

Como las expansiones en series de Fourier, los polinomios no pueden describir eficientemente características locales sin usar una  $K$  muy grande. Estos tienden a ajustar muy bien el centro de los datos pero su conducta en los extremos no es muy buena.

Aunque las expansiones de las derivadas de los polinomios son muy simples de calcular, rara vez son buenos como estimadores de las verdaderas derivadas a causa de las oscilaciones, las cuales son altas cuando tenemos ajuste de polinomios de un grado alto.

## 2.4.3 Bases Splines

Con seguridad las bases de Fourier y polinomiales han sido sobre-usadas en trabajos de aplicación. Su falta de habilidad para caracterizar las propiedades locales nos lleva al desarrollo de *splines polinomiales*. Los splines son funciones construidas al juntar polinomios suavemente en valores  $\tau_k$ , llamados nodos. El número de estos nodos es denotado por  $K_1 + 1$ , donde los nodos de los extremos definen el intervalo  $\Upsilon$  sobre el cual la estimación toma lugar y  $K_1 - 1$ , es el número de nodos interiores. Entre dos nodos adyacentes, un spline polinomial es un polinomio de grado fijo  $K_2$ , pero en un nodo interior, dos polinomios adyacentes son requeridos para coincidir en los valores de un número fijo de derivadas, generalmente  $K_2 - 1$ . Esto implica que un spline de grado 0 es una función escalón discontinua en los nodos, un spline de grado 1 es una función lineal poligonal y un spline de grado 3 es una cuadrática con primera derivada continua. La continuidad de las dos primeras derivadas de un spline cúbico quiere decir que la curva es visualmente suave.

Los splines combinan el cómputo fácil, con la capacidad para cambiar su comportamiento local y gran flexibilidad. Los splines se representan generalmente como sigue:

$$\phi_k(t) = (t - \tau_k)_+^{K_2},$$

donde  $u_+ = u$  si  $u \geq 0$  y 0 en otro caso, y donde sólo los puntos interiores son usados. Esto es llamada la base de potencia truncada. Esta base tiene que ser aumentada por la base monomial  $t^k$ ,  $k = 0, \dots, K_2$ . para producir un spline polinomial completo de grado  $K$ . El número total de funciones base es  $K = K_1 + K_2$ , o el grado de los polinomios más el número de nodos interiores.

El problema con estas bases es que pueden producir singularidades en los productos de matrices  $\Phi'\Phi$ , cuando hay pocos nodos. Es mejor usar los llamados B-splines, los cuales tienen *soportes compactos*, es decir, son cero casi donde sea excepto sobre un intervalo finito. En el caso de splines cúbicos, cada B-spline es un spline cúbico con soporte sobre el intervalo  $[\tau_{k-2}, \tau_{k+2}]$ , y con un soporte más corto en los extremos. El soporte compacto implica que  $\Phi'\Phi$  tenga una *estructura de banda facilitando el cómputo de valores suavizados*. El cómputo de todos los valores suavizados es del orden  $O(K)$

Algunas veces puede existir la duda de donde situar los nodos  $\tau_l$ . Aunque algunas aplicaciones fuertemente sugieren en dónde ubicar los nodos, sin embargo, la elección algunas veces puede ser hasta cierto punto arbitraria. Por otro lado, algunas aproximaciones capitalizan la libertad de elección de la localización de los nodos, empezando con un conjunto denso de estos y después ir eliminando los que no sean necesarios por un algoritmo similar a la técnica de elección de variables usado en regresión múltiple. (Friedman y Silverman, 1989).

## 2.5 Ponderación Local

En un sentido muy amplio, el suavizamiento es la quinta esencia de la Estadística: remover la variabilidad en los datos que no tenga causas asignables y entonces hacer más aparente los aspectos sistemáticos de los datos. Suavizamiento por ponderación local es un método, el cual nos ayudará a establecer lo anterior.

Ponderación local es un método no paramétrico de estimación de la función  $x$ . Supongamos que deseamos estimar el valor de la función  $x(t)$  para alguna  $t \in [0, 1]$ . Si  $x$  es una función continua, entonces los valores  $t_i/s$  cercanos a  $t$  implican que  $x(t_i/s)$  están cercanos a  $x(t)$ . Esto sugiere que promediando los  $y_i/s$  correspondientes a  $t_i/s$  cercanos a  $t$ , producirán un estimador insesgado de  $x(t)$  aproximadamente. Ponderación local tiene el efecto benéfico de reducir la variabilidad proveniente de los términos de error.

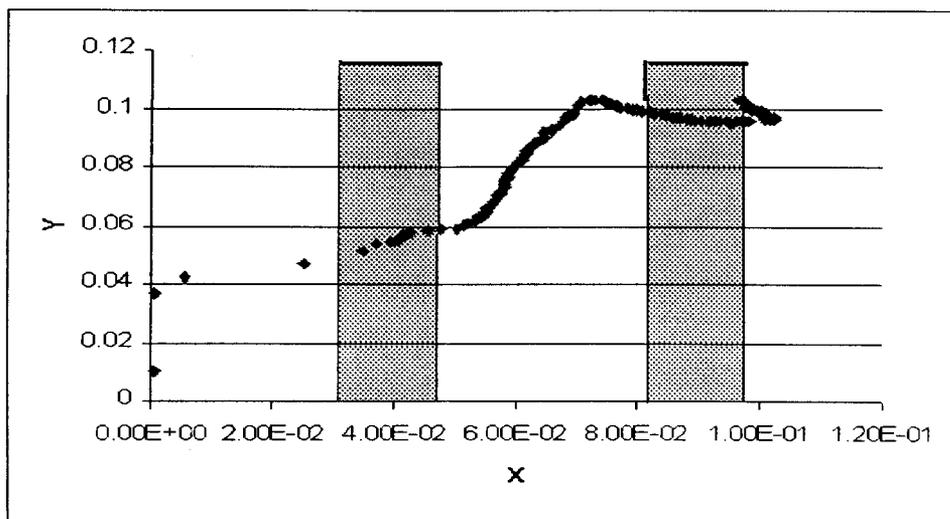


Figura 2.1. Ventanas para datos que serán suavizados

Para cada  $t$ , consideremos el intervalo  $[t - h, t + h]$ , donde  $h$  es un número positivo pequeño. Ahora formamos una “ventana”, es decir dos rectas que son paralelas al eje  $Y$  y que se intersectan con el eje  $T$  en  $t - h$  y  $t + h$ . (ver Figura 2.1). Las ventanas son las zonas sombreadas. Ahora consideremos los pares  $(t_j, y_j)$  que estén dentro de la ventana, y promediamos todos los  $y_j/s$  de los pares antes mencionados. Este promedio es el estimado del valor de la función  $x(t)$ . La ventana se puede mover hacia la izquierda o derecha, de tal manera, que se pueda computar un estimado en cualquier punto  $t$ . El estimado resultante de  $x$  es llamado algunas veces el *estimado ventana* o *promedio móvil*.

### 2.5.1 Suavizamiento por kernel

Una versión más sofisticada de ponderación local hace uso de *estimadores kernel*, también llamados *suavizadores kernel*. Aquí, el promedio computado que se analizó de una manera

breve anteriormente, es reemplazado por una suma ponderada. Típicamente, pesos grandes son asignados a  $y_i$ 's cuyos  $x_i$ 's son los más cercanos al punto que se quiere estimar ( $t$ ). Existen al menos tres versiones de estimadores kernel, a los cuales nos referiremos en las siguientes páginas. Primero consideraremos el estimador Nadaraya-Watson (Nadaraya, 1964 y Watson, 1964). Se define como

$$\hat{x}_h^{NW}(t) = \frac{\sum_{j=1}^n y_j K\left(\frac{t-t_j}{h}\right)}{\sum_{j=1}^n K\left(\frac{t-t_j}{h}\right)}, \quad 0 \leq t \leq 1, \quad (2.10)$$

donde  $K$  es una función llamada kernel. Esta función kernel  $K(u)$  está diseñada para tener la mayoría de su masa concentrada cerca del cero, y decaer rápidamente ó incluso desaparecer para  $|u| \geq 1$ . Tres kernels son usados comúnmente:

- uniforme* :  $K(u) = 0.5, \quad |u| \leq 1, \quad 0$  en otro caso,
- cuadrático* :  $K(u) = 0.75(1 - u^2), \quad |u| \leq 1, \quad 0$  en otro caso
- Gausiano* :  $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$ .

La cantidad  $h$  que está involucrada en la expresión (2.10) es llamado *ancho de banda* o *parámetro de suavizamiento* y controla la suavidad de  $\hat{x}_h^{NW}(t)$ , de la misma forma que el ancho de la ventana controla la suavidad del promedio móvil. Generalmente el kernel gaussiano es usado en el estimador Nadaraya-Watson. El estimador ventana es un caso particular del estimador Nadaraya-Watson utilizando el kernel uniforme.

Es de valor considerar a otros dos tipos de suavizadores kernel. Introducido por Priestley y Chao (1972), definimos el suavizador de la siguiente manera

$$\hat{x}_h^{PC}(t) = \frac{1}{h} \sum_{i=1}^n (t_i - t_{i-1}) y_i K\left(\frac{t - t_i}{h}\right). \quad (2.11)$$

Un tipo de suavizador similar, generalmente conocido como el estimador Gasser-Müller (1979), es

$$\widehat{x}_h^{GM}(t) = \frac{1}{h} \sum_{i=1}^n y_i \int_{s_{i-1}}^{s_i} K\left(\frac{x-u}{h}\right) du, \quad (2.12)$$

donde  $s_0 \doteq 0$ ,  $s_i = \frac{t_i + t_{i+1}}{2}$ ,  $i = 1, \dots, n-1$ , y  $s_n = 1$ . El estimador Gasser-Müller también puede ser escrito como

$$\frac{1}{h} \int_0^1 y_n(u) K\left(\frac{x-u}{h}\right) du, \quad (2.13)$$

donde  $y_n(\cdot)$  es la función constante

$$y_n(u) = \sum_{i=1}^n y_i I_{[s_{i-1}, s_i)}(u).$$

$I$  la función indicadora. En otras palabras el estimado Gasser-Müller es la convolución de  $y_n(\cdot)$  con  $\frac{K(\cdot/h)}{h}$ .

Una consecuencia de (2.13) es que el estimado Gasser-Müller tiende a la función  $y_n(\cdot)$  cuando el ancho de banda tiende a cero. Contrastando, el estimado Nadaraya-Watson no está bien definido si la  $h$  es suficientemente pequeña cuando  $K$  tiene soporte finito. Así el estimado tiende a ser mucho más inestable para  $h$  pequeña, que el estimado de Gasser-Müller.

Notemos que cada uno de los estimadores de los cuales hemos hablado en esta sección pueden ser escritos en la forma

$$\widehat{x}(t) = \sum_{i=1}^n w_{i,n}(t) y_i \quad (2.14)$$

para una función de peso en particular  $w_{i,n}(t)$ . Para asegurar la consistencia de  $\widehat{x}(t)$  es necesario que  $\sum_{i=1}^n w_{i,n}(t) = 1$ . Para el estimador Nadaraya-Watson esta condición es satisfecha debido a la manera en la cual se construyó  $\widehat{x}_h^{NW}(t)$ . Analizando la suma de pesos del estimador Gasser-Müller (el caso del estimador Priestley-Chao es esencialmente el mismo) tenemos que

$$\begin{aligned} \sum_{i=1}^n \frac{1}{h} \int_{s_{i-1}}^{s_i} K\left(\frac{t-u}{h}\right) du &= \frac{1}{h} \int_0^1 K\left(\frac{t-u}{h}\right) du \\ &= \int_{(t-1)/h}^{t/h} K(v) dv. \end{aligned}$$

Luego la función kernel debe ser seleccionada de tal manera que su integral sea 1. Entonces, la suma de pesos de kernel se aproximará a 1 tanto como  $h$  se haga pequeña relativa al  $\min(t, 1-t)$ . Si  $\int K(u)du = 1$  y  $K$  se hace cero fuera de  $(-1,1)$ , entonces la suma de pesos kernel es exactamente 1 siempre que  $h \leq \min(t, 1-t)$ . Esto explica por qué kernels con soporte finito son generalmente usados con los estimadores Gasser-Müller y Priestley-Chao. Cuando  $t$  está en  $[0, h)$  ó  $(1-h, 1]$ , los pesos no suman uno, incluso si  $K$  tiene soporte  $(-1,1)$ . Esto es la primera indicación de lo que llamamos *efectos de frontera*.

Si hay falta de información a priori de la función  $x$ , parece intuitivo que  $K$  debe de ser simétrica y tener un máximo único en cero. Una manera muy popular de asegurar que estas dos condiciones ocurran, además de que  $\int K(u)du = 1$ , es hacer  $K$  igual a una función de densidad unimodal y simétrica alrededor del cero. Haciendo esto, se garantiza un estimado positivo, si los datos son positivos; esto último es una propiedad atractiva cuando se conoce que  $x \geq 0$ .

Si queremos que el estimador kernel tenga propiedades atractivas para el error cuadrático medio, es importante seleccionar  $K$  de tal manera que

$$\int [K(u)] u du = 0, \quad \int K^2(u) du < \infty, \quad \int [K(u)] u^2 du < \infty. \quad (2.15)$$

### 2.5.2 Estimadores por funciones base localizadas

Las ideas de los estimadores kernel y estimadores funciones base pueden, en un sentido, ser combinados para producir estimadores de funciones base localizadas, los cuales incluyen una

gran clase de estimadores de funciones y derivadas. La idea básica es extender el criterio de mínimos cuadrados

$$\sum_{j=1}^n \left[ y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2,$$

con el objetivo de dar pesos locales a los errores como sigue:

$$\sum_{j=1}^n w_j(t) \left[ y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2, \quad (2.16)$$

que en términos matriciales se expresa como:

$$SMSSSE_t(y | c) = (y - \Phi c)' W(t) (y - \Phi c),$$

donde  $W(t)$  es una matriz diagonal que contiene los valores  $w_j(t)$  en su diagonal.

Minimizando  $SMSSSE_t$  sobre  $c$  se tiene que

$$c = [\Phi' W(t) \Phi]^{-1} \Phi' W(t) y.$$

Sustituyendo en la expresión  $\hat{x}(t) = \sum_{k=1}^K c_k \phi_k(t)$  da un estimador de suavizador lineal de

la forma  $\hat{x}(t) = \sum_{k=1}^K S_j(t) y_j$  con

$$S(t) = W(t) \Phi [\Phi' W(t) \Phi]^{-1} \phi(t) \quad (2.17)$$

donde  $\phi(t)$  es el vector con elementos  $\phi_k(t)$ .

Debido a que la base tiene únicamente que aproximar un segmento limitado de los datos alrededor de  $t$ , la base puede hacer un mejor trabajo en aproximar las características de los datos y al mismo tiempo operar bien con una  $K$  pequeña, es decir, con pocas funciones base, pero puede haber un costo computacional alto.

### 2.5.3 Representación polinomial local

Se escoge el valor estimado de la curva  $\hat{x}(t)$  para minimizar el criterio de mínimos cuadrados localizado

$$SMSSSE_t(y | c) = \sum_{j=1}^n \text{Kern}_h(t_j, t) \left[ y_j - \sum_{e=1}^L c_e (t - t_j)^e \right]^2. \quad (2.18)$$

Haciendo  $L = 0$  se recobra al estimador de Nadaraya-Watson. Para el valor de la función y  $L$  de sus derivadas pueden ser estimadas por las correspondientes derivadas del polinomio localmente ajustado en  $t$ . Si el orden de la derivada es  $m$ ,  $L$  debe ser de preferencia  $m + 2$

Expansiones lineales locales dan buenos resultados cuando requerimos únicamente un estimado del valor de la función.

## 2.6 Representación de Datos Penalizando la Rugosidad

Esta es una tercera opción para aproximar datos discretos por una función. El enfoque de penalización de la rugosidad o de regularización, conserva las ventajas de las técnicas de suavizamiento por funciones base y por expansión local, pero evita sus limitaciones

La expansión por funciones base trabaja bien si las funciones base tienen esencialmente las mismas características que el proceso que generó los datos.

Las técnicas de suavizamiento por kernel estándar y de ajuste de polinomios locales son eficientes y fáciles de entender en sus algoritmos. Ofrecen control de la suavidad de la aproximación, pero son soluciones raramente óptimas en un problema estadístico explícito tal como el de minimizar un error cuadrático total; además su carácter heurístico las hace de difícil generalización.

Como el enfoque de expansión en funciones base, los métodos que penalizan la rugosidad están basados en un enunciado explícito de lo que una representación suave de los datos

esté tratando de hacer, pero la necesidad de tener una representación suave está expresada explícitamente en el criterio a ser optimizado. Además, este enfoque puede ser aplicado a otros problemas de suavizamiento diferentes al de estimación de una curva  $x$ , a partir de observaciones  $x(t_i)$  para ciertos puntos  $t_i$ .

Finalmente, desarrollos computacionales recientes en suavizamiento mediante la penalización de la rugosidad, implican cálculos del orden  $O(n)$ .

Este tipo de representación está íntimamente ligada con splines suavizante, que se analizará en la sección 2.7. Primero consideraremos cómo la regularización trabaja en el caso funcional más simple cuando la meta es estimar una función  $x$  sobre la base de que tenemos un vector de datos  $y$ .

Este método estima una curva  $x$  de observaciones que satisfacen el modelo (2.1), haciendo explícito dos posibles metas en la estimación de la función. Deseamos que la curva estimada tenga un buen ajuste a los datos, por ejemplo hablando en términos de la suma de cuadrados de residuales  $\sum_j \{y_j - x(t_j)\}^2$ . Por otro lado, no deseamos que el ajuste sea demasiado bueno, si el resultado es una  $x$  que sea excesivamente variable hablando localmente.

Estas metas pueden ser vistas, en algún sentido, como las correspondientes a los dos elementos del principio básico de la estadística:

$$\text{Error cuadrático medio} = \text{Sesgo}^2 + \text{Varianza muestral}.$$

De acuerdo al modelo (2.1), cada  $y_j$  es insesgado de  $x(t_j)$ ; una curva así estimada, debe tener mucha variabilidad; pero si sacrificamos un poco el insesgamiento, un poco más suave será la curva estimada. Requiriendo que el estimado varíe sólo un poco de un valor a otro valor, estaremos tomando información de los valores de datos vecinos, y así expresando nuestra fe en la regularidad de la función subyacente  $x$  que estemos tratando de estimar.

El error cuadrático medio captura, lo que se podría llamar pobreza de estimación, y puede ser reducido de forma dramática a veces, sacrificando algo de sesgo para reducir la varianza muestral; ésta es una razón clave para imponer suavidad sobre la curva por estimar.

La filosofía de este método es permitir una clase más grandes de funciones  $x$ , pero cuantificando la rapidez de la variación local de  $x$  y haciendo un balance explícito entre regularidad y bondad de ajuste a los datos, el cual corresponde a un balance implícito entre varianza y sesgo.

Una medida popular de la rugosidad de una función se computa integrando el cuadrado de su segunda derivada

$$PEN_2(x) = \int \{D^2x(s)\}^2 ds = \|D^2x\|^2 \quad (2.19)$$

Este criterio cuantifica la curvatura total en  $x$ , o alternativamente, el grado en el cual  $x$  se desvía de una línea recta. Entonces podemos definir una suma de cuadrados de residuales penalizada como:

$$PENSSE_\lambda(x | y) = \sum_j \{y_j - x(t_j)\}^2 + \lambda \times PEN_2(x). \quad (2.20)$$

Nuestro estimado de la función es obtenido al encontrar la función  $x$  que minimiza  $PENSSE_\lambda(x)$  sobre el espacio de las funciones  $x$  para la cual  $PEN_2(x)$  esté definida, es decir, el espacio de las funciones con segunda derivada cuadrado-integrable.

El parámetro  $\lambda$  es un parámetro de suavidad que mide la tasa de intercambio entre el ajuste de los datos y la variabilidad de la función  $x$ . Entre más grande  $\lambda$ , más grande el peso que se le dará a la rugosidad de la estimación. Para  $\lambda$  pequeña la curva tiende a ser más y más variable, ya que hay menos penalidad sobre su rugosidad. Cuando  $\lambda \rightarrow 0$  la curva se

aproxima a una interpolación de los datos satisfaciendo  $y_j = x(t_j) \quad \forall j$ . La curva  $x$  resultante del uso de  $PEN_2$  puede mostrarse (Green y Silverman, 1994) que es un spline cúbico con nodos en los  $t_j$ 's. Si la derivada de mayor orden requerida es  $m$ , uno debería penalizar sobre derivadas de orden  $m + 2$ . Por ejemplo si  $m = 2$   $PEN_4 = \int_{\gamma} \{D^4 x(s)\}^2 ds = \|D^4 x\|^2$ .

## 2.7 Propiedades de Splines

Dada la importancia de los splines cúbicos, profundicemos en algunas de sus propiedades. En esta sección explicaremos como la curva  $\hat{x}$  que minimiza la suma de cuadrados penalizada (2.20), resulta ser un spline cúbico natural. Anteriormente hablamos acerca de lo que es un spline, en el método de funciones base, así mismo hicimos una breve explicación de B-splines. Nuestra atención en esta sección se enfocará en los splines cúbicos. Una función  $x$  sobre  $[a, b]$ , es un spline cúbico si

- 1) en cada uno de los intervalos  $(a, \tau_1), (\tau_1, \tau_2), \dots, (\tau_n, b)$ ,  $x$  es un polinomio cúbico
- 2) las piezas polinomiales se juntan en los puntos  $\tau_i$  de tal forma que  $x$  en sí misma y sus primeras y segundas derivadas son continuas en cada  $\tau_i$  y así en todo el intervalo  $[a, b]$ .

Hay muchas formas de representar un spline cúbico, al parecer; pero una forma obvia es dar los coeficientes de cada cacho o pieza cúbica que conforma el spline cúbico, es decir,

$$x(t) = d_i (t - \tau_i)^3 + c_i (t - \tau_i)^2 + b_i (t - \tau_i) + a_i \text{ para } \tau_i \leq t \leq \tau_{i+1} \quad (2.21)$$

donde los  $\tau_i$ 's son los nodos en el intervalo  $[a, b]$ , y  $a_i, b_i, c_i, d_i$  son constantes dadas; cabe recalcar que se define  $\tau_0 = a$  y  $\tau_{n+1} = b$ .

Se dice que un spline cúbico sobre un intervalo  $[a, b]$ , es un *spline cúbico natural* (SCN) si su segunda y tercera derivada son cero en  $a$  y  $b$ . Estas condiciones son llamadas *condiciones naturales de frontera*. Estas últimas implican que  $d_0 = c_0 = d_n = c_n = 0$ , entonces  $x$  es lineal sobre los dos extremos del intervalo  $[a, \tau_1]$  y  $[\tau_n, b]$ .

Ahora pasaremos a representar un SCN en forma matricial, lo cual nos facilitara el cómputo y la representación matemática. Supongamos que  $x$  es un SCN con nodos  $\tau_1 < \dots < \tau_n$ . Se define

$$x_i = x(\tau_i) \text{ y } \gamma_i = D^2x(\tau_i) \text{ para } i = 1, \dots, n.$$

Sean  $\mathbf{x}$  el vector  $(x_1, \dots, x_n)'$  y  $\boldsymbol{\gamma}$  el vector  $(\gamma_2, \dots, \gamma_{n-1})'$ . Nota que las entradas  $\gamma_i$  del vector  $\boldsymbol{\gamma}$  de longitud  $n-2$  están numeradas empezando en  $i = 2$ , lo cual simplificará algunos cálculos más adelante. Los vectores  $\mathbf{x}$  y  $\boldsymbol{\gamma}$  especifican la curva  $x$  completamente, y es posible dar una fórmula explícita en términos de estos dos vectores. Existen condiciones suficientes y necesarias para que los vectores representen genuinamente un spline cúbico natural sobre una sucesión de nodos dada.

La condición depende de dos matrices  $\mathbf{Q}$  y  $\mathbf{R}$  las cuales definiremos a continuación. Sea  $h_i = \tau_{i+1} - \tau_i$ , para  $i = 1, \dots, n-1$ . Definamos a  $Q$  como una matriz de  $n \times (n-2)$  con entradas  $q_{ij}$ , donde  $i = 1, \dots, n$  y  $j = 2, \dots, n-1$ , dadas por

$$q_{j-1,j} = h_{j-1}^{-1}, \quad q_{jj} = -h_{j-1}^{-1} - h_j^{-1}, \quad \text{y } q_{j+1,j} = h_j^{-1}$$

para  $j = 2, \dots, n-1$ , y  $q_{ij} = 0$  para  $|i-j| \geq 2$ .  $\mathbf{Q}$  es bandeada.

La matriz simétrica  $\mathbf{R}_{(n-2) \times (n-2)}$  con elementos  $r_{ij}$ , para  $i, j = 2, \dots, n-1$ , dados por

$$r_{ii} = \frac{1}{3}(h_{i-1} + h_i) \text{ para } i = 2, \dots, n-1$$

$$r_{i,i+1} = r_{i+1,i} = \frac{1}{6}h_i \text{ para } i = 2, \dots, n-1$$

y  $r_{ii} = 0$  para  $|i-j| \geq 2$ .  $\mathbf{R}$  también es bandeada

La matriz  $\mathbf{R}$  es estrictamente diagonal dominante, es decir, que  $|r_{ii}| > \sum_{j \neq i} |r_{ij}|$  lo cual implica que  $\mathbf{R}$  es estrictamente definida positiva por álgebra lineal. Definamos la matriz  $\mathbf{K}$  como

$$\mathbf{K} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}'. \tag{2.22}$$

Con estos elementos en mano podemos enunciar el siguiente resultado

**Teorema 2.1** *Los vectores  $\mathbf{x}$  y  $\boldsymbol{\gamma}$  especifican un SCN  $x$  si y sólo si la condición*

$$\mathbf{Q}'\mathbf{x} = \mathbf{R}\boldsymbol{\gamma} \quad (2.23)$$

*es satisfecha. Si esto ocurre entonces el término de penalización satisface*

$$\int_a^b D^2x(t)dt = \boldsymbol{\gamma}'\mathbf{R}\boldsymbol{\gamma} = \mathbf{x}'\mathbf{K}\mathbf{x} \quad (2.24)$$

Para la demostración vea Green and Silverman (1994).

Ahora contando con estos elementos podemos afirmar que  $PENSSE_\lambda(x | y)$  en (2.20) tiene un mínimo único dado por  $\mathbf{x} = (\mathbf{I} + \lambda\mathbf{K})^{-1}\mathbf{y}$ , el cual define un único spline cúbico natural, apoyándonos en el siguiente teorema

**Teorema 2.2** *Supongamos  $n \geq 3$  y que tenemos una sucesión de puntos  $a < \tau_1 < \dots < \tau_n < b$ . Dados los datos  $y_1, \dots, y_n$ , y un parámetro de suavidad estrictamente positivo  $\lambda$ , sea  $\hat{x}$  un SCN con nodos en  $\tau_1, \dots, \tau_n$  para el cual  $\mathbf{x} = (\mathbf{I} + \lambda\mathbf{K})^{-1}\mathbf{y}$ . Entonces para cualquier  $x$  en  $\mathcal{L}^2[a, b]$ ,*

$$PENSSE_\lambda(\hat{x} | y) \leq PENSSE_\lambda(x | y)$$

*con igualdad únicamente si  $x$  y  $\hat{x}$  son iguales. Para la demostración vea Green and Silverman (1994).*

Finalmente enunciaremos el algoritmo de Reinsch (1967) para encontrar el spline suavizante.

Si

$$(\mathbf{I} + \lambda\mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}')\mathbf{x} = \mathbf{y} \quad (2.25)$$

reordenamos (2.25) de tal manera que se obtiene

$$\mathbf{x} = \mathbf{y} - \lambda \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}' \mathbf{x};$$

ahora sustituimos  $\mathbf{Q}' \mathbf{x} = \mathbf{R} \boldsymbol{\gamma}$  y simplificamos para obtener una fórmula explícita para  $\mathbf{x}$  en términos de  $\mathbf{y}$  y  $\boldsymbol{\gamma}$

$$\mathbf{x} = \mathbf{y} - \lambda \mathbf{Q} \boldsymbol{\gamma}. \quad (2.26)$$

Otra vez utilizando la igualdad para obtener (2.26) tenemos

$$\mathbf{Q}' \mathbf{y} - \lambda \mathbf{Q}' \mathbf{Q} \boldsymbol{\gamma} = \mathbf{R} \boldsymbol{\gamma}$$

obteniendo la ecuación para  $\boldsymbol{\gamma}$

$$(\mathbf{R} + \lambda \mathbf{Q}' \mathbf{Q}) \boldsymbol{\gamma} = \mathbf{Q}' \mathbf{y}. \quad (2.27)$$

Esta ecuación es el núcleo del algoritmo. Contrastando (2.25) con (2.26), puede ser resuelto en tiempo lineal usando técnicas de matrices con estructura de banda.

Finalmente haciendo una descomposición de Cholesky para la matriz  $\mathbf{R} + \lambda \mathbf{Q}' \mathbf{Q}$  obtenemos

$$\mathbf{R} + \lambda \mathbf{Q}' \mathbf{Q} = \mathbf{L} \mathbf{D} \mathbf{L}'$$

donde  $\mathbf{D}$  es una matriz diagonal estrictamente positiva y  $\mathbf{L}$  es una matriz triangular inferior. Así el algoritmo de Reinsch puede ser operado. Debido a que las matrices en el algoritmo son bandeadas, cada paso se hace en cálculos del orden  $O(n)$ .

## 2.8 Propiedades Estadísticas de Suavizadores

Los suavizadores analizados en las secciones anteriores proveen descripciones muy útiles acerca de los datos. Sin embargo, cuando usamos estos suavizadores para estimar a la función subyacente, es de importancia analizar sus propiedades estadísticas. El enfoque de suavizamiento utilizado en los programas desarrollados por Ramsay y Silverman es el de penalización de la rugosidad.

Silverman (1984) mostró que existe una relación entre el estimado por suavizamiento (penalizando la rugosidad) y un estimado particular por kernel. De la ecuación (2.25) se puede pensar que

$$\hat{x}(s) = n^{-1} \sum_{i=1}^n y_i w_{i,n}(s)$$

para cada  $\tau_i$ ,  $w_{i,n}(s)$ , la función de peso, es la curva estimada que resultará si  $y_i = n$  o  $y_j = 0$ ,  $j \neq i$ . En este sentido  $w_{i,n}(s)$  puede ser pensada como una versión suavizada de una función  $\delta$  en  $\tau_i$ . Esta  $w_{i,n}(s)$  depende obviamente de los  $\tau_i$ 's y del parámetro de suavizamiento  $\lambda$  pero no depende de  $y_i$ 's, por construcción. Se demostró que suponiendo que  $n$  es grande, que los puntos de diseño  $\tau_i$ 's tienen una densidad local  $f(\tau)$ , tal que  $s$  no está cercano a la orilla del intervalo en el que los datos se ubican y que  $\lambda$  no es muy grande ni muy chica,

$$w_{i,n}(s) \approx \frac{1}{f(\tau)} \frac{1}{h(\tau)} K\left(\frac{s - \tau}{h(\tau)}\right)$$

con

$$K(u) = \frac{1}{2} \exp\left(-\frac{|u|}{\sqrt{2}}\right) \operatorname{sen}\left(\frac{|u|}{\sqrt{2}} + \frac{\pi}{4}\right)$$

y

$$h(\tau) = \lambda^{1/4} n^{-1/4} f(\tau)^{-1/4}.$$

Analizando estas expresiones se aprecia que el suavizamiento por spline es aproximadamente un método de suavizamiento por convolución. El estimador Gasser-Müller como se puede apreciar en (2.12), es un método de suavizamiento por convolución también.

En esta sección plantaremos resultados acerca del error cuadrático medio de los estimadores Gasser-Müller (Hart, 2000), esto se debe a la equivalencia explicada en el principio de la sección.

Anteriormente se había comentado de las propiedades que deben tener las funciones kernel para el estimador Gasser-Müller, expresión (2.15), para que éste a su vez tenga propiedades

atractivas para el error cuadrático medio del mismo. Se ha comentado también que el ancho de banda de este estimador afecta sus propiedades estadísticas. Generalmente, el sesgo de un estimador kernel es más pequeño cuando el ancho de banda lo es. Desafortunadamente, cuando el ancho de banda tiende a ser pequeño también tiene el efecto de incrementar la varianza del estimador. Una meta principal en la estimación por kernel es encontrar el ancho de banda que logre un balance adecuado entre sesgo y varianza. Hay algunos métodos para escoger  $h$ , uno de estos métodos será analizado más adelante.

En esta sección supondremos que los datos  $y_1, \dots, y_n$  satisfacen el modelo (2.1), además supondremos que los puntos de diseño  $t_1, t_2, \dots, t_n$  son generados por una función de densidad  $f$ , positiva y continua en el sentido de Lipschitz, (se dice que una función  $g$  es *Lipschitz continua* sobre el intervalo  $[a, b]$  si existe una constante  $C$  tal que  $|g(u) - g(v)| \leq C |u - v|$ ,  $\forall u, v \in [a, b]$ ), con lo que

$$t_j = Q\left(\frac{j - \frac{1}{2}}{n}\right), j = 1, \dots, n$$

donde  $Q(u) = F^{-1}(u)$  y  $F(x) = \int_0^x f(u)du$ , con  $0 \leq u \leq 1$ , es decir que los  $t_j$ 's son cuantiles de la función de distribución  $F$ .

Ahora nuestra atención se enfocará en el error cuadrático medio (ECM) para el estimador Gasser-Müller. Para simplificar notación denotaremos al estimador Gasser-Müller  $\hat{x}_h^{GM}(t)$  por  $\hat{x}_h(t)$ . Además supondremos las siguientes condiciones para el kernel  $K$ :

1.  $K$  tiene soporte  $(-1, 1)$ .
2.  $K$  es Lipschitz continua.
3.  $\int_{-1}^1 K(u)du = 1$ .
4.  $\int_{-1}^1 K(u)udu = 0$ .

Definimos las constantes  $J_K$  y  $\sigma_K^2$  como

$$J_K = \int_{-1}^1 K(u)^2 du$$

$$J_K = \int_{-1}^1 K^2(u) du \quad (2.28)$$

y

$$\sigma_K^2 = \int_{-1}^1 u^2 K(u) du. \quad (2.29)$$

El segundo momento de  $K$  es denotado por  $\sigma_K^2$  conforme a la notación usual, pero notemos que  $\sigma_K^2$  podría ser negativo si suponemos la cuatro condiciones anteriores, permitiendo a  $K$  que tome valores negativos.

A continuación enunciaremos un teorema clave en las propiedades estadísticas del estimador Gasser-Müller, donde obtendremos expresiones para la varianza, el sesgo y el error cuadrático medio, el cual nos dará la pauta para obtener el balance requerido entre sesgo y varianza.

**Teorema 2.3.** Supongamos que el modelo (2.1) y sus condiciones adicionales mencionadas anteriormente se cumplen, y sea  $t \in (0, 1)$ . Suponga además que  $x$  tiene dos derivadas continuas. Entonces, cuando  $n \rightarrow \infty$ ,  $h \rightarrow 0$  y  $nh \rightarrow \infty$ , la varianza, el sesgo y por ende el error cuadrático medio del estimador Gasser-Müller son respectivamente

$$V(t; h) = \frac{\sigma^2}{nh} \frac{1}{f(t)} J_K + O(n^{-1}) + O(nh)^{-1} \quad (2.30)$$

$$E[\hat{x}(t)] - x(t) = \frac{h^2}{2} D^2 x(t) \sigma_K^2 + o(h^2) + O(n^{-1}) \quad (2.31)$$

$$M(t; h) = \frac{\sigma^2}{nh} \frac{1}{f(t)} J_K + \frac{h^4}{4} [D^2 x(t)]^2 \sigma_K^4 + o(h^4) + O(n^{-1}) + O(nh)^{-2}. \quad (2.32)$$

Demostración ver Hart (2000) teoremas 3.1 y 3.2 y corolario 3.1.

De (2.30) - (2.32), es claro que cuando  $h$  se hace pequeña, el sesgo al cuadrado decrece, pero la varianza se incrementa.

Si se escoge a  $h$  para minimizar a  $M(t; h)$  en sus primeros dos términos,

$$D \left[ \frac{\sigma^2}{nh} \frac{1}{f(t)} J_K + \frac{h^4}{4} [D^2 x(t)]^2 \sigma_K^4 \right] = 0$$

implica que

$$h_n = \left( \frac{\sigma^2 J_K}{f(t) [D^2 x(t)]^2 \sigma_K^4} \right)^{1/5} n^{-1/5} \quad (2.33)$$

y así

$$M(t; h_n) \approx 1.25 J_K^{4/5} \sigma_K^{4/5} |D^2 x(t)|^{2/5} \left( \frac{\sigma^2}{f(t)} \right)^{4/5} n^{-4/5}. \quad (2.34)$$

Con ello note que como  $h_n$  es proporcional a  $\sigma^{2/5}$ , entre más ruidosos los datos, mayor será el ancho de banda óptimo. También se aprecia que el ancho de banda óptimo es más pequeño en puntos  $t$  donde la curvatura de  $x$  sea grande. De este modo se aprecia que, como era de esperarse, la varianza y la curvatura son factores competitivos.

También se aprecia que  $h_n$  es decreciente en  $n$ ; así dado que el ancho de banda debe ser lo suficientemente pequeño para detectar ciertas características de  $x$ , habrá un tamaño muestral por abajo del cual será imposible detectar tales características aún cuando el ancho de banda sea el “óptimo”.

Seleccionando el ancho de banda óptimo, existen varias posibilidades para tener una función Kernel adecuada; pero como explica Hart (2000), la selección de la función kernel *no* es ni cercanamente tan importante como la selección del ancho de banda.

Ahora entrando propiamente a la inferencia estadística del estimador analizado anteriormente, consideremos la cantidad

$$\frac{\hat{x}(t) - x(t)}{\sqrt{V(\hat{x}(t))}} = Z_n + B_n$$

dónde

$$Z_n = \frac{\hat{x}(t) - E[\hat{x}(t)]}{\sqrt{V(\hat{x}(t))}}$$

$$B_n = \frac{E[\hat{x}(t)] - x(t)}{\sqrt{V(\hat{x}(t))}}.$$

Hart (2000) muestra que, una vez seleccionado el ancho de banda óptimo (2.33), para cada  $n$

$$Z_{n,h_n} \xrightarrow{D} N(0, 1),$$

es decir,  $Z_{n,h_n}$  converge en distribución a una normal con media 0 y varianza 1 cuando  $n \rightarrow \infty$ .

Por lo tanto para construir intervalos de confianza para  $x(t)$  hay varias posibilidades, una de ellas es:

$$\hat{x}_h(t) \pm Z_{\alpha/2} \hat{\sigma} \left[ \sum_{i=1}^n w_i^2(t, h) \right]^{1/2}$$

y es tal que no converge el nivel a  $(1 - \alpha)$ , debido a que la probabilidad de cobertura es

$$\Phi(Z_{\alpha/2} - B_{nh}) + \Phi(Z_{\alpha/2} + B_{nh}) - 1,$$

donde  $\Phi$  es una función de distribución acumulativa de una normal con media 0 y varianza 1.

Sin embargo, considerando  $E[\hat{x}_h(t)]$  como la parte estimable de  $x(t)$ , y así tratando a  $E[\hat{x}(t)]$  como el parámetro de interés, este intervalo puede ser usado para obtener un intervalo asintóticamente válido para  $E[\hat{x}(t)]$ , una versión “suavizada” de  $x(t)$ ; si lo que interesa es la forma de  $x(t)$  este enfoque es adecuado.

Existen otras alternativas, ver detalles, por ejemplo, en Hart (2000).

Otro enfoque es considerar una banda de confianza simultánea para la función entera  $x$ . Un número de métodos han sido propuestos para construir tales bandas, en donde se toma en cuenta el sesgo, igual que en el caso de intervalos puntuales, para garantizar la validez de los intervalos correspondientes (Hart, 2000).

Con estos resultados, los cuales hacen ver la importancia de la elección del ancho de banda, en la estimación por kernel, es el tiempo para ocuparnos de otro parámetro importante en el ajuste de la curva  $x$ , pero estimando penalizando la rugosidad, es la elección del parámetro de suavidad  $\lambda$ , que juega el rol equivalente al ancho de banda en la estimación por kernel.

Existen dos filosofías de aproximación a la pregunta de la elección del parámetro de suavidad. La primera aproximación es ver la libre elección del parámetro de suavidad como una ventaja del procedimiento. Variando  $\lambda$ , características de los datos obtenidos a partir de esta variación sobre diferentes “escalas” y pueden ser exploradas, y si un estimado es necesitado al último, puede ser elegido subjetivamente. Esta aproximación subjetiva es en realidad la más utilizada en la práctica. La otra forma de ver la elección, por cierto opuesta a la primera, es que existe la necesidad de un método automático que use los datos para escoger el valor de  $\lambda$ . Es más claro usar la palabra *automático* que la palabra *objetivo* para nombrar a tales métodos, porque, como en la mayoría de los procedimientos estadísticos, existen decisiones arbitrarias que se involucran en la elección del método mismo.

Los métodos automáticos no necesariamente deben ser usados de una manera no crítica; ellos pueden ser usados como un punto inicial de estimación. En el contexto anterior puede ser bueno que exista una preferencia para usar el mismo valor del parámetro de suavidad a través de diferentes conjunto de datos a fin de lograr comparaciones, o simplemente porque un valor en particular es conocido debido a la experiencia, con el cual se pueden obtener buenos resultados.

Existen un gran número de diferentes procedimientos automáticos disponibles. Probablemente el más conocido sea validación cruzada. A continuación describiremos brevemente una versión modificada de este método, el cual llamaremos *validación cruzada*

*generalizada* (GCV, pos sus iniciales en inglés), el cual es un método popular para escoger el parámetro de suavidad.

Recordaremos el procedimiento de validación cruzada y al mismo tiempo denotaremos algunos parámetros. Sea  $\hat{x}^{(-i)}(t; \lambda)$  la curva estimada que minimiza

$$\sum_{j \neq i} \{y_j - x(t_j)\}^2 + \lambda \int D^2 x(t). \quad (2.35)$$

La cualidad de  $\hat{x}^{(-i)}$  como un predictor sobre una nueva observación puede ser valorada, analizando qué tan bien el valor  $\hat{x}^{(-i)}(t_i)$  predice a  $y_j$ . Debido a que la elección de la observación omitida es arbitraria, la eficacia global del procedimiento con el parámetro de suavidad  $\lambda$  puede ser cuantificada por la función score validación cruzada

$$CV(\lambda) = n^{-1} \sum_{i=1}^n \{y_j - \hat{x}^{(-i)}(t_j; \lambda)\}^2. \quad (2.36)$$

La idea básica de validación cruzada es escoger el valor de  $\lambda$  que minimice  $CV(\lambda)$ .

De (2.36) se podría pensar que es necesario resolver  $n$  problemas diferentes de suavizamiento, a fin de encontrar  $n$  curvas  $\hat{x}^{(-i)}$ . Sin embargo, no es el caso, se tiene una simplificación a partir del Teorema 2.2, renombrando los valores computados del spline  $\hat{x}$ , el cual depende linealmente de los datos  $y_j$  a través de la ecuación

$$\mathbf{x} = A(\lambda)\mathbf{y} \quad (2.37)$$

donde la matriz  $A(\lambda)$  es llamada *matriz sombrero* debido a que mapea el vector de valores observados  $y_j$  a sus “valores predichos”  $\hat{x}(t_j)$ , así tendremos una manera económica de computar (2.36), debido a que este score satisface

$$CV(\lambda) = n^{-1} \sum_{i=1}^n \left( \frac{y_i - \hat{x}(t_i)}{1 - A_{ii}(\lambda)} \right) \quad (2.38)$$

donde  $\hat{x}$  es el spline suavizante estimado a partir del conjunto de datos completo con el parámetro de suavidad  $\lambda$ .

Con esta forma de estimación del score, conociendo  $A_{ii}(\lambda)$ , los elementos de la diagonal principal de  $A(\lambda)$ , éste puede ser fácilmente computado de los residuales  $y_i - \hat{x}(t_i)$ , los cuales son obtenidos de la estimación de la curva  $x$  a partir del conjunto de datos completo.

La idea básica de GCV es reemplazar los factores  $1 - A_{ii}(\lambda)$  de la ecuación (2.38) por su valor promedio  $1 - n^{-1}tr[A(\lambda)]$ . Así que análogamente los scores GCV son

$$GCV(\lambda) = n^{-1} \frac{\sum_{i=1}^n \{y_i - \hat{x}(t_i)\}^2}{\{1 - n^{-1}tr[A(\lambda)]\}^2}, \quad (2.39)$$

es decir, la suma de cuadrados de residuales sobre  $\hat{x}$  dividida por un factor de corrección de  $n \{1 - n^{-1}tr[A(\lambda)]\}^2$ . De la misma forma que validación cruzada la elección del parámetro de suavidad está determinada por la minimización de la función  $GCV(\lambda)$  sobre  $\lambda$ .

Una de las razones por las cuales se introdujo GCV fue computacional. Usando una expresión alternativa para la traza de la matriz sombrero como la suma de sus valores propios, es posible encontrar la traza de  $A(\lambda)$  sin encontrar todos los elementos de su diagonal. Supongamos que la matriz  $\mathbf{QR}^{-1}\mathbf{Q}'$  tiene valores propios  $\omega_v$ ; entonces se sigue de (2.25) y (2.37) que  $A(\lambda)$  tiene eigenvalores  $(1 + \lambda\omega_v)^{-1}$ , de este modo tendremos que

$$n \{1 - n^{-1}tr[A(\lambda)]\}^2 = n \left( 1 - n^{-1} \sum_{v=1}^n \frac{1}{1 + \lambda\omega_v} \right)^2. \quad (2.40)$$

Una vez que los valores propios  $\omega_v$  han sido computados, el score GCV puede ser encontrado a partir del cuadrado de la suma de residuales, para *cualquier*  $\lambda$ , llevando a cabo el cómputo fácil de (2.39).

Esto motiva a una conexión con regresión clásica, computando “grados de libertad equivalentes”, que dan una indicación del número de parámetros efectivos, en algún sentido, que son ajustados para cualquier valor particular del parámetro de suavidad. Supongamos, por un momento, que se ajustó una curva  $x$  en funciones base, es decir tiene la forma  $\hat{x}(t) = \sum_{j=1}^K c_j \phi_j(t)$ , vía mínimos cuadrados. Además si suponemos que los parámetros son identificables, sobre la base que tenemos datos observados, la matriz sombrero  $A$  es entonces una proyección sobre un espacio de dimensión  $K$ , el número de parámetros ajustados, así como su traza es igual a  $K$ . Así los grados de libertad,  $K$ , son iguales a la traza de  $A$ , mientras los grados de libertad para el ruido,  $n - K$ , son iguales a  $\text{tr}(I - A)$ .

Si ahora regresamos al caso del spline suavizante, análogamente, definimos *los grados de libertad para el ruido* (EDF) como

$$EDF = \text{tr} \{I - A(\lambda)\} \quad (2.41)$$

donde, ahora,  $A(\lambda)$  es la matriz sombrero asociada con el spline suavizante con parámetro de suavidad  $\lambda$ . Los grados de libertad equivalentes para el ruido toman valores desde 0 cuando  $\lambda = 0$  (interpolación, matriz sombrero es igual a la identidad) a  $n - 2$  cuando  $\lambda = \infty$  (regresión lineal). Esto se sigue inmediatamente de las definición de score GCV que puede ser escrita en la forma

$$GCV(\lambda) = n \times \frac{\text{suma de cuadrados de residuales}}{(\text{grados de libertad equivalentes})^2}$$

Con estos elementos en mano podemos proseguir para poder estimar la varianza del error en nuestro modelo (2.1), al igual que en regresión paramétrica, donde esta varianza se estima dividiendo la suma de cuadrados de residuales por los grados de libertad del error, se hace una analogía en el contexto de penalización de la rugosidad, dividiendo la suma de cuadrados de residuales por sus grados de libertad equivalentes. Así si el parámetro de suavidad es  $\lambda$ , se tiene que el estimador de la varianza es

$$\hat{\sigma}_\lambda^2 = \frac{\sum_i \{y_i - \hat{x}(t_i)\}^2}{\text{tr} \{I - A(\lambda)\}} \quad (2.42)$$

donde  $\hat{x}$  es el estimador resultante de la penalización de la rugosidad con parámetro de suavidad  $\lambda$ . Se puede mostrar que en el caso donde la función verdadera  $x$  es una recta, el estimador  $\hat{\sigma}_\lambda^2$  es un estimador insesgado de  $\sigma$  para toda  $\lambda$ . Esto motiva a usar el término de “grados de libertad equivalentes” para  $tr \{I - A(\lambda)\}$ .

Es así como hemos analizado los distintos enfoques para estimar la curva  $x$ , los estimadores resultantes de cada método tiene propiedades estadísticas muy interesantes y las cuales nos guían sobre el método que debemos elegir para hacer una buena estimación, es claro también que el cómputo fácil, la estructura de los datos y el contexto en el que estemos trabajando son de ayuda para escoger el método. Teniendo todos estos elementos requeridos para representar a  $x$  de la mejor forma posible, empezaremos con el análisis estadístico de las curvas, siguiendo la metodología antes mencionada.

## 3 Análisis de Componentes Principales

### 3.1 Introducción

Análisis de componentes principales (ACP) de datos funcionales es una técnica muy importante a considerar. Este análisis, es una generalización del caso multivariado, es decir, estamos pasando de espacios de dimensión finita a espacios de dimensión infinita. Después de haber hecho análisis descriptivos de un conjunto de datos funcionales sin mayor estructura a priori conocida, si se quisiera encontrar fuentes de variación importantes en los procesos que generen los datos funcionales, el Análisis de Componentes Principales es una herramienta, realmente admirable, para descubrir fuentes de variación.

Como en el caso multivariado, donde la interpretación directa de la matriz de varianzas y covarianzas, así como la de correlación, puede ser complicada y no siempre podemos obtener una presentación fácil y comprensible de la estructura de la variabilidad en los datos observados, igualmente ocurre con las funciones de varianza, covarianza y correlación de los datos funcionales. Así el análisis de componentes principales nos provee la oportunidad de examinar la estructura de varianza y covarianza de una manera que es más informativa y complementa, incluso puede reemplazar, el análisis directo de la función de varianza y covarianza.

### 3.2 Análisis de Componentes Principales Funcionales

Ahora pasaremos de lo discreto a lo continuo  $x_{ij} \rightarrow x_i(s)$ , si antes se sumaba, ahora se integra, entonces las combinaciones lineales serán:

$$f_i = \int \beta(s)x_i(s)ds = \langle \beta, x_i \rangle. \quad (3.1)$$

En el primer paso de ACP funcional, la función de peso  $\beta$  es escogida de tal manera que maximiza

$$N^{-1} \sum_i f_{i1}^2 = N^{-1} \sum_i \langle \beta, x_i \rangle^2 \quad (3.2)$$

sujeto a la restricción  $\|\beta\|^2 = 1$  donde  $\|\beta\| = \int \beta(s)^2 ds$ .

Denotemos tal función de peso obtenida por  $\xi_1$ . Las funciones de peso  $\xi_m$  con  $m \geq 2$ , son requeridas también y deben cumplir la restricción de ortogonalidad, *i.e.*  $\langle \xi_k, \xi_m \rangle = 0$   $k < m$ , sobre los pasos subsecuentes. Cada función de peso tiene la tarea de definir los modos de variación más importantes en las curvas.

A los componentes principales funcionales también se les llama armónicos, esto es debido a que así se le llama en Física a los componentes principales de variación en curvas en general. En la Figura 3.1 mostramos los cuatro armónicos del ejemplo del clima en las estaciones canadienses, concluyendo de la siguiente forma: aunque  $\xi_1$  es positivo a través de todo el año, los pesos situados en las temperaturas de invierno son cerca de cuatro veces mayor que los de verano. Esto significa que la mayor variabilidad se encuentra en las temperaturas de invierno, además de que el porcentaje de variación es de 89.3, nos dice que tan dominante es este tipo de variación.

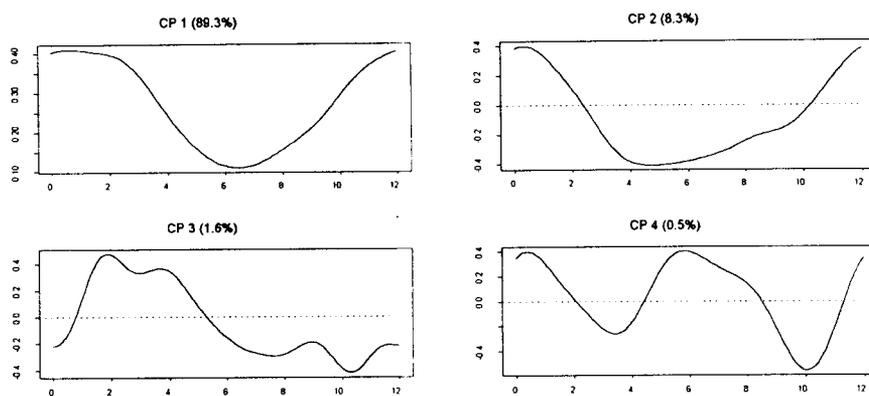


Figura 3.1. Primeras cuatro funciones componentes principales estimadas de la expansión de los datos de clima canadiense en 12 términos de la serie de Fourier.

### 3.2.1 Bases empíricas ortonormales

Otra forma de motivar el análisis de componentes principales funcionales se presenta de la siguiente manera:

Se quiere encontrar un conjunto que contenga exactamente  $K$  funciones ortonormales  $\xi_m$  tales que la expansión de cada curva en términos de estas funciones (base)

aproximen a la curva tanto como sea posible. Debido a que estas funciones base son ortonormales se sigue que la aproximación (expansión) será de la forma:

$$\tilde{x}_i(t) = \sum_{k=1}^K f_{ik} \xi_k(t) \quad (3.3)$$

donde

$$f_{ik} = \langle x_i, \xi_k \rangle.$$

Un criterio de ajuste para cada curva puede ser considerado el de integrar al error cuadrático medio, es decir,

$$\|x_i - \tilde{x}_i\|^2 = \int [x(s) - \tilde{x}(s)]^2 ds, \quad (3.4)$$

y una medida global de aproximación es:

$$PCASSE = \sum_{i=1}^N \|x_i - \tilde{x}_i\|^2 \quad (3.5)$$

Luego el problema se reduce a encontrar la base que minimiza (3.5). La respuesta es el mismo conjunto de funciones de peso que maximizan (3.2); por esta razón estas funciones  $\xi_m$  se les llama en algunos contextos funciones ortonormales empíricas, debido a que están determinadas por los datos.

### 3.2.2 Eigenanálisis

Veamos el caso multivariado:

Sea  $\mathbf{X}$  una matriz  $N \times p$  con valores  $x_{ij}$  (centrados) y el vector  $\xi$  de longitud  $p$  que contiene los pesos de la combinación lineal, entonces el criterio para encontrar el primer componente principal puede ser escrito, utilizando (3.5), como

$$\max_{\xi' \xi = 1} N^{-1} \xi' X' X \xi = \max_{\xi' \xi = 1} \xi' V \xi. \quad (3.6)$$

con  $V = N^{-1} X' X$ .

Usando cálculo el problema de maximización es resuelto encontrando la solución con el valor propio  $\rho$  más grande de la ecuación propia

$$V \xi = \rho \xi. \quad (3.7)$$

Llevando esto al caso funcional la ecuación propia resultante es:

$$\int v(s, t) \xi(t) dt = \langle v(s, \cdot), \xi \rangle = \rho \xi(s) \quad (3.8)$$

donde  $v(s, t)$  es el elemento  $(s, t)$  de  $V$ , es decir,  $v(s, t) = N^{-1} \sum_{i=1}^N x_i(s) x_i(t)$ . Se puede cambiar  $N$  por  $N - 1$  pero el análisis no muestra cambios substanciales.

El lado izquierdo de (3.8) es una transformación  $V$  de la función de peso  $\xi$  con el kernel de la transformación  $v$  definido por:

$$V \xi = \int v(\cdot, t) \xi(t) dt \quad (3.9)$$

Esta transformación integral opera sobre  $\xi$ , entonces podemos expresar la ecuación-propia directamente como  $V \xi = \rho \xi$  donde  $\xi$  es una función-propia y ya no un vector propio.

**Resumiendo:** se encuentra que ACP está definido como la búsqueda de un conjunto de funciones de peso ortogonales y normalizados  $\xi_m$ . ACP funcional puede ser expresado como el problema del eigenanálisis del operador covarianza  $V$ , definido al usar la función covarianza  $v$  como el kernel de una transformación integral.

### 3.2.3 Interpretación de ACP

Tres métodos nos pueden ayudar a la interpretación del análisis de componentes principales:

1) Graficar componentes como perturbaciones de la media. Para construir esta gráfica es necesario escoger un múltiplo de la función componente principal a graficar. Tal múltiplo denotado por  $C$  puede ser

$$C^2 = T^{-1} \|\widehat{\mu} - \bar{\mu}\|$$

donde  $\bar{\mu} = T^{-1} \int \widehat{\mu}(t) dt$ , es decir  $C$  es la raíz de la diferencia cuadrada entre  $\widehat{\mu}$  y  $\bar{\mu}$ .

Es entonces apropiado graficar  $\widehat{\mu}$  y  $\widehat{\mu} \pm 0.2C\widehat{\xi}_j$ , (Ramsay y Silverman, 1997) donde se ha escogido la constante 0.2 ya que da una fácil interpretación de resultados. Puede ser útil ajustar el valor 0.2 subjetivamente en cada caso. De cualquier manera es útil usar la misma constante para todas las funciones componentes principales graficados en cualquier caso particular.

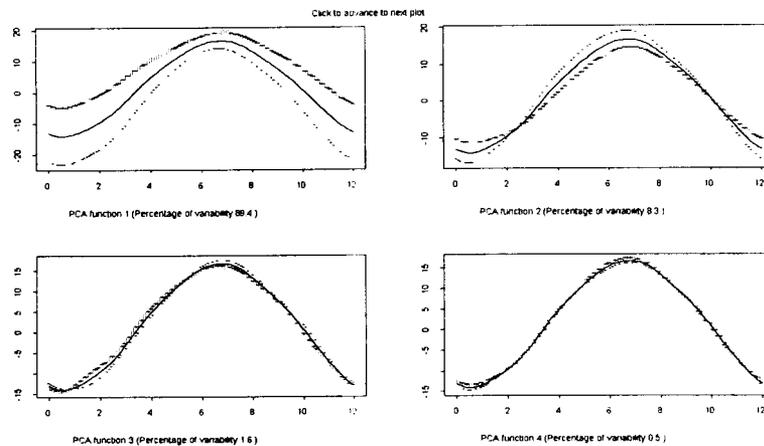


Figura 3.2 Cuatro componentes principales del clima canadiense con su porcentaje de variación explicada

La Figura 3.2 muestra la gráfica de los componentes principales de las estaciones del clima canadiense. La curva sólida es la temperatura media, las curvas con cruces y guiones corresponden a los efectos de sumar y restar un múltiplo de

componente principal, respectivamente. Definitivamente esto resulta más fácil de interpretar y se puede obtener información más precisa.

2) Graficar los “scores”. Un aspecto importante de ACP es el escrutinio de los valores  $f_{im}$  de cada curva sobre cada componente, resultante de la expansión (3.1). En la Figura 3.3 se muestran tales scores para el caso de las estaciones meteorológicas de Canadá.

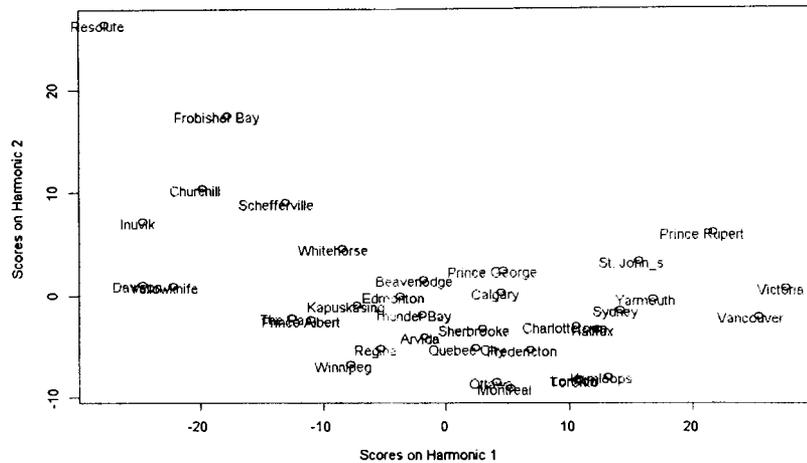


Figura 3.3. Escores de las estaciones del clima sobre los dos primeros CP. La localización de cada estación es mostrada mediante sus nombres

3) Rotar los componentes principales. Se pueden hacer rotaciones a los componentes principales para ganar interpretabilidad. Si denotamos por  $\xi$  a  $(\xi_1, \dots, \xi_k)'$  donde  $\xi_i$  es una función, entonces un conjunto ortonormal definido por:

$$\psi = T\xi, \tag{3.10}$$

donde  $T$  es cualquier matriz ortonormal de orden  $K$ , es decir,  $T'T = TT' = I$ . De una perspectiva geométrica, el vector de funciones  $\psi$  es una rotación rígida de  $\xi$ . Por supuesto que después de esta rotación no se puede esperar que  $\psi_1$  sea el componente de variación más grande, pero el punto es que las funciones base ortonormales  $\psi_1, \dots, \psi_K$  son efectivas en aproximar las curvas originales en  $K$  dimensiones tan bien como la contraparte no rotada.

Un criterio de rotación muy útil en este caso, es la rotación VARIMAX. Sea  $B$  una matriz  $K \times n$  que representa las primeras  $K$  funciones componentes principales  $\xi_1, \dots, \xi_K$ . Por el momento, suponer que  $B$  tiene como renglón  $m$ -ésimo los valores equiespaciados en el intervalo  $\Upsilon$ , la matriz  $A = \{a_{mj}\}$  de valores de las funciones base rotadas están dadas por:

$$A = TB$$

La estrategia VARIMAX para escoger la matriz de rotación ortonormal  $T$ , es maximizar la variación en los valores  $a_{mj}^2$  vistos estos como un vector. Debido a que  $T$  es una matriz de rotación, la suma global de estos cuadrados permanecerá igual no importando que rotación representen en términos algebraicos

$$\sum_m \sum_j a_{mj}^2 = \text{tr}(A'A) = \text{tr}(B'T'TB) = \text{tr}(B'B). \quad (3.11)$$

Intuitivamente hablando, la maximización de la varianza de los  $a_{mj}^2$  puede ocurrir únicamente si estos valores tienden a ser relativamente grandes o relativamente cercanos a cero.

Otra manera de usar varimax es hacer que  $B$  contenga los coeficientes para la expansión de cada  $\xi_m$  en términos de una base  $\phi$  de  $n$  funciones, entonces se rotan los coeficientes de la expansión de cada  $\xi_m$ .

### 3.3 Análisis de Componentes Principales Funcionales Regularizados

Si nosotros queremos incorporar suavidad dentro de nuestro procedimiento de ACP funcionales, entonces tendremos otra meta, controlar la rugosidad de los componentes principales estimados  $\xi$ , para que no sea demasiado grande. La clave de la penalización a la rugosidad es hacer explícito este posible conflicto. En el contexto de que estimando la función componente principal, existe un balance entre maximizar la varianza muestral funcional  $\langle \xi, V\xi \rangle$ , y cuidar la penalidad de la rugosidad  $PEN_2(\xi)$  de que sea demasiado

grande. Este balance es controlado, como se ha visto antes, por un parámetro de suavidad  $\lambda \geq 0$ , el cual regula la importancia del término de penalidad de la rugosidad.

Una manera de penalizar la varianza muestral  $\langle \xi, V\xi \rangle$ , es dividirla por  $\{1 + \lambda PEN_2(\xi)\}$ , la **varianza muestral penalizada**,

$$PCAPSV = \frac{\langle \xi, V\xi \rangle}{\|\xi\|^2 + \lambda PEN_2(\xi)}. \quad (3.12)$$

Haciendo uso de que  $\|D^2\xi\|^2 = \langle \xi, D^4\xi \rangle$  se tiene que:

$$PCAPSV = \frac{\langle \xi, V\xi \rangle}{\langle \xi, (1 + \lambda D^4)\xi \rangle}. \quad (3.13)$$

El máximo de  $PCAPSV$  es la eigenfunción  $\xi$  que resuelve la eigenecuación funcional:

$$V\xi = \rho(1 + \lambda D^4)\xi. \quad (3.14)$$

Generalmente no sólo es de interés estimar el primer componente principal funcional, también los otros componentes lo son. Una buena manera de estimar todos los componentes principales funcionales, es encontrar todos los valores-propios y funciones-propias (3.14). Supongamos que  $\xi_j$  es la eigenfunción correspondiente al  $j$ -ésimo eigenvalor mas grande. Debido a que las  $\xi_j$ 's serán usadas como estimados de las funciones componentes principales, los normalizamos de tal manera que satisfagan que  $\|\xi_j\|^2 = 1$ .

La función componente principal  $\xi_j$  maximiza la varianza penalizada, sujeta a  $\|\xi_j\|^2 = 1$  y una ortogonalidad modificada, de tal manera que se cumpla que:

$$\langle \xi_j, \xi_k \rangle + \lambda \langle D^2\xi_j, D^2\xi_k \rangle = 0 \quad \text{para } k = 1, \dots, j-1. \quad (3.15)$$

El uso de la condición de ortogonalidad modificada significa que podemos encontrar los estimados de todas las eigenfunciones requeridas resolviendo el eigenproblema generalizado.

### 3.3.1 Encontrar ACP Regularizados en la Práctica

#### Caso Periódico

Supongamos que  $\Upsilon$  es el intervalo  $[0,1]$  y que se satisfacen condiciones periódicas de frontera para todas las funciones que consideraremos, en particular  $x_i(s)$ , para toda  $i$ . Sea  $\{\phi_v\}$  las funciones de la serie de Fourier definidas en la sección 1.4. Para cada  $j$ , definimos  $\varpi_{2j-1} = \varpi_{2j} = 2\pi j$ . Dada cualquier función periódica  $x$ , podemos expandirla por series de Fourier con coeficientes  $c_v = \langle x, \phi_v \rangle$ , de tal manera que

$$x(s) = \sum_v c_v \phi_v(t) = c' \phi(s) \quad (3.16)$$

usando  $D^2$  tendremos que:

$$D^2 x(s) = - \sum_v \varpi_v^2 c_v \phi_v \quad \text{con} \quad \|D^2 x\|^2 = \sum_v \varpi_v^4 c_v^2. \quad (3.17)$$

Si  $V$  es la matriz de varianzas y covarianzas de los vectores de coeficientes y  $S$  la matriz diagonal con entradas  $S_{vv} = (1 + \lambda \varpi_v^4)^{-1/2}$ .  $S$  corresponde al operador de suavidad.

Sea  $y$  el vector de coeficientes de cualquier curva componente principal  $\xi$ , es decir,

$$\xi(s) = \sum_v y_v \phi_v(t) = y' \phi(s). \quad (3.18)$$

En términos de la serie de Fourier la eigenecuación generalizada (3.14) puede ser escrita como:

$$Vy = \rho S^{-2} y \quad (3.19)$$

llegando a la expresión:

$$(SVS)(S^{-1}y) = \rho(S^{-1}y) \quad (3.20)$$

El encontrar los eigenvectores y eigenvalores en (3.20) corresponde precisamente, a llevar a cabo un ACP no suavizado de los datos suavizados  $Sc_i$ . Si  $u$  es un eigenvector

de  $SVS$  con eigenvalor  $\rho$ , entonces es aparente que cualquier múltiplo de  $Su$  es una solución de (3.19) para la misma  $\rho$ . Debido a que se requiere que  $\|y\|^2 = 1$ , renormalizamos, hacemos  $y = \frac{Su}{\|Su\|}$ . El componente principal funcional  $\xi$  que corresponde a  $y$  es computado de (3.18).

Poniendo estos pasos juntos tenemos el siguiente procedimiento para llevar a cabo el análisis de componentes principales suavizados de los datos originales:

1. Computar los coeficientes  $c_i$  para la expansión de cada función muestral  $x_i$  en términos de la base  $\phi$ .
2. Aplicar a estos coeficientes el operador de suavidad  $S$ .
3. Llevar a cabo un ACP estándar sobre los coeficientes suavizados  $Sc_i$ .
4. Aplicar el operador de suavidad  $S$  a los eigenvectores resultantes  $u$ , y renormalizarlos, de tal manera que los vectores resultantes  $y$  tengan norma uno.
5. Computar la función componente principal  $\xi$  de (3.18)

### Caso no periódico

Ahora enfocaremos nuestra atención en el caso no periódico, donde las expansiones de Fourier no son apropiadas debido a las condiciones de frontera. Supongamos que  $\{\phi_v\}$  es una base adecuada para el espacio de las funciones suaves sobre el  $[0, 1]$ . Sea  $c_i$  el vector de coeficientes de la función de datos  $x_i(s)$  en la base  $\{\phi_v\}$ . Sea  $V$  la matriz de varianzas y covarianzas de los vectores  $c_i$ .

Se define  $J$  como la matriz  $\int \phi\phi'$ , cuyos elementos son  $\langle \phi_j, \phi_k \rangle$  y  $K$  la matriz cuyos elementos son  $\langle D\phi_j, D\phi_k \rangle$ . De este modo la varianza muestral penalizada se puede escribir de la siguiente manera:

$$PCAPSV = \frac{\langle \xi, V\xi \rangle}{\|\xi\|^2 + \lambda \|D^2\xi\|^2} = \frac{y'JVJy}{y'Jy + \lambda y'Ky} \quad (3.21)$$

y la eigenecuación correspondiente a (3.14) está dada por:

$$JVJy = \rho(J + \lambda K)y. \quad (3.22)$$

Haciendo una factorización donde  $LL' = J + \lambda K$  y definiendo  $S = L^{-1}$ . Podemos encontrar una matriz adecuada  $L$  por una Descomposición en Valores Singulares (DVS) o por una Descomposición de Cholesky, en cualquier caso  $L$  es una matriz triangular inferior. La ecuación (3.22) puede ser escrita como:

$$(SJVJS)'(L'y) = \rho L'y. \quad (3.23)$$

Ahora podemos trabajar a través de las mismas etapas que en el caso periódico, teniendo cuidado en el uso de la matriz  $S$  o  $S'$  en cada etapa. El algoritmo obtenido es el siguiente:

1. Computar los coeficientes  $c_i$  para la expansión de cada función muestral  $x_i$  en términos de la base  $\phi$ .
2. Resolver  $Ld_i = Jc_i$  para cada  $i$  encontrar los vectores  $SJc_i = d_i$
3. Llevar a cabo un ACP estándar sobre los vectores de coeficientes  $d_i$ .
4. Aplicar el operador de suavidad  $S'$  a los eigenvectores resultantes  $u$ , resuelve  $L'y = u$  para cada caso, y normalizar para que de este modo los vectores resultantes  $y$  cumplan con que  $y'Jy = 1$ .
5. Computar la función componente principal de (3.18).

Si se usa una base B-spline y definimos  $L$  por medio de una factorización de Cholesky, entonces las matrices  $J, K$  y  $L$  son todas matrices bandeadas, y usando rutinas apropiadas de Álgebra Lineal, todos los cálculos resultan extremadamente económicos.

### 3.4 Detalles Computacionales

En esta sección nos encargaremos, propiamente de los algoritmos computacionales desarrollados por Ramsay y Silverman, los cuales pueden ser encontrados en la red, para correr dichos algoritmos se puede usar los programas S-plus y Matlab, en este proyecto se utilizó el primero y todo lo que tenga que ver con estos detalles computacionales están referidos a S-plus.

El algoritmo computacional sigue los pasos descritos en la última sección para el caso no periódico, usando la base con la cual estamos expandiendo nuestras funciones muestrales, misma base que se ocupa para las funciones componentes principales y se penaliza la rugosidad sobre toda la expansión de nuestras funciones de datos.

De este modo podemos definir los siguientes pasos para computar las funciones componentes principales y scores, además de las gráficas de los mismos, siguiendo el ejemplo de los datos de labio. (Los nombres en negritas significa que son algoritmos propios de S-plus o desarrollados por Ramsay y Silverman):

```
# ----- Datos de Entrada -----  
  
lipmat <- matrix(scan("lip.dat", 0), 51, 20)  
liptime <- seq(0,1,.02)  
  
# ----- Crear el Objeto Funcional -----  
  
# usar 31 funciones base para splines de orden 6  
  
nbasis <- 31  
norder <- 6  
  
lipbasis <- create.bspline.basis(c(0,1), nbasis, norder)  
  
lipfd <- data2fd(lipmat, liptime, lipbasis, argnames=c("Normalized time",  
"Replications", "mm"))  
  
# ----- Llevar a cabo un ACP y graficar resultados -----
```

```
lippca.fd <- pca.fd(lipfd, nharm=3, lambda=1e-6)
```

```
par(mfrow=c(1,1),pty="m")
```

```
plot.pca.fd(lippca.fd)
```

```
lipeigvals <- lippca.fd[[2]]
```

Obteniendo los resultados vistos en la Figura 3.4

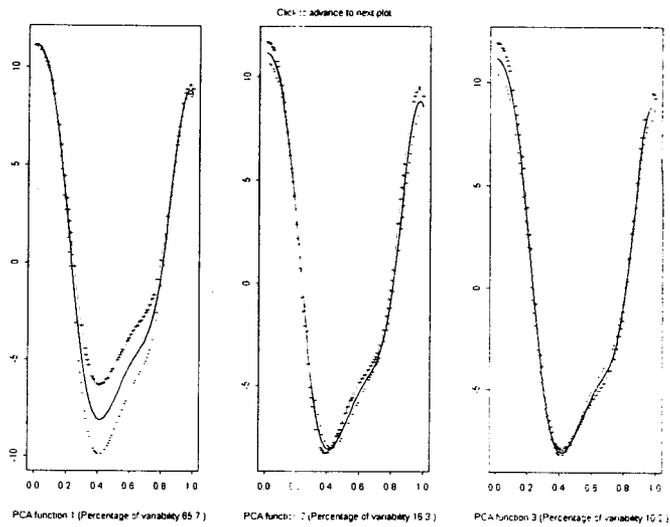


Figura 3.4. Tres funciones CP de datos de labio.

## 4 Registro de Datos Funcionales

### 4.1 Introducción

Las técnicas en análisis funcional de datos (Ramsay & Silverman, 1997) pueden ser empleadas para estudiar la variación en las funciones muestrales  $x_i$ ,  $i = 1, \dots, N$ , y sus derivadas.

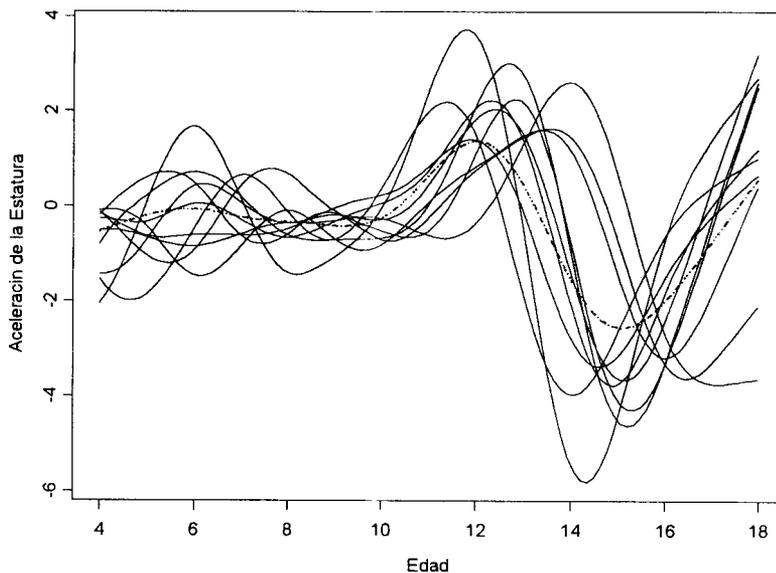


Figura 4.1. Curvas de aceleración para el crecimiento humano

En la práctica estas funciones son algunas veces consecuencia de un proceso de suavizamiento preliminar, aplicado a los datos funcionales, y en otros casos las funciones propiamente son a las que se le aplican estas técnicas. En cualquier caso, queremos estudiar la variación que hay dentro de cada función y entre las funciones; las estadísticas

descriptivas funcionales serán de interés como lo son en los métodos multivariados, tales como, promedios, varianzas, y correlaciones, al igual que el análisis de componentes principales, como hemos visto en los capítulos anteriores.

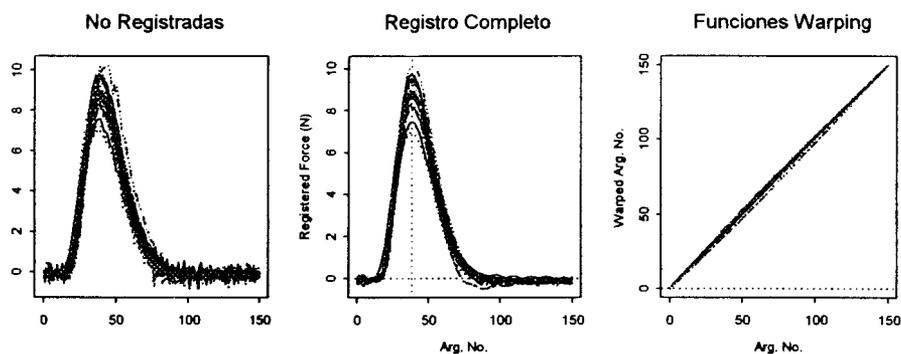


Figura 4.2. Curvas suavizadas de los datos de “pinch force”, con registro y sus funciones “warping”

Sin embargo, los datos funcionales tienden a mostrar dos tipos de variación, llamados en este ámbito, variación de amplitud y variación de fase. La Figura 4.1 ilustra que si se combinan estas dos clases de variación, pueden frustrar hasta el análisis más simple de curvas replicadas. Como se explicó anteriormente estas curvas se caracterizan por la gran desaceleración en la infancia, la cual es seguida por una pequeña aceleración al terminar la misma, y entonces la dramática aceleración-desaceleración se encuentra en la etapa de la pubertad, llegando a cero en la edad adulta. Es claro que la edad en la que ocurren estas características varían de niño a niño. Si ignoramos esta variación en el tiempo y computamos la función media (curva punteada en la Figura 4.1) puede resultar que la aceleración promedio sea un estimado que no recupere las características de las curvas observadas: la curva media tiene menos variación durante la pubertad que cualquier curva en particular, además de que en esta curva media la duración del crecimiento durante la edad de la pubertad es mayor que cada curva en lo individual.

Veamos otro ejemplo; el primer panel de la Figura 4.2 presenta un problema de registro particularmente común. Son las 20 curvas del ejemplo de Pinch Force, descrito anteriormente. Debido a que el tiempo inicial para cada curva fue arbitrario, es esencial entonces, encontrar un origen común en el tiempo, a fin de combinar información entre las funciones muestrales.

Esos ejemplos ilustran que la métrica rígida del tiempo físico puede ser directamente irrelevante a las dinámicas internas de varios sistemas de la vida real. Más aún, puede existir una clase de escala en el tiempo fisiológico o meteorológico que relaciona la no linealidad con el tiempo físico y varía de caso a caso. Por ejemplo el crecimiento humano está ignorando factores externos, consecuencia de una sucesión de eventos hormonales que no pasan en la misma edad en cada niño, y también tiene una tasa variable sobre el crecimiento de un niño en particular. Finalmente en el caso de “pinch force” las contracciones del músculo no ocurren exactamente al mismo tiempo de una presión del dedo a otra.

Hablando formalmente, los valores  $x_i(t_j)$  de dos o más funciones pueden diferir debido a dos tipos de variación. La primera es la más familiar, *variación de amplitud*, o variación vertical debido al hecho de que dos funciones  $x_1$  y  $x_2$  pueden diferir simplemente en puntos en el tiempo en los cuales pueden ser comparados. Pero también pueden existir *variación de fase*, en el sentido que  $x_1$  y  $x_2$  no deberían ser comparados en un tiempo fijo  $t$ , sino en tiempos  $t_1$  y  $t_2$  en los cuales los dos procesos son esencialmente comparables en sus estados, de tal manera que las curvas tienen características comparables en esos tiempos. Por ejemplo, la intensidad del crecimiento humano en la etapa de la pubertad de dos niños, debe ser comparada en sus respectivos picos de velocidad de crecimiento definidos por  $Dx_1(t_1) = Dx_2(t_2) = 0$ , más que en cualquier edad fija.

#### 4.1.1 Formulación del problema de registro

El problema de registrar curvas puede ser planteado formalmente de la siguiente manera. Sean  $N$  funciones  $x_i$  definidas en intervalos  $[0, T_i]$ ; para datos periódicos se puede trabajar sobre intervalos de la forma  $[\delta, T_i + \delta]$ , para cualquier  $\delta$ . En general

denotaremos al intervalo estándar  $[0, T_0]$ . Sea  $h_i(t)$  una transformación del tiempo  $t$  para el caso  $i$  con dominio  $[0, T_0]$  para datos no periódicos,  $[\delta, T_0 + \delta]$  para datos periódicos. El hecho de que los tiempos de los eventos permanezcan en el mismo orden después de ser transformados, sin considerar los tiempos de escala, implica que  $h_i$ , deba ser estrictamente creciente, es decir,  $h_i(t_1) > h_i(t_2)$  para  $t_1 > t_2$ , esto implica que  $h_i$  sea invertible y por lo tanto podemos resolver siempre la ecuación  $y = h_i(t)$  para  $t$  dado  $y$ . Además  $h_i(t)$  tiene que satisfacer las condiciones de frontera  $h_i(0) = 0$  y  $h_i(T_0) = T_i$ . Algunas veces se podría necesitar que  $h_i(t)$  sea una función suave, en el sentido de que sea diferenciable un cierto número de veces.

Sea  $x_0(t)$  una función objetivo fija definida sobre  $[0, T_0]$ , en el sentido de que las características de las curvas  $x_i$  serán alineadas en algún sentido con las características de  $x_0$ . Entonces podemos proponer el modelo

$$x_i[h_i(t)] = x_0(t) + \varepsilon_i(t) \quad \text{ó} \quad x_i \circ h_i = x_0 + \varepsilon_i, \quad (4.1)$$

donde  $\varepsilon$  es pequeño relativo a  $x_i$  y centrado en cero. Si a  $x_0$  la definimos a través de valores discretos  $x_{j0}$ ,  $j = 1, \dots, n$ , entonces nuestro modelo (4.1) se convierte en:

$$x_i[h_i(t_j)] = x_{j0} + \varepsilon_{ij}. \quad (4.2)$$

Debido al supuesto de que  $\varepsilon$  es pequeño relativo a  $x_i$ , este modelo postula que diferencias mayores en forma entre la función objetivo y una función específica son debido únicamente a la variación de fase.

Un modelo más complejo para variación inter-curvas, que combinan variación de fase y amplitud podría ser

$$x_i[h_i(t)] = A_i(t)x_0(t) + \varepsilon_i(t), \quad (4.3)$$

donde  $A_i(t)$  es una función que modula la amplitud; así la versión discreta de (4.3) será

$$x_i[h_i(t_j)] = A_i(t_j)x_{j0} + \varepsilon_{ij}. \quad (4.4)$$

Supongamos, ahora, que se ha podido identificar estas  $N$  funciones  $h_i(t)$ , adjetivadas “warping”. Podemos calcular las funciones *registradas*  $x_i^*(t)$  de la siguiente forma:

1. Calcular valores de la función inversa  $h_i^{-1}(t)$  para una fina malla de valores de  $t$ ,  $i = 1, \dots, N$ .
2. Calcular los valores de  $x_i(t)$  para la misma malla de valores de  $t$ .
3. Representar la relación entre  $h_i^{-1}(t)$  y  $x_i(t)$  como una función usando las mismas técnicas usadas para encontrar las funciones no registradas  $x_i$ .

La tarea de registro, entonces, es estimar las funciones de tiempo “warping”  $h_i$ .

## 4.2 Tipos de Registro

Encontramos dos técnicas de registro de suma importancia, la primera la llamamos *registro landmark*, que requiere la localización de un número de características visibles en cada curva a ser registrada. Estas características deben ser localizadas de una manera no ambigua, y el número de características y curvas debe ser modesto; este método es muy sencillo de usar y entender.

El segundo método involucra a la curva completa y por lo que es operable cuando las características de las curvas son difíciles de encontrar o se tiene un gran número de éstas o un gran número de curvas.

### 4.2.1 Alineación Landmark

Es el proceso de *alinear* curvas mediante la identificación del tiempo de características comunes sobresalientes en las curvas. Estas características son algunas veces picos o valles, si la derivada de una curva puede ser fácilmente evaluada, entonces estos dos elementos pueden ser encontrados en el tiempo  $t$  tal que la derivada evaluada en  $t$  es cero.

Los que es esencial en la alineación landmark es que la  $l$ -ésima característica, de  $L$  de interés, para la curva  $x_i(t)$  sea localizable en un único tiempo  $t_{0l}$ . Podemos designar el

comienzo y el fin de las curvas como características con tiempos 0 y  $T_i$ , respectivamente, e indicar la sucesión entera de los tiempos característicos  $t_{il}$ ,  $l = 0, \dots, L + 1$ , donde  $L$  es el número de características dentro del interior del intervalo  $[0, T_i]$ .

El registro de una curva  $x_i(t)$  se reduce a la tarea de encontrar una función “warping”  $h_i(t)$ , tal que

$$x_i[h_i(t_{il})] = x_0(t_{0l}), \quad l = 0, \dots, L + 1. \quad (4.5)$$

Usando esta estrategia, las curvas son alineadas transformando el tiempo, de tal manera, que los eventos característicos de las curvas ocurran en los mismos valores de los tiempos ya transformados. Comparaciones entre tiempos característicos pueden también ser hechos usando los tiempos transformados correspondientes.

La interpretación de las funciones “warping” se hace comparando la curva  $h_i(t_j)$  versus  $t_j$  con la recta de 45°, si la curva está por arriba de la recta, indica que los tiempos característicos para la curva están después que los de la curva de referencia u objetivo.

Sin embargo, registro landmark puede presentar algunos problemas: los eventos característicos puede que no estén presentes todos en todas las curvas; los estimados de los tiempos característicos pueden ser difíciles de obtener, y pueden ser ambiguos. Además, construir algoritmos para identificar las características de las curvas puede ser un ejercicio engañoso, y los usuarios pueden preferir identificar estas características visualmente. Por ejemplo tenemos el caso de “lip data”, donde tenemos dos puntos críticos muy claros, los cuales son un mínimo y un “codo” del lado derecho de las curvas, los resultados se muestran en la Figura 4.3.

Ahora en la siguiente sección analizaremos dos enfoques distintos para representar o estimar a las funciones “warping”.

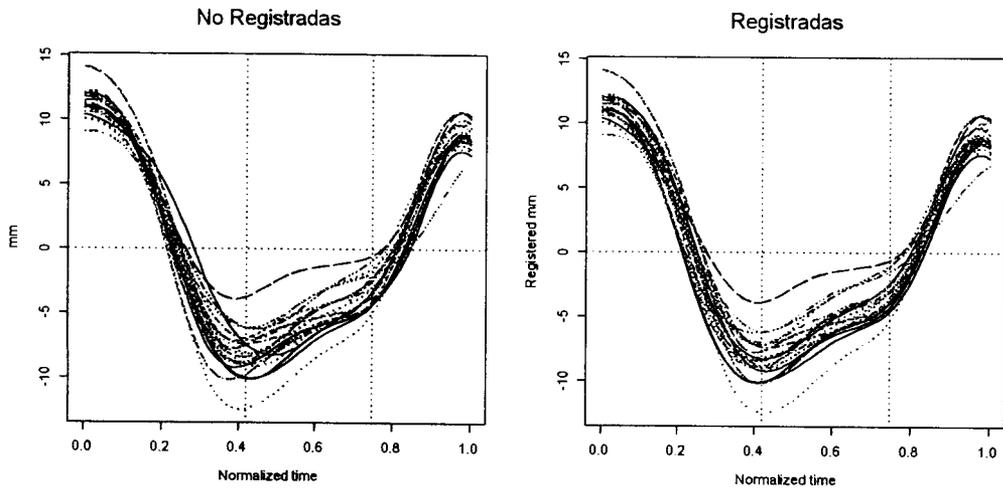


Figura 4.3. Curvas no registradas (derecha) y curvas registradas (izquierda) de “lip data”

#### 4.2.2 Funciones “Warping”

La funciones “warping”, como se ha dicho, marcan la pauta para hacer el registro de curvas; existen varias técnicas que se han desarrollado para estimar este tipo de transformaciones monótonas. Éstas tienden a ser algorítmicamente muy complejas o involucran algún compromiso de flexibilidad. Por ejemplo, Ramsay (1998) estimó transformaciones monótonas de los datos tomando combinaciones lineales de splines monótonos de regresión, los coeficientes fueron restringidos a ser no negativos, implicando que el procedimiento de ajuste optimice bajo restricciones con desigualdades lineales. Además, con la restricción que tienen los coeficientes de ser positivos es suficiente para asegurar la monotonidad, pero no necesario, de tal manera que siempre hay la posibilidad de tener un mejor ajuste, permitiendo a los coeficientes poder ser negativos mientras la monotonidad se siga cumpliendo.

Técnicas de transformación monótona son especialmente útiles en aplicaciones donde las variables observadas son índices, más que simple mediciones, es decir, ellas no tienen

propiedades métricas intrínsecas, pero pueden ser usadas para reflejar relaciones de orden entre objetos, a las cuales, sus valores han sido asignados.

El desarrollo que se describirá a continuación para estimar las funciones “warping”, es muy conveniente hablando computacionalmente, ya que este procedimiento estima una función estrictamente monótona, arbitraria y dos veces diferenciable, definida sobre un intervalo semicerrado por la izquierda.

Utilizaremos la siguiente notación para el operador de integración:

$$D^{-1}h(t) = \int_0^t h(s)ds.$$

Ahora nuestra atención se enfocará en la clase de funciones que sean monótonas, que  $\ln(Dh)$  es diferenciable y  $D\{\ln(Dh)\} = \frac{D^2h}{Dh}$  es Lebesgue cuadrado integrable. Estas condiciones aseguran que la función sea estrictamente monótona ( $Dh > 0$ ) y que su primera derivada sea suave y acotada casi dondequiera. De este modo tendremos el siguiente teorema, el cual establece que esta clase de funciones son identificadas con una simple ecuación diferencial lineal.

**Teorema 4.1.** Cada función  $h$ , tal que  $\ln(Dh)$  es diferenciable y  $D\{\ln(Dh)\} = \frac{D^2h}{Dh}$  es Lebesgue cuadrado integrable, se puede representar como

$$h = \beta_0 + \beta_1 D^{-1}\{\exp(D^{-1}w)\} \quad (4.6)$$

o como una solución de una ecuación diferencial lineal homogénea

$$D^2h = wDf \quad (4.7)$$

donde  $w$  es una función Lebesgue cuadrado integrable con  $\beta_0$  y  $\beta_1$  constantes arbitrarias. (Demostración ver Ramsay, 1998).

La función coeficiente  $w = D\{\ln(Dh)\} = \frac{D^2h}{Dh}$ , mide la curvatura relativa de la función “warping” en el sentido que evalúa el tamaño de la curvatura  $D^2h$  relativa a la pendiente  $Dh$ .

El caso especial de  $w = \alpha$  implica que  $h(x) = C_0 + C_1 \exp(\alpha x)$ , por lo tanto las funciones exponenciales tienen curvatura relativa constante, y si  $w = 0$  define una función lineal. De este modo valores pequeños o iguales a cero de  $w$  corresponden a funciones lineales hablando localmente y donde sea que haya valores grandes, corresponden a regiones donde se forma una curva.

La representación (4.6) también tiene una interpretación, que puede ocasionalmente sugerir un modelo plausible. Expandiendo  $Dh(t)$  alrededor de un valor fijo  $t_0$ , obtenemos

$$Dh(t) \approx Dh(t_0)\{1 + w(t_0)(t - t_0)\},$$

luego,  $w(t)$  induce un cambio proporcional local en la velocidad  $Dh(t)$ , actuando como un múltiplo de  $t - t_0$ , es decir define un sistema en el cual la velocidad está siendo actualizada constantemente por una transformación multiplicativa.

La clase de funciones como las de la expresión (4.6), puede ser extendida a grados de diferenciabilidad  $j > 0$ , reemplazando  $D^{-1}w$  por  $D^{-(j-1)}w$ . El operador  $M = D^{-1} \exp$  puede ser llamado el operador de monotonicidad y opera sobre una función monótona creciente y de un orden mayor de diferenciabilidad.

Ahora consideraremos el problema de suavidad motivado por el modelo  $y_i = h(t_i) + \epsilon_i$  donde los valores  $\epsilon_i$  suponemos que son independientes e idénticamente distribuidos con media cero y varianza  $\sigma^2$ , y la función toma valores dentro del intervalo  $[0, T]$ .

Consideremos el criterio de ajuste siguiente

$$F_\lambda(y|w) = N^{-1} \sum_i \{y_i - \beta_0 - \beta_1 m(t_i)\}^2 + \lambda \int_0^T w^2(t) dt \quad (4.8)$$

donde

$$m(t) = \{D^{-1} \exp(D^{-1}w)\} (t). \quad (4.9)$$

El primer término es el criterio de ajuste por mínimos cuadrados que es común en el spline suavizante, excepto que los parámetros de regresión lineal  $\beta_0$  y  $\beta_1$  son esenciales a causa de que  $m(0) = 0$  y  $Dm(0) = 1$ . El primer término en la ecuación (4.8) se podría generalizar agregando variables de peso para las observaciones, y otras funciones de pérdida, tales como la log-verosimilitud negativa, que podría ser apropiada en ciertas aplicaciones.

El segundo término de la derecha de la ecuación (4.8) corresponde al término de penalización o regularización, el cual tiene alguna característica de la norma de la segunda derivada usada en el spline cúbico suavizante, pero el rol que juega el denominador en  $w = \frac{D^2 h}{Dh}$ , es importante también, debido a que queremos el balance entre regularización y ajuste. Bajo la condición de frontera  $Dh = 0$ . De la ecuación (4.6) tenemos que el caso límite cuando  $\lambda \rightarrow \infty$  es una recta.

Un estimado de  $\sigma^2$  es

$$\hat{\sigma}^2 = (N - 2)^{-1} N^{-1} \sum_i \{y_i - \beta_0 - \beta_1 m(t_i)\}^2, \quad (4.10)$$

el cual ignora el efecto de ajustar  $w$  a los datos.

La ventaja principal que proporciona la representación (4.6) es el cambiar el proceso de estimación, de un problema de encontrar una función  $f$  con ciertas restricciones a un problema de computar una función  $w$  libre de restricciones. Debido a esto último,  $w$  puede ser definida como una combinación lineal de algunos conjuntos de funciones base  $\phi_k$ ,  $k = 1, \dots, K$ , esto es apropiado para el problema en el cual nos encontramos. Sea  $\Phi_k = D^{-1}\phi_k$ , y sean  $\phi$  y  $\Phi$  los vectores  $(\phi_1, \dots, \phi_K)^T$  y  $(\Phi_1, \dots, \Phi_K)^T$  respectivamente, de tal manera que  $w(t) = c^T \phi(t)$  donde  $c$  es el vector de coeficientes que define la combinación lineal. Entonces la función de ajuste tiene la forma

$$\hat{y}(t) = \beta_0 - \beta_1 m(t) = \beta_0 - \beta_1 D^{-1} [\exp \{c^T \Phi(t)\}]. \quad (4.11)$$

Así el criterio (4.8) tiene que ser minimizado con respecto al vector de coeficientes  $c$  y los coeficientes de regresión  $\beta_0$  y  $\beta_1$ .

Ramsay (1998) desarrolló un algoritmo de 2 etapas para la minimización de la expresión (4.8). Empezando con un estimado inicial  $c^{(0)}$ , el cual puede ser un vector de ceros, estimados de  $\beta_0^{(0)}$  y  $\beta_1^{(0)}$  por regresión lineal. Entonces en cualquier iteración  $v > 0$  para la cual  $\beta_j^{(v-1)}$  y  $c^{(v-1)}$  son estimados de la iteración previa, primero optimice con respecto a  $c$  aplicando el procedimiento de Gauss-Jordan para problemas de mínimos cuadrados no lineales, obteniendo  $c^{(v)}$  y luego computando  $\beta_j^{(v)}$  por regresión lineal. El procedimiento de Gauss-Jordan requiere que el vector de actualización

$$\delta^{(v)} = c^{(v)} - c^{(v-1)}$$

sea la solución de la ecuación lineal

$$\mathbf{R}^{(v-1)}\delta^{(v)} = -s^{(v-1)}$$

donde

$$\mathbf{R} = N^{-1}\beta_1^2\mathbf{X}^T\mathbf{X} + \lambda\mathbf{K}, \quad (4.12)$$

$\mathbf{X}$  es una matriz de  $N \times K$  y tiene renglones

$$x(t_i) = \frac{\partial m(t_i)}{\partial c} = \int_0^{t_i} \Phi(s) \exp\{c^T \Phi(s)\} ds,$$

y  $\mathbf{K}$  es una matriz simétrica de orden  $K$ , que se define como

$$\mathbf{K} = \int_0^T \phi(s)\phi^T(s)ds, \quad (4.12)$$

el vector  $s$  de longitud  $K$  tiene valores

$$s = -N^{-1}\beta_1\mathbf{X}^T r + \lambda\mathbf{K}c \quad (4.13)$$

y, finalmente, donde  $r$  es el vector de longitud  $N$ , el cual contiene los residuales  $r_i = y_i - \beta_0 - \beta_1\hat{m}(t_i)$ . La tasa de convergencia de estas iteraciones es únicamente lineal, pero es aceptablemente rápido, y se obtiene iterando cuatro o cinco veces generalmente, según Ramsay (1998). La Figura 4.4 nos muestra las funciones “warping” para el caso de “lip data”, las cuales están expandidas en 6 funciones base con splines cúbicos.

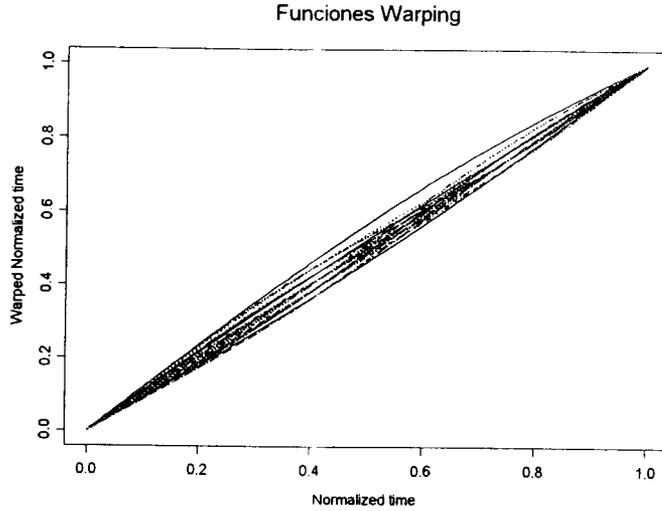


Figura 4.4. Funciones “warping” para el registro landmark para los datos de labio

### 4.2.3 Estimación de la función “warping” por suavizamiento

Este es un enfoque alternativo al anterior, donde se pueden estimar las funciones “warping” para el registro landmark únicamente.

Siguiendo la idea de suavizamiento por funciones base y penalizando la rugosidad encontramos la funciones “warping” de la siguiente forma:

Sean  $t_1, t_2, \dots, t_l$  los  $l$  landmarks para la curva  $x$  y sean  $t_{01}, t_{02}, \dots, t_{0l}$  los  $l$  landmarks para la función objetivo, denotando por  $h$  la función “warping” asumiendo que  $x$  está definida en el intervalo  $[0, T]$ , tenemos que  $h$  debe cumplir con lo siguiente:

$$h(0) = 0$$

$$h(T) = T$$

$$h(t_{0i}) = t_i$$

para  $i = 1, \dots, l$ .

Así siguiendo el modelo (2.1) y el criterio (2.20) para  $h$ , es decir

$$PENSSE_\lambda(h | t) = \sum_j \{t_i - h(t_{0i})\}^2 + \lambda \times PEN_2(h),$$

se encuentra la  $h$  que minimice el criterio y obtenemos la función “warping” y así para cada curva.

#### 4.2.4 Registro continuo

Se pueden registrar dos curvas,  $x_0$  y  $x_1$ , optimizando alguna medida global de similitud entre ellas, usando las curvas completas en el proceso. Considera la siguiente matriz de covarianzas entre  $x_0$  y  $x_1$ .

$$\begin{bmatrix} \int x_0^2(t)dt & \int x_0(t)x_1(t)dt \\ \int x_0(t)x_1(t)dt & \int x_1^2(t)dt \end{bmatrix}. \quad (4.14)$$

Supongamos ahora que los valores de las curvas pueden ser analizados como un conjunto de puntos, el cual puede ser sujeto a un análisis de componentes principales. Si los valores de las curvas son proporcionales, entonces tal análisis produciría únicamente un componente principal distinto de cero (Ramsay, 2000). Esto es, únicamente uno de los 2 eigenvalores de la matriz (4.14) será distinto de cero. Es por esta razón que podemos escoger una función “warping”  $h_i(t)$  de tal manera que minimice el logaritmo del eigenvalor más pequeño de la matriz (4.14), es decir,  $h_i(t)$  se encuentra con resolver

$$\min_h F(h_i) = \min_h \log \mu_2 \begin{bmatrix} \int x_0^2(t)dt & \int x_0(t)x_1[h_i(t)]dt \\ \int x_0(t)x_1[h_i(t)]dt & \int x_1^2[h_i(t)]dt \end{bmatrix}. \quad (4.15)$$

Sin embargo, observamos aquí que este criterio trabaja aún mejor si reemplazamos los valores de la funciones por sus primeras derivadas, o derivadas de orden mayor si pueden éstas ser estimadas de manera estable. Esto es debido a que las derivadas tienden a oscilar más rápido que las propias funciones, y también varían alrededor del cero, luego entonces, la medida del eigenvalor más pequeño es más sensible a funciones que puede diferir únicamente por la variación a la amplitud. En la Figura 4.5 se pueden observar el registro de cuatro estaciones canadienses y se puede comparar con el panel izquierdo en el cual están las curvas sin registrar, haciendo uso del registro global (4.15).

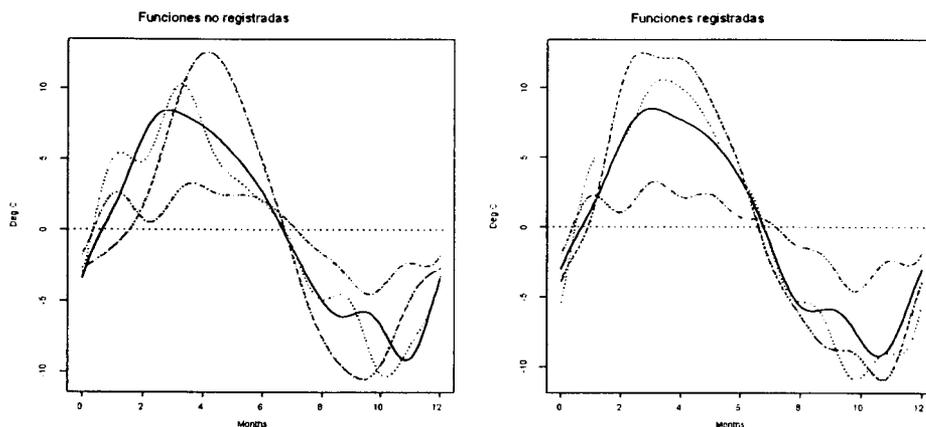


Figura 4.5. Cuatro estaciones canadienses registradas

La búsqueda de las funciones “warping”  $h$  se realiza en la misma familia utilizada para el caso del registro landmark visto en la subsección anterior, criterio (4.8)

### 4.3 Detalles Computacionales

Enseguida mostraremos dos rutinas para registro continuo y registro landmark, con las cuales se verá lo poderoso que es el registro continuo y lo útil que es el registro landmark, como se puede observar en la Figura 4.3 para “lip data”, en la cual nos podemos apoyar para hacer mejores comparaciones entre las curvas registradas.

Este es un pequeño ejercicio mostrado en Ramsay (2000) con algunas modificaciones. Sea  $x(t) = \sqrt{2}[\text{sen}(2\pi t) + \text{cos}(2\pi t)]$  la función a ser registrada y sea  $x_0(t) = \text{sen}(2\pi t)$  la función objetivo. Estas dos funciones tienen una diferencia de fase de  $\frac{1}{8}$ , además de que  $x(t)$  tiene un máximo de 2 mientras que  $x_0(t)$  tiene un máximo de uno. Así que tendremos la siguiente rutina en S-plus:

Primero crearemos mallas de valores de la función objetivo y la función a registrar.

```
x <- seq(0,1,0.01)
y <- sqrt(2)*(sin(2*pi*x) + cos(2*pi*x))
y0 <- sin(2*pi*x)
```

Ahora haremos el objeto funcional expandiendo en base de Fourier con 91 funciones base.

```
basisfd <- create.fourier.basis(c(0,1), 99)
fd <- data2fd(y, x, basisfd)
fd0 <- data2fd(y0, x, basisfd)
```

Creamos un objeto funcional para la función “warping”

```
wbasis <- create.bspline.basis(c(0,1), nbasis=6, norder=4)
coef <- matrix(0, 6, 1)
Wfd0 <- create.fd(coef, wbasis)
```

Y ahora registramos indicando la periodicidad de los datos

```
result <- registerfd(fd, fd0, Wfd0, periodic=T)
```

Obteniendo el asombroso resultado de la Figura 4.6

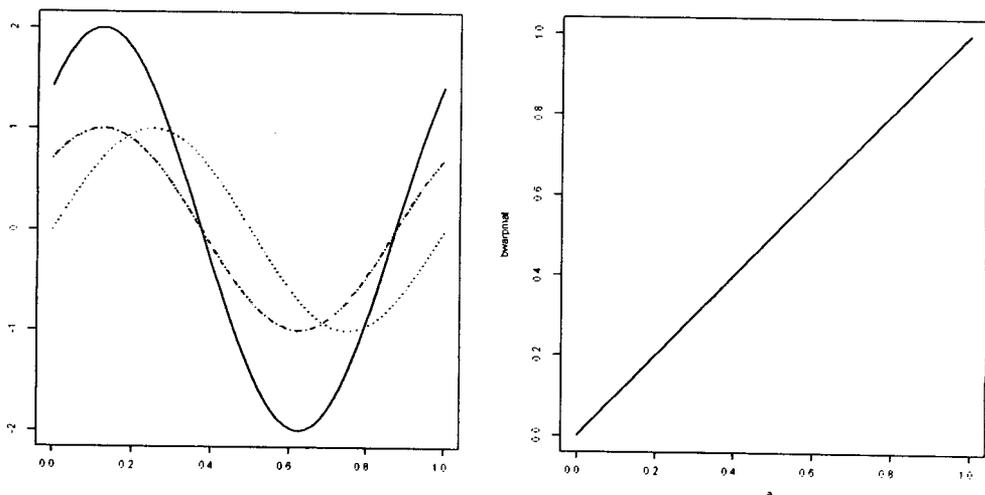


Figura 4.6. Resultado del ejercicio descrito. En negro curva por registrar, en rojo curva registrada y en azul función objetivo. Panel derecho función warping

La curva en negro es la curva a ser registrada y la curva azul la función objetivo y la curva roja es el resultado del registro. Con esto nos podemos dar cuenta como el registro global *puede* reducir de una manera drástica la variabilidad en el rango de las curvas a ser registradas. El panel derecho de la Figura 4.6 corresponde a la función “warping”, la cual da como resultado una recta, haciendo que los puntos críticos de la curva a ser registrada, en el registro, se alcancen en el mismo rango que los de la función objetivo.

Pasaremos ahora a la rutina de registro landmark para “lip data”, las curvas registradas y no registradas se pueden observar en la Figura 4.3, así como las funciones “warping” en la Figura 4.4

```
# ----- Datos -----
lipmat <- matrix(scan("lip.dat", 0), 51, 20)
liptime <- seq(0,1,.02)

# ----- Creación del objeto funcional -----
# use 31 order 6 splines so we can look at acceleration

nbasis <- 31
norder <- 6
lipbasis <- create.bspline.basis(c(0,1), nbasis, norder)
lipfd <- data2fd(lipmat, liptime, lipbasis,
                argnames=c("Normalized time", "Replications", "mm"))

# ----- Aplicando un poco de suavidad -----

lipfd <- smooth.fd(lipfd, Lfd=4, lambda=1e-12)

# ----- Registro de curvas -----
# Primero creamos una base para las funciones “warping”
# Tiene orden cuatro (spline cúbico) y 2 nodos interiores
# NBASIS = NORDER + # nodos interiores

nbasis <- 6
```

```

norder <- 4
breaks <- c(0,lipmeanmarks,1)
warpbasis <- create.bspline.basis(c(0,1), nbasis, norder, breaks);
lipreglist <- landmarkreg(lipfd, lipmeanfd, lipmarks, lipmeanmarks, warpbasis)
lipregfd <- lipreglist$regfd
lipwarpfd <- lipreglist$warpfd

```

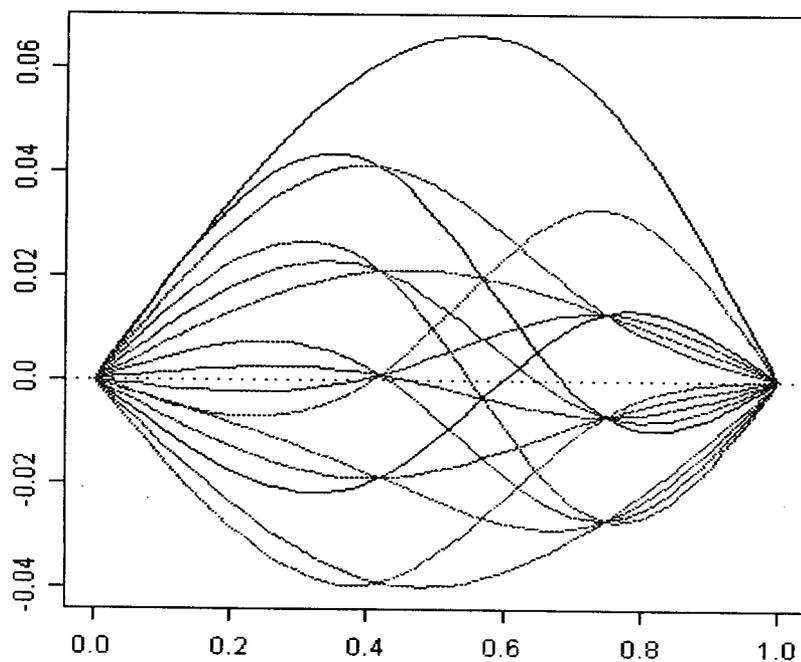


Figura 4.7 Funciones de deformación de los datos de labios

La Figura 4.7 son las funciones de deformación, es decir  $h_i(t) - t$ , la cual nos da una mejor impresión de la complejidad de las funciones “warping”.

## 5 Aplicaciones

En este capítulo aplicaremos los elementos de AFD revisados en los capítulos anteriores a un conjunto de datos en tecnología de alimentos, específicamente del maíz. En nuestros días es de vital importancia poder caracterizar éste, ya que en México es uno de los alimentos de mayor importancia, ya sea como consumo o como insumo. El maíz puede ser clasificado dependiendo de los siguientes aspectos: genéticos, botánicos, físicos y químicos.

Como se explica en Mauricio-Sánchez (2001), el grano de maíz está constituido principalmente por almidón; este almidón tiene múltiples aplicaciones en la industria alimentaria, por ejemplo, como espesante, como formador de gel o como estabilizante en emulsiones. Con la formación de gel o la llamada así, *gelatinización*, se pierde irreversiblemente la cristalinidad dentro de los gránulos de almidón; con la gelatinización, los alimentos formulados con almidón se modifican en su textura, viscosidad y capacidad de retención de agua.

Para estudiar las propiedades de la gelatinización de almidones de maíz se utilizan los llamados *perfiles térmicos*. Los perfiles térmicos se obtienen a partir de un voltaje fijo, variando la temperatura en el maíz y midiendo la corriente en amperes. El comienzo de la gelatinización comienza, matemáticamente, cuando la primera derivada de la corriente con respecto a la temperatura es igual a cero

$$\frac{\partial \text{Corriente}}{\partial \text{Temperatura}} = 0;$$

denotemos por  $T_0$  el valor de temperatura donde lo anterior se cumple; posterior al comienzo de la gelatinización, el pico de gelatinización se caracteriza matemáticamente cuando se alcanza un mínimo

$$\frac{\partial \text{Corriente}}{\partial \text{Temperatura}} = 0, \quad \frac{\partial^2 \text{Corriente}}{\partial \text{Temperatura}^2} > 0;$$

denotemos por  $T_p$  el valor de temperatura donde lo anterior se cumple.

Los datos fueron obtenidos a partir de 38 tipos de maíces los cuales son: Arroam, Natel-1, Natel-9, Palme-8, Palome-9, ConMich, Con-O, Con-Mr, Con-H, Vande9-1, Vande9-2, Dtziba-Q, Dtziba-V, Vande9-3, Tehua9-1, Tehua9-2, Bolita, Connor-H, Cel9-1, Cr-Chih-Ch, Blanso, Cr-Chih-Son, Onave-Son, Cel9-2, Arro, Con-M2, Tabli8-1, Connor-A, Chapal-1, Arro, Chapal-2, Con-P, Con-T, Con-M1, Con-M2, Tabli8-1, Chalq9, Connor-Q, Connor-A, Dulnor9-Son, Dulnor8-Sin, Onave7-Qro. La representación gráfica en su forma discreta de estos perfiles se puede observar en la Figura 5.1.

Nuestro objetivo en esta investigación, *en coordinación con el CINVESTAV Unidad Querétaro*, es describir las fuentes de variación más importantes en el proceso de gelatinización entre diferentes razas de maíces, vía estadísticas descriptivas, componentes principales y registro. Se utilizó la representación por medio de funciones base, usando las del tipo B-spline, con treinta funciones base, splines de orden tres y 18 nodos. 28 perfiles se usaron para el análisis, ya que de los 38 maíces seleccionados sólo 28 cumplieron con la característica de tener los dos puntos críticos de interés (landmarks). Estos maíces son: Chapal-1, Chapal-2, Con-P, Con-T, Con-M1, Chalq9, Connor-Q, Dulnor9-Son, Dulnor8-Sin, Onave7-Qro (ver Figura 5.12, al final del capítulo). De esta forma extraímos los 10 perfiles de todo el conjunto de datos, para poder comparar sobre el mismo conjunto los perfiles sin registro, con registro landmark y con registro continuo.

## 5.1 Perfiles sin Registrar

En el primer panel de la Figura 5.2 se muestran los 28 perfiles originales, pero ya suavizados, es decir, como objetos funcionales. En la Figura 5.12, también se pueden observar los maíces suavizados individualmente. Destaca el perfil térmico del maíz Con-H que muestra el perfil donde se inicia la gelatinización a una menor temperatura, alcanzando el pico de gelatinización con el menor incremento de temperatura. Considerando a todos los perfiles, se puede apreciar que a mayor temperatura de inicio de la gelatinización, mayor la temperatura del pico de gelatinización, es decir se aprecia que un tipo de variación en estos perfiles es el debido a un corrimiento de temperaturas.

Existen tres curvas donde el comienzo de la gelatinización lo alcanza en un amperaje muy

alto, estos tipos de maíz son Dtziba-Q, Con-M2 y Bolita. Todos los demás maíces varían mucho donde alcanzan sus puntos críticos, sin embargo estos tres destacan.

Los perfiles van subiendo rápidamente con la temperatura en el amperaje hasta alcanzar el punto donde comienza la gelatinización y se puede observar que las funciones decrecen no muy rápidamente hasta llegar al pico de gelatinización terminando algunas curvas con un incremento en los amperes y otras con un decaimiento, sin embargo el maíz Dtziba-V muestra una gran variabilidad hacia el final del proceso, con una caída pronunciada, para después crecer rápidamente, algo similar pasa con Dtziba-Q, Vande9-3, Tehua9-1 y Tehua9-2.

Para poder comparar más fácilmente los amperajes que los perfiles muestran se hace necesario registrar las curvas.

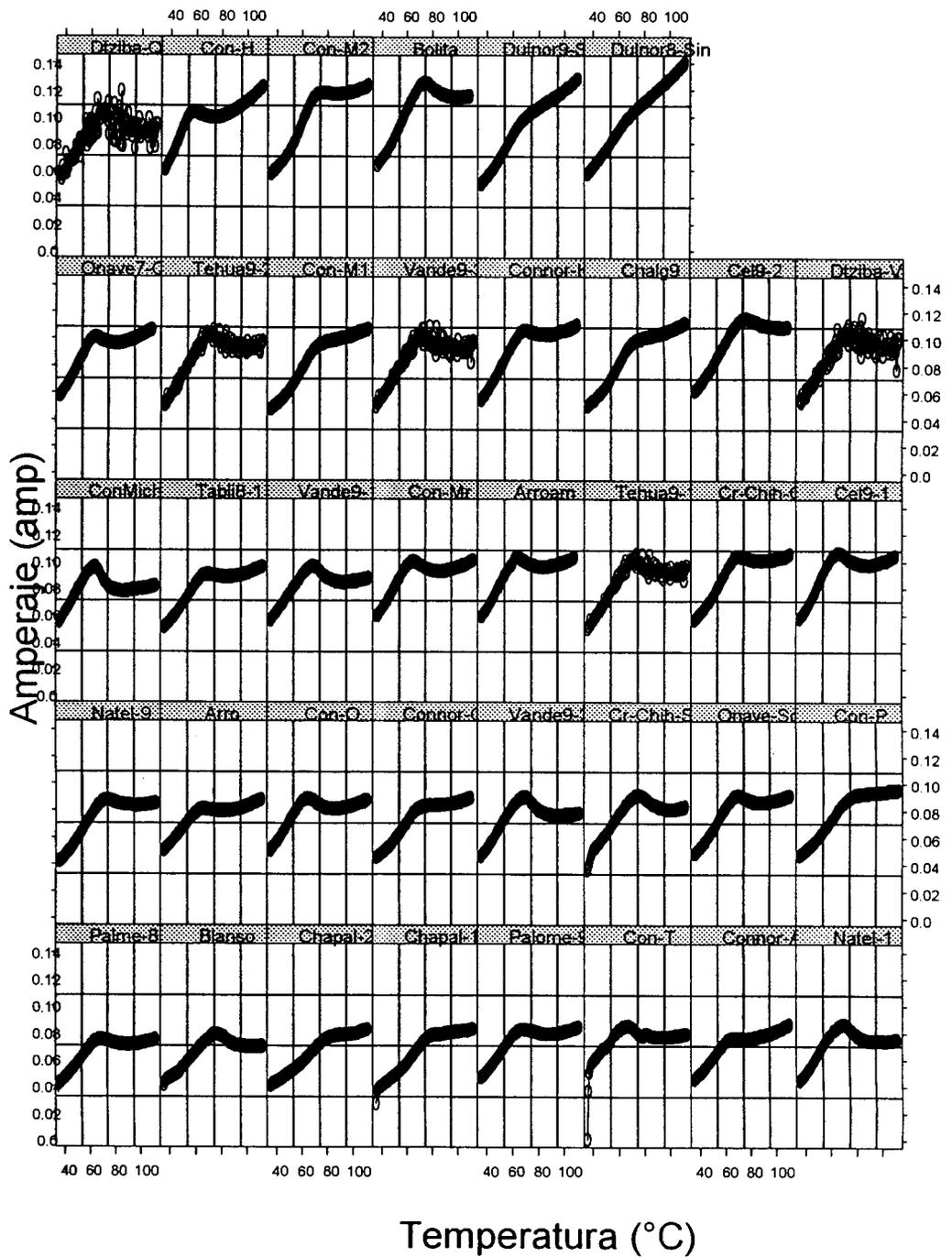


Figura 5.1. Los 28 perfiles representados en forma discreta

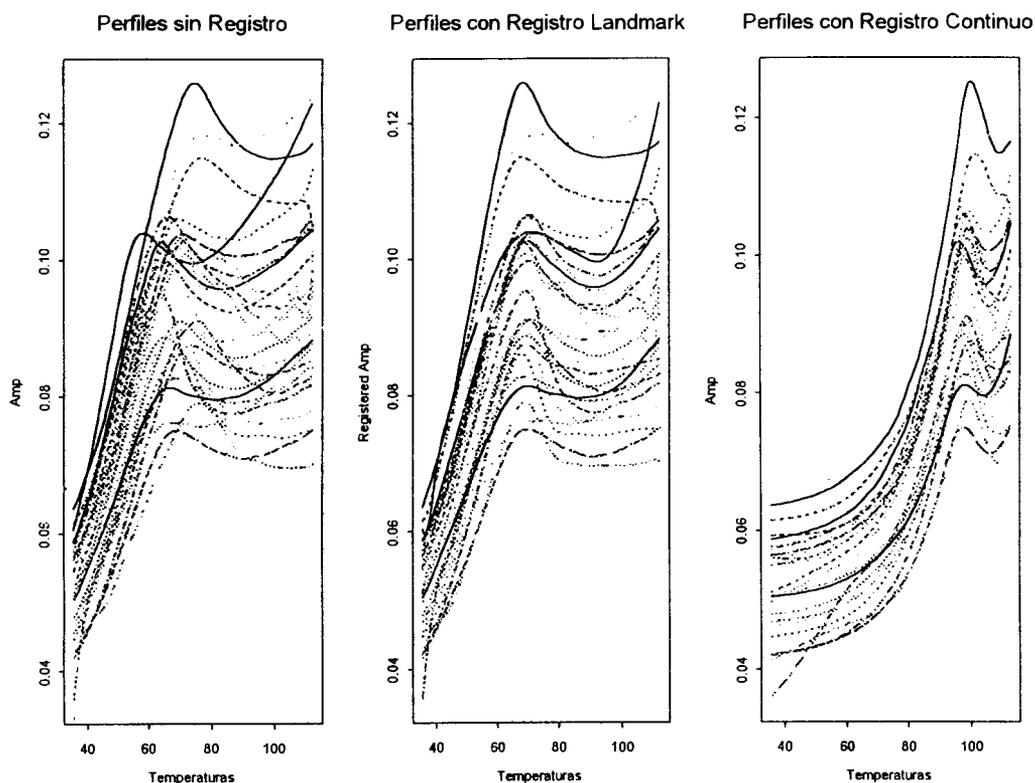


Figura 5.2. Perfiles suavizados sin registro, con registro landmark y con registro continuo

## 5.2 Perfiles con Registro Landmark

En el panel central de la Figura 5.2 se muestran los perfiles registrados por el criterio de landmark (2 landmark en nuestro caso), utilizando funciones “warping” expandidas por seis funciones base del tipo B - spline de orden 4; Para realizar este registro, se identificaron las dos temperaturas de los landmarks de interés; éstas se muestran en el Cuadro 5.1. Como se mencionó anteriormente se tuvo el problema de que existían 10 curvas que no tenían landmarks, debido a esto se extrajeron las 10 curvas, antes mencionadas, del análisis. Computando sus derivadas se puede observar fácilmente que estas curvas son atípicas ya que las derivadas nunca se hacen cero, las Figuras 5.3 y 5.4 nos muestran dos ejemplos de estas derivadas. El alineamiento logrado por registro landmark, se debe a un corrimiento en la temperatura fundamentalmente en el inicio y en el pico de gelatinización, como era de esperarse, permitiendo una mejor interpretación de la variación por amplitud de los perfiles. Se pueden apreciar perfiles que inician la gelatinización alrededor de los 12 amperes (Bolita, Con-M2 : Dtziba-Q), alrededor de los 10 amperes y alrededor de los 8 amperes.

<i>Maíz</i>	$T_0$	$T_p$
Arroam	62.903	84.914
Natel-1	70.240	102.923
Natel-9	72.908	98.921
Palme-8	67.572	91.584
Palome-9	68.239	89.583
ConMich	62.903	88.249
Con-O	64.237	88.916
Con-Mr	63.570	86.915
Con-H	56.900	74.909
Vande9-1	68.239	94.252
Vande9-2	68.906	98.921
Dtziba-Q	69.573	94.252
Dtziba-V	71.574	97.587
Vande9-3	70.907	95.586
Tehua9-1	70.907	96.253
Tehua9-2	70.907	99.588
Bolita	74.242	98.254
Connor-H	68.239	90.250
Cel9-1	64.237	88.916
Cr-Chih-Ch	37.572	87.582
Blanso	75.908	105.588
Cr-Chih-Son	73.575	98.254
Onave-Son	67.572	88.916
Cel9-2	76.910	98.921
Arro	65.571	86.915
Con-M2	73.575	94.914
Tabli8-1	67.572	88.249
Connor-A	64.904	85.913

Cuadro 5.1. Temperaturas de inicio ( $T_0$ )  
y de pico de la gelatinización ( $T_p$ )

La Figura 5.13 al final del capítulo muestra las curvas registradas individualmente; existe un grado de deformación muy pequeño en los perfiles, lo que nos dice que se hizo más un corrimiento que una deformación propiamente, para alinear a los puntos críticos.

### 5.3 Perfiles con Registro Continuo

El panel extremo izquierdo de la Figura 5.2 se presentan los perfiles registrados con el criterio continuo con funciones “warping” expandidas por 20 funciones base B - spline de orden 4; los perfiles registrados muestran una gran deformación y con respecto al inicio y al pico de

gelatinización el registro global muestra resultados insatisfactorios de alineamiento; esto es natural dado que de entrada no se señala a landmarks de interés porque la técnica responde a características globales y no locales de los perfiles. Así si el interés radica en comparar landmarks específicos debe usarse el registro landmark; el registro global es más natural de aplicarse cuando se desea registrar perfiles en los que no hay características comunes específicas de interés por parte del investigador.

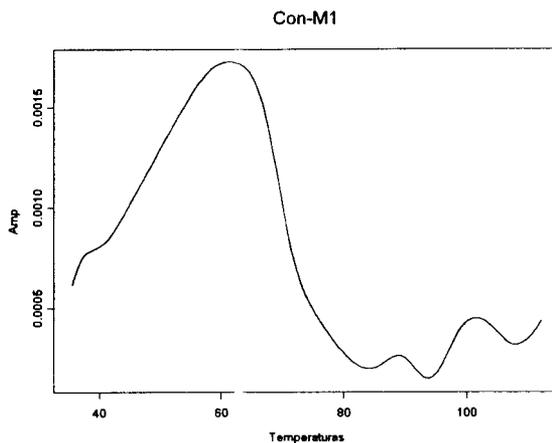


Figura 5.3 Perfil derivada del maíz Con-M1 (no considerado en el análisis).

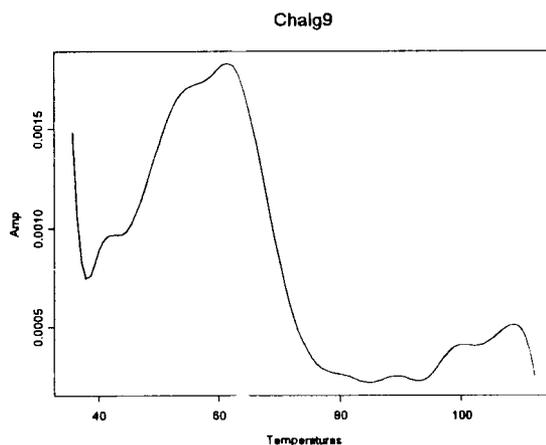


Figura 5.4. Perfil derivada del maíz Chalq9 (no considerado en el análisis).

## 5.4 Estadísticas Descriptivas

Continuando con el análisis, ahora compararemos los perfiles promedio, desviación estándar y correlación, para extraer la mejor información posible.

El perfil promedio sin registrar, ver Figura 5.5 primer panel, proporciona un perfil en el que se observa que la diferencia en amperaje entre el estado de inicio y el estado de pico de la gelatinización es más pequeña, que la misma diferencia en el perfil promedio de perfiles con registro landmark. En el caso de perfiles promedio de landmark el inicio de la gelatinización se logra en un amperaje menor de 0.098, el cual es el amperaje del estado de inicio del promedio con registro continuo, aunque la diferencia es casi nula, mientras que el estado de inicio para el perfil de las curvas sin registro se alcanza en un nivel de 0.095 amperes. Refiriéndonos al estado de pico de gelatinización la diferencia entre el perfil promedio con registro landmark y el perfil promedio sin registro es muy pequeña, este punto se alcanza en 0.089 amperes, para el promedio landmark, mientras que para el promedio sin registro se llega a 0.09 amperes.

El perfil promedio de perfiles registrados continuamente, dada una pobre alineación en el pico de gelatinización *no* será confiable del todo.

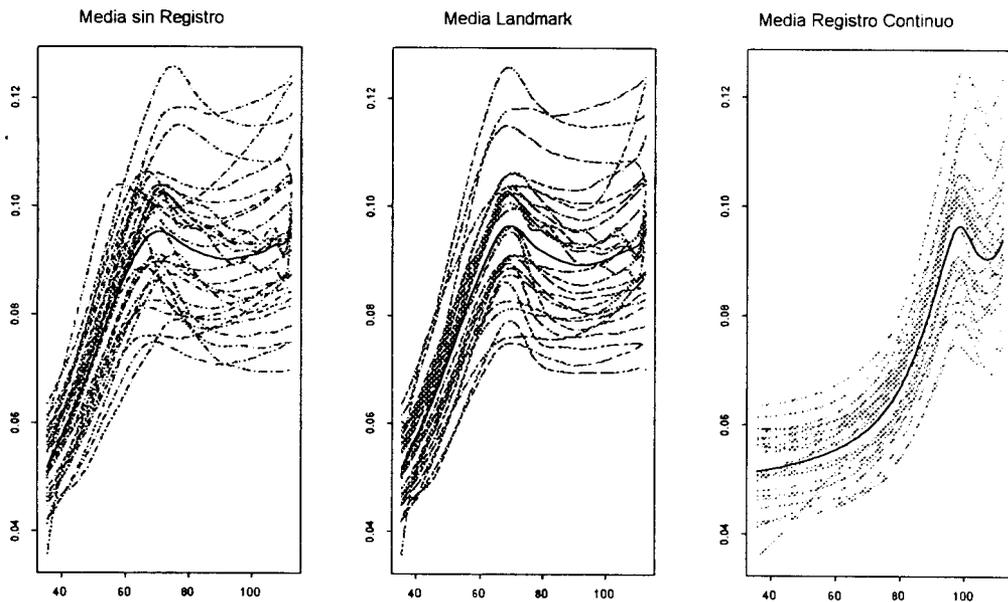


Figura 5.5. Perfiles promedio con su respectivo conjunto de curvas sin registro, con registro landmark y con registro continuo.

Los tres perfiles muestran la variabilidad que puede haber en el proceso, ver Figura 5.6, en azul desviación estándar de todos los perfiles, en negro desviación estándar landmark y en rojo desviación estándar registro global. Los tres tienen una

característica muy importante en común, se aprecia que la variabilidad es creciente hasta los 12 amperes, y desde ahí la variabilidad prácticamente se mantiene constante, es decir, desde el estado de inicio hasta el estado de pico de gelatinización tenemos un decaimiento en las desviaciones de los perfiles. Estas etapas tienen un desfazamiento, debido a que tenemos los tres tipos de análisis involucrados. Esto último nos puede guiar acerca de la variabilidad que se tiene en este proceso, sin embargo con un análisis de componentes principales podremos ver si en verdad en esta etapa existe una fuente de variación importante.

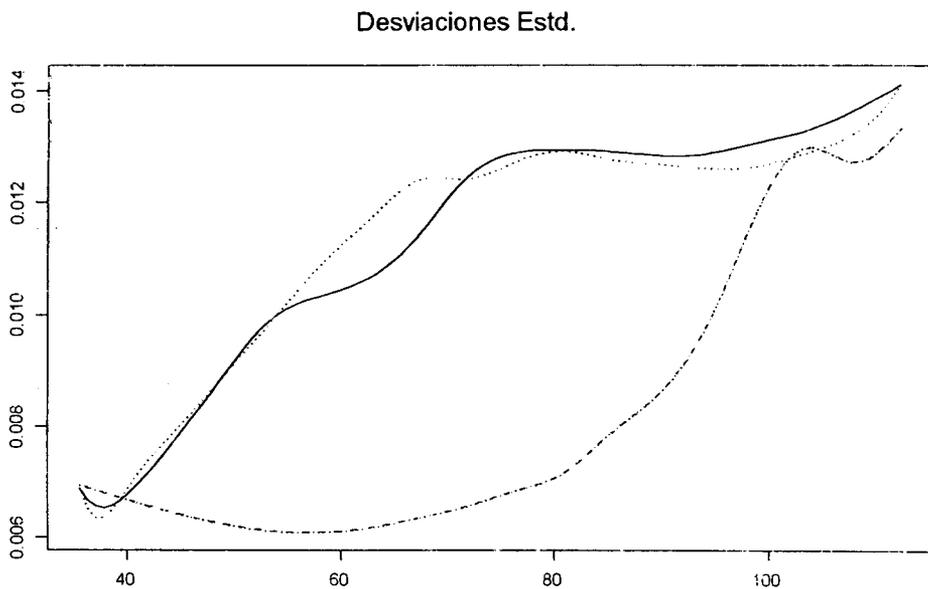


Figura 5.6. Desviaciones estándar de los perfiles sin registrar (azul), con registro landmark (negro) y con registro continuo (rojo)

La función de correlación de los perfiles no registrados en la Figura 5.7 primer panel, muestra que entre  $30^{\circ}C$  y  $80^{\circ}C$  de temperatura la asociación entre dos mediciones decrece más rápidamente en la medida que la diferencia en temperaturas crece: fuera de este rango de temperaturas por abajo de  $30^{\circ}C$  y por arriba de  $80^{\circ}C$  la asociación no decae tan rápidamente al crecer la diferencia de temperaturas. Esto es importante debido a que entre las temperaturas antes mencionadas se da la tasa máxima de crecimiento del amperaje. Leyendo la función correspondiente a los perfiles landmark, ver Figura 5.7 segundo panel, se aprecia una correlación fuerte entre mediciones de amperaje aún con diferencias importantes en temperaturas, sin embargo el decaimiento se sigue dando en la etapa donde se alcanza la

tasa de amperaje máxima, de una manera más pronunciada se puede ver en la función de correlación de los perfiles con registro global, Figura 5.7 tercer panel, donde entre  $90^{\circ}C$  y  $120^{\circ}C$  decae rápidamente la correlación hasta llegar a 0.75, recordemos que en este registro es entre estas etapas donde el amperaje tiene su tasa máxima. Las tres funciones coinciden en que fuera de la zona de baja correlación, esta última es muy alta.

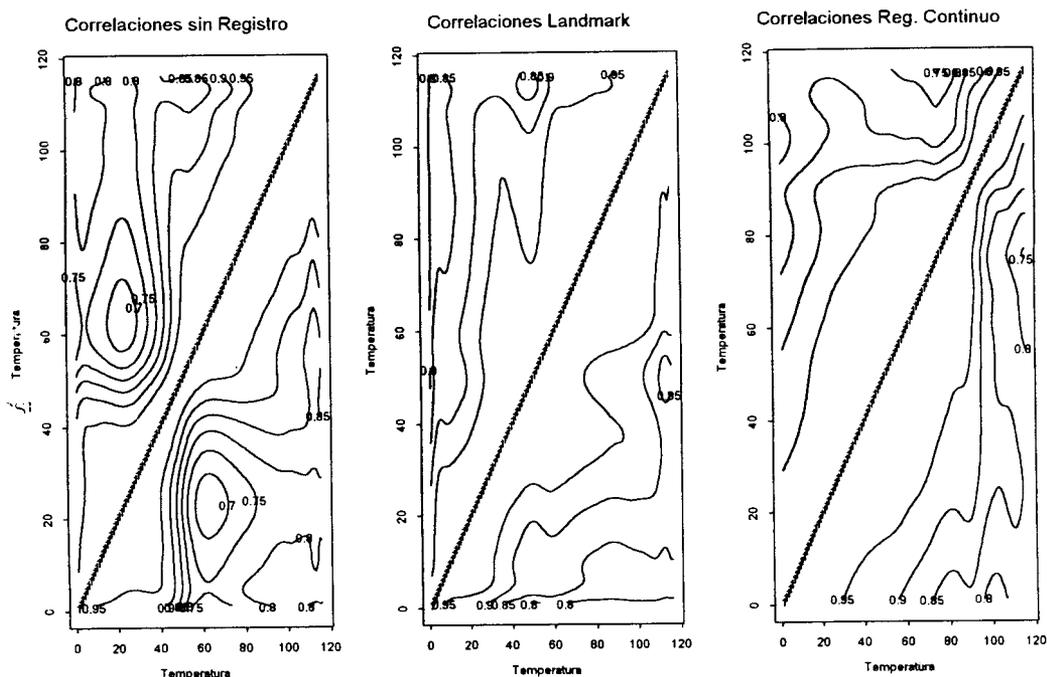


Figura 5.7. Correlaciones de los perfiles sin registro (izquierda), con registro landmark (centro) y con registro continuo (derecha).

## 5.5 Análisis de Componentes Principales

Respecto al análisis por componentes principales, en la Figura 5.8 se muestran los múltiplos de los cuatro primeros componentes principales generados como fue explicado anteriormente. En el caso de perfiles con registro landmark se aprecia que la fuente primordial de variabilidad se debe a las diferencias globales versus el perfil promedio; el segundo componente principal funcional se debe a la variación en el inicio de la gelatinización, disminuyendo posteriormente pero volviendo a mostrar variabilidad al inicio del pico de gelatinización. Los resultados de la rotación VARIMAX, en el caso de perfiles registrados por landmark muestran básicamente las mismas fuentes de variación por medio de los componentes rotados 1 y 4. Los scores

de los maíces respecto a los dos primeros componentes principales sin rotar, ver Figura 5.10, muestran que Con-M2, con un peso muy importante en los dos primeros componentes principales, y Con- Mich, con un peso muy importante en el segundo componente principal; Con - M2 muestra un perfil en el que la diferencia en amperaje es muy pequeña entre  $T_0$  y  $T_p$  y Con-Mich con una diferencia muy grande en amperaje entre  $T_0$  y  $T_p$ ; esto se puede verificar en la Figura 5.12 al final del capítulo. Lo mismo sucede Cel9-2 y Bolita, al mostrar un peso importante en el segundo componente principal, se evidencian como perfiles con puntos de inicio gelatinización muy tardíos respecto al resto de los perfiles.

Debido a la mayor representatividad del registro landmark sobre el proceso estudiado, resulta más informativo interpretar sus componentes principales, pero se puede observar algunas similitudes con los componentes principales sin registro.

En conclusión podemos destacar dos fuentes de variación muy importantes en este proceso, las cuales se encuentran en el estado de inicio y en el estado de pico de la gelatinización, confirmando las suposiciones antes mencionadas.

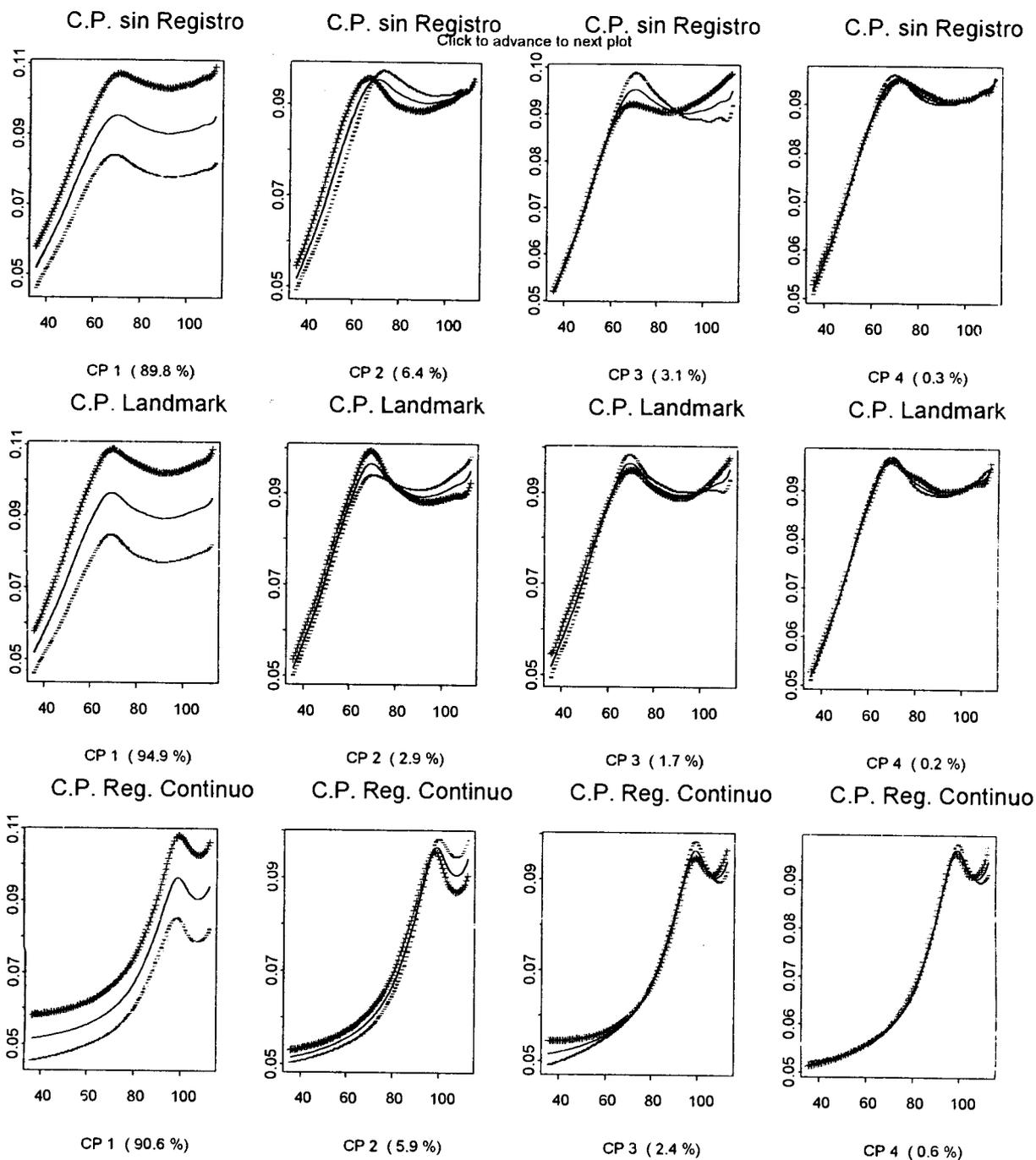


Figura 5.8. Componentes principales de los perfiles correspondientes a un tipo de registro y sin registro.

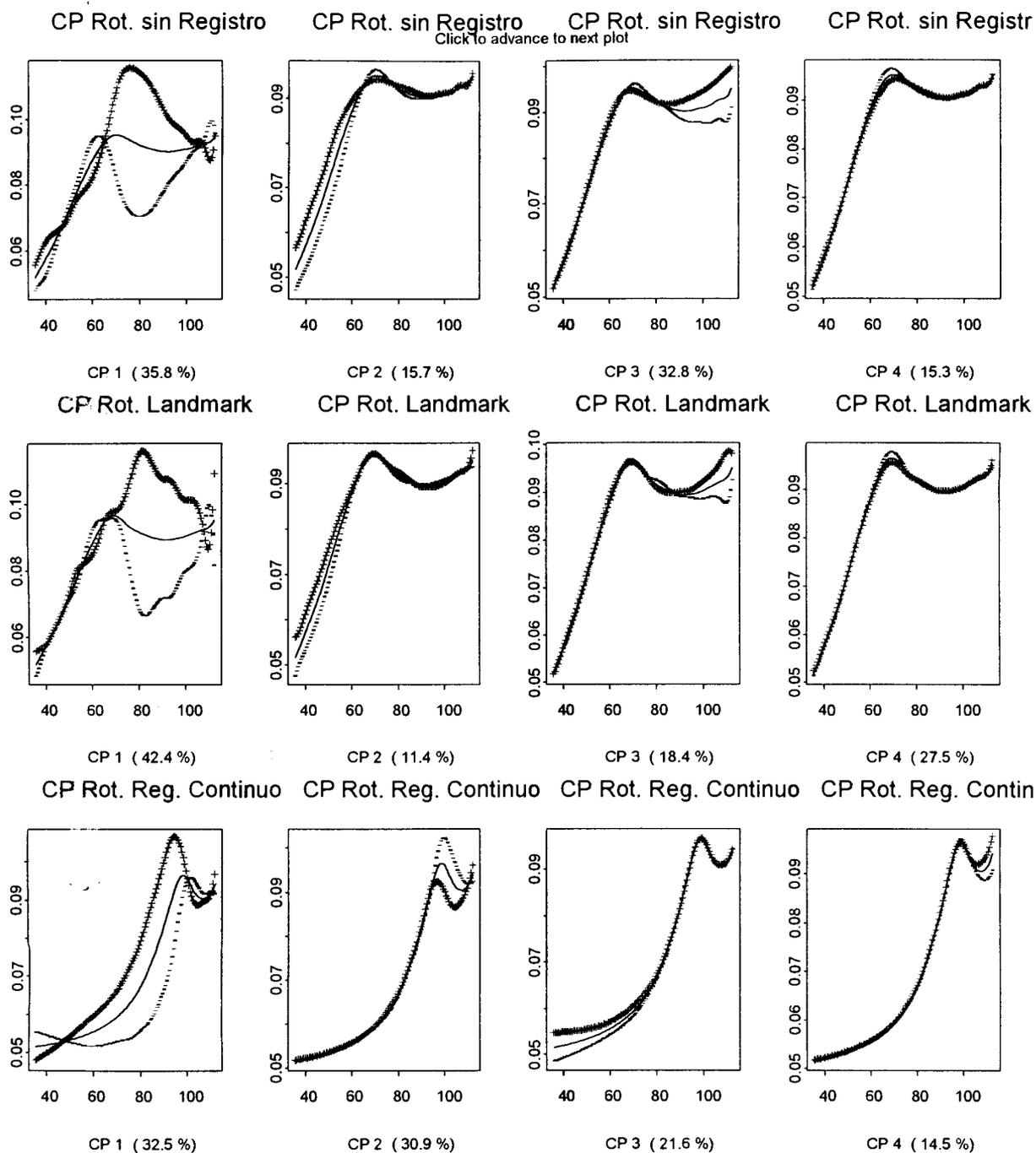


Figura 5.9. Componentes principales rotados de los perfiles correspondientes a un tipo de registro y sin registro

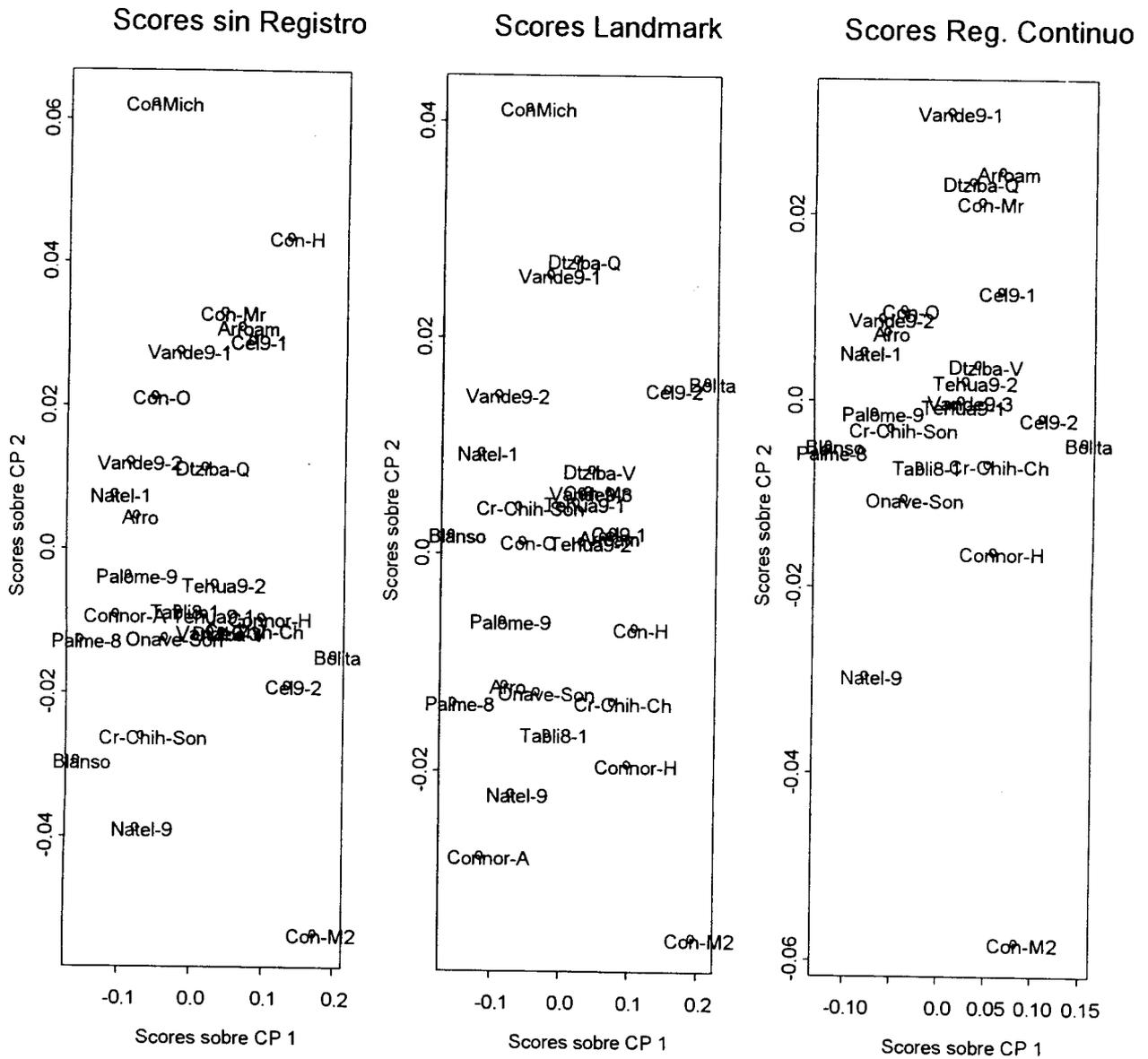


Figura 5.10 Scores de los perfiles sin registrar, con registro landmark y con registro continuo respectivamente.

Scores Rot. sin Registro

Scores Rot. Landmark

Scores Rot. Reg. Contin

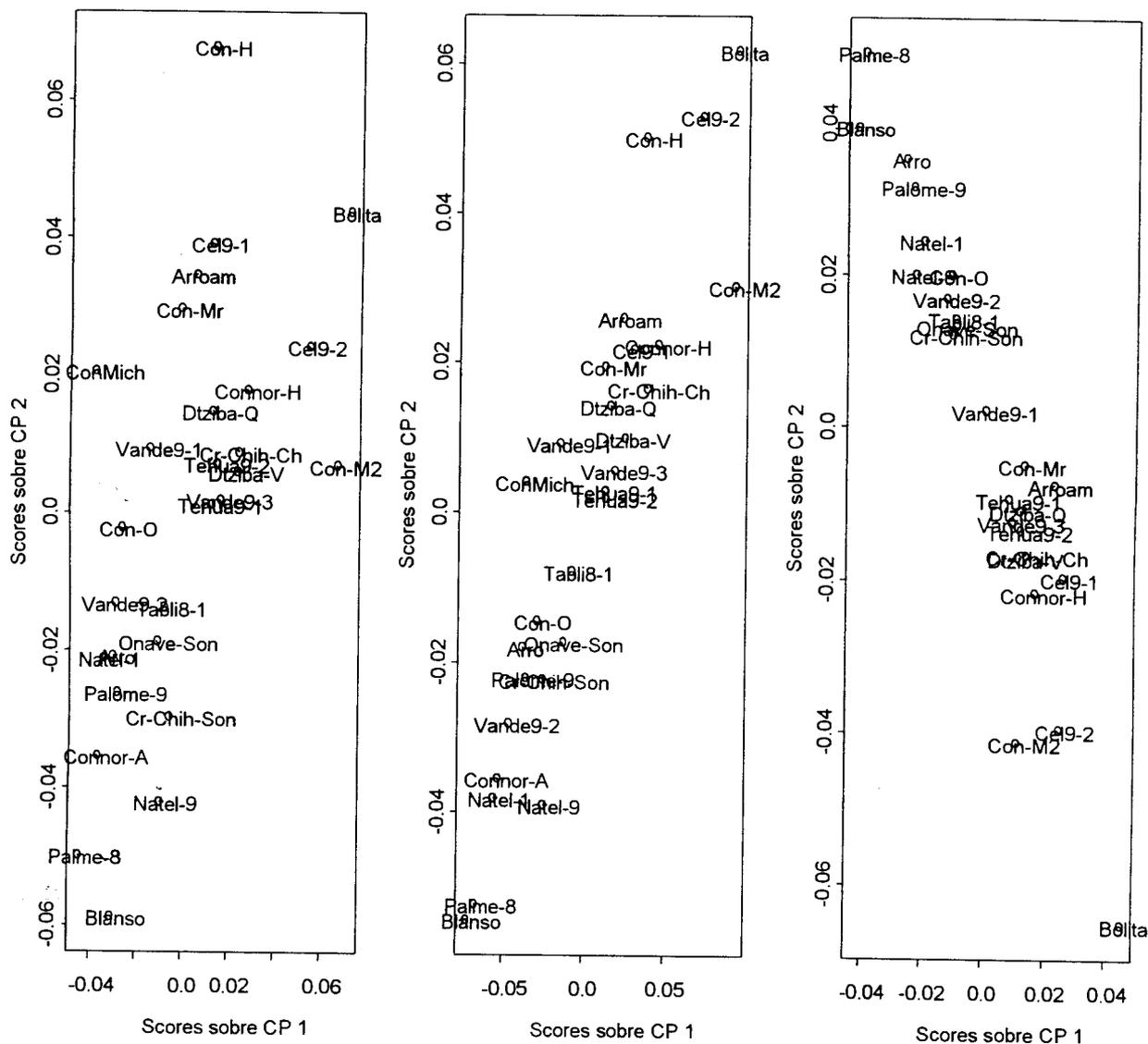


Figura 5.11. Scores rotados correspondientes a los perfiles sin registrar, con registro landmark y con registro continuo

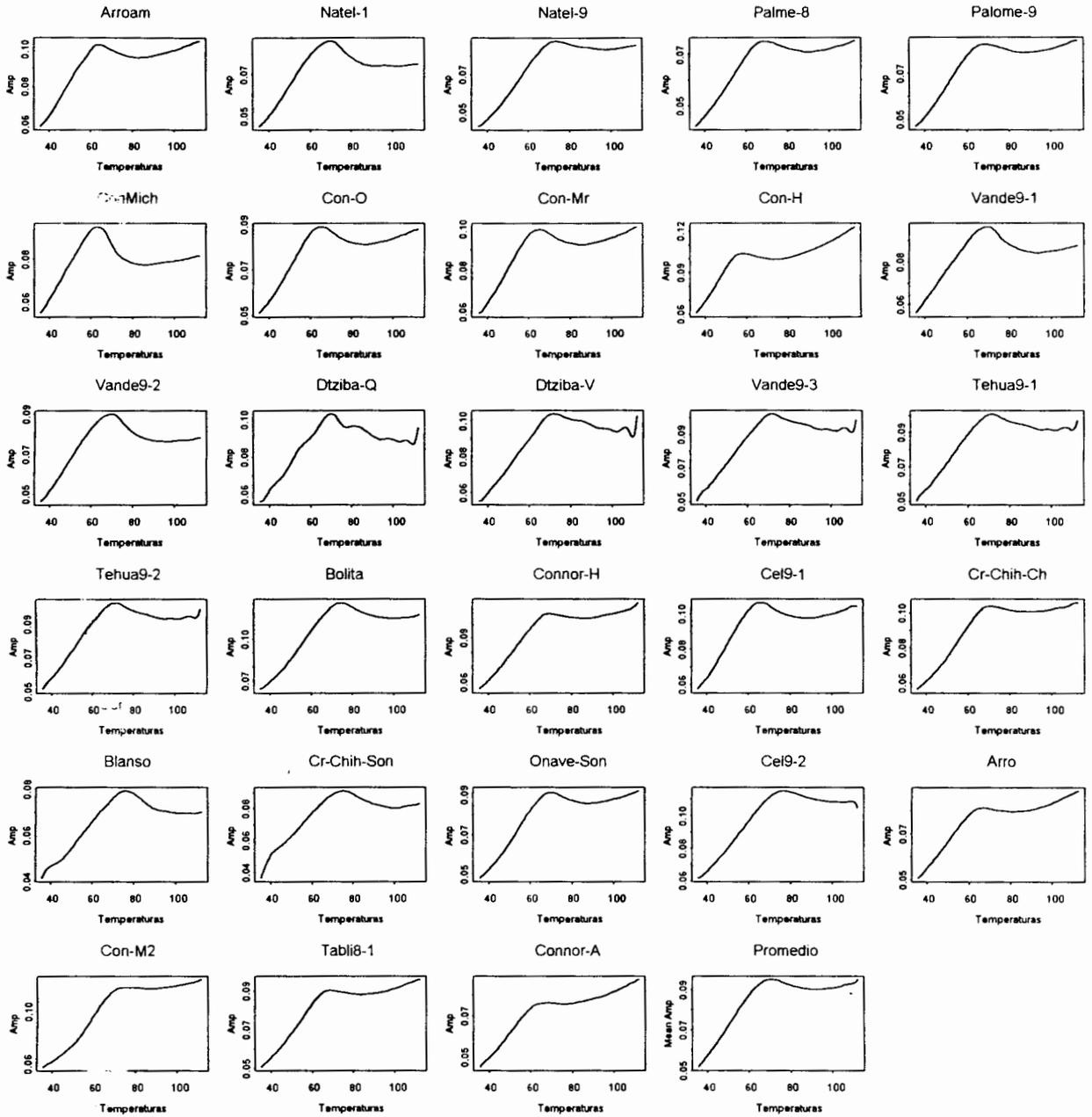


Figura 5.12. Los 28 perfiles suavizados con sus nombres y su promedio

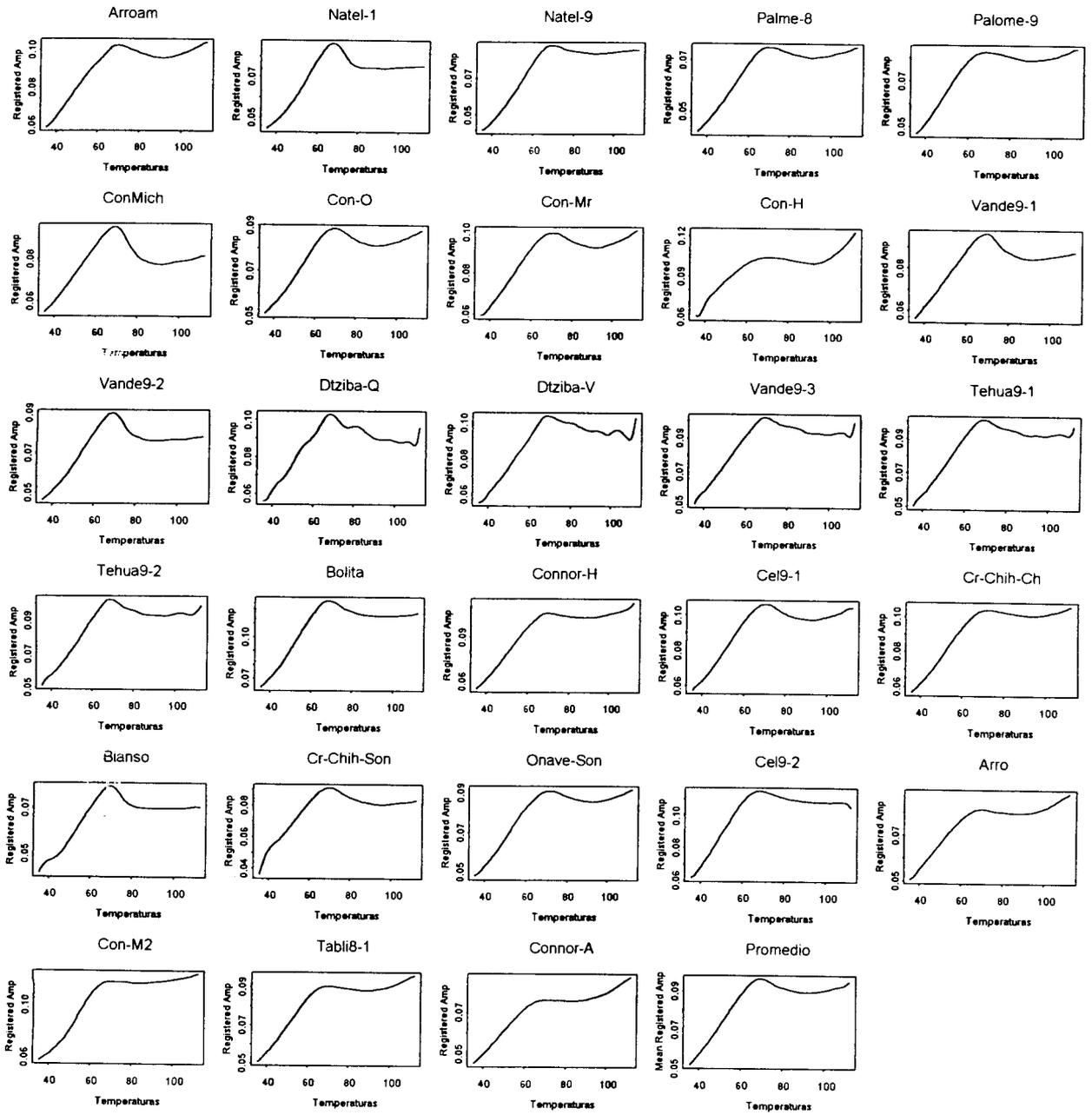


Figura 5.13. Los 28 perfiles suavizados con registro landmark y su promedio

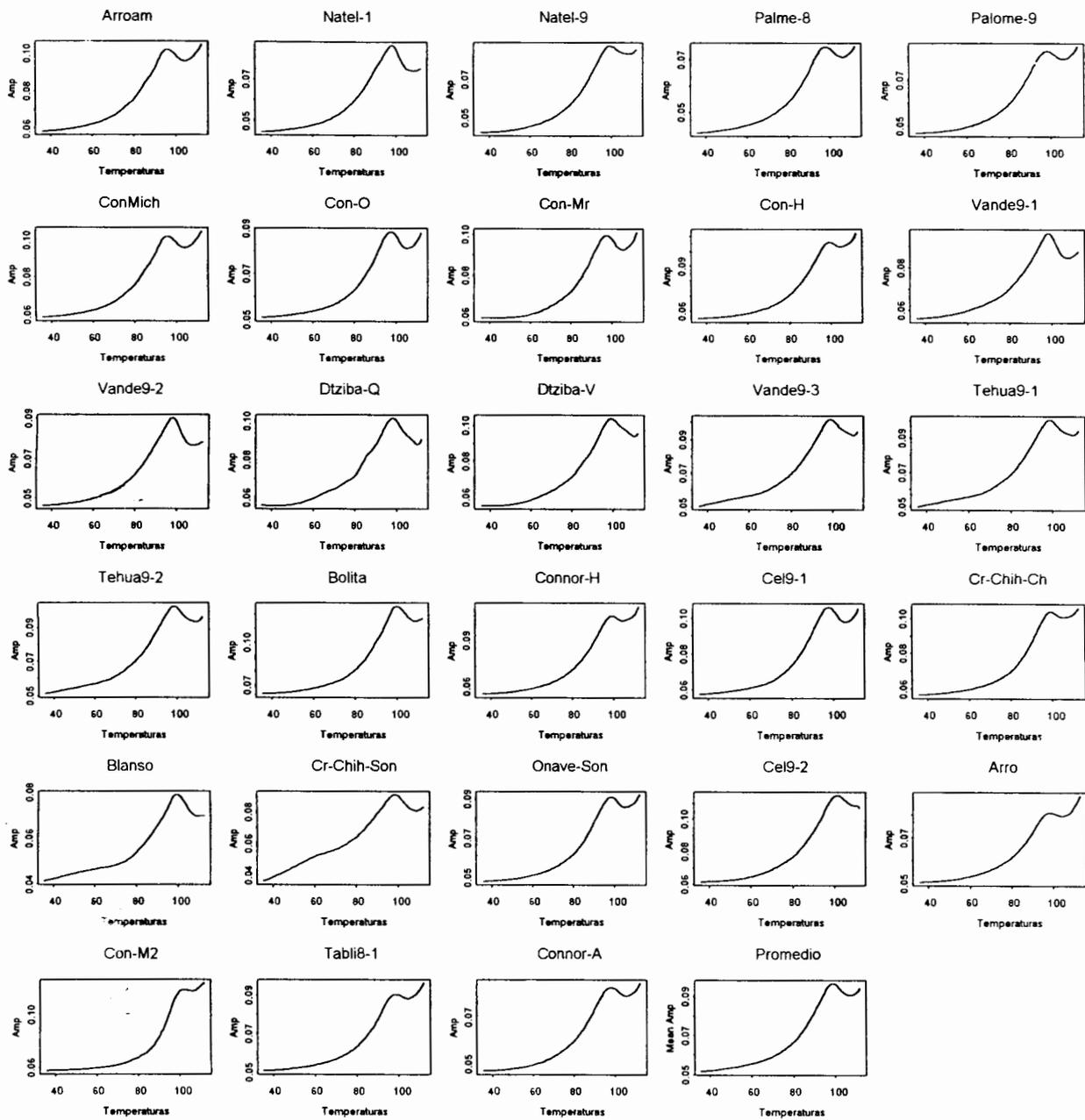


Figura 5.14. Los 28 perfiles suavizados con registro continuo y su promedio

De este modo concluimos nuestra aplicación habiendo encontrado las fuentes de variación importantes en el proceso de gelatinización. Definitivamente registrando los datos se ganan representaciones mas fidedignas del proceso, como lo fue con registro landmark, el cual proporcionó mucha información. El registro global no obtiene una buena representación sobre todo en el estado de pico de la gelatinización.

## 6 Conclusiones

Como se ha podido analizar en el transcurso de este trabajo, *análisis funcional de datos*, es un enfoque alternativo para analizar datos de mediciones repetidas en el tiempo. Este tipo de análisis, tiene una gran capacidad para extraer la mayor información posible del monitoreo de un proceso, cualquiera que éste sea, además de que se tiene una gran gama de estimadores para las funciones que se requieren conocer con diversas propiedades, las cuales hacen atractivo para un usuario utilizar este tipo de enfoque.

Métodos de suavizamiento tales como suavizamiento por kernel y suavizamiento polinomial local, son fáciles de entender, además de que tienen excelentes características computacionales. El parámetro  $h$  (ancho de banda) juega un rol de vital importancia, ya que permite que el ajuste se adapte a la variación de la curvatura. El problema que se tiene con este tipo de suavizadores es su inestabilidad cerca de las fronteras de los intervalos sobre los que se está trabajando, aunque suavizamiento polinomial local trabaja substancialmente mejor que suavizamiento por kernel en este sentido. Es así, que para las expansiones por funciones base es de mucho valor considerar empatar la elección de las funciones base a las características conocidas de los datos, especialmente en regiones donde los datos están situados asimétricamente alrededor de un punto de interés, como puede ser cerca de las fronteras. Por lo tanto el método de penalización de la rugosidad rescata las mejores características de las técnicas anteriormente mencionadas y mejora substancialmente sus debilidades, aunado a las excelentes propiedades de los estimadores resultantes de este método, resulta, sin lugar a dudas, uno de los mejores en su tipo.

El análisis de componentes principales funcionales, como se ha visto anteriormente, es simplemente una generalización del caso multivariado, pero con la ventaja de una más fácil interpretación en los resultados obtenidos, los cuales gráficamente son de muchísima ayuda, además de que computacionalmente este algoritmo es muy estable.

Un método para la ayuda del análisis de datos funcionales, es el registro de datos, el cual enfatiza o hace más representativas las características a estudiar en el monitoreo de un proceso en particular. Registro de datos es una técnica que se debe manejar con cuidado, por ejemplo, el registro global es más natural de aplicarse cuando se desea registrar perfiles en los que no hay características comunes específicas de interés por parte del investigador, si

se hace caso omiso de esta indicación, los resultados del análisis pueden ser poco confiables en su interpretación.

En cuanto a la aplicación hecha en el capítulo 5, podemos observar la gran representación, que el registro landmark, hizo del proceso de gelatinización del maíz, a diferencia del análisis de los datos sin registro y con registro continuo, describiendo claramente las fuentes de variación más importantes en el proceso de gelatinización, es decir, las diferencias entre los maíces dentro de este proceso se dieron principalmente en los estados de inicio y pico de la gelatinización, así como el período entre estos dos. También se señalaron características que hacen a algunos maíces diferentes de los demás.

## 7 Bibliografía

- Falkner, F. (1960) *Child Development: An international method of study*. Basel: Karger.
- Friedman, J.H. and Silverman, B. W. (1989) Flexible parsimonious smoothing and additive modeling (with discussion and response). *Journal of the American Statistical Association*, 90, 1179-1188.
- Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models, A roughness penalty approach*. CHAPMAN&HALL/CRC.
- Gasser, Th. y Müller, H.G. (1979). Kernel estimation of regression functions. *Smoothing techniques for Curve Estimation* (Th. Gasser and M. Rosenblatt, eds.), Springer Lecture Notes in Mathematics N0. 757, Springer-Verlag, Berlin, pp. 23-68.
- Hart, Jeffrey D. (2000). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer-Verlag, New York.
- Kneip, A. y Gasser, T. (1992). Statistical tools to analyze data representing a sample of curves. *The Annals of Statistics*, 20, 3, 1266-1305.
- Mauricio Sánchez, R.A. (2001). Caracterización Físicoquímica, Térmica y Eléctrica de Razas Mexicanas de Maíz y Evaluación de sus Posibles Usos en la Industria Alimentaria. *Tesis de Maestría en Ciencia y Tecnología de los Alimentos*, Universidad Autónoma de Querétaro, Facultad de Química. Programa de Posgrado en Alimentos del Centro de la República.
- Nadaraya, E.A. (1964). On estimating regression. *Theory Probab. Appl.*, 9, 141-142.

- Priestley, M.B. y Chao, M.T. (1972). Non-parametric function fitting. *J. Roy. Statist. Soc. Ser. B* 34, 385-392.
- Ramsay, J.O., Bock, R.D. y Gasser, T. (1995) Comparison of height acceleration curves in the Fels, Zurichn and Berkeley growth data. *Ann. Hum. Biol.*, 22, 413-426.
- Ramsay, J.O. y Silverman, B.W. (1997). *Functional Data Analysis*. Springer-Verlag. New York.
- Ramsay, J.O. y Dalzell, C.J. (1991). Some tools for Functional Data Analysis. *Journal of the Royal Statistical Society, B*, 3, 539-572.
- Ramsay, J.O. (1995). Some tools for the multivariate analysis of functional data. *In Recent Advances in Descriptive Multivariate Analysis*. Edited by Wojtek J. Krazanoswski. Claredon Press, Oxford.
- Ramsay, J.O. (1998). Curve Registration. *Journal of the Royal Statistical Society, B*, 60, 365-375.
- Ramsay, J.O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society, B*, 60, part 2, 365-375.
- Ramsay, J.O. (2000). *A Guide to Curve Registration*. FTP [ego.psych.mcgill.ca/pub/ramsay](http://ego.psych.mcgill.ca/pub/ramsay)
- Ramsay, J.O. y Silverman, B.W. (1998). *S-PLUS Functions for FDA*. FTP [ego.psych.mcgill.ca/pub/ramsay](http://ego.psych.mcgill.ca/pub/ramsay)
- Ramsay, J.O. and Li Xiaochun, (1998). Curve Registration. *Journal of the Royal Statistical Society, B*, 60, part 2, 351-363.

Ramsay, J.O., Wang, X. y Flanagan, R. (1995). A functional data analysis of the pinch force of human fingers. *Appl. Statist.*, 44, 17-30

Reinsch, C. (1967). Smoothing by spline functions. *Numer. Math.*, 10, 177-83

Rice, J.A. y Silverman, B.W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, B.* 53. 233-243.

Silverman, B. W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.*, 12, 898-916.

Watson, G.S. (1964). Smooth regression analysis. *Sankhya Ser. A*, 26, 359-372