



*Modelos estadísticos espacio-temporales  
en perimetría*

*Tesis doctoral presentada por  
M<sup>a</sup> Victoria Ibáñez Gual*

*y dirigida por  
Amelia Simó Vidal*

*Departamento de Matemáticas  
Escuela Superior de Tecnología y Ciencias Experimentales  
UNIVERSITAT JAUME I.- Julio 2003*



Dña. AMELIA SIMÓ VIDAL, profesora titular del Departamento de Matemáticas de la Universitat Jaume I,

CERTIFICA:

que la presente memoria “*Modelos estadísticos espacio temporales en perimetría*”, ha sido realizada bajo su dirección, en el Departamento de Matemáticas de la Universitat Jaume I por Dña. M<sup>ra</sup> Victoria Ibáñez Gual, y constituye su tesis para optar al grado de Doctor por la Universitat Jaume I.

Y para que así conste, en cumplimiento de la legislación vigente, presenta la referida memoria ante el Departamento de Matemáticas de la Universitat Jaume I de Castellón, a 16 de julio de 2003.

LA DIRECTORA

Amelia Simó Vidal



*A mis padres,  
a Jesús y a Héctor.*



## Agradecimientos

Esta memoria resume el trabajo de investigación realizado durante los últimos 5 años, y su desarrollo no habría sido posible sin la colaboración y apoyo de familiares, compañeros y amigos. A todos vosotros, muchas gracias.

Ante todo quiero agradecer a Amelia Simó, su paciencia y generosidad, y el que me haya ofrecido a lo largo de estos años no sólo su guía y sus conocimientos científicos, sino también su amistad. Sin su ayuda, esta memoria no sería hoy una realidad.

A la Dra. Lucía Martínez-Costa, debo agradecerle que nos hablara por primera vez del Glaucoma, que nos permitiera ver un campímetro, y que nos proporcionara los primeros datos, con los que empezamos a trabajar. Gracias también por estar siempre dispuesta a ayudarnos con todas las dudas “oftalmológicas” que nos han sido saliendo en este tiempo.

Agradecer también al Dr. Alfonso Antón el habernos proporcionado gran parte de la base de datos que ha servido de base a este trabajo, y también por su asesoramiento en el trabajo.

Al profesor Peter Diggle, quisiera agradecerle su interés, amabilidad y dedicación, durante los meses que pasé en el departamento de “Medical Statistics” de la Universidad de Lancaster. Quisiera hacer extensivo el agradecimiento a todas las personas que encontré tanto en Lund como en Lancaster, por acogerme tan bien y por ofrecerme tanto su ayuda profesional, como su amistad personal, que en muchos casos aún seguimos conservando.

Este trabajo se ha desarrollado en el Departamento de Matemáticas de la

Universitat Jaume I. Quisiera agradecer a todos y cada uno de sus miembros el apoyo y la ayuda que constantemente me han ofrecido. Este agradecimiento va especialmente dirigido a mis compañeros de área, y un poquito más especialmente aún a Irene. Gracias a todos por crear un clima de trabajo tan agradable.

Por último no podía faltar el agradecimiento a Guille, Toni, Xexo, Vicent, Raquel, y al resto de “culebras” por tantos y tantos momentos y tantos y tantos cafés como hemos compartido durante estos años, ... y por tantos y tantos mails y talks intercambiados cuando he estado fuera. En este agradecimiento deben entrar también Suso, Héctor, Vero y Patri, Mila, Jorge, Domi, Brígida, Gemma y Gabi. Gracias a todos por estar ahí, y por compartir todo esto conmigo.

# Índice general

<b>Parte I</b>	<b>Introducción</b>	<b>21</b>
<b>1.</b>	<b>Introducción</b>	<b>23</b>
<b>2.</b>	<b>Glaucoma</b>	<b>29</b>
2.1.	Introducción . . . . .	30
2.2.	Análisis del campo visual . . . . .	34
2.3.	SITA . . . . .	36
2.4.	Resultados de la perimetría . . . . .	40
2.5.	Características generales de los campos visuales de diferentes poblaciones . . . . .	44
<b>Parte II</b>	<b>Geoestadística</b>	<b>47</b>
<b>3.</b>	<b>Geoestadística. Fundamentos teóricos y metodología</b>	<b>49</b>
3.1.	Introducción. . . . .	50
3.2.	Modelizando los datos . . . . .	53
3.2.1.	Sobre la estructura de la media . . . . .	53
3.2.2.	Sobre la estructura de la covarianza . . . . .	57
3.3.	Kriging . . . . .	67
3.4.	Análisis de validación cruzada . . . . .	76

3.5. Problemas que encontramos . . . . .	78
3.5.1. Geoestadística cuando no hay estacionariedad espacial	79
<b>4. Modelización espacio temporal de CV de pacientes sanos</b>	<b>81</b>
4.1. Introducción . . . . .	83
4.2. Análisis exploratorio espacial . . . . .	85
4.2.1. Notación . . . . .	86
4.2.2. Modelización de la media . . . . .	89
4.2.3. Estructura de correlación en los residuos . . . . .	93
4.2.4. Ajuste del modelo espacial . . . . .	97
4.3. Análisis exploratorio temporal . . . . .	98
4.4. Modelización espacio temporal conjunta . . . . .	103
4.4.1. Modelos de covarianza separables . . . . .	106
4.4.2. Modelos de covarianza no separables . . . . .	106
4.4.3. Estimación . . . . .	109
4.4.4. Comprobando los residuos . . . . .	110
4.4.5. Validación cruzada . . . . .	111
4.5. Posibles aplicaciones del modelo . . . . .	115
4.5.1. Construcción de intervalos de confianza para identi- ficar posiciones “enfermas” . . . . .	116
4.5.2. Predicción . . . . .	116
4.5.3. Simulación . . . . .	117
<b>5. Simulación de CV</b>	<b>121</b>
5.1. Introducción . . . . .	122
5.1.1. Notación . . . . .	126
5.2. Simulación no condicional . . . . .	126
5.2.1. Simulación directa . . . . .	127
5.2.2. Simulación no condicional de campos visuales . . . . .	129

<i>Contenidos</i>	11
-------------------	----

5.3. Simulación condicional . . . . .	134
5.3.1. Primer método.- Método secuencial gaussiano . . . . .	136
5.3.2. Segundo método.- Basado en Kriging Simple . . . . .	136
5.3.3. Simulación condicional de campos visuales . . . . .	137
<b>6. Modelo Geoestadístico para pacientes con Glaucoma</b>	<b>147</b>
6.1. Introducción . . . . .	148
6.2. Análisis espacial . . . . .	149
6.3. Buscando homogeneidad en la varianza espacial . . . . .	152
6.4. Modelización espacio temporal . . . . .	156
6.5. Bondad del modelo . . . . .	162
6.5.1. Validación cruzada . . . . .	162

**Parte III Series temporales y Campos Aleatorios de Markov**  
**165**

<b>7. Series temporales multivariantes y C.A.M.</b>	<b>167</b>
7.1. Introducción. . . . .	169
7.2. Series temporales multivariantes estacionarias. . . . .	172
7.2.1. Procesos autoregresivos multivariantes (Procesos <i>VAR</i> ).173	
7.2.2. Procesos de medias móviles multivariantes (Procesos MA). . . . .	195
7.2.3. Procesos ARMA multivariantes. . . . .	198
7.3. Series multivariantes no estacionarias. . . . .	199
7.3.1. Introducción. . . . .	199
7.3.2. Procesos estacionarios en tendencia . . . . .	200
7.3.3. Procesos con raíces unitarias . . . . .	202
7.3.4. Comparación: Procesos estacionarios en tendencia y procesos integrados . . . . .	210

7.3.5. Procesos con cambios de régimen (o con cambios estructurales) . . . . .	213
7.4. CAM y algoritmos de estimación de parámetros . . . . .	219
7.4.1. Introducción . . . . .	219
7.4.2. Campos aleatorios de Markov. Conceptos básicos . . . . .	222
7.4.3. Verosimilitud de Montecarlo . . . . .	227
7.4.4. Algoritmos basados en la pseudoverosimilitud . . . . .	229
7.4.5. Algoritmos basados en el gradiente estocástico . . . . .	230
7.4.6. Estudio comparativo . . . . .	232
<b>8. Estimación bayesiana de la variable estatus</b> . . . . .	<b>241</b>
8.1. Introducción . . . . .	243
8.1.1. Conjunto de datos y notación . . . . .	244
8.1.2. Objetivos . . . . .	245
8.2. Rasgos característicos de los campos visuales . . . . .	247
8.3. Modelización de los umbrales condicionados a los estatus . . . . .	248
8.4. Modelización a priori de los estatus . . . . .	253
8.5. Distribución conjunta . . . . .	258
8.6. Estimación de parámetros . . . . .	260
8.6.1. Paso 2. Estimación de los parámetros de la distribución a priori de los estatus . . . . .	262
8.6.2. Paso 3. Estimación de los parámetros de la distribución de los umbrales condicionados a los estatus . . . . .	262
8.6.3. Paso 4. Estimación de los estatus . . . . .	267
8.6.4. Simulación de los estatus . . . . .	268
8.7. Resultados y conclusiones . . . . .	271
<b>9. Modelización de CV con la metodología de series temporales</b>	<b>275</b>
9.1. Introducción . . . . .	277

<i>Contenidos</i>	13
9.2. Modelización de los umbrales condicionados a los estatus . . .	278
9.3. Modelización de los estatus . . . . .	278
9.4. Verosimilitud . . . . .	280
9.5. Estimación de parámetros. Búsqueda del método apropiado .	282
9.5.1. Paso 1. Simulación de los estatus . . . . .	283
9.5.2. Paso 2. Estimación de los parámetros de la distribución de los estatus . . . . .	284
9.5.3. Paso 3. Estimación de los parámetros de la distribución de los umbrales condicionados a los estatus . . . . .	285
9.6. Predicción. . . . .	287
9.7. Bondad del ajuste . . . . .	288
<b>Parte IV Conclusiones, comentarios y continuación del trabajo</b>	<b>293</b>
<b>10. Conclusiones</b>	<b>295</b>
Bibliografía . . . . .	301



# Índice de figuras

2.1. Anatomía del ojo. . . . .	31
2.2. (a) Campímetro “Octopus”, (b) Campímetro “Humphrey”. . .	35
2.3. Resultado de una campimetría. Además de los datos del pa- ciente y los índices de fiabilidad incluye: a) gráfico de rep- resentación numérica de los resultados; b) representación en escala de grises; c) y e) diagramas de desviación total; d) y f) diagramas de patrón de desviación, y g) índices globales. . . .	42
3.1. (a) Modelos de la familia Bessel, (b) Modelos de la familia Cauchy, (c) Modelos de la familia exponencial, (d) Modelo Gausiano. . . . .	61
3.2. (a) Modelo esférico, (b) Familia Mattérn. . . . .	62
4.1. a) y b) Medias y varianzas de los datos no transformados. Colores más oscuros indican valores más bajos. c) Varianza vs distancia al centro del campo visual. d) Varianza vs media. . .	87
4.2. a) y b) Medias y varianzas de los datos tras la transformación de Box-Cox. Colores más oscuros indican valores más bajos. c) Varianza vs distancia al centro del CV. d) Varianza vs media.	88
4.3. Variograma direccional empírico. . . . .	94

4.4. Variograma espacial empírico. La línea horizontal es una estimación de la varianza total. . . . .	96
4.5. Variogramas temporales empíricos calculados: a) en la posición número 16, b) en la posición número 37. . . . .	100
4.6. (a) Variograma temporal empírico medio. La línea horizontal es una estimación de la varianza total. (b) Ajuste de distintas funciones paramétricas al variograma temporal empírico. . . .	101
4.7. Variograma espacio-temporal empírico. . . . .	105
4.8. Estimación de $f(x)$ . . . . .	111
4.9. (a) Media empírica de los residuos estandarizados en cada posición. (b) Variograma espacio temporal empírico calculado sobre los residuos estandarizados. . . . .	112
4.10. En e) tenemos la predicción del CV de un paciente sano a la edad de 70.5 años. En a), b), c) y d) vemos las observaciones que disponíamos de este paciente a las edades de 66.8, 68.2, 68.7 y 69.4 años. . . . .	118
4.11. Comparación entre la clasificación de posiciones como “normales” o “enfermas”. En a), c) y e) hemos utilizado intervalos de confianza construidos a partir de nuestro modelo, y en b),d) y f) vemos la clasificación dada por el SITA para los mismos CV, ambos trabajando a una confianza del 95 %. . . . .	119
5.1. Simulaciones de tres diferentes pacientes de la misma edad. En a), g), d) los tres tienen 42 años, en b), h) y e) tienen 64, y en c), i) y f) tienen 76. . . . .	133

5.2. Las líneas continuas muestran las envolventes superior e inferior del variograma espacial normalizado calculado a partir de los 56 pacientes de nuestra base de datos. Con líneas discontinuas tenemos los variogramas normalizados calculados a partir de los campos visuales simulados de la figura 5.1. . . . 135

5.3. En a) tenemos el CV observado en un paciente de 49 años. En b), c) y d) mediante simulación condicionada hemos simulado CV a partir del CV que tenemos en a) con un retardo temporal de 1 año entre dos simulaciones consecutivas. . . . . 140

5.4. Estimación Kriging de la figura 5.3 a), a las edades de 50, 51 y 52 años. . . . . 141

5.5. Simulaciones no condicionales de un proceso aleatorio gaussiano  $Z$  de media 0 y matriz de covarianzas  $V$  detallada en la ecuación 5.7. El retardo temporal entre dos simulaciones consecutivas es de 1 año. . . . . 142

5.6. Estimaciones Kriging de la figura 5.5. . . . . 143

5.7. En a) vemos el CV “real” de un paciente de 63 años. En b), c) y d) hemos simulado condicionalmente CV a partir del primero, con un retardo temporal de 1 año entre dos simulaciones consecutivas. . . . . 144

5.8. En a) vemos el CV “real” de un paciente de 58 años. En b), c) y d) hemos simulado condicionalmente CV a partir del primero, con un retardo temporal de 1 año entre dos simulaciones consecutivas. . . . . 145

- 5.9. Comparación de CV “reales” y “simulados”. En a), b), c) y d) vemos 4 CV “reales”, pertenecientes al mismo paciente a distintas edades. En e), f), y g) vemos tres CV que han sido simulados condicionalmente a partir de a), al igual que en h), i) y j), a las mismas edades que tenía el paciente en, b), c) y d). 146
- 6.1. En (a) vemos la media empírica de las observaciones recogidas en cada posición, (en rojo tenemos la media para las posiciones enfermas, y en azul la de las posiciones sanas ). Las posiciones han sido enumeradas del 1 al 52 siguiendo el orden natural. En (b) tenemos un gráfico análogo para las varianzas empíricas. Las varianzas para las posiciones enfermas (en rojo) son también las que presentan valores más bajos. . . . . 151
- 6.2. Mapas de medias y varianzas en posiciones sanas y en posiciones enfermas. . . . . 153
- 6.3. En (a) vemos la media empírica de los residuos y en (b) tenemos la varianza empírica, también de los residuos. . . . . 155
- 6.4. Estimación de la superficie suave  $f(x)$  que aparece en la formulación de la media (ecuación 6.4). . . . . 157
- 6.5. Variograma espacio temporal empírico. . . . . 158
- 6.6. Estimación de  $f(x)$  de la media de la distribución espacio temporal conjunta. . . . . 161
- 6.7. (a)Media empírica de los residuos estandarizados en cada posición. (b) Variograma espacio temporal empírico calculado sobre los residuos estandarizados. . . . . 163
- 7.1. Ejemplos de las imágenes utilizadas en nuestros experimentos, generadas a partir de la distribución (7.50) con  $\alpha = 0,6$  . . . . 237

7.2. Añadimos un proceso de ruido blanco con varianza 0,8 a las imágenes de la figura 7.1 . . . . . 237

7.3. Añadimos un proceso de ruido blanco con varianza 1 a las imágenes de la figura 7.1 . . . . . 237

7.4. Ejemplos de las imágenes utilizadas en nuestros experimentos, generadas a partir de la distribución (7.50) con  $\alpha = 0,3$  . . . . . 238

7.5. Añadimos un proceso de ruido blanco con varianza 0,8 a las imágenes de la figura 7.4 . . . . . 238

7.6. Añadimos un proceso de ruido blanco con varianza 1 a las imágenes de la figura 7.4 . . . . . 238

8.1. Ilustración de los resultados obtenidos. En a), b), c) y d) tenemos 4 CV observados en un mismo paciente con diferencia de un año entre observaciones consecutivas. En e), f), g) y h) vemos las estimaciones obtenidas para la variable “estatus”. En negro representamos las posiciones estimadas como “enfermas”. 273

9.1. En a), b), c), d) y e) vemos 5 campos visuales observados en un paciente, en cinco exploraciones consecutivas. A partir de ellas predecimos CVs de este paciente. En f) vemos la predicción un año después de la última observación, y en g) dos años después. 289

9.2. Algunas de las correlaciones estimadas para los residuos de un paciente del que tenemos 9 exploraciones.  $r(k, l, i)$  indica el  $(k, l)$ -ésimo elemento de la matriz de correlación de retardo  $i$  (ver apartado 7.2.1). . . . . 291



## Parte I

# Introducción



# Capítulo 1

## Introducción

El desarrollo de gran parte de los modelos y métodos estadísticos ha ido ligado al deseo de estudiar aplicaciones específicas dentro de diversos ámbitos científicos. El presente trabajo también surgió con el objetivo de resolver diversos problemas que se plantean dentro del ámbito de la oftalmología, todos ellos ligados con un conjunto de datos y con una aplicación muy concreta: el estudio del glaucoma. El glaucoma es una enfermedad ocular muy extendida entre la sociedad actual, que se caracteriza por producir una pérdida de visión gradual en el paciente, pudiendo en sus estadios más avanzados llegar a producir la ceguera del paciente.

Para diagnosticar y evaluar el desarrollo de la enfermedad, los oftalmólogos basan buena parte sus diagnósticos en el análisis de unos mapas, llamados campos visuales, que les informan de la intensidad de visión que tiene el paciente en un conjunto de puntos de su retina. Por ello, gran parte de este trabajo está dedicado a estudiar modelos estadísticos espacio temporales que nos permitan caracterizar la evolución espacio-temporal de los campos visuales tanto de pacientes sanos, como de pacientes que padecen glaucoma.

Los modelos espacio temporales, empezaron a plantearse y a estudiarse a partir de la abundancia de aplicaciones que desde diversos ámbitos aplicados demandaron estas técnicas. Problemas relacionados con cuestiones meteorológicas [90, 176, 177], con contaminación ambiental [76, 58, 86], con el análisis

de problemas ecológicos y de recursos mediambientales [40], con problemas relativos a la epidemiología y a las ciencias de la salud, o relacionados con la construcción de mapas de enfermedades [200] entre otros, demandaron el desarrollo de este tipo de modelos.

El primer paso a la hora de empezar a desarrollar modelos espacio-temporales, fue el intentar adaptar a este contexto, modelos ya conocidos y ampliamente utilizados. Así empezó a estudiarse cómo modelizar fenómenos espacio-temporales generalizando tanto los métodos propios de la estadística espacial (Geoestadística, Cadenas de Markov, Métodos MCMC, ...), como los métodos ya desarrollados en la teoría de series temporales (Modelos ARMA, ARIMA, Modelos en espacio de estados, Filtro de Kalman,...).

En algunas ocasiones, aunque se disponía de observaciones recogidas en un dominio espacio temporal, se intentaba seguir trabajando con modelos puramente espaciales o puramente temporales. En estos casos, como Kyr-  
iakidis y Journal [122] explican, el análisis se enfocaba a menudo hacia la construcción (modelización o en general estudio) de mapas espaciales para cada instante de tiempo con el objetivo de compararlos y detectar persistencias o cambios en los patrones espaciales en el tiempo [69, 185]. Alternativamente el análisis podía enfocarse hacia la modelización de series temporales, definiendo una serie temporal para todas las observaciones recogidas en cada posición espacial [16]. Cuando existe correlación entre las distintas series temporales, esta se refleja permitiendo una correlación espacial entre los parámetros de dichas series temporales [76].

Otra aproximación de las que se propusieron para modelizar fenómenos espacio temporales, fue la de considerar cualquier fenómeno espacio temporal como la realización de un campo aleatorio definido en  $\mathbb{R}^3$  (dimensión espacial=2+1=dimensión temporal). Este método resultó no ser en muchas ocasiones adecuado, debido a las diferencias existentes entre el dominio es-

pacial y el temporal [171]. Las diferencias entre los dos dominios son muchas, por citar unas cuantas podemos destacar que los datos están ordenados de forma natural en el tiempo, y que esta ordenación no existe en el espacio. Además los datos en el tiempo y en el espacio se miden en escalas distintas, que no pueden ser comparadas, y normalmente no se dispone de la misma cantidad de información en ambos dominios. Es más habitual tener datos mucho más densos en el tiempo que en el espacio, debido al coste que puede suponer el aumentar el número de “observatorios” espaciales, siendo habitualmente el coste mínimo a la hora de tomar repeticiones de las observaciones en el tiempo. Por último decir que los datos pueden mostrar periodicidad temporal y no estacionariedad espacial, lo cual imposibilita también el trabajar de forma conjunta en  $\mathfrak{R}^3$ .

En este trabajo vamos a hacer una revisión de la teoría de las series temporales multivariantes, y de la metodología geoestadística, y utilizaremos modelos y técnicas estadísticas espacio temporales extraídas de ambas metodologías para resolver una serie de problemas que se plantean en el estudio del glaucoma.

En el capítulo 2 vamos a explicar brevemente qué es el glaucoma, y nos centraremos en las características de los campos visuales.

A lo largo de la tesis, trabajaremos con bases de datos en las que tendremos campos visuales de un conjunto de pacientes, que han sido obtenidos en la consulta oftalmológica. Nuestros datos van a ser en principio algo distintos a los datos de los que normalmente se dispone en cualquier estudio espacio temporal, ya que para nosotros cada observación es un mapa completo, y tenemos mapas repetidos en el tiempo para cada paciente. Esa es la peculiaridad de nuestro conjunto de datos, el disponer de observaciones (mapas) en el tiempo para un grupo de pacientes.

Como hemos comentado anteriormente, varias son las perspectivas bajo

las cuales analizar datos espacio temporales. En este trabajo nos planteamos el análisis de los campos visuales bajo dos perspectivas distintas, y sus correspondientes metodologías. Por eso hemos dividido la tesis en dos partes que corresponden a las dos metodologías utilizadas.

En la primera parte de la tesis, vamos a trabajar utilizando la metodología propia de la Geoestadística. Por ello en el capítulo 3 revisamos los principios teóricos de la Geoestadística, que necesitaremos en los capítulos 4, 5 y 6. El lector familiarizado con las técnicas geoestadísticas puede saltar la lectura del capítulo 3, y pasar directamente a las aplicaciones.

En el capítulo 4 nos planteamos el construir un modelo para caracterizar la distribución espacio temporal de campos visuales de pacientes sanos. Una vez construido el modelo lo utilizaremos para construir intervalos de confianza que nos permitan clasificar cada una de las posiciones de un campo visual como “normal” o “afectada por el glaucoma”, para realizar predicciones, y en el capítulo 5 para realizar simulaciones. En el capítulo 6, generalizaremos el modelo para que describa tanto campos visuales de pacientes sanos, como de pacientes afectados por la enfermedad.

En la segunda parte de la tesis cambiaremos la metodología y empezaremos a trabajar con los métodos propios de las series temporales multivariantes, desde el punto de vista clásico. En el capítulo 7 veremos la teoría de las series temporales multivariantes que necesitaremos más tarde.

Los dos capítulos teóricos que contiene la tesis, no pretenden recoger toda la teoría propia de la Geoestadística y de las series temporales multivariantes, sino centrarse en aquellos aspectos teóricos que nos son necesarios a la hora de trabajar en los distintos problemas planteados.

El lector familiarizado con la teoría clásica de series temporales multivariantes puede omitir la lectura del capítulo 7, y pasar directamente a los capítulos 8 y 9 en los que planteamos dos problemas. En el capítulo 8, bajo

un enfoque bayesiano planteamos el problema de la estimación de la variable “estatus” en cada posición. Esta variable nos permite saber si cada una de las posiciones testeadas es “sana” o “enferma”.

En el capítulo 9 el objetivo que planteamos está más próximo al perseguido en la primera parte de la tesis en la que utilizamos la metodología geoes-tadística, ya que pretendemos construir un modelo espacio temporal conjunto que nos permita caracterizar la distribución espacio temporal de campos visuales de pacientes que sufren glaucoma, y utilizar el modelo obtenido para realizar predicciones.

Destacar por último que el software hemos utilizado en este trabajo ha sido: el R en la parte de geoestadística, y el Matlab, para la parte de series temporales.



# Capítulo 2

## Glaucoma

### Contenidos

---

2.1. Introducción . . . . .	30
2.2. Análisis del campo visual . . . . .	34
2.3. SITA . . . . .	36
2.4. Resultados de la perimetría . . . . .	40
2.5. Características generales de los campos visuales de diferentes poblaciones . . . . .	44

---

## 2.1. Introducción

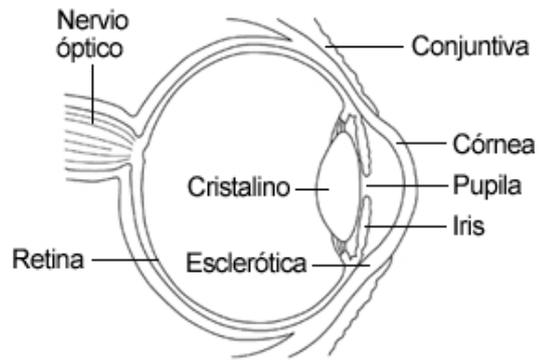
Glaucoma es el nombre con el que se conoce a una enfermedad ocular que puede llegar a dañar seriamente la función visual, y que actualmente se sitúa entre las principales causas de ceguera en todo el mundo. Esta enfermedad suele caracterizarse por llevar habitualmente asociada una elevación de la presión intraocular, que con el tiempo puede llegar a producir daños irreversibles en las fibras del nervio óptico.

Las fibras del nervio óptico se deterioran cuando la presión intraocular se eleva por encima de un nivel, que varía de unos individuos a otros. Si la presión elevada se mantiene durante mucho tiempo, o alcanza cifras exageradamente altas, las fibras se pueden dañar de forma irreparable, y la pérdida de visión se hace irreversible, pudiendo llegar a producir una ceguera total. Por lo tanto, un diagnóstico precoz, iniciar el tratamiento correcto y establecer las pautas de seguimiento adecuadas, constituyen las armas más eficaces para evitar estas graves consecuencias.

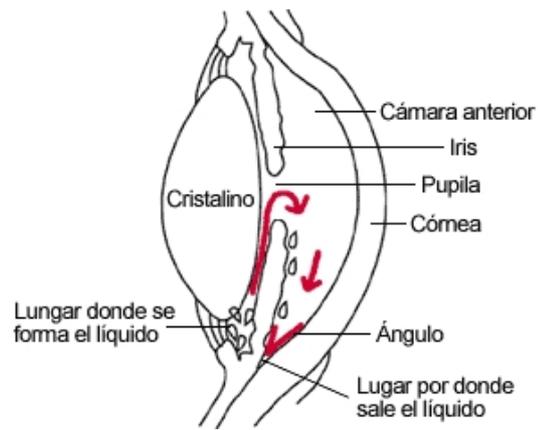
El glaucoma es una enfermedad que suele afectar a los dos ojos, y aunque se sabe que suele ir asociado con una alta presión ocular, se desconocen las causas exactas que lo provocan. Hasta ahora no se ha encontrado una relación directa entre una elevada presión intraocular y la presencia de glaucoma.

Por el interior del ojo circula un líquido, encargado de la nutrición de sus estructuras internas. Este líquido, denominado humor acuoso, tiene un sistema de producción y otro de evacuación (ver figuras 2.1 (a) y (b)). El equilibrio entre estos dos sistemas, permite mantener prácticamente constante la presión intraocular. Si algo falla en estos mecanismos, y entra más líquido en el ojo del que puede salir, la presión se eleva y el nervio óptico comienza a dañarse.

Los oftalmólogos son capaces de diagnosticar varias decenas de glaucomas diferentes, pero aquí vamos a esbozar brevemente sólo los tipos más



(a)



(b)

Figura 2.1: Anatomía del ojo.

frecuentes:

1. Glaucoma primario congénito: Se produce como consecuencia de un desarrollo defectuoso de las vías de salida del humor acuoso, ya desde las primeras semanas o meses de vida.
2. Glaucoma crónico de ángulo abierto: Es el más frecuente (aproximadamente las 3/4 partes de los diagnosticados). Se produce por el deterioro progresivo del sistema de eliminación del humor acuoso, deterioro mucho más pronunciado al que de forma natural se produce con la edad. Se presenta muy lentamente, sin producir síntomas que la persona que lo sufre, sea capaz de detectar.
3. Glaucoma agudo o de ángulo cerrado: Esta forma de glaucoma se presenta bruscamente, con gran dolor y brusca disminución de la visión, pudiendo ir acompañado de náuseas, vómitos, etc... Se produce por el cierre brusco de las vías de eliminación del humor acuoso. Debido a la forma especial del ojo de estas personas, el ángulo a través del cual se ha de eliminar este líquido, es excesivamente estrecho y en determinadas circunstancias, las paredes de este ángulo pueden ponerse en contacto, obstruyendo por completo el paso. Esto trae como consecuencia la rapidísima elevación de la presión y el intensísimo dolor.

Los principales grupos de riesgo están formados por personas de avanzada edad, que tienen antecedentes familiares de glaucoma, que padecen miopía o diabetes, personas que han seguido tratamientos prolongados con corticosteroides, que padecen enfermedades cardiovasculares y aquellas que han sufrido traumatismos o intervenciones quirúrgicas oculares. Las personas que están en alguna de estas circunstancias, deben pasar una revisión oftalmológica anual. Si concurren varios de estos factores, es posible que las revisiones

deban ser a más corto plazo. En cada revisión, se suelen incluir las siguientes exploraciones:

1. Tonometría.- Medida de la presión ocular con la ayuda de un tonómetro.
2. Oftalmoscopia o exploración del fondo de ojo, para comprobar si existe algún tipo de daño en el nervio óptico.
3. Gonioscopia para comprobar, en caso de sospecha de glaucoma, a qué tipo pertenece.
4. Campimetría o perimetría. Consiste en la exploración del campo visual, y es una prueba imprescindible para confirmar el diagnóstico de la enfermedad y establecer el tratamiento adecuado.

Como las posibilidades del tratamiento son mayores cuanto más precozmente se realice el diagnóstico, las revisiones periódicas son de gran importancia. El tratamiento tiene como objetivo conservar la visión y el campo visual tal y como están en el momento del diagnóstico, pues hoy por hoy resulta imposible regenerar las fibras del nervio óptico que ya estén atrofiadas.

La única forma de evitar la progresión del daño en el nervio óptico, es manteniendo la presión intraocular en cifras normales. Por ello, tras el diagnóstico el oftalmólogo debe optar entre un tratamiento médico (colirios hipotensores oculares, que se aplicarán ya indefinidamente, una o varias veces al día), y un tratamiento quirúrgico, (intervenciones con láser, o intervenciones quirúrgicas propiamente dichas), dependiendo por una parte del tipo de glaucoma, y por otra de la mayor o menor gravedad de la enfermedad en el momento del diagnóstico.

De todas formas, independientemente del tipo de tratamiento que se siga, el paciente de glaucoma requiere vigilancia y seguimiento. Un glaucoma puede mantenerse durante muchos años sin problemas con un tratamiento

determinado, pero en un momento concreto puede necesitar el cambio a otro tratamiento más adecuado.

En la mayor parte de los casos, especialmente los diagnosticados precozmente, el paciente puede desarrollar sus actividades con toda normalidad, sin ningún tipo de limitaciones para su trabajo, lectura etc.

Hemos dicho que una de las pruebas médicas más concluyentes a la hora de diagnosticar glaucoma es la campimetría o exploración del campo visual. Por ello, ahora que conocemos un poco por encima las características de la enfermedad, vamos a centrarnos en ver cuales son los datos de que dispone el oftalmólogo para diagnosticarla y controlar su seguimiento. En la sección 2.2 veremos cómo funciona un campímetro (perímetro o analizador de campo), y qué información proporciona. Estudiaremos el funcionamiento de los campímetros Humphrey y del paquete estadístico SITA.

## **2.2. Análisis del campo visual**

La perimetría es una herramienta clave para el diagnóstico del glaucoma, y su automatización ha hecho posible que se desarrollen varios dispositivos de interpretación asistida por ordenador.

Los campímetros o analizadores de campos visuales más utilizados en las consultas oftalmológicas son el analizador de campo Humphrey y el perímetro Octopus 500 (<http://www.humphrey.com>; <http://www.octopus.ch>), (ver figura 2.2).

Todos los campos visuales de la base de datos que vamos a analizar, han sido adquiridos con un campímetro de la casa Humphrey, por ello vamos a estudiar únicamente las propiedades de este campímetro y los resultados que presenta al finalizar cada test.

Cualquier campímetro consta fundamentalmente de una pantalla y un ordenador. La pantalla muestra un fondo con luminosidad constante de 31.5



(a)

(b)

Figura 2.2: (a) Campímetro “Octopus”, (b) Campímetro “Humphrey”.

apostilbs (asb), y desde este fondo se irán “disparando” estímulos luminosos de intensidad variable hacia la retina del paciente. Estos estímulos tienen un tamaño estándar, y su intensidad puede variar de 0,08 a 10,000 asb. Cada campímetro lleva implementado un software que define la “estrategia de test” que se va a realizar. Estos paquetes incorporan mayoritariamente herramientas estadísticas, y han ido evolucionando rápidamente, buscando siempre obtener la máxima información del estado del campo en el menor tiempo posible. Los paquetes estadísticos que se han implementado en los campímetros Humphrey, de más antiguo a más reciente son los llamados: FAST-PAC [81], STATPAC [194, 93], SITA [15, 14, 12], SITA Fast [13]. Existen numerosos trabajos médicos que realizan estudios clínicos para evaluar y

comparar estos métodos. Entre ellos destacamos los siguientes: [204, 170].

Todos los campos visuales que analizaremos han sido tomados en campímetros Humphrey y utilizando el paquete SITA, por lo que en adelante, al hablar del funcionamiento del campímetro, realmente nos estaremos refiriendo a cómo trabaja el SITA.

Antes de someterse a un análisis de campo visual, y para que los resultados sean válidos, el paciente debe fijar su mirada en un punto de fijación. Para controlar esta fijación, los campímetros Humphrey, “bombardean” periódicamente la mancha ciega con estímulos luminosos. La mancha ciega es una zona de la retina sin visión, por lo que respuesta positivas a estos estímulos indicarán una fijación pobre.

Existen dos tipos básicos de perimetrías: la perimetría estática y la kinética. Se diferencian en que en la primera, la posición desde la que se emite el estímulo no se mueve durante el periodo de exposición, y el tiempo de exposición de los estímulos no varía, mientras que en la kinética los estímulos se van moviendo durante su exposición, normalmente hacia el centro de la pantalla. Los campímetros utilizan la perimetría estática, y en adelante nos referiremos siempre a esta modalidad.

### **2.3. SITA**

Se conoce con el nombre de SITA (Swedish Interactive Threshold Algorithms) a una familia de algoritmos desarrollados por un grupo de investigadores suecos encabezados por el doctor A. Heijl [15]. Estos algoritmos determinan la estrategia de test que se aplicará al paciente. Al hablar de estrategia de test nos referimos al número de localizaciones de la retina que se testearán, a definir el orden en que se visitarán, a cómo se elegirá el primer valor del estímulo que se presente en cada localización, etc, ... En esta sección

Cuadro 2.1: Rejilla con la localización de los puntos que el campímetro Humphrey con el programa umbral 24-2, testea en un análisis de campo visual

			1	2	3	4			
			5	6	7	8	9	10	
		11	12	13	14	15	16	17	18
19	20	21	22	23	24	25		26	
27	28	29	30	31	32	33		34	
	35	36	37	38	39	40	41	42	
		43	44	45	46	47	48		
			49	50	51	52			

veremos la estrategia de test definida por el SITA.

Dentro del SITA existen varios programas. Todos los campos visuales que analizaremos han sido adquiridos con el programa 24 – 2, que testea 52 puntos de la retina del paciente (ver su disposición en la tabla 2.1). Estas posiciones están situadas sobre un grid regular, siendo el distanciamiento entre dos posiciones de  $6^\circ$  (vertical y horizontalmente). Los dos puntos en blanco corresponden con las localizaciones de la mancha ciega (o punto ciego).

El objetivo de la campimetría es obtener el umbral de visión en cada uno de los puntos testeados. El umbral de visión (que en adelante llamaremos simplemente umbral), se define como la intensidad del estímulo con una probabilidad de respuesta del 50 %, y se utilizará para cuantificar pérdidas de función visual.

El umbral en cada localización se estimaba originalmente mediante un método de “escalera de luminancia” en el que el test escogía un nivel inicial para el estímulo, y lo presentaba en una localización. Si se producía una respuesta, el nivel de intensidad (medido en  $d = -dB$ , decibelios negativos), decrecía 4 unidades, y en caso contrario se incrementaba 4 unidades.

Este proceso se iba repitiendo hasta que se producía al primer reverso (i.e. la primera respuesta distinta a las anteriores). Entonces el test continuaba igual, pero con incrementos o decrementos de 2 unidades en vez de 4. (A esta estrategia se la llama estrategia 4-2). La intensidad del último estímulo percibido por el paciente se utilizaba como estimación del umbral en ese punto.

El SITA, aprovechando los estudios realizados durante los años 80 sobre campos visuales de pacientes sanos y de pacientes con glaucoma, modifica esta estrategia para hacer el test más rápido. En [154], Olsson y Rootzen obtuvieron las distribuciones a posteriori de la variable umbral en cada punto, dadas las respuestas de los pacientes incluidas en las bases de datos. El pico de cada distribución proporciona el estimador máximo a posteriori (MAP) del umbral en dicho punto, y esta estimación será mejor, cuanto menor sea la amplitud de la distribución. Al realizar un nuevo test, las nuevas respuestas (tanto positivas como negativas) se introducen en el modelo, y las distribuciones de probabilidad a posteriori en cada punto se recalculan (después de cada respuesta producida en cada punto, al testear las localizaciones vecinas, y al finalizar toda la prueba). A partir de estas nuevas distribuciones de probabilidad a posteriori se obtendrán nuevos estimadores MAP de los valores del umbral en cada punto. La influencia del modelo a posteriori de partida irá disminuyendo gradualmente, conforme se vaya introduciendo más y más datos en el modelo (ver [15] para una explicación más detallada).

El test empieza evaluando el valor del umbral en cuatro puntos primarios (fijos, siempre los mismos), y a partir de ellos calcula los niveles de intensidad iniciales para los puntos adyacentes. La última intensidad percibida en puntos vecinos, sirve para calcular los valores iniciales de los nuevos puntos, aún no testeados.

La estimación MAP permite la estimación de los errores de medida en

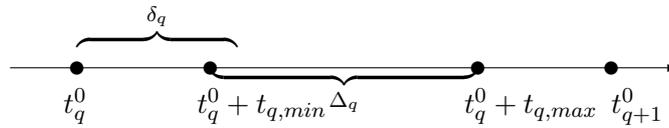
tiempo real. El test se puede detener en aquellos puntos en los que la estimación del error cometido es pequeño en comparación con un nivel de precisión pre-establecido, llamado ERF (error related factor). Acabaremos el test, si hemos conseguido en ese punto al menos una respuesta positiva, y el error cometido es menor que el ERF. En caso de que el error sea mayor que el definido por el ERF al finalizar una escalera, se realizará otra escalera completa con dos reversos, siguiendo la estrategia del "Full Treshold Strategy".

Además de determinar los umbrales, el SITA proporciona al oftalmólogo tres índices que le permitirán valorar la fiabilidad de los resultados obtenidos. Estos índices son:

- Índice de pérdida de fijación: A lo largo del test, el campímetro presenta estímulos en la mancha ciega. Si el paciente responde, se registra como fijación perdida. Si el valor de fijaciones perdidas es superior al veinte por ciento, hay razones para preocuparse sobre la fiabilidad de los resultados que se están obteniendo, y el símbolo XX aparece a continuación de tal valor, para llamar la atención del explorador.
- Error falso positivo: En las estrategias de test anteriores (FASTPAC y STATPAC), para detectar lo que definieron como respuestas falso positivas, el proyector maniobraba como si presentara un estímulo, pero no lo hacía. Si el paciente respondía, se registraba un error falso positivo. Un número alto de falsos positivos vuelve a indicar poca fiabilidad en los resultados obtenidos. Una de las principales "mejoras" que aporta el *SITA* frente a paquetes anteriores, es que reduce el tiempo de realización del test, y para ello utiliza una técnica distinta para calcular el error falso positivo.

Divide en tres partes el tiempo que transcurre entre dos exposiciones consecutivas de estímulos: una primera parte desde el momento de la

exposición, hasta que ha transcurrido el tiempo necesario para que pueda ser percibido ( $[t_q^0, t_q^0 + t_{q,min}]$ ); una segunda parte que llama tiempo (o ventana) de escucha y es el periodo durante el cual esperamos que la respuesta se produzca  $[t_q^0 + t_{q,min}, t_q^0 + t_{q,max}]$ ; y una tercera parte que va desde que acaba el tiempo razonable de espera, hasta que se expone un nuevo estímulo. ( $[t_q^0 + t_{q,max}, t_{q+1}^0]$ )



Para modelizar el número de respuestas falso positivas producidas, se utiliza una distribución Poisson que tiene en cuenta en su definición el número de respuestas obtenidas fuera del intervalo de escucha, y la longitud de este intervalo.

- Error falso negativo: Un estímulo mucho más intenso que el umbral, se presenta en un área donde la sensibilidad ha sido ya determinada. Si el paciente no responde, se registra un falso positivo. Un número alto de falsos positivos indica fatiga o falta de atención del paciente.

Para calcular el coste del algoritmo (tiempo) se almacenan todos los tiempos de respuesta de los pacientes. Sus medias y desviaciones típicas nos servirán para ajustar repetidamente la duración de las ventanas de respuesta. (Conforme el test vaya discurriendo, tendremos ventanas más robustas, y los ajustes serán menos frecuentes).

## 2.4. Resultados de la perimetría

Al finalizar cada test el campímetro presenta al oftalmólogo los índices de fiabilidad: fijaciones perdidas; número de errores falsos negativos, y errores

falsos positivos. Si estos índices sugieren que los resultados no son fiables, el test no servirá para diagnosticar y la prueba deberá repetirse.

Junto con los índices de fiabilidad presenta también los datos del paciente, el ojo que se está explorando, el tiempo consumido en su realización y el número de estímulos totales mostrados al paciente. En la figura 2.3 podemos ver una salida completa del campímetro tras la realización de un test perimétrico. Los resultados directos del test se imprimen en dos formatos gráficos:

1. Representación numérica (figura 2.3 a): Muestra el umbral de sensibilidad en todos los puntos explorados. Los puntos en los que se encuentra una sensibilidad de al menos 5 db. menor de lo esperado, son explorados una segunda vez y el valor hallado se representa entre paréntesis. Este doble “mapeo” permite verificar contestaciones inesperadas y diferenciar entre errores del paciente y defectos del campo verdaderos. Generalmente el segundo valor puede ser considerado más exacto que el primero. Este será el tipo de datos con el que trabajaremos en esta tesis.
2. Representación gráfica en escala de grises (figura 2.3 b): Da una idea inmediata del tamaño y seriedad de los defectos que pueden existir en el campo visual.

Por último se presentan una serie de índices y gráficos estadísticos basados en comparar los resultados que se obtienen en el test con los obtenidos en estudios empíricos realizados sobre un extenso grupo de sujetos normales. La comparación con los resultados de este estudio permite al campímetro analizar la “normalidad” (ausencia de enfermedad), en cada uno de los puntos testeados. Tenemos entonces en la salida de la perimetría cuatro gráficos que realizan este análisis “punto a punto”, y dos índices numéricos globales.

Los **índices globales** son:

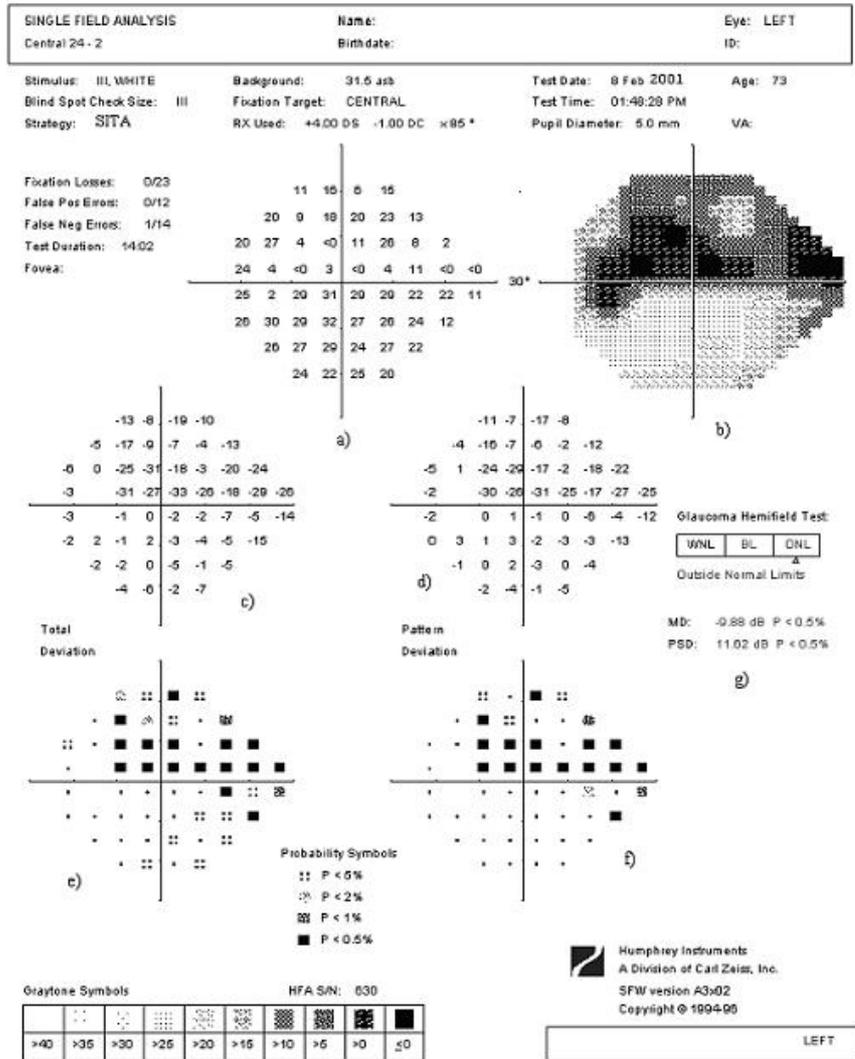


Figura 2.3: Resultado de una campimetría. Además de los datos del paciente y los índices de fiabilidad incluye: a) gráfico de representación numérica de los resultados; b) representación en escala de grises; c) y e) diagramas de desviación total; d) y f) diagramas de patrón de desviación, y g) índices globales.

1. Desviación media (DM). Desviación media ponderada del campo de referencia normal.

$$DM = \left\{ \frac{1}{n} \sum_{i=1}^n \frac{(x_i - N_i)}{s_{1i}^2} \right\} / \left\{ \frac{1}{n} \sum_{i=1}^n \frac{1}{s_{1i}^2} \right\}$$

siendo:

$x_i$  y  $N_i$  el umbral medido, y el umbral del campo de referencia normal en el punto  $i$ .

$s_{1i}^2$  la varianza de las medidas del campo normal en el punto  $i$ .

$n$  el número total de puntos del test (quitando la mancha ciega).

Una disminución en el valor de este índice indicará que los defectos del campo visual van en aumento, o de forma más general, que hay una pérdida generalizada de sensibilidad.

2. Desviación estándar del modelo (DSM). Desviación estándar ponderada de las diferencias puntuales entre el campo medido y el campo de referencia central.

$$DSM^2 = \left\{ \frac{1}{n} \sum_{i=1}^n s_{1i}^2 \right\} * \left\{ \frac{1}{n-1} \sum_{i=1}^n n \frac{(x_i - N_i - DM)^2}{s_{1i}^2} \right\}$$

Es una medida del grado en que la forma del campo del paciente se desvía de lo normal, para su edad. Una DSM baja, indica una colina de la visión uniforme. Una DSM alta indica una colina irregular que será debida a una variabilidad en las respuestas del paciente, o a irregularidades en el campo visual.

Límites de predicción de normalidad, han sido calculados para estos índices visuales. Si obtenemos algún valor fuera de estos límites, el ordenador nos proporciona también el nivel de significatividad de la desviación.

Veamos ahora los **diagramas de desviación total**. El primero de estos diagramas (figura 2.3 c) consta de valores numéricos y expresa la diferencia

en decibelios, entre los resultados del test del paciente y lo que corresponde a una persona normal, de su misma edad (i.e.  $x_i - N_i$ ). El segundo de los diagramas (figura 2.3 e) utiliza símbolos para indicar la probabilidad que existe de que esos resultados estén dentro de la normalidad.

Por último, los **diagramas de patrón de desviación** (figura 2.3 d y 2.3 f), son similares a los esquemas anteriores, pero plasmando posibles cambios del campo en conjunto, causados por ejemplo, por cataratas o pupilas pequeñas (i.e.  $x_i - N_i - DM$ ).

Obviamente, si tras examinar a un paciente debe decidirse si su resultado es normal o si no lo es, lo primero que debemos determinar es qué se entiende por un resultado normal. Por ello, se han realizado estudios empíricos sobre un numeroso grupo de personas sanas, y sobre otro grupo de enfermos de glaucoma. A partir de ellos se han construido dos bases de datos, y dos distribuciones empíricas que permiten estudiar la distribución de los umbrales a lo largo de todo el campo visual y el establecimiento de límites de confianza para los valores esperados en cada punto.

## 2.5. Características generales de los campos visuales de diferentes poblaciones

Para construir modelos fiables que informaran sobre si los valores umbral encontrados en una determinada posición son propios de personas sanas o de personas enfermas, ha sido necesario estudiar de las propiedades típicas de los campos visuales tanto de personas normales como de personas con glaucoma. Estos estudios han permitido concluir que:

1. Los valores esperados para los umbrales, decrecen continua y linealmente con la edad, y este decrecimiento no es homogéneo en todo el

## 2.5. Características generales de los campos visuales de diferentes poblaciones 15

campo visual, sino que es más rápido en la zona “midperipheral” que en la central [94].

2. Los resultados de las primeras exploraciones perimétricas a las que se somete un paciente pueden no ser muy fiables, debido a la existencia de un efecto aprendizaje. Los resultados de las primeras campimetrías suelen mejorarse tras la tercera o cuarta exploración, y no porque la visión del paciente haya mejorado, sino porque el paciente ya conoce el funcionamiento del test [95]. Por ello a la hora de analizar campos visuales se sugiere eliminar de la base de datos los dos primeros tests realizados por cada paciente.
3. No hay influencia de la edad en la variabilidad de los umbrales [153].
4. La sensibilidad decrece en los pacientes con glaucoma, por lo que se necesitará un estímulo más intenso para lograr la misma probabilidad de una respuesta afirmativa.
5. Los defectos en el campo visual debidos a la existencia de glaucoma se van desarrollando siguiendo las fibras nerviosas (del nervio óptico y en general de la retina) [153].



Parte II

Geoestadística



## Capítulo 3

# Metodología Geoestadística

### Contenidos

---

<b>3.1. Introducción.</b>	<b>50</b>
<b>3.2. Modelizando los datos</b>	<b>53</b>
3.2.1. Sobre la estructura de la media	53
3.2.2. Sobre la estructura de la covarianza	57
<b>3.3. Kriging</b>	<b>67</b>
<b>3.4. Análisis de validación cruzada</b>	<b>76</b>
<b>3.5. Problemas que encontramos</b>	<b>78</b>
3.5.1. Geoestadística cuando no hay estacionariedad espacial	79

---

### 3.1. Introducción.

El término Geoestadística fue acuñado por Georges Matheron y sus colaboradores en la Escuela de Minas de Fontainebleau [142, 143] en Francia, para referirse a una metodología que empezaban a desarrollar, y que buscaba dar respuesta a problemas de predicción espacial surgidos en la industria minera [119]. Las ideas de este grupo de investigadores se desarrollaron durante mucho tiempo con una terminología y un estilo propios, y fueron los trabajos paralelos de Matérn [140, 141] y de Whittle [203], junto con los libros de Ripley [168] y de Cressie [38] quienes permitieron la difusión de estas técnicas y su integración con el resto de técnicas de la estadística espacial.

Todo estudio geoestadístico parte de un conjunto de observaciones  $\{z_i\}_{i \in \{1, \dots, n\}}$  tomadas sobre localizaciones (en principio espaciales)  $\{x_i\}_{i \in \{1, \dots, n\}}$ , contenidas en una región continua del espacio  $D \subset \mathfrak{R}^2$ , y considera las observaciones como una realización de un proceso estocástico  $\{Z(x) : x \in D\}$ , (el cual se considera a su vez una realización parcial del proceso estocástico  $\{Z(x) : x \in \mathfrak{R}^2\}$ ).

A menudo se considera a cada observación  $Z_i$  como el valor en la posición  $x_i$  de un proceso subyacente  $\{S(x) : x \in \mathfrak{R}^2\}$  (proceso que realmente se quiere estudiar y que no es directamente observable) afectado por un ruido. En este caso el modelo geoestadístico deberá caracterizar el proceso estocástico  $S(x)$  y definir también un modelo estadístico para las medidas  $Z = (Z(x_1), Z(x_2), \dots, Z(x_n))$  condicionado a  $\{S(x) : x \in \mathfrak{R}^2\}$ .

El modelo que clásicamente se ha utilizado en los estudios geoestadísticos se basa en considerar:

1. Que  $\{S(x) : x \in \mathfrak{R}^2\}$  es un proceso gaussiano, con media  $\mu$ , varianza  $\sigma^2$  y función de correlación  $\rho(x, x') = \text{Corr}(S(x), S(x'))$ .
2. Que condicionadas a  $\{S(x) : x \in \mathfrak{R}^2\}$ , las observaciones  $z_i$  son rea-

lizaciones de v.a.  $Z(x_i)$  gaussianas y mutuamente independientes, con media condicional  $E[Z(x_i) | S(\cdot)] = S(x_i)$ , y varianza condicional  $\tau^2$ .

Estas dos suposiciones equivalen a modelizar:

$$Z(x_i) = S(x_i) + \epsilon_i \quad : \quad i = 1, \dots, n$$

siendo  $\{S(x) : x \in \mathfrak{R}^2\}$  como en la suposición 1 anterior, y donde las  $\epsilon_i \sim N(0, \tau^2)$  *i.i.d.*

Si definimos  $Z = (Z(x_1), \dots, Z(x_n))$ , y  $S = (S(x_1), \dots, S(x_n))$ , estas suposiciones equivalen a modelizar  $S \sim N(\mu \cdot 1_n, \sigma^2 R)$  siendo  $1_n$  un vector de  $n$  unos, y  $R = (r_{i,j})$  con  $r_{i,j} = \rho(x_i, x_j)$ , y como consecuencia:

$$Z \sim N(\mu \cdot 1_n, \sigma^2 R + \tau^2 I) \quad \text{siendo } I \text{ la matriz identidad } n \times n.$$

Apartándose un poco de esta doble suposición de normalidad, en [50], Diggle, Tawn y Moyeed acuñan la frase “Geoestadística basada en el modelo” para referirse a una aproximación Geoestadística a problemas en los que  $S$  es un proceso gaussiano, pero en los que  $Z_i | S$  ya no es gaussiana.

Nosotros vamos, en principio, a centrarnos en estudiar la metodología de la Geoestadística tradicional. Para ello seguiremos suponiendo que las dos distribuciones son gaussianas (a menos que en algún momento digamos explícitamente lo contrario).

Los objetivos “clásicos” de la Geoestadística son: la estimación de los parámetros del modelo estocástico de los datos, y la predicción.

En cuanto a la predicción, puede interesarnos predecir  $Z$  en una posición  $x_0$  no observada, o predecir  $\{S(x)\}$ . Si el objetivo es predecir  $S(x)$ , la predicción se puede realizar en toda la región de interés,  $D$ , con la intención de hacer mapas de esta variable, o sólo en algún punto aislado  $x_0 \in D$ . En cualquier caso, las predicciones se realizan a partir de las observaciones  $Z(x_1), \dots, Z(x_n)$ .

En el capítulo 4 plantearemos una modelización espacio temporal, en la que nos interesará predecir el valor de la variable observada, en posiciones no monitorizadas.

La metodología geoestadística clásica se ha extendido al contexto espacio temporal, pudiendo encontrar actualmente trabajos geoestadísticos espacio temporales en diversas disciplinas científicas ([198, 90, 86, 20, 169, 77, 138, 36, 22] ...) e incluso libros dedicados íntegramente a la geoestadística espacio temporal [34].

La generalización de la Geoestadística al caso espacio temporal, supone ahora que las observaciones

$$\{Z(x, t) : x \in D, t \in T\}$$

con  $D \subset R^2, T \subset R$ , se han recogido en un dominio espacio temporal. Este proceso puede suponerse también como resultado de haber añadido ruido gaussiano a un proceso  $\{S(x_i, t_i)\}$ , subyacente, no observable.

Siguiendo con los estándares de la Geoestadística clásica, modelizaremos la distribución espacio temporal de las observaciones, como una distribución gaussiana. Esta distribución quedará perfectamente caracterizada al definir sus momentos de primer y segundo orden. En la sección 3.2 veremos cómo hacerlo.

Para realizar predicciones, la Geoestadística ha desarrollado algoritmos, que proporcionan el predictor lineal óptimo (lineal en las observaciones), y que reciben el nombre general de métodos (o algoritmos) kriging en reconocimiento del trabajo realizado en este campo por Danie Krige [119]. (Comentaremos alguno de estos algoritmos en la sección 3.3).

## 3.2. Modelizando los datos

Como acabamos de ver, la Geoestadística clásica trabaja con procesos gaussianos, que quedan perfectamente determinados por su esperanza y su función de covarianza.

En este apartado vamos a ver cómo construir el modelo, es decir cómo a partir de las observaciones  $Z(x_1, t_1), \dots, Z(x_n, t_n)$  determinar la estructura de la media y de la covarianza,

$$E(Z(x, t)) = \mu(x, t) = ? \quad ; \quad Cov(Z(x_1, t), Z(x_2, q)) = ?$$

y cómo estimar los parámetros del modelo.

### 3.2.1. Sobre la estructura de la media

El primer paso de todo estudio geoestadístico suele ser la realización de un análisis exploratorio de los datos. En este análisis se busca qué covariables resultan influyentes en las observaciones y pueden entrar a formar parte en una formulación paramétrica de la media.

Este análisis exploratorio deberá realizarse tanto en el dominio espacial como en el dominio temporal, ya que si los datos muestran comportamientos cíclicos, periódicos, o tendencias, estos rasgos deberán plasmarse en la formulación de la media a través de funciones periódicas.

A menudo la media del proceso se estimará por mínimos cuadrados, y será restada de las observaciones para seguir trabajando con los residuos, o se estimará al tiempo que se realicen las predicciones mediante el método “kriging universal” que comentaremos en la sección 3.3.

### Modelos aditivos

Al aplicar las técnicas de la Geoestadística, en los capítulos 4 y 6, nos encontraremos con un problema que puede ocurrir cuando buscamos la estruc-

tura de la media, y es que no podamos explicitar la media paramétricamente, por no disponer de suficiente información (suficientes covariables). Resolveremos esta situación recurriendo a una formulación semiparamétrica de la media. En particular, utilizaremos los modelos aditivos (AM), introducidos por Hastie y Tibshirani [91, 92].

Las técnicas de regresión no paramétricas son herramientas estadísticas muy populares, utilizadas tanto en análisis exploratorios como a la hora de construir modelos. La utilidad de estas técnicas empezó a probarse en problemas de regresión univariante, pero ha sido en el caso de regresión multivariante donde realmente presentan mayores ventajas frente a las técnicas de regresión puramente paramétricas.

Los modelos aditivos son una generalización del modelo de regresión lineal. Vamos a establecer una notación, y veremos en qué sentido se da esta generalización.

Supongamos que tenemos una variable respuesta  $Y$  y  $p$  variables explicativas  $(X_1, X_2, \dots, X_p)$ . La herramienta estándar para estudiar la dependencia entre la variable dependiente  $Y$ , y las independientes  $X_1, X_2, \dots, X_p$  es el modelo de regresión lineal (múltiple si  $p > 1$ ), donde:

$$Y = \alpha + X_1\beta_1 + \dots + X_p\beta_p + \epsilon$$

con  $E(\epsilon) = 0$  y  $var(\epsilon) = \sigma^2$ . Este modelo está haciendo una suposición muy fuerte sobre cómo es la dependencia de  $E(Y)$  sobre cada una de las variables, ya que estamos afirmando que ésta dependencia es lineal en cada una de las variables predictoras.

Una de las principales ventajas del modelo lineal es que, una vez ajustado, podemos examinar los efectos de cada variable predictiva sobre la respuesta de forma separada, con ausencia de interacciones. Los modelos aditivos van a retener esta propiedad, ya que se van a caracterizar por ser aditivos en los efectos de los predictores.

Un modelo aditivo se define como:

$$Y = \alpha + \sum_{j=1}^p f_j(X_j) + \epsilon \quad (3.1)$$

donde al igual que en el modelo lineal, los errores  $\epsilon$  son independientes de los  $X_j$ ,  $E(\epsilon) = 0$  y  $var(\epsilon) = \sigma^2$ . En este modelo, las  $f_j$  son funciones normalmente univariantes, una para cada predictor. (En el caso multivariante se supone además que los  $\epsilon_i$  son independientes entre sí).

Destacar que este modelo retiene uno de los rasgos interpretativos más importantes del modelo lineal: la variación de la superficie respuesta fijando los valores de todos los predictores excepto de uno sólo dependerá de los valores del predictor que queda por fijar, y en ningún caso dependerá de los valores que hayan tomado los otros. (Para estudiar más detalladamente tanto los modelos aditivos como los modelos aditivos generalizados consultar [91, 92].)

Para no imponer restricciones demasiado severas en la formulación del modelo, en la ecuación 3.1 las funciones  $f_j(\cdot)$  suelen ser funciones suaves de las covariables. La ventaja de utilizar estos modelos es que no estamos imponiendo ninguna estructura paramétrica rígida para relacionar las covariables con la variable respuesta, sino que se deja a los datos que den forma a esta relación.

El método más general para estimar modelos aditivos, consiste en estimar cada una de las funciones  $f_j$  mediante un suavizador (“smoother”) arbitrario. Los suavizadores más comunes son los “smoothing splines” y los “kernel smoothers”. Existen varias aproximaciones para estimar estas superficies suaves. Las dos más utilizadas son la estimación por el método “backfitting” y técnicas basadas en la integración.

El método backfitting se presentó en [24]. Es un algoritmo iterativo que permite ajustar un modelo aditivo utilizando mecanismos de ajuste similares

a los de regresión. Su motivación es la siguiente: si el modelo 3.1 es correcto, entonces

$$E[Y - \alpha - \sum_{j \neq k} f_j(X_j) \mid X_k] = f_k(X_k).$$

A partir de aquí podemos calcular iterativamente todas las funciones  $f_j$  a partir de los datos y del suavizador  $S_j$  adecuado. El algoritmo iterativo es el siguiente:

1. Inicializar:  $\alpha = \bar{Y}$ ,  $f_j = f_j^0 \forall j$
2. Iterar: Para  $j = 1, \dots, p, 1, \dots, p, \dots$

$$f_j = S_j(y - \alpha - \sum_{k \neq j} f_k \mid x_j),$$

(siendo  $S_j(y \mid x_j)$  un suavizamiento de la respuesta  $y$  en función del predictor  $x_j$ . Por tanto, al reajustar  $f_j$  quitamos de  $y$  los efectos de todas las otras variables, y ajustamos la función resultante al suavizar los residuos respecto a la variable predictiva  $x_j$ ).

3. Repetir el paso 2 hasta que las funciones  $f_j$  no cambien.

Aunque es de uso dominante, las propiedades asintóticas de las estimaciones obtenidas por el método backfitting no se empezaron a desarrollar hasta muy recientemente [156, 115]. De hecho, aunque en algunas ocasiones ha sido utilizado con buenos resultados [151], las propiedades asintóticas para datos relacionados no han sido todavía desarrolladas. En [32] se recomienda su uso sólo en caso de variables no muy fuertemente correlacionadas. Para compensar esta deficiencia, Linton y Nielsen [129] propusieron un método de ajuste no iterativo basado en la integración marginal de estimaciones de kernel estándar para modelos aditivos. Este método fue extendido unos años después para modelos aditivos generalizados [128]. Para ver un estudio comparativo entre los dos métodos recomendamos consultar [188] y [148].

Hay estudios previos en los que se han combinado las nociones de la Geoestadística con las de la modelización aditiva, como por ejemplo el trabajo de Kelsall y Diggle [116].

### 3.2.2. Sobre la estructura de la covarianza

Antes de ver cómo construir una función de covarianzas, necesitamos ver una serie de definiciones y caracterizaciones que nos facilitarán el trabajo. Diremos que:

**Definición 2.1** *Un campo aleatorio  $Z(x, t)$  es estrictamente estacionario si su distribución de probabilidad es invariante ante traslaciones, es decir si dados dos vectores cualesquiera  $(Z(x_1, t_1), Z(x_2, t_2), \dots, Z(x_n, t_n))$  y  $(Z(x_1 + h, t_1 + u), Z(x_2 + h, t_2 + u), \dots, Z(x_n + h, t_n + u))$  tienen la misma función de distribución multivariante.*

Una condición menos exigente que la estacionariedad estricta, es la estacionariedad de segundo orden.

**Definición 2.2** *Un proceso aleatorio espacio-temporal  $Z(x, t)$  es estacionario de segundo orden, o estacionario en sentido amplio si se cumple que:*

- $E(Z(x, t)) = \mu(x, t) = \mu \quad (\text{cte}) \quad \forall (x, t) \in \mathbb{R}^2 \times \mathbb{R}$  y además:
- $\text{cov}(Z(x, t), Z(y, q)) = C(x - y; t - q) \quad \forall (x, t), (y, q) \in \mathbb{R}^2 \times \mathbb{R}$

*Si además  $C(x - y; t - q)$  es función simplemente de la distancia entre las posiciones, i.e. de  $(\|x - y\|; |t - q|)$  diremos que además de estacionario el proceso es isotrópico en el espacio y en el tiempo.*

**Definición 2.3** *Diremos que un proceso estocástico es ergódico si la dependencia entre observaciones tiende a 0 al aumentar la separación entre ellas (tanto en el sentido espacial como temporal).*

La ergodicidad es una propiedad necesaria para poder estimar las características de un proceso a partir de una única realización ya que si el proceso no es ergódico, al aumentar el tamaño muestral no se adquiere información adicional por ser todas las observaciones muy dependientes entre sí. En adelante supondremos siempre procesos estacionarios ergódicos.

Otra propiedad muy importante de las funciones de covarianza es la siguiente:

**Definición 2.4** *Se dice que una función  $C(-, -)$  es definida positiva si para todo  $(x_1, t_1), (x_2, t_2), \dots, (x_n, t_n) \in \mathfrak{R}^2 \times \mathfrak{R}$ ,  $\lambda_1, \dots, \lambda_n \in \mathfrak{R}$  y para todo entero positivo  $n$  satisface:*

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i - x_j; t_i - t_j) \geq 0$$

Si  $Z(\cdot, \cdot)$  es un campo aleatorio tal que  $Var(Z(x, t)) < \infty \forall (x, t) \in \mathfrak{R}^2 \times \mathfrak{R}$ , con el hecho de que

$$var\left(\sum_{i=1}^n \lambda_i Z(x_i, t_i)\right) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Cov(Z(x_i, t_i), Z(x_j, t_j))$$

es no negativa, se comprueba que la función de covarianza

$$Cov(Z(x_i, t_i), Z(x_j, t_j)) = E\{[Z(x_i, t_i) - E(Z(x_i, t_i))][Z(x_j, t_j) - E(Z(x_j, t_j))]\}$$

es una función definida positiva.

Luego la función de covarianza de un proceso estocástico estacionario de segundo orden es siempre definida positiva.

La pregunta que se plantea ahora es si cualquier función definida positiva  $C(\cdot, \cdot)$  se corresponde con la función de covarianzas de un proceso estocástico estacionario de segundo orden. La respuesta es que sí, y se puede deducir a partir del teorema de Bochner [19] (ver también [38], pág84).

Por lo tanto, la clase de funciones definidas positivas en  $\mathfrak{R}^2 \times \mathfrak{R}$  y la clase de funciones de covarianzas de procesos estacionarios en  $\mathfrak{R}^2 \times \mathfrak{R}$  son

idénticas, por lo que buscar funciones de covarianza equivaldrá a buscar funciones definidas positivas.

Para determinar la estructura de la covarianza en la práctica, garantizando que la función de covarianzas sea definida positiva, a menudo se escoge esta función entre una familia de funciones paramétricas que ya sabemos que son definidas positivas, es decir se asume que

$$\text{cov}(Z(x_i, t_i), Z(x_i + h, t_i + u)) = C^0(h, u)$$

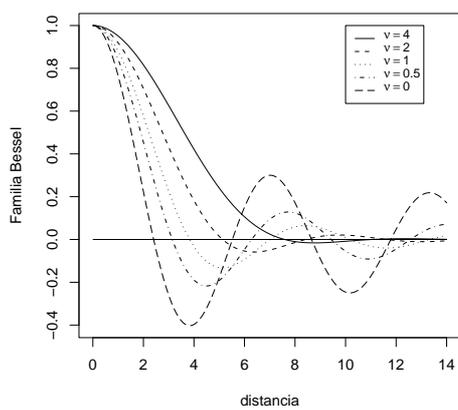
siendo  $C^0$  una función definida positiva de una determinada familia, y dependiendo de una serie de parámetros a determinar. Los modelos de covarianza puramente espacial, y los modelos de covarianza temporal han sido ampliamente estudiados [140, 38, 179], y se dispone de una amplia lista de modelos que a priori podemos utilizar para modelizar las funciones de covarianza, con la garantía de que la función de covarianza resultante será definida positiva. (Podemos ver algunos de ellos en la tabla 3.1 y los gráficos 3.1 y 3.2).

Si resulta complicado demostrar que una función espacial o temporal es definida positiva [179], la cosa se complica más a la hora de buscar modelos de covarianza espacio temporales válidos. Por ello, muchos autores empezaron a investigar la forma de combinar modelos espaciales y modelos temporales válidos para obtener modelos de covarianza espacio temporales, dando lugar a los llamados modelos de covarianza espacio temporal separables.

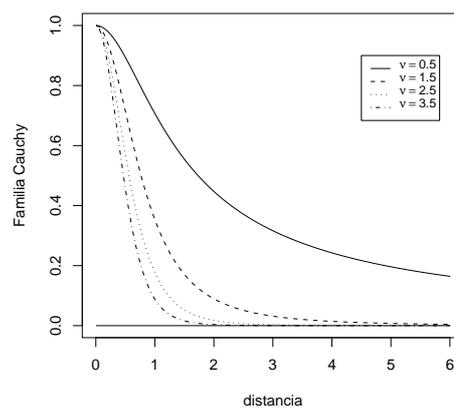
**Definición 2.5** *Una función de covarianza espacio-temporal  $C(h, u)$  (siendo  $h \in R$  la distancia espacial entre las posiciones, y  $u \in R$  la distancia temporal) es separable si se puede descomponer en una operación (suma o producto) entre una función de covarianza puramente espacial y de otra puramente temporal.*

Cuadro 3.1: Familias de funciones definidas positivas.

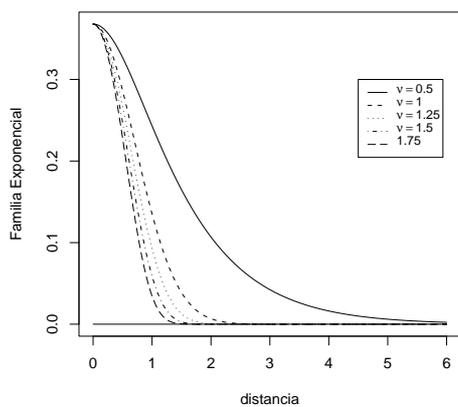
Familia de Bessel	$\rho_\nu(h) = 2^\nu \Gamma(\nu + 1) h^{-\nu} J_\nu(h)$ <p>Es definida positiva en <math>R^d</math> para <math>\nu \geq \frac{d-2}{2}</math></p>
Familia de Cauchy	$\rho_\nu(h) = (1 + h^2)^{-\nu}$ <p>que es definida positiva en <math>R^\infty</math></p>
Modelo exponencial	$\rho(h) = \exp\{-(h/\phi)^k\}$ <p>con <math>\phi &gt; 0</math>, y <math>0 &lt; k &lt; 2</math></p>
Coseno	$\rho(h) = \cos(h)$ <p>modelo válido únicamente en <math>R^1</math></p>
Modelo esférico	$\rho(h) = \begin{cases} 1 - \frac{3}{2}h + \frac{1}{2}h^3 & 0 \leq h \leq 1 \\ 0 & \text{resto} \end{cases}$ <p>Función definida positiva en <math>R^3</math></p>
Modelo gaussiano	$\rho(h) = e^{-h^2}$
Nugget effect	$\rho(h) = \begin{cases} 0 & \text{si } h = 0 \\ 1 & \text{en otro caso} \end{cases}$
Clase de Whittle-Matérn o de Basset	$\rho(h) = 2^{1-\nu} \Gamma(\nu)^{-1} h^\nu K_\nu(h) \text{ con } \nu > 0.$ <p>Funciones definidas positivas en <math>R^\infty</math>. Permiten especificar el grado de diferenciabilidad subyacente. Si <math>\nu = (2n + 1)/2</math> entonces <math>\rho(\cdot)</math> es <math>2n</math> veces diferenciable.</p>
Hole effect	$\rho(h) = h^{-1} \text{sen}(h)$



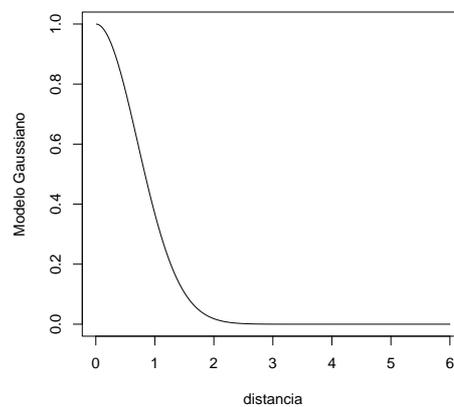
(a)



(b)



(c)



(d)

Figura 3.1: (a) Modelos de la familia Bessel, (b) Modelos de la familia Cauchy, (c) Modelos de la familia exponencial, (d) Modelo Gaussiano.

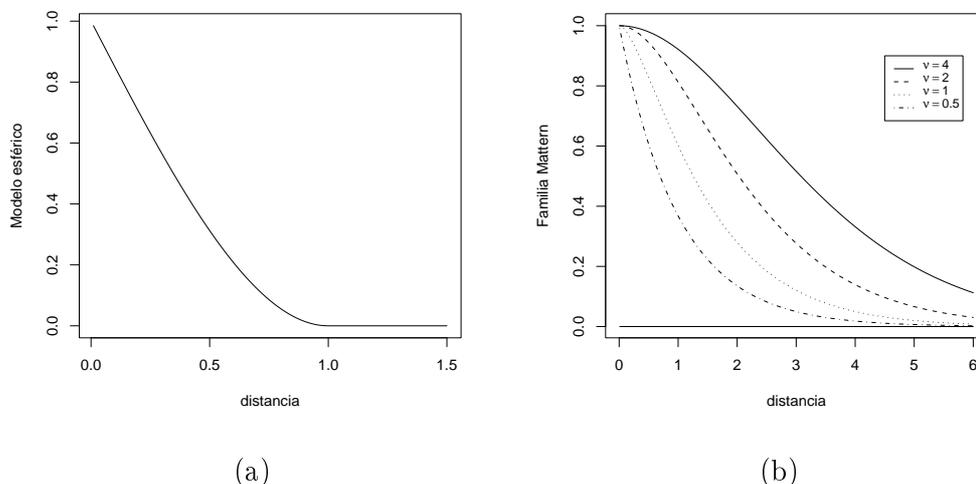


Figura 3.2: (a) Modelo esférico, (b) Familia Matérn.

Uno de los modelos de covarianza separables más utilizados en la literatura es el **modelo producto**, que define

$$C(h, u) = C^1(h) \cdot C^2(u)$$

siendo  $C^1$  y  $C^2$  funciones definidas positivas, ( $C^1$  actuando como función de covarianza de un modelo puramente espacial, y  $C^2$  actuando como la función de covarianza de un modelo puramente temporal). Este modelo ha sido utilizado en muchas aplicaciones prácticas, muchas veces atendiendo más a su simplicidad que a su conveniencia a la hora de ajustar los datos.

La definición (2.5) implica que dados dos retardos espaciales  $h_1$  y  $h_2$  cualesquiera:

$$\left. \begin{array}{l} C(h_1, u) = C^1(h_1) \cdot C^2(u) \\ C(h_2, u) = C^1(h_2) \cdot C^2(u) \end{array} \right\} \rightarrow C(h_1, u) \propto C(h_2, u) \quad \forall u \in R$$

Por lo tanto la estructura de correlación temporal es la misma para cualquier separación espacial, y análogamente podemos ver que también la correlación

espacial continua teniendo la misma forma para los diferentes retardos temporales.

Otro tipo de separabilidad consiste en descomponer la covarianza espacio-temporal como suma de una función de covarianza puramente espacial y de otra puramente temporal (Cressie et al. [40]). A este tipo de modelo separable se le conoce como **modelo lineal**

$$C(h, u) = C^1(h) + C^2(u)$$

pero en este caso la función de covarianza resultante no siempre será definida positiva (Rouhani et al [171]).

En [179] podemos encontrar criterios necesarios y suficientes para construir funciones definidas positivas; propiedades para combinarlas de forma que las funciones resultantes sigan siendo definidas positivas, y comentarios sobre los errores más comunes al trabajar con este tipo de funciones.

En este trabajo se ve por ejemplo bajo qué condiciones una combinación lineal finita de funciones definidas positivas con escalares no negativos,  $C = a_1C_1 + a_2C_2$ , da lugar a una función definida positiva. Comprueba que al multiplicar dos funciones de covarianza válidas siempre obtenemos una nueva función de covarianzas válida, y que un modelo  $\varphi$  válido en  $R^d$  es válido también en  $R^{d-1}$  (aunque el recíproco ya no sea cierto).

A pesar de la simplicidad de los modelos separables, estos resultan en ocasiones bastante limitados. Por ello en los últimos años la demanda de modelos de covarianza espacio-temporal no separables ha ido en aumento. En este campo cabe destacar varios trabajos. En primer lugar y siguiendo un orden cronológico, el de Jones y Zhang [106], quienes utilizan ecuaciones en derivadas parciales, para construir familias de densidades espectrales que implícitamente determinan, aunque no de forma cerrada, las funciones de covarianza buscadas. En un trabajo posterior, Cressie y Huang [39] desarrollan una metodología que ya no se basará en ecuaciones diferenciales, y que

resultará bastante más sencilla. Basándose en el teorema de Bochner [19], Cressie y Huang demostraron que si podemos expresar:

$$C(h; u) \equiv \int e^{ih'\omega} \rho(\omega; u) k(\omega) d\omega$$

siendo  $\rho(-, -)$  y  $k(-)$  dos funciones verificando que:

1.  $\forall \omega \in \mathfrak{R}^d$ ,  $\rho(\omega, -)$  es una función de autocorrelación,  
 $\int \rho(\omega; u) du < \infty$  y  $k(\omega) > 0$
2.  $\int k(\omega) d\omega < \infty$

se cumple que  $C$  es una función de covarianzas espacio temporal, estacionaria, continua y válida.

Funciones de covarianza así generadas serán, en general, funciones de covarianza no separables, aunque algunas funciones de covarianza separables podrán ser formuladas como casos particulares de estas.

En un trabajo posterior, Gneiting [66] da una caracterización aún más general para los modelos de covarianza espacio temporal no separables. Antes de verla necesitamos una definición.

**Definición 2.6** Una función continua  $\varphi(t)$  definida para  $t > 0$  o  $t \geq 0$  se dice que es completamente monótona si posee derivadas de todos los órdenes  $\varphi^{(n)}$  y además:

$$(-1)^n \varphi^{(n)}(t) \geq 0$$

Gneiting demuestra que dada  $\varphi(t)$ ,  $t > 0$  cualquier función completamente monótona,  $\psi(t)$ ,  $t > 0$  cualquier función positiva con derivada completamente monótona, y  $\sigma^2 > 0$ , entonces:

$$C(h, u) = \frac{\sigma^2}{\psi(\|u\|^2)^{\frac{d}{2}}} \varphi\left(\frac{\|h\|^2}{\psi(\|u\|^2)}\right), \quad (h, u) \in \mathfrak{R}^d \times \mathfrak{R}$$

es un modelo de covarianza espacio-temporal válido.

Pensando que estos modelos en general son no separables, Gneiting propone el asociar  $\varphi(t)$  y  $\psi(t)$  con la estructura espacial y temporal de los datos, respectivamente.

Existen más trabajos en los que se plantean (o derivan a partir de los ya existentes), nuevos modelos de covarianza espacio temporal, demostrándose también su validez. Entre estos trabajos destacamos los trabajos de Christakos y Hristopulos [35], de De Cesare, Myers y Posa [27, 28], y de Ma [135, 134, 133], pero nosotros nos quedaremos con los dos trabajos que acabamos de comentar, el de Cressie y Huang [39], y el de Gneiting [66]. Entre los dos trabajos nos van a permitir definir una gran variedad de funciones de covarianza espacio temporal válidas, y nos basaremos en ellos en el capítulo 4 para determinar la estructura de covarianza en la modelización de campos visuales.

Una vez tenemos claro que podemos optar por una gran variedad de modelos para la estructura de covarianza, ante un problema particular debemos ver cómo quedarnos con uno u otro.

A menudo, en vez de trabajar directamente con la función de covarianzas, en Geoestadística se trabaja con el variograma, ya que resulta más fácil de manejar.

**Definición 2.7** *Dado un proceso estocástico  $Z(x, t)$  se define el variograma como:*

$$\gamma(h, u) = \frac{1}{2} \text{Var}[\{Z(x, t) - Z(x + h, t + u)\}], \quad h > 0$$

Como

$$\begin{aligned} \text{Var}[\{Z(x, t) - Z(x + h, t + u)\}] &= \text{Var}[Z(x, t)] + \text{Var}[Z(x + h, t + u)] - \\ &\quad - 2\text{Cov}[Z(x, t), Z(x + h, t + u)] \end{aligned}$$

si el proceso es estacionario, el variograma se simplifica a:

$$\gamma(h, u) = C(0, 0) - C(h, u) \quad (3.2)$$

En [67], Gneiting et al. analizan la equivalencia de distintos resultados teóricos al trabajar con variogramas o con funciones de covarianza.

Los variogramas son funciones condicionalmente definidas negativas, es decir:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i - x_j; t_i - t_j) \leq 0$$

$$\forall (x_1, t_1), (x_2, t_2), \dots, (x_n, t_n) \in \mathfrak{R}^d \times \mathfrak{R} \quad \text{y} \quad \forall \lambda_1, \dots, \lambda_n \in \mathfrak{R} : \sum_{i=1}^n \lambda_i = 0$$

y al trabajar con procesos estacionarios, como  $E[Z(x, t)]$  es constante, podemos formular el variograma como:

$$\gamma(h, u) = \frac{1}{2} E[\{Z(x, t) - Z(x + h, t + u)\}^2]$$

Un estimador no paramétrico del variograma es el variograma empírico, definido como:

$$\hat{\gamma}(h, u) = \frac{1}{2 |N(h, u)|} \sum_{N(h, u)} \{Z((x_i, t_i) - Z(x_j, t_j))\}^2 \quad (3.3)$$

con  $N(h, u) = \{(x_i, t_i), (x_j, t_j) : \|x_i - x_j\| = h \text{ y } |t_i - t_j| = u\}$ , y siendo  $|N(h, u)|$  el cardinal de  $N(h, u)$ .

Una vez calculado el variograma empírico, podemos dibujar la nube de puntos de los valores obtenidos, en función de  $h$  y de  $u$ . Como  $\hat{\gamma}(h, u) \approx C(0, 0) - C(h, u)$ , podemos buscar modelos paramétricos para la función de covarianza y estimar sus parámetros.

Para la estimación de los parámetros del modelo podemos utilizar cualquiera de los métodos tradicionales de estimación. Cressie ([39]) recomienda utilizar

el método de mínimos cuadrados ponderados, y minimizar:

$$W(\theta) = \sum_h \sum_u |N(h, u)| \left\{ \frac{\hat{\gamma}(h, u)}{\gamma(h, u/\theta)} - 1 \right\}^2$$

sobre todos los posibles valores de los parámetros,  $\theta$ . De todas formas este método no es único, y cualquier otro método de estimación de parámetros puede ser utilizado.

### 3.3. Kriging

Dentro del contexto de la Geoestadística, se conoce con el nombre de kriging a una familia de algoritmos de regresión por mínimos cuadrados generalizados que, a partir de un conjunto de observaciones  $Z(x_i, t_i)_{i \in \{1, \dots, n\}}$ , proporcionan el predictor lineal óptimo para la variable  $Z$  en una nueva posición  $(x_0, t_0)$ . Estos algoritmos reciben el nombre de kriging en reconocimiento del trabajo realizado en este campo por Danie Krige [119].

Existen varios métodos de kriging, pero todos comparten la misma filosofía, minimizar la varianza del error de la estimación, con la restricción de que el estimador obtenido  $Z^*$  sea insesgado, es decir minimizar:

$$\sigma_E^2(x_0, t_0) = \text{Var}\{Z^*(x_0, t_0) - Z(x_0, t_0)\}$$

con la restricción:  $E(Z^*(x_0, t_0) - Z(x_0, t_0)) = 0$ .

Para poder obtener este predictor óptimo, el proceso  $Z(x, t)$  deberá satisfacer la condición de regularidad:

$$\text{Var}(Z(x, t)) < \infty \quad \forall (x, t) \in \mathbb{R}^2 \times \mathbb{R}, \quad t \geq 0$$

ya que en este caso tanto la función media  $\mu(x, t) = E(Z(x, t))$  como la covarianza entre dos observaciones cualesquiera

$$K(x, y; t, q) \equiv \text{cov}(Z(x, t), Z(y, q)); \quad x, y, \in \mathbb{R}^2, \quad t > 0, \quad q > 0$$

existen y están bien definidas.

En adelante llamaremos  $Z^*(x_0, t_0)$  al predictor kriging de  $Z(x_0, t_0)$ , que será aquella combinación lineal de las observaciones

$$Z^*(x_0, t_0) = \mu(x_0, t_0) + \sum_{i=1}^n a_i (Z(x_i, t_i) - \mu(x_i, t_i)) \quad (3.4)$$

que minimice el error de predicción cuadrático medio. Para que el estimador sea insesgado deberemos exigir que  $\sum_{i=1}^n a_i = 1$ .

Los métodos kriging han sido ampliamente estudiados y aplicados en el contexto de la estadística espacial [38], y espacio-temporal [39].

Existen varios tipos de kriging, y dependiendo de las suposiciones que hagamos sobre la media  $\mu(x, t)$  de  $Z(x, t)$ , trabajaremos con una variedad o con otra.

### Distintas variantes de kriging

Dependiendo de las suposiciones que hagamos sobre la tendencia  $\mu(x, t)$ , vamos a distinguir tres variantes de kriging:

1. **Kriging Simple (SK)** Considera la media conocida y constante a través de todo el dominio espacio temporal. Es decir:  $\mu(x, t) = \mu \forall (x, t) \in \mathfrak{R}^2 \times \mathfrak{R}$ . En este caso el estimador (3.4) será de la forma:

$$Z^*(x_0, t_0) = \sum_{i=1}^n a_i Z(x_i, t_i) + \left(1 - \sum_{i=1}^n a_i\right) \mu$$

Este estimador siempre será insesgado, por lo tanto no necesitaremos imponer la restricción de que  $\sum_{i=1}^n a_i = 1$ , y la expresión a minimizar

será:

$$\begin{aligned}
\sigma_E^2(x_0, t_0) &= \text{Var}\{Z^*(x_0, t_0) - Z(x_0, t_0)\} = \\
&= \text{Var}\left(\mu + \sum_{i=1}^n a_i(Z(x_i, t_i) - \mu) - Z(x_0, t_0)\right) = \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}\left(Z(x_i, t_i) - \mu, Z(x_j, t_j) - \mu\right) - \\
&\quad - 2 \sum_{i=1}^n a_i \text{Cov}\left(Z(x_i, t_i) - \mu, Z(x_0, t_0) - \mu\right) + \\
&\quad + \text{Var}\left(Z(x_0, t_0) - \mu\right)
\end{aligned} \tag{3.5}$$

Si definimos  $C_{i,j} := \text{Cov}(Z(x_i, t_i), Z(x_j, t_j))$ , la expresión a minimizar será:

$$\sigma_E^2(x_0, t_0) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j C_{i,j} - 2 \sum_{i=1}^n a_i C_{i,0} + \text{Var}\left(Z(x_0, t_0)\right).$$

Calculando las derivadas parciales de esta expresión respecto a  $a_i$ , e igualando a 0:

$$\sum_{j=1}^n a_j C_{i,j} - C_{0,i} = 0 \quad \forall i = 1, \dots, n \tag{3.6}$$

y si llamamos:  $\theta = (a_1, a_2, \dots, a_n)'$ ,  $C_0 = (C_{1,0}, C_{2,0}, \dots, C_{n,0})'$ , y

$$\Sigma = \begin{pmatrix} \text{Var}(Z(x_1, t_1)) & C_{1,2} & \cdots & C_{1,n} \\ C_{2,1} & \text{Var}(Z(x_2, t_2)) & \cdots & C_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ C_{x_n,1} & C_{x_n,2} & \cdots & \text{Var}(Z(x_n, t_n)) \end{pmatrix}$$

la solución del sistema anterior será:

$$\theta = \Sigma^{-1} C_0 \tag{3.7}$$

Una vez tenemos ya los valores de los pesos  $a_i$ , ya estamos en condiciones de realizar la predicción.

Otra cantidad que puede resultar de interés es la varianza del error de predicción. Para calcularla, sustituimos (3.6) en (3.5), con lo que:

$$\begin{aligned}\sigma_E^2(x_0, t_0) &= \sum_{i=1}^n a_i C_{0,i} - 2 \sum_{i=1}^n a_i C_{0,i} + C_{0,0} \\ \sigma_E^2(x_0, t_0) &= C_{0,0} - \sum_{i=1}^n a_i C_{0,i}\end{aligned}$$

2. **Kriging Ordinario (OK)** Esta variante de kriging se utiliza cuando desconocemos el valor de la media, sabemos que no es constante en todo el área de estudio, pero podemos considerarla constante localmente, i.e. considerarla constante ( $\mu = \mu_0$ ) en una vecindad centrada en el punto en el que queremos predecir. El dominio de estacionariedad de la media se restringe en este caso a una vecindad local, centrada en la localización  $(x_0, t_0)$  en la que queremos predecir.

Sea  $N(x_0, t_0)$  el conjunto de puntos vecinos a  $(x_0, t_0)$  en el que consideramos la media constante. Si suponemos que  $N(x_0, t_0) = \{(x_i, t_i) : n_0 \leq i \leq n_1\}$ , el predictor kriging (ec. 3.4) quedará:

$$\begin{aligned}Z^*(x_0, t_0) &= \mu_0 + \sum_{i=n_0}^{n_1} a_i (Z(x_i, t_i) - \mu_0) = \\ &= \sum_{i=n_0}^{n_1} a_i Z(x_i, t_i) + (1 - \sum_{i=n_0}^{n_1} a_i) \mu_0\end{aligned}$$

Para que el estimador sea insesgado exigiremos que  $\sum_{i=n_0}^{n_1} a_i = 1$ , y entonces el estimador kriging quedará:

$$Z^*(x_0, t_0) = \sum_{i=n_0}^{n_1} a_i Z(x_i, t_i)$$

Por tanto, si el único objetivo es la predicción, no necesitaremos estimar la media del proceso.

Definimos un multiplicador de Lagrange,  $m$ , para garantizar la insesgadez del estimador. Para minimizar el ECM de la predicción, mimizaremos:

$$\begin{aligned}
& \sigma_E^2(x_0, t_0) - 2m \left( \sum_{i=n_0}^{n_1} a_i - 1 \right) = \\
& = Var\{Z^*(x_0, t_0) - Z(x_0, t_0)\} - 2m \left( \sum_{i=n_0}^{n_1} a_i - 1 \right) = \\
& = Var \left( \sum_{i=n_0}^{n_1} a_i Z(x_i, t_i) - Z(x_0, t_0) \right) - 2m \left( \sum_{i=n_0}^{n_1} a_i - 1 \right) = \\
& = \sum_{i=n_0}^{n_1} \sum_{j=n_0}^{n_1} a_i a_j Cov \left( Z(x_i, t_i) - \mu_0, Z(x_j, t_j) - \mu_0 \right) - \\
& - 2 \sum_{i=n_0}^{n_1} a_i Cov \left( Z(x_i, t_i) - \mu_0, Z(x_0, t_0) - \mu_0 \right) + \\
& + Var \left( Z(x_0, t_0) - \mu_0 \right) - 2m \left( \sum_{i=n_0}^{n_1} a_i - 1 \right)
\end{aligned}$$

Si llamamos  $C_{i,j} := Cov(Z(x_i, t_i), Z(x_j, t_j))$ , la ecuación a minimizar quedará:

$$\sum_{i=n_0}^{n_1} \sum_{j=n_0}^{n_1} a_i a_j C_{i,j} - 2 \sum_{i=n_0}^{n_1} a_i C_{i,0} + Var \left( Z(x_0, t_0) - \mu_0 \right) - 2m \left( \sum_{i=n_0}^{n_1} a_i - 1 \right)$$

Calculando las derivadas parciales respecto a  $a_i$  y a  $m$  e igualando a 0, obtendremos el siguiente sistema de ecuaciones:

$$\begin{cases} \sum_{j=n_0}^{n_1} a_j C_{i,j} - C_{0,i} - m = 0 & \forall i = n_0, \dots, n_1 \\ \sum_{i=n_0}^{n_1} a_i = 1 \end{cases} \quad (3.8)$$

Si definimos  $\theta = (a_{n_0}, \dots, a_{n_1}, -m)'$ ;  $C_0 = (C_{n_0,0}, \dots, C_{n_1,0}, 1)'$  y

$$\Sigma = \begin{pmatrix} \text{Var}(Z(x_{n_0}, t_{n_0})) & C_{n_0, n_0+1} & \cdots & C_{n_0, n_1} & 1 \\ C_{n_0+1, n_0} & \text{Var}(Z(x_{n_0+1}, t_{n_0+1})) & \cdots & C_{n_0+1, n_1} & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ C_{n_1, n_0} & C_{n_1, n_0+1} & \cdots & \text{Var}(Z(x_{n_1}, t_{n_1})) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}$$

podremos expresar el sistema (3.8) matricialmente como:

$$\begin{aligned} \Sigma \theta &= C_0 \quad \text{es decir:} \\ \theta &= \Sigma^{-1} C_0 \end{aligned} \quad (3.9)$$

El sistema kriging tendrá solución única siempre que el sistema (3.9) sea compatible determinado. En este caso, la varianza del error de predicción será:

$$\begin{aligned} \sigma_E^2(x_0, t_0) &= \sum_{i=n_0}^{n_1} \sum_{j=n_0}^{n_1} a_i a_j \text{Cov} \left( Z(x_i, t_i), Z(x_j, t_j) \right) - \\ &- 2 \sum_{i=n_0}^{n_1} a_i \text{Cov} \left( Z(x_i, t_i), Z(x_0, t_0) \right) + \text{Var} \left( Z(x_0, t_0) \right) = \\ &= \sum_{i=n_0}^{n_1} a_i (C_{0,i} + m) - 2(C_{0,0} + m) + C_{0,0} \end{aligned}$$

$$\sigma_E^2(x_0, t_0) = \sum_{i=n_0}^{n_1} a_i C_{0,i} - C_{0,0} - m$$

En ocasiones, en lugar de predecir el valor de la variable que estamos estudiando en una nueva posición, nos puede interesar el estimar y hacer mapas de medias. Estamos suponiendo que la media es desconocida pero constante en la vecindad del punto de predicción. Para estimar la media  $\mu_0^*$ , de  $Z$  en el punto  $(x_0, t_0)$ , consideramos una combinación lineal de las variables aleatorias definidas en la vecindad de  $(x_0, t_0)$ ,

$N(x_0, t_0) := \{(x_{n_0}, t_{n_0}), (x_{n_0+1}, t_{n_0+1}), \dots, (x_{n_1}, t_{n_1})\} :$

$$\mu_0^* = \sum_{i=n_0}^{n_1} a_i Z(x_i, t_i)$$

Para garantizar la insesgadez del estimador, exigimos de nuevo que  $\sum_{i=n_0}^{n_1} a_i = 1$ . La varianza del error será ahora:

$$\sigma_E^2(x_0, t_0) = Var(\mu_0^* - \mu_0) = Var(\mu_0^*) + Var(\mu_0) - 2Cov\{\mu_0^*, \mu_0\}$$

Si en la ecuación anterior consideramos que  $\mu_0$  es una tendencia determinística

$$\sigma_E^2(x_0, t_0) = Var(\mu_0^*) = \sum_{i=n_0}^{n_1} \sum_{j=n_0}^{n_1} a_i a_j Cov(Z(x_i, t_i), Z(x_j, t_j))$$

y la ecuación a minimizar será:

$$\sum_{i=n_0}^{n_1} \sum_{j=n_0}^{n_1} a_i a_j Cov(Z(x_i, t_i), Z(x_j, t_j)) - 2m \left( \sum_{i=n_0}^{n_1} a_i - 1 \right)$$

que lleva al siguiente sistema de ecuaciones

$$\begin{cases} \sum_{j=n_0}^{n_1} a_j Cov(Z(x_i, t_i), Z(x_j, t_j)) - m = 0 & \forall i \in \{n_0, \dots, n_1\} \\ \sum_{i=n_0}^{n_1} a_i = 1 \end{cases} \quad (3.10)$$

Notar que los pesos  $a_i$  no van a depender del punto en el que vamos a estimar, sino de las observaciones que consideremos para hacer la predicción. Si queremos predecir la media utilizando kriging ordinario y el mismo conjunto de observaciones, en dos localizaciones espaciales distintas, obtendremos el mismo valor.

### 3. Kriging con un modelo de tendencia (KT) o Kriging universal

Considera la media desconocida, y permite que varíe dentro de cada vecindad situada en el área de estudio.

La tendencia se modeliza como una combinación de funciones  $f_k(x, t)$  de las coordenadas, o de variables explicativas asociadas con los datos en  $(x, t)$ :  $\mu(x_0, t_0) = \sum_{k=0}^K c_k f_k(x_0, t_0)$ , siendo  $f_k(\cdot, \cdot)$  funciones conocidas, y  $c_k$  nuevas constantes a estimar.

En este caso, el estimador kriging será de la forma:

$$\begin{aligned} Z^*(x_0, t_0) &= \mu(x_0, t_0) + \sum_{i=1}^n a_i (Z(x_i, t_i) - \mu(x_i, t_i)) = \\ &= \sum_{k=0}^K c_k f_k(x_0, t_0) + \sum_{i=1}^n a_i \left( Z(x_i, t_i) - \sum_{k=0}^K c_k f_k(x_i, t_i) \right) = \\ &= \sum_{i=1}^n a_i Z(x_i, t_i) + \sum_{k=0}^K c_k \left( f_k(x_0, t_0) - \sum_{i=1}^n a_i f_k(x_i, t_i) \right) \end{aligned}$$

Al trabajar con kriging universal, vamos a buscar un predictor uniformemente insesgado ([38] pág 152), es decir, vamos a exigir que:

$$\sum_{i=1}^n a_i f_k(x_i, t_i) = f_k(x_0, t_0)$$

ya que entonces podremos expresar el estimador kriging, como función únicamente de los datos, es decir:

$$Z^*(x_0, t_0) = \sum_{i=1}^n a_i Z(x_i, t_i)$$

Además, en este caso el predictor será insesgado, ya que:

$$\begin{aligned} E\left(Z^*(x_0, t_0) - Z(x_0, t_0)\right) &= \sum_{i=1}^n a_i E(Z(x_i, t_i)) - E(Z(x_0, t_0)) = \\ &= \sum_{i=1}^n a_i \sum_{k=0}^K c_k f_k(x_i, t_i) - \sum_{k=0}^K c_k f_k(x_0, t_0) = \\ &= \sum_{k=0}^K c_k \left( \sum_{i=1}^n a_i f_k(x_i, t_i) - f_k(x_0, t_0) \right) = 0. \end{aligned}$$

La minimización de la varianza del error de predicción, teniendo en cuenta todas las restricciones que hemos impuesto, supone minimizar:

$$E\left(\sum_{i=1}^n a_i Z(x_i, t_i) - Z(x_0, t_0)\right)^2 - \sum_{k=0}^K m_k \left(\sum_{i=1}^n a_i f_k(x_i, t_i) - f_k(x_0, t_0)\right).$$

Si, de nuevo, llamamos  $C_{i,j} := Cov(Z(x_i, t_i), Z(x_j, t_j))$ , la ecuación a minimizar quedará:

$$\begin{aligned} &\sum_{i=1}^n \sum_{j=1}^n a_i a_j C_{i,j} - 2 \sum_{i=1}^n a_i C_{i,0} + Var\left(Z(x_0, t_0)\right) - \\ &- \sum_{k=0}^K m_k \left(\sum_{i=1}^n a_i f_k(x_i, t_i) - f_k(x_0, t_0)\right) \end{aligned}$$

derivando respecto a todos los parámetros  $(a_1, \dots, a_n, m_1, \dots, m_K)$ , e igualando las derivadas a 0 obtendremos el siguiente sistema:

$$\begin{cases} \sum_{j=1}^n a_j C_{i,j} - C_{0,i} - \sum_{k=0}^K m_k f_k(x_i, t_i) = 0 & \forall i = 1, \dots, n, \\ \sum_{i=1}^n a_i = 1, \\ \sum_{i=1}^n a_i f_k(x_i, t_i) = f_k(x_0, t_0), \end{cases}$$

y al resolverlo obtendremos los valores de todos los pesos.

Teniendo en cuenta las  $n$  primeras ecuaciones del sistema anterior, es fácil comprobar que la varianza del error de predicción para esta variante de kriging es:

$$\begin{aligned}
\sigma_E^2(x_0, t_0) &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov} \left( Z(x_i, t_i), Z(x_j, t_j) \right) - \\
&- 2 \sum_{i=1}^n a_i \text{Cov} \left( Z(x_i, t_i), Z(x_0, t_0) \right) + \text{Var} \left( Z(x_0, t_0) \right) = \\
&= \sum_{i=1}^n a_i \left( C_{0,i} + \sum_{k=0}^K m_k f_k(x_i, t_i) \right) - 2 \left( C_{0,0} + \sum_{k=0}^K m_k f_k(x_0, t_0) \right) + \\
&+ C_{0,0}.
\end{aligned}$$

$$\sigma_E^2(x_0, t_0) = \sum_{i=1}^n a_i C_{0,i} - C_{0,0} - \sum_{k=0}^K m_k f_k(x_0, t_0).$$

Esta última variante de kriging es la más frecuente en las modelizaciones de fenómenos físicos. Las funciones  $f_k(x, t)$  son funciones conocidas, muchas veces sugeridas directamente por la física del fenómeno estudiado. Por ejemplo, una serie de funciones seno y coseno puede usarse para modelizar una tendencia periódica de algún componente.

La intención de este apartado era la de revisar las tres técnicas de kriging más importante, aunque nosotros sólo utilizaremos una de ellas, el kriging simple, en capítulos posteriores. Para un estudio más exhaustivo de estas técnicas recomendamos el capítulo 5 del libro de Goovaerts [68], o [38] entre otros. Allí se pueden encontrar comparaciones entre los distintos métodos, y las ventajas y desventajas de utilizar cada uno.

### 3.4. Análisis de validación cruzada

Un análisis de validación cruzada es un método que nos permite comprobar las suposiciones que hemos hecho tanto en el modelo (i.e. el tipo de función que hemos ajustado al variograma, sus parámetros, el tamaño de la

vecindad a la hora de hacer el kriging, etc,...) como sobre los datos (es decir, para detectar datos atípicos, “outliers”).

En un análisis de validación cruzada, empezamos extrayendo un valor muestral  $Z(x_i, t_i)$  de la base de datos, y utilizando el resto de observaciones estimamos el valor de la variable en la posición “eliminada”. Denotaremos por  $\hat{Z}(x_i, t_i)_{-i}$  a la estimación obtenida para la variable en la posición  $(x_i, t_i)$  a partir de todas las observaciones de la base de datos exceptuando a la  $i$ -ésima.

Repetimos este procedimiento para todas las observaciones de la base de datos. Quitamos una cada vez, y estimamos con las restantes.

En cada caso, la diferencia  $Z(x_i, t_i) - \hat{Z}(x_i, t_i)_{-i}$  mide si la observación en el punto  $(x_i, t_i)$  se asemeja a la obtenida a partir de las observaciones tomadas en los puntos vecinos.

Si la media de los errores de validación cruzada está próxima a 0,

$$\frac{1}{n} \sum_{i=1}^n \left( Z(x_i, t_i) - \hat{Z}(x_i, t_i)_{-i} \right) \cong 0$$

podremos decir que no hay un sesgo aparente. Valores positivos de esta media podrían indicar que sistemáticamente estamos estimando por defecto, y valores negativos de la media indicarán que estamos estimando en todas las posiciones por exceso.

Por otra parte, la varianza del error de predicción (la varianza del kriging)  $\sigma^2(x_i, t_i)_{-i}$  representa el error de predicción del modelo cuando predecimos en la posición  $(x_i, t_i)$  a partir de las observaciones recogidas en el resto de posiciones. Dividiendo el error de validación cruzada por  $\sigma(x_i, t_i)_{-i}$  podremos comparar la magnitud del error actual y del error de predicción:

$$\frac{Z(x_i, t_i) - \hat{Z}(x_i, t_i)_{-i}}{\sigma(x_i, t_i)_{-i}}$$

Si el modelo que hemos adoptado y las estimaciones obtenidas para los

parámetros son las adecuadas, deberá verificarse que

$$\frac{1}{n} \sum_{i=1}^n \frac{\left( Z(x_i, t_i) - \hat{Z}(x_i, t_i)_{-i} \right)^2}{\sigma^2(x_i, t_i)_{-i}} \cong 1$$

i.e. el error de estimación tiene que ser, en media, igual al error predicho por el modelo.

### 3.5. Problemas que hemos encontrado al trabajar desde el punto de vista de la Geoestadística

En el capítulo 4 vamos a plantear la modelización de la distribución espacio temporal de campos visuales de pacientes sanos. Uno de los objetivos que nos plantearemos con esta modelización será la predicción de campos visuales en instantes de tiempo en los que no dispongamos de observaciones.

Partiremos por tanto de un conjunto de observaciones que consideraremos una realización de una variable espaciotemporal  $Z(x, t)$ , que descompondremos como suma de una media (o tendencia) y una parte residual  $Z(x, t) = \mu(x, t) + \epsilon(x, t)$ .

Entonces, siguiendo la metodología Geoestadística deberíamos:

1. Estimar la tendencia del fenómeno,  $\hat{\mu}(x, t)$ .
2. Estimar los residuos:  $\hat{\epsilon}(x, t) = Z(x, t) - \hat{\mu}(x, t)$ .
3. Buscar un modelo de covarianza para estos residuos.
4. Aplicar el algoritmo kriging adecuado para para predecir en posiciones no monitorizadas, aunque también podríamos haberlo utilizado para estimar la media (ya que será desconocida, aunque aún no sabemos si será constante o con una forma paramétrica determinada).

5. Utilizar la varianza de la predicción como una estimación de la precisión.
6. Realizar un análisis de validación cruzada para analizar la bondad del modelo.

Al intentar seguir estos pasos generales, veremos que nos encontramos con una serie de problemas. El primero ha sido el no poder hablar de procesos estacionarios (ver sección 3.5.1), problema que hemos resuelto aplicando una transformación a los datos.

Al realizar un estudio exploratorio espacial de los datos para formular un modelo para la media espacial, nos hemos encontrado con una estructura bastante complicada para la media, que ha impedido que pudiéramos formularla paramétricamente y que hayamos tenido que recurrir a una formulación semiparamétrica.

El no tener una estructura paramétrica para la media, y el no poder considerarla constante, ni siquiera en vecindades locales, hace que no podamos utilizar ningún algoritmo kriging para estimarla. Por ello buscaremos algoritmos que nos permitan estimar todos los parámetros del modelo, y para la predicción consideraremos la media conocida, y utilizaremos kriging simple.

Por lo tanto vamos a trabajar mezclando la metodología propia de la estadística clásica (máxima verosimilitud, backfitting,...) con la metodología geoestadística.

### 3.5.1. Geoestadística cuando no hay estacionariedad espacial

Hasta ahora hemos trabajado suponiendo estacionariedad en los datos, pero ¿qué ocurre cuando esta condición no se satisface? Sampson y Guttorp [175] desarrollaron una técnica para estimar la estructura de la covarianza en el caso de no estacionariedad espacial, y la aplicaron en la modelización de datos de concentración de Ozono en una determinada región de Arizona.

Otros autores han desarrollado trabajos en los que se proponen modelos no estacionarios y no isotrópicos. Entre ellos destacamos el trabajo de Guttorp et al [76], quienes utilizaron la técnica de deformación propuesta por Sampson y Guttorp en [175] para capturar la no isotropía espacial en series en las que previamente eliminan la dependencia temporal. Los trabajos de Higdon [98, 97] en los que utiliza convoluciones y una metodología bayesiana para aproximar el modelo temporal y la no estacionariedad espacial, y el trabajo de Lavine y Lozier, quienes propusieron un modelo Bayesiano para modelizar temperaturas en una zona del Océano Atlántico [125]. En otro trabajo, Sampson et al [174] revisan las distintas técnicas utilizadas para trabajar cuando no se tiene isotropía espacial. Por último, citar un trabajo de Cressie y Majure [40], quienes plantean la modelización geoestadística de un proceso espacio temporal, con heterogeneidad en la varianza. Esta referencia nos ha servido como punto de apoyo al desarrollar el problema que abordamos en el capítulo 6.

## Capítulo 4

# Modelización espacio temporal de CV de pacientes sanos

### Contenidos

---

<b>4.1. Introducción</b>	<b>83</b>
<b>4.2. Análisis exploratorio espacial</b>	<b>85</b>
4.2.1. Notación	86
4.2.2. Modelización de la media	89
4.2.3. Estructura de correlación en los residuos	93
4.2.4. Ajuste del modelo espacial	97
<b>4.3. Análisis exploratorio temporal</b>	<b>98</b>
<b>4.4. Modelización espacio temporal conjunta</b>	<b>103</b>
4.4.1. Modelos de covarianza separables	106
4.4.2. Modelos de covarianza no separables	106
4.4.3. Estimación	109
4.4.4. Comprobando los residuos	110
4.4.5. Validación cruzada	111
<b>4.5. Posibles aplicaciones del modelo</b>	<b>115</b>
4.5.1. Construcción de intervalos de confianza para identificar posiciones “enfermas”	116

4.5.2. Predicción . . . . .	116
4.5.3. Simulación . . . . .	117

---

## 4.1. **Introducción**

En este capítulo vamos a trabajar en la caracterización de la distribución espacio temporal de campos visuales de pacientes sanos. Nuestro objetivo final será la caracterización de campos visuales (CV) de pacientes con glaucoma, pero para entender el comportamiento de CV en pacientes enfermos nos ha parecido adecuado empezar estudiando su comportamiento en pacientes sanos, y dejar la generalización del modelo que obtengamos para considerar también a los enfermos, para el capítulo 6.

Como comentábamos en el capítulo 2, el análisis de campos visuales (CV) tras la realización de perimetrías, es una herramienta clave para el diagnóstico y seguimiento del glaucoma. La información contenida en un campo visual ayuda al oftalmólogo en el diagnóstico de la enfermedad, y en el control de su evolución.

Al hablar de modelización espacio temporal, no nos referimos a la incidencia de la enfermedad en un país o región geográfica, sino a su distribución espacial en la retina del paciente que sufre la enfermedad. Como comentamos en el capítulo 2, un campo visual es un mapa numérico que proporciona información sobre el umbral de visión de un paciente en 52 puntos de su retina. Estos puntos están situados sobre un grid regular, y son puntos contiguos separados 6 grados (vertical y horizontalmente). (Ver capítulo 2 para recordar la disposición espacial de los puntos testeados en el grid).

En este capítulo, y en el capítulo 6, abordaremos la modelización de la distribución espacio temporal de los campos visuales utilizando la metodología propia de la Geoestadística, y distinguiendo claramente dos etapas: la modelización de CV de pacientes sanos en este capítulo y la extensión del modelo al caso de pacientes con glaucoma en el capítulo 6.

Aunque vamos a trabajar bajo un punto de vista geoestadístico, vamos a utilizar un tipo de “Geoestadística basada en el modelo”, queriendo de-

cir con ello que además de utilizar métodos exploratorios para identificar la estructura de la media y de la matriz de varianzas covarianzas propios de todo análisis geoestadístico, y además de utilizar kriging para realizar predicciones en posiciones no monitorizadas, utilizaremos las propiedades de la distribución gaussiana para estimar los parámetros del modelo, y para construir intervalos de confianza para pacientes sanos, (que podremos utilizar después para clasificar una posición observada en un CV de cualquier paciente como “normal” o “anormal”). También podremos utilizar el modelo para predecir futuras observaciones, y para simular campos visuales de pacientes sanos. Esta última aplicación (la simulación), podría ser muy útil en el estudio del glaucoma debido a la gran dificultad existente para recoger grandes conjuntos de datos. Un test perimétrico resulta aburrido y engorroso para el paciente. Además, la evolución de la enfermedad es muy lenta, por lo que no es posible conseguir conjuntos de datos grandes en un breve período de tiempo.

Resulta difícil encontrar bases de datos de CV en el tiempo para un grupo de pacientes, pero la tarea se complica aún más si nos restringimos a observaciones de pacientes sanos. Como comentábamos en el capítulo 2, en la mayoría de los hospitales públicos españoles, si no se trata de pacientes que requieran una vigilancia especial, las campimetrías se realizan a cada paciente con una periodicidad de un año, y es difícil convencer a pacientes sanos para que vuelvan año tras año para repetir las exploraciones.

En nuestra base de datos tenemos 131 pacientes sanos con una única observación, 38 pacientes sanos con 2 observaciones, 9 con 3 observaciones y 8 con 4. Todas estas observaciones satisfacen unos requisitos mínimos de fiabilidad ( menos del 20 % de pérdidas de fijación, porcentaje de respuestas falso negativas inferior al 33 %, y porcentaje de respuestas falso positivas inferior al 33 %). Estos criterios son los mismos que se han exigido en otros

trabajos ([153]).

Para empezar a trabajar, en la sección 4.2 vamos a plantear un análisis exploratorio espacial de los datos. Para ello, deberemos quedarnos aleatoriamente con una de las observaciones de cada paciente, y nuestro objetivo será modelizar la media, varianza y la estructura de covarianza espacial. En la sección 4.3 plantearemos un análisis exploratorio puramente temporal, para ello consideraremos sólo aquellos pacientes con medidas repetidas, restaremos de las observaciones la media espacial estimada en 4.2, y estudiaremos la correlación temporal de los residuos. En la sección 4.4 tendremos en cuenta los resultados de las secciones anteriores para construir un modelo espacio temporal conjunto. Finalmente en la sección 4.5 y en el capítulo 5, veremos algunas de las posibles aplicaciones del modelo.

## 4.2. **Análisis exploratorio espacial**

El análisis de datos exploratorio (EDA [195]), resulta una herramienta fundamental como primer paso para un análisis espacial. Para poder realizarlo, necesitamos tener un conjunto de datos independientes, y para ello escogemos aleatoriamente una observación de cada paciente.

Como un primer paso en este análisis exploratorio, nos planteamos el cálculo de la media y de la varianza de las intensidades observadas en cada posición del CV. Una vez calculadas, dibujamos mapas con estos valores. Así en la figura 4.1 podemos ver un primer mapa con el valor de la media obtenida en cada posición, y otro para la varianza.

Analizando estos mapas y las figuras 4.1 c) y 4.1 d) en las que representamos la media y la varianza respectivamente de cada posición en función a su distancia al centro del CV, podemos ver que existe una fuerte relación entre los valores de la media y de la varianza de cada posición, con esta distancia. Las posiciones más periféricas son las que presentan valores más bajos para

la media, y más altos para la varianza. Además media y varianza están entre sí fuertemente correlacionadas.

Para romper esta fuerte relación existente entre media y varianza necesitamos transformar los datos. Hemos probado con la transformación “raíz cuadrada” y con la transformación “logarítmica”, pero ninguna de las dos ha sido capaz de destruir la fuerte dependencia. Al plantearnos la posibilidad de realizar una transformación de Box-Cox, vimos que la más adecuada era la transformación con  $\lambda = 4$ , es decir:

$$y_i = \frac{x_i^4 - 1}{4}$$

Tras transformar los datos, vemos que ya no existe correlación entre media y varianza (figura 4.2), y que podemos asumir homogeneidad en la varianza, así pues en adelante, vamos a trabajar con los datos transformados por una transformación de Box-Cox con  $\lambda = 4$ .

#### 4.2.1. Notación

Tras la transformación de Box-Cox llamamos  $Y_{i,j}$  al valor observado en el CV realizado al paciente número  $i$  en la posición número  $j$ , e  $Y_i = (Y_{i,1}, Y_{i,2}, \dots, Y_{i,52})$  al vector con las observaciones de las 52 posiciones del campo visual del  $i$ -ésimo paciente.

Para trabajar bajo una metodología geoestadística, vamos a buscar un modelo en el que podamos expresar:

$$Y_i = \mu_i + \epsilon_i$$

con  $\mu_i = (\mu_{i,1}, \mu_{i,2}, \dots, \mu_{i,52})$ ,  $\epsilon_i = (\epsilon_{i,1}, \epsilon_{i,2}, \dots, \epsilon_{i,52})$ ,  $\epsilon_i \sim MVN(0, V)$  variables aleatorias independientes e idénticamente distribuidas que siguen una distribución normal con una matriz de varianzas covarianzas  $V$  que deberemos determinar.

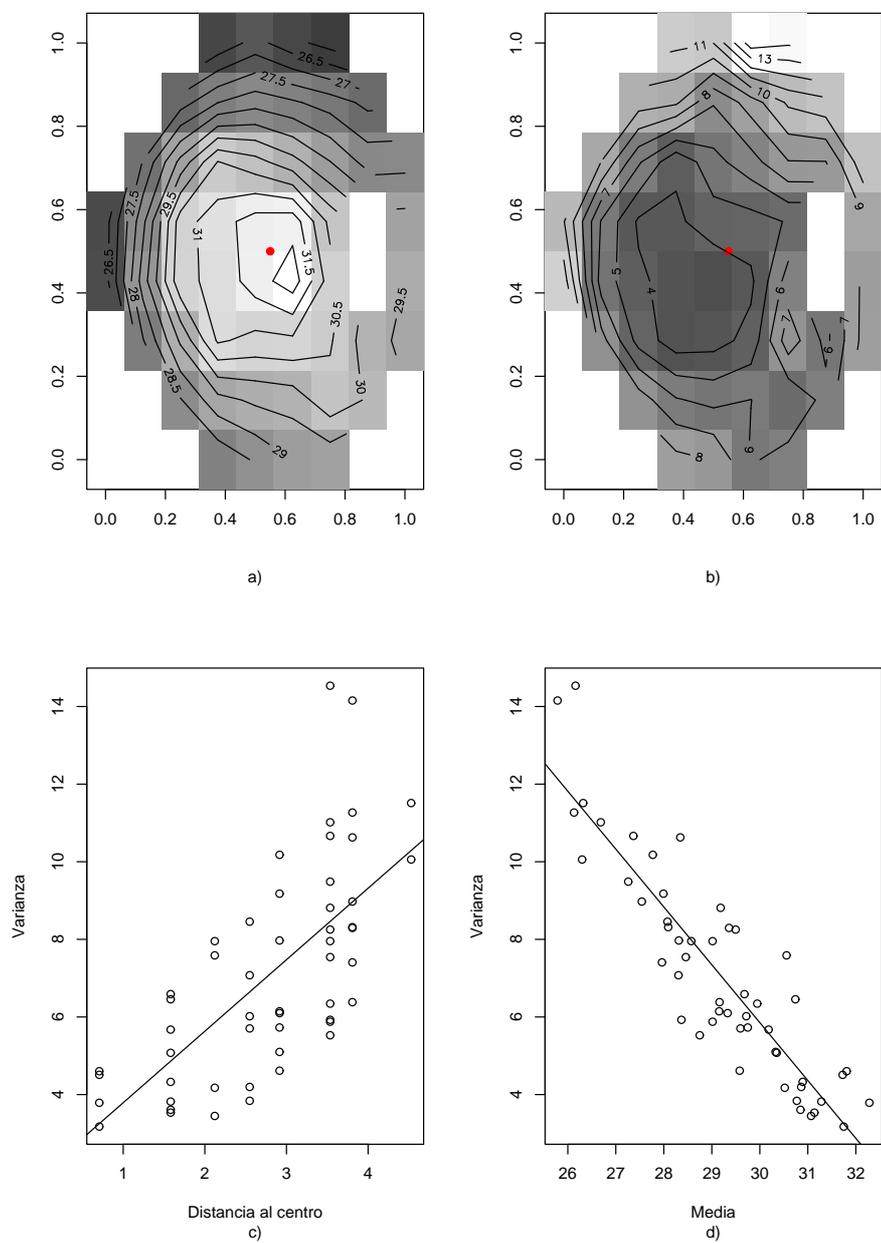


Figura 4.1: a) y b) Medias y varianzas de los datos no transformados. Colores más oscuros indican valores más bajos. c) Varianza vs distancia al centro del campo visual. d) Varianza vs media.

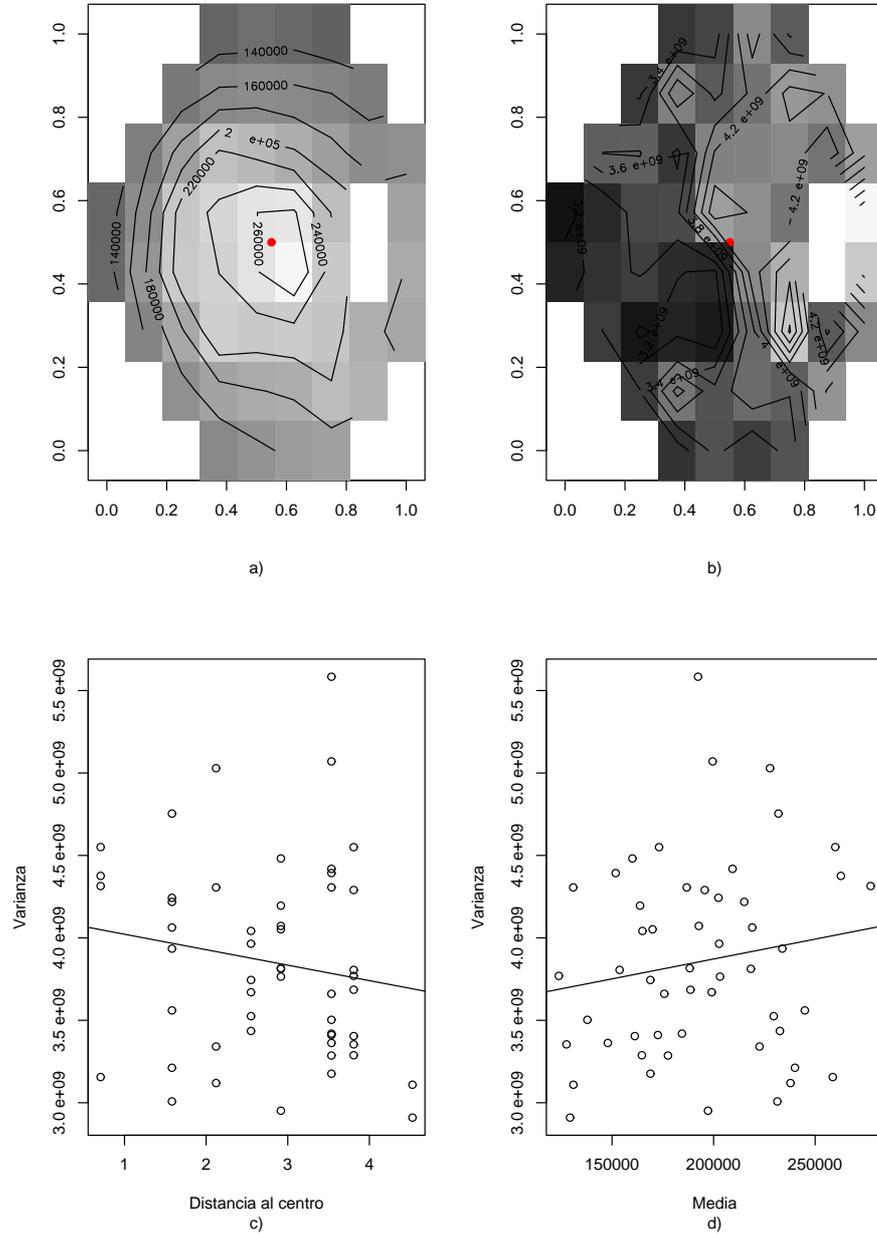


Figura 4.2: a) y b) Medias y varianzas de los datos tras la transformación de Box-Cox. Colores más oscuros indican valores más bajos. c) Varianza vs distancia al centro del CV. d) Varianza vs media.

### 4.2.2. Modelización de la media

Como primer paso para la construcción del modelo espacial, empezamos buscando un modelo para la media, y como es habitual en estos casos, vamos a buscar un conjunto de covariables adecuado, y a plantear una modelización paramétrica.

#### Modelización paramétrica

Para modelizar la media, buscamos el conjunto de covariables que puedan ser significativas.

Llamamos  $d_j$  a la distancia euclídea de la posición  $j$ -ésima al centro del campo visual. Empíricamente, en el gráfico 4.2 podemos ver que existe una relación lineal (aunque no demasiado fuerte) entre las dos variables. Las posiciones más periféricas tienen valores más bajos para la media. Además de esta relación lineal entre la media y la distancia al centro del campo visual, empíricamente, en 4.2 podemos ver que existe un efecto que podríamos llamar “de la fila”, ya que las intensidades de la media son mayores en las filas superiores del campo visual que en las inferiores (el hemisferio superior presenta medias más altas que el inferior).

También la edad es una covariable a tener en cuenta, ya que en la práctica oftalmológica se sabe que la agudeza visual disminuye con la edad. Hemos comprobado que esta disminución es lineal, y tras realizar varios análisis de la varianza (ANOVAS) hemos podido concluir que la pendiente de las rectas es la misma en todos los puntos del campo visual. La recta que explica la disminución de la agudeza visual como función de la edad, tendrá en cada punto una ordenada en el origen distinta, pero todas ellas tendrán la misma pendiente.

Una vez determinadas las covariables a tener en cuenta, realizamos un análisis de regresión múltiple, para comprobar que las tres covariables son

significativas, y que podemos formular la media espacial como:

$$\mu_{i,j} = \beta_0 + \beta_1 d_j + \beta_2 r_j + \beta_3 a_i \quad i = 1, \dots, N \quad j = 1, \dots, 52$$

siendo  $a_i$  la edad del paciente número  $i$ ;  $r_j$  la fila en la que se encuentra la posición número  $j$ ;  $\beta_0, \beta_1, \beta_2$  y  $\beta_3$  los parámetros del modelo, y  $N$  el número total de pacientes sanos en el estudio.

En cuanto a la covariable “fila”, decir que hemos numerado las filas con números consecutivos del 1 al 8, empezando a numerar por la fila más superior del campo visual.

A partir de esta formulación podemos expresar:

$$\mu_i = \begin{pmatrix} \mu_{i,1} \\ \mu_{i,2} \\ \vdots \\ \mu_{i,52} \end{pmatrix} = \begin{pmatrix} 1 & d_1 & r_1 & a_i \\ 1 & d_2 & r_2 & a_i \\ 1 & d_3 & r_3 & a_i \\ \vdots & \vdots & \vdots & \vdots \\ 1 & d_{52} & r_{52} & a_i \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} := X_i \beta \quad (4.1)$$

y expresar vectorialmente el modelo como:

$$Y_i = X_i \beta + \epsilon_i \quad (4.2)$$

con  $\epsilon_i \sim MVN(0_{52}, V)$  variables aleatorias independientes e idénticamente distribuidas con media igual a 0, y varianza  $V, \forall i \in \{1, \dots, N\}$ .

Una vez formulado el modelo, podemos estimar el valor del parámetro  $\beta$  por mínimos cuadrados, y comprobar la bondad del ajuste, realizando tests sobre los residuos ( $r_i = y_i - X_i \hat{\beta}$ ).

Comprobamos dos hipótesis sobre los residuos; la hipótesis de que la media sea igual a 0 en todas las posiciones espaciales, y la hipótesis de independencia entre residuos de pacientes distintos. Ambas hipótesis han sido rechazadas, lo que nos lleva a concluir que el modelo que hemos construido para la media no es suficientemente descriptivo. El que los residuos de

pacientes distintos sean no independientes nos lleva a concluir que existen covariables comunes a todos los individuos que no han sido introducidas en el modelo todavía. Además, analizando los residuos vemos que existen posiciones que son sistemáticamente sobre estimadas, y que otras son siempre estimadas por defecto.

Deberíamos por tanto plantearnos el introducir nuevas covariables en la modelización, pero el problema es que no disponemos de más información, y por lo tanto no podemos introducir nuevas covariables para completar la descripción paramétrica de la media. Esto nos ha llevado a plantearnos la posibilidad de utilizar una aproximación semiparamétrica, y en particular a utilizar la metodología de los modelos aditivos, desarrollada por Hastie y Tibshirani en 1990 [92], que comentamos brevemente en el capítulo 3.

### **Aproximación semiparamétrica. Modelos aditivos**

La metodología de los modelos aditivos (AM) [92] permite la construcción de modelos más flexibles y formular la estructura de la media como:

$$\mu_i = X_i\beta + f(x) \quad \forall i \in \{1, \dots, N\} \quad (4.3)$$

con  $x = (x_1, x_2, \dots, x_{52})$ , siendo  $x_j$  las coordenadas cartesianas de la posición número  $j$ ;  $f(-)$  una superficie suave de  $x$ , y  $X_i\beta$  la componente paramétrica de la estructura de la media (como en la ecuación (4.1)). (Ver el capítulo 3 para tener una visión de los modelos aditivos, o consultar directamente [92].)

Siguiendo la teoría de Hastie y Tibshirani [92], y la aplicación de Kelsall y Diggle [116], utilizaremos kernels gaussianos para modelizar las funciones suaves  $f(-)$ , y un algoritmo backfitting para ajustar el modelo.

La mayor parte de la investigación reciente sobre backfitting, se centra en estudiar la convergencia del algoritmo [24], y las propiedades asintóticas de

los estimadores [155]. La literatura sobre regresión no paramétrica con datos dependientes, es escasa. Entre los trabajos existentes destacamos [32, 151], que justifican el uso de algoritmos tipo backfitting en situaciones en que la dependencia serial no es muy fuerte. Casi todos ellos se dirigen al trabajo con series temporales.

Nosotros utilizaremos el método backfitting para estimar los parámetros del modelo, y una vez tengamos el modelo completamente construido, comprobaremos su bondad.

Veamos los pasos del algoritmo que vamos a seguir:

**Algoritmo 2.1 .**

1. Inicializar  $\hat{f}(x) = 0$ , y estimar por mínimos cuadrados el valor de  $\beta$  en:  $Y_i = X_i\beta$ .
2. Calcular  $Z_i = Y_i - X_i\hat{\beta}$ .
3. Calcular las varianzas empíricas de las observaciones en cada posición, y definir los pesos  $w_j$  como la inversa de estas varianzas,  $\forall j \in \{1, \dots, 52\}$ .
4. Realizar una regresión kernel ponderada de  $Z$  sobre  $x$ , con pesos  $w_j$ ,

$$\hat{f}(x_l) = \sum_i \sum_{j=1}^{52} w_j K_h(x_l - x_j) z_i(j) / \sum_i \sum_{j=1}^{52} w_j K_h(x_l - x_j), \quad (4.4)$$

$\forall l \in \{1, \dots, 52\}$ , siendo  $K_h(x) = \frac{1}{h^2} K(x/h)$ , y  $K(x)$  el kernel gaussiano estándar, es decir:

$$K(x) = \frac{1}{2\pi} \cdot \exp\left(-\frac{1}{2} \|x\|^2\right),$$

$$\hat{f}(x) = (\hat{f}(x_1), \dots, \hat{f}(x_{52})).$$

5. Definir  $T_i = Y_i - \hat{f}(x)$ , y estimar por mínimos cuadrados los valores de  $\beta$  en:

$$T_i = X_i\beta.$$

6. Repetir los pasos 2-5 hasta obtener la convergencia de las estimaciones.

Para elegir la amplitud de banda (bandwidth) del kernel de regresión, (i.e. para elegir los valores de  $h$  para la ecuación 4.4), utilizamos el método de validación cruzada. Para ello cada vez que llegamos al paso (4.) y tenemos que estimar la superficie suave, elegimos el valor de  $h$  que minimiza:

$$CV(h) = \frac{1}{52 \cdot N} \sum_{i=1}^N \sum_{j=1}^{52} w_j \cdot \{Y_{i,j} - \hat{f}_i^{-j}(x_j)\}^2, \quad (4.5)$$

siendo  $\hat{f}_i^{-j}(x_j)$  la estimación de  $f(x_j)$  utilizando como amplitud de banda  $h$ , y utilizando todos los datos menos el par  $(Z_{i,j}, x_j)$ .

Para minimizar la ecuación (4.5), calculamos  $CV(h)$  para un número finito de valores de  $h$  en un rango apropiado, y nos quedamos con aquel valor de  $h$  que haya proporcionado el valor más pequeño para  $CV(h)$ .

### 4.2.3. Estructura de correlación en los residuos

Una vez estimada la media espacial, la restamos de todas las observaciones para obtener los residuos.

Contrastamos empíricamente la hipótesis de homogeneidad en la varianza en los 52 puntos del campo visual (figura 4.2), y dibujamos un mapa de correlaciones direccionales [49] (figura 4.3) para garantizar (empíricamente) la isotropía de la estructura de correlación.

Para dibujar este mapa, consideramos todas las parejas de puntos que están situados a la misma distancia y con la misma dirección, es decir, fijamos un par de valores  $(i, j)$  y cogemos todas las parejas de puntos de coordenadas

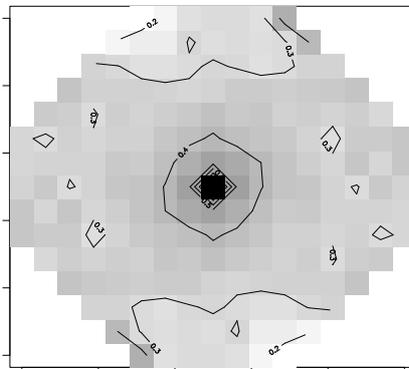


Figura 4.3: Variograma direccional empírico.

$(x, y)$  y  $(x+i, y+j)$ , calculamos la correlación entre cada una de estas parejas, y finalmente la media de todas las correlaciones. Representamos el valor que hemos obtenido en la coordenada  $(i, j)$  del gráfico que hemos llamado mapa de correlaciones direccionales (fig 4.3). En caso de isotropía debemos obtener un gráfico en el que todos los puntos equidistantes al centro tengan el mismo valor, tal y como ocurre en nuestro caso.

Tras haber comprobado las hipótesis de homogeneidad e isotropía, dibujamos el variograma espacial empírico para detectar las fuentes de variación presentes en nuestro conjunto de datos (ver [48] para una explicación más detallada de las distintas fuentes de variabilidad).

En la figura 4.4 hemos dibujado el variograma empírico, y podemos detectar la presencia de tres fuentes de variabilidad: un efecto aleatorio reflejando la variabilidad entre pacientes, la correlación serial, y un término de error de medida. Por lo tanto a la hora de formular la varianza deberemos tenerlas en cuenta y escribir:

$$\text{Var}(\epsilon) = \nu^2 \cdot J + \sigma^2 \cdot H + \tau^2 \cdot I \quad (4.6)$$

siendo  $J$  una matriz  $52 \times 52$  con todos sus valores iguales a 1,  $H = (\rho_{i,j})$  la matriz de correlaciones espaciales, e  $I$  la matriz identidad (en este caso  $52 \times 52$ ).

Como hemos probado que podemos trabajar bajo la hipótesis de isotropía, utilizaremos una función de la familia de las funciones de Matérn para modelizar la correlación espacial:

$$\rho(u, \phi) = (2^{\kappa-1}\Gamma(\kappa))^{-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi)$$

con  $\Gamma()$  denotando a la función  $\Gamma$ , y siendo  $K_{\kappa}$  la función de Bessel modificada de orden  $\kappa$ .

Esta familia es válida para  $\phi > 0$  y  $\kappa > 0$ . Fue introducida por B. Matérn en 1960 y publicada algunos años después ([141]).

Como casos particulares de las funciones de Bessel tenemos la función de correlación exponencial (en el caso particular de que  $\kappa = 0,5$ ), y la función de correlación gaussiana (cuando  $\kappa \rightarrow \infty$ ). Uno de los rasgos más atractivos de esta familia es que el parámetro  $\kappa$  controla la diferenciabilidad del proceso de una forma directa: la parte entera de  $\kappa$  da el número de veces que el proceso subyacente es diferenciable (mean-square differentiable). Por lo tanto la función de correlación de Matérn es una de las funciones de correlación más flexibles para el uso general. (Ya habíamos comentado las funciones de Matérn en el capítulo 3, y remitimos a [190] a todos a aquellos interesados en estudiarlas de forma más detallada).

Para estimar todos los parámetros que aparecen en la formulación de la varianza, ajustamos la ecuación 4.6 a nuestro variograma espacial empírico utilizando el método de mínimos cuadrados y considerando una función de Matérn para la correlación espacial. Este ajuste nos permite tener una primera especificación del modelo ya con todos los parámetros estimados. De todas formas vamos a quedarnos simplemente con las modelizaciones de todas las estructuras, y vamos a buscar un método que nos permita estimar

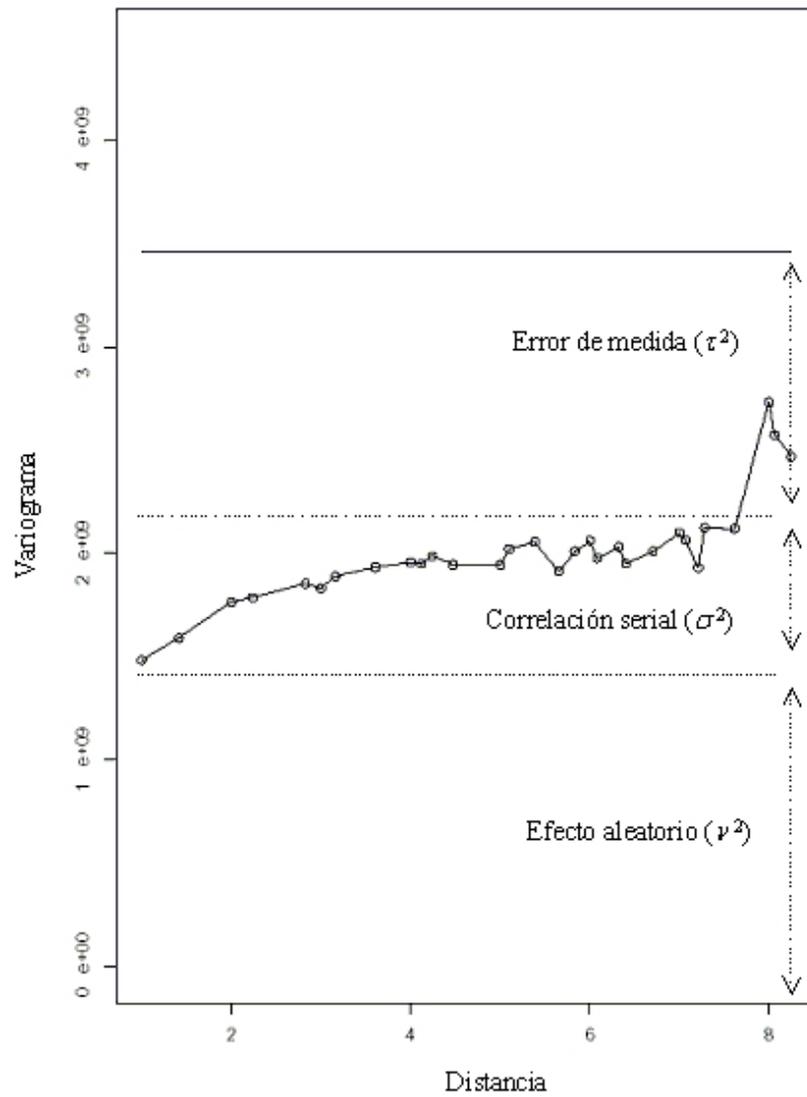


Figura 4.4: Variograma espacial empírico. La línea horizontal es una estimación de la varianza total.

conjuntamente todos los parámetros de modelo. Para ello vamos a modificar el algoritmo que nos ha permitido estimar la media espacial tal y como se explica a continuación.

#### 4.2.4. Ajuste del modelo espacial

Una vez determinadas las estructuras de la media y de la varianza, nos planteamos el estimar conjuntamente todos los parámetros del modelo. Para obtener estos estimadores, combinamos el algoritmo de máxima verosimilitud con el algoritmo backfitting. Tendremos entonces que modificar el algoritmo 2.1 que quedará como sigue:

##### Algoritmo 2.2 .

1. Inicializar  $\hat{f}(x) = 0$ , y estimar los parámetros de:

$$Y_i = X_i\beta + \epsilon_i$$

por máxima verosimilitud, con  $X_i\beta$  como en la ecuación (4.1), y  $\epsilon_i$  una variable aleatoria gaussiana de media 0 y varianza dada por la ecuación (4.6).

2. Calcular  $Z_i = Y_i - X_i\hat{\beta}$ .
3. Definir los pesos  $w_j$  como la inversa de las varianzas estimadas,  $\forall j \in \{1, \dots, 52\}$ .
4. Realizar una regresión kernel ponderada de  $Z$  sobre  $x$  con pesos  $w_j$  como en el algoritmo 2.1.  $\hat{f}(x) = (\hat{f}(x_1), \dots, \hat{f}(x_{52}))$ .
5. Definir  $T_i = Y_i - \hat{f}(x)$ , y ajustar por máxima verosimilitud los parámetros del modelo:

$$T_i = X_i\beta + \epsilon_i.$$

Cuadro 4.1: Estimaciones de los parámetros del modelo 4.2

$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
344377.31	-38626.64	1358.14	-898.12
$\sigma^2$	$\nu^2$	$\tau^2$	$\phi$
1384625939	1497356676	595927637	0.9367

6. Repetir los pasos 2-5 hasta lograr la convergencia de las estimaciones.

Utilizamos de nuevo el método de validación cruzada para elegir el valor de  $h$  a utilizar en el paso 4 del algoritmo.

Dada la complejidad de la función de Matérn, fijamos  $\kappa = 1$  y estimamos el resto de los parámetros ( $\beta_0, \beta_1, \beta_2, \beta_3, \nu^2, \sigma^2, \tau^2, \phi$ ) siguiendo los pasos del algoritmo anterior. Repetimos la estimación fijando otros valores para  $\kappa$ , pero vemos que no hay variaciones sustanciales en los valores que obtenemos para la verosimilitud.

En la tabla 4.1 podemos ver las estimaciones de todos los parámetros del modelo espacial, cuando en la función de Matérn fijamos  $\kappa = 1$ .

### 4.3. Análisis exploratorio temporal

Una vez construido el modelo espacial, pasamos a explorar la estructura temporal de los datos. Como en la sección anterior hemos estimado la media espacial, vamos a restársela a todas las observaciones de nuestra base de datos, y estudiaremos la correlación temporal de los residuos.

Para llevar a cabo el análisis temporal, quitamos de la base de datos a todos aquellos pacientes de los que tenemos una única observación y trabajamos con los 55 pacientes de los que tenemos observaciones repetidas. Identificamos la edad del paciente en el momento de someterse al test con el índice “tiempo”.

Nuestras observaciones son más densas en el espacio que en el tiempo, ya que tenemos muy pocas repeticiones de cada paciente. Si cogemos un único paciente y nos fijamos en una de las posiciones espaciales, vemos que no tenemos suficiente información para poder realizar inferencias a partir de ahí. Esta “pérdida de información” se suple por el hecho de tener 55 repeticiones independientes de la misma medida (de cada serie temporal).

La primera cuestión que nos aparece inmediatamente es si podemos asumir un mismo modelo temporal sobre todas las posiciones espaciales. Recordar que para cada paciente tenemos 52 series temporales, una sobre cada localización espacial, entonces ... ¿podemos suponer un único modelo para las 52 series temporales? Para comprobar empíricamente esta hipótesis construimos sobre cada posición un “variograma medio” de todas las series temporales que tenemos sobre ella (es decir calculamos un variograma para cada una de las series temporales y luego calculamos la media de los variogramas, promediando sobre retardos iguales). Al comparar los variogramas (figura 4.5) vemos que el comportamiento es bastante parecido, y podemos suponer un modelo temporal común para las 52 series temporales.

Por lo tanto vamos a adoptar tres suposiciones: que para cada paciente tenemos 52 series temporales muy cortas con media igual a 0; que existe una estructura temporal común para modelizar las 52 series temporales, y que la estructura de la correlación es isotrópica.

Para modelizar la correlación, empezamos dibujando el variograma temporal empírico medio. Calculamos en cada posición el variograma temporal de cada paciente por separado, y luego promediamos todos los variogramas. Podemos ver el gráfico resultante en la figura 4.6 (a).

En ella detectamos de nuevo la presencia de las tres fuentes de aleatoriedad: un efecto aleatorio reflejando la variabilidad existente entre distintos pacientes, la correlación temporal para medidas del mismo individuo, y un

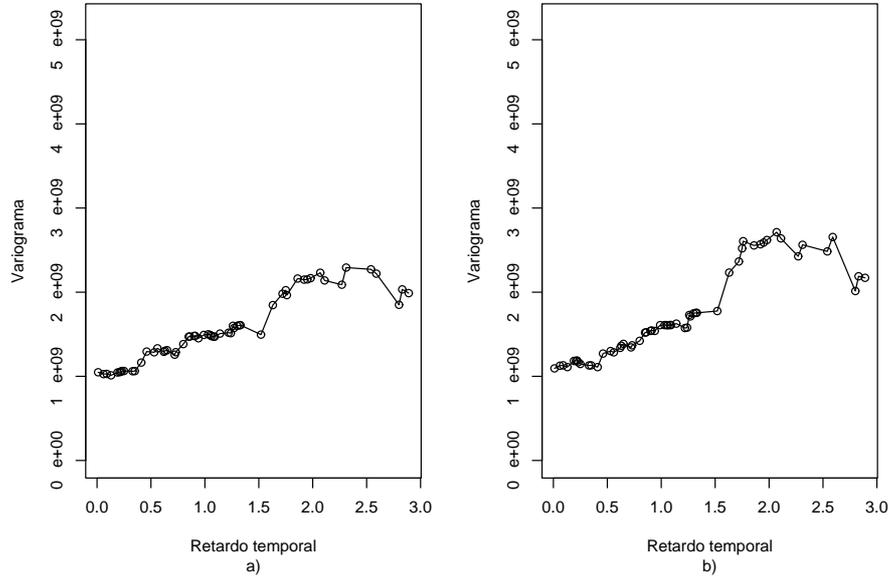


Figura 4.5: Variogramas temporales empíricos calculados: a) en la posición número 16, b) en la posición número 37.

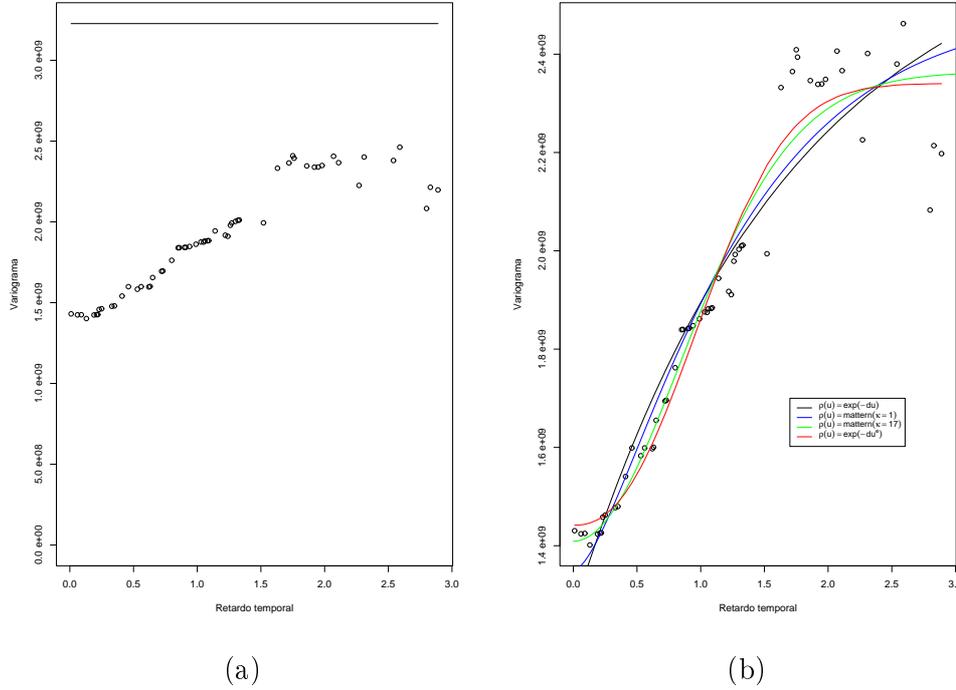


Figura 4.6: (a) Variograma temporal empírico medio. La línea horizontal es una estimación de la varianza total. (b) Ajuste de distintas funciones paramétricas al variograma temporal empírico.

término de error de medida. Por lo tanto la matriz de varianzas-covarianzas volverá a ser:

$$\text{Var}(\epsilon) = \nu^2 J + \sigma^2 H + \tau^2 I \quad (4.7)$$

siendo  $J$  una matriz con todos sus elementos iguales a 1,  $H = (\rho_{i,j})$  la matriz de correlaciones temporales, e  $I$  la matriz identidad.

Sobre el variograma empírico,  $\gamma(u) = \tau^2 + \sigma^2(1 - \rho(u))$ , podemos ajustar por mínimos cuadrados varios modelos paramétricos. Por probar varios modelos, vamos a ajustar dos funciones de Matérn (con dos valores distintos para  $\kappa$ ) y dos funciones exponenciales (una función exponencial, y una

Cuadro 4.2: Resultados de ajustar modelos paramétricos al variograma espacial empírico.

$\rho(u)$	$matern(u, \phi, \kappa)$	$matern(u, \phi, \kappa)$	$exp(-d \cdot u)$	$exp(-d \cdot u^e)$
	$\tau^2 = 1.350e+09$	$\tau^2 = 1.41e+09$	$\tau^2 = 1.27e+09$	$\tau^2 = 1.44e+09$
	$\sigma^2 = 1.15e+09$	$\sigma^2 = 9.55e+08$	$\sigma^2 = 1.41e+09$	$\sigma^2 = 8.99e+08$
	$\phi = 8.42e-01$	$\phi = 1.51e-01$	$d = 5.86e-01$	$d = 6.29e-01$
	$k = 1$	$k = 17$		$e = 2.36e+00$
RSE	94510000	81240000	105500000	79040000
GL	55	55	55	54

función exponencial potencial). Estudiaremos el ajuste tanto gráfica (figura 4.6 (b)) como numéricamente. En la tabla 4.2 vemos las estimaciones de los parámetros (por mínimos cuadrados), el error residual estándar y los grados de libertad de los diferentes ajustes.

En vez de utilizar el método de mínimos cuadrados para estimar los parámetros, hubiéramos podido optar también por maximizar la cuasiverosimilitud, procedimiento más complicado que el de mínimos cuadrados, pero que proporciona estimaciones más precisas. En este caso hubiéramos necesitado utilizar la información sobre la media y como Barry et al [10] sugieren, en la práctica la mejora en las estimaciones obtenidas suele ser poca. Como nosotros estamos simplemente explorando varios modelos y viendo qué modelo paramétrico podemos utilizar para modelizar la correlación temporal, nos quedamos con el método de estimación de parámetros que nos parece más sencillo. Después, cuando ya tengamos clara la estructura temporal, construiremos un modelo espacio temporal conjunto, y volveremos a estimar conjuntamente todos los parámetros del modelo por máxima verosimilitud.

Notar que la función exponencial es un caso particular de función de Matérn (con  $\kappa = 0,5$ ), y que hubiéramos podido no considerarla en el estudio

ya que es también un caso particular de la función exponencial potencial (para  $e = 1$ ), sin embargo la incluimos ya que se trata de un modelo muy utilizado.

Como podemos ver, obtenemos el mejor ajuste al considerar la función exponencial potencial. Sin embargo Diggle y Ribeiro [49] nos hablan de esta función y nos avisan de que aunque estos modelos suelen ajustar de forma razonable (cualitativamente hablando) la estructura de correlación empírica de un conjunto de datos, a menudo las predicciones basadas en este modelo son poco robustas a pequeñas desviaciones del modelo asumido. Por lo tanto a la hora de construir el modelo espacio temporal conjunto, tendremos en cuenta el modelo exponencial potencial para modelizar la correlación temporal, pero no olvidaremos los otros.

#### 4.4. Modelización espacio temporal conjunta

Una vez analizadas por separado la estructura espacial y la temporal vamos a plantearnos la modelización espacio temporal conjunta.

Volvemos a introducir la notación ya que ahora en cada observación tendremos que señalar tanto su posición espacial como temporal. Llamamos  $Y_{i,j,t}$  a la observación número  $t$  del paciente número  $i$  en la posición número  $j$ , y denotamos por  $Y_{i,\cdot,t} = (Y_{i,1,t}, Y_{i,2,t}, \dots, Y_{i,52,t})$  al vector con los 52 valores obtenidos en el test número  $t$  del paciente número  $i$ ,  $i \in \{1, \dots, N\}$ ,  $t \in \{1, \dots, n_i\}$ , y por

$$Y_i = (Y_{i,1,t_1}, Y_{i,2,t_2}, \dots, Y_{i,52,t_1}, \dots, Y_{i,1,t_{n_i}}, \dots, Y_{i,52,t_{n_i}}).$$

Al igual que en la sección anterior, identificamos el “tiempo”  $t$  con la edad del paciente, y nuestro objetivo sigue siendo el modelizar:

$$Y_{i,\cdot,t} = \mu_{i,\cdot,t} + \epsilon_{i,\cdot,t} \quad (4.8)$$

con  $\epsilon_{i,j,t}$  variable gaussiana, con:

$E[\epsilon_{i,j,t}] = 0 \forall (i, j, t)$ ,  $Cov(\epsilon_{i,j,t}, \epsilon_{i,k,m}) = f(\|j - k\|, |t - m|)$ . Necesitamos por tanto un modelo para la estructura de la media y otro modelo para la covarianza espacio temporal. Seguiremos suponiendo que la covarianza es estacionaria tanto en el espacio como en el tiempo.

Con respecto a la media, seguiremos considerando la aproximación semi-paramétrica que hemos construido en la sección 4.2.2, es decir:

$$\mu_{i,j,t} = \beta_0 + \beta_1 d_j + \beta_2 r_j + \beta_3 a_i + f(x_j)$$

con  $f(x_j)$  una superficie suave dependiendo de la posición.

Para modelizar la covarianza, dibujamos el variograma espacio temporal conjunto empírico (figura 4.7), cuyo cálculo viene dado por la siguiente expresión:

$$\hat{\gamma}(h, u) = \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{2 |N(h, u)|} \sum_{(j,k,t_1,t_2) \in N(h,u)} (Y_{i,j,t_1} - Y_{i,k,t_2})^2 \right]$$

Ahora debemos buscar un modelo paramétrico adecuado para ajustar esta superficie. Como ya comentamos en el capítulo 3, el principal problema que existe a la hora de modelizar una función de covarianzas, es que debemos garantizar que sea definida positiva, condición bastante complicada de demostrar. Por ello normalmente se opta por ajustar al variograma empírico el modelo que resulte más apropiado.

Ya hemos hablado en el capítulo 3 de las funciones de covarianza separables y de las no separables. Vamos a ajustar al variograma empírico varios modelos de covarianza, tanto separables como no separables, y luego nos quedaremos con el que proporcione mejor ajuste.

Entre los modelos de covarianza separables [169, 171, 122, 39] consideraremos sólo modelos multiplicativos, i.e. aquellos en los que  $C(h, u) = C^1(h) \cdot C^2(u)$  siendo  $C^1(-)$  un modelo de covarianza espacial válido, y  $C^2(-)$  un

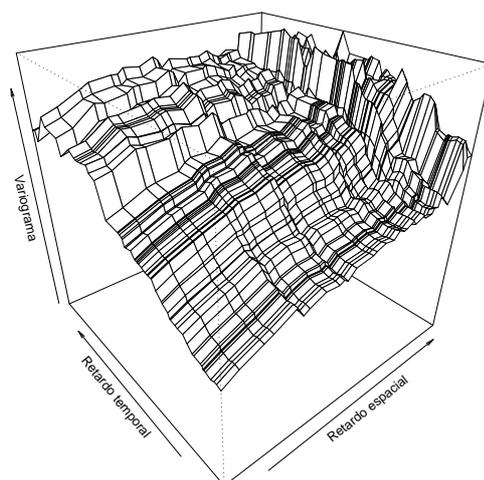


Figura 4.7: Variograma espacio-temporal empírico.

modelo de covarianza temporal válido. En cuanto a los modelos no separables nos basaremos en los modelos propuestos en los trabajos de Cressie y Huang [39] y de Gneiting [66].

Como una primera aproximación, ajustaremos tanto modelos separables como no separables al variograma espacio temporal empírico, y estimaremos sus parámetros así como los de la media espacial, combinando el método de mínimos cuadrados con el algoritmo backfitting. Esta estrategia nos permitirá ver cual es el modelo paramétrico más adecuado.

Una vez hayamos encontrado “el mejor” modelo paramétrico para la covarianza, buscaremos estimaciones más precisas de todos parámetros del modelo, y para ello volveremos a estimar los parámetros utilizando el método de máxima verosimilitud conjuntamente con el algoritmo backfitting.

#### 4.4.1. Modelos de covarianza separables

Una función de covarianza separable, es del tipo:

$$C(h, u/\theta) = C^1(h/\theta_1)C^2(u/\theta_2)$$

(siendo  $C^1(h/\theta_1)$  una función de correlación espacial, y  $C^2(u/\theta_2)$  la temporal). Entonces:

$$\gamma(h, u) = \tau^2 + \sigma^2(1 - \rho_1(h)\rho_2(u))$$

En el análisis exploratorio (sección 4.2) hemos visto que podíamos ajustar la correlación espacial utilizando una función de Matérn. Después, en el análisis exploratorio temporal 4.3, hemos ajustado tres funciones distintas a la estructura de correlación temporal: una función de Matérn, y dos funciones exponenciales. Combinamos todas estas funciones para obtener modelos de correlación espacio-temporal válidos, y los ajustamos por mínimos cuadrados al variograma espacio temporal empírico. Los resultados se pueden ver en la tabla 4.3, y vemos que el mejor ajuste se obtiene al combinar dos funciones de Matérn.

#### 4.4.2. Modelos de covarianza no separables

Como ya hemos comentado, para construir modelos de covarianza no separables, nos basaremos en los trabajos de Cressie y Huang [39] y de Gneiting [66].

Cuadro 4.3: Resultados de ajustar modelos de covarianza separables al variograma espacio temporal empírico . Llamamos  $\rho_1(h) = \text{matern}(h, \phi_1, \kappa_1)$ .

	$\rho_1(h)$ $\rho_2(u) = \text{matern}(u, \phi_2, \kappa_2)$	$\rho_1(h)$ $\rho_2(u) = \exp(-d \cdot u)$	$\rho_1(h)$ $\rho_2(u) = \exp(-d \cdot u^e)$
	$\tau^2 = 1,20e + 09$	$\tau^2 = 1,01e + 09$	$\tau^2 = 1,12e + 09$
	$\sigma^2 = 1,74e + 09$	$\sigma^2 = 2,08e + 09$	$\sigma^2 = 1,79e + 09$
	$\phi_1 = 1,10e + 00$	$\phi_1 = 3,56e + 00$	$\phi_1 = 3,15e + 00$
	$\phi_2 = 6,42e - 01$	$\kappa_1 = 2$	$\kappa_1 = 2$
	$\kappa_1 = 14$	$d = 7,20e - 01$	$d = 7,79e - 01$
	$\kappa_2 = 1,25$		$e = 1,47e + 00$
RSE	193200000	197700000	195900000
GL	1736	1736	1735

En [39], Cressie y Huang dan una serie de condiciones que deben tenerse en cuenta a la hora de construir funciones definidas positivas, y por lo tanto, para construir funciones de covarianza espacio temporales válidas, que ya comentamos en el capítulo 3.

Además del desarrollo teórico, presentan ejemplos de familias espacio temporales. Veremos si podemos utilizar alguna de ellas en nuestra modelización.

Para elegir la función más adecuada, nos fijamos en que nuestro variograma espacial empírico es cóncavo, y consideramos los dos modelos de entre los expuestos por Cressie y Huang que cumplen esta condición. Estos modelos son:

- **Cressie 1**

$$C(h, u | \theta) = \frac{\sigma^2}{(a | u | + 1)} \exp\left\{ - \frac{b^2 \| h \|^2}{a | u | + 1} \right\},$$

$\theta = (a, b, \sigma^2)$ , siendo  $a \geq 0$  un parámetro de escala temporal, y  $b \geq 0$

un parámetro de escala espacial.

En este caso:

$$\gamma(h, u) = \tau^2 + \sigma^2 \cdot \left( 1 - \frac{1}{a |u| + 1} \exp \left\{ - \frac{b \|h\|^2}{a |u| + 1} \right\} \right)$$

▪ **Cressie 2**

$$C(h, u | \theta) = \frac{\sigma^2(a |u| + 1)}{\{(a |u| + 1)^2 + b^2 \|h\|^2\}^{\frac{3}{2}}}$$

$\theta = (a, b, \sigma^2)$ , con  $a \geq 0$  un parámetro de escala temporal, y  $b \geq 0$  un parámetro de escala espacial.

Entonces:

$$\gamma(h, u) = \tau^2 + \sigma^2 \cdot \left( 1 - \frac{a |u| + 1}{\{(a |u| + 1)^2 + b^2 \|h\|^2\}^{\frac{3}{2}}} \right)$$

En un artículo reciente [66], Gneiting da una caracterización más general de estructuras de covarianza no separables, que también comentamos en el capítulo 3. En él propone construir una función definida positiva:

$$C(h, u) = \frac{\sigma^2}{\psi(|u|^2)} \varphi \left( \frac{\|h\|^2}{\psi(|u|^2)} \right), (h, u) \in \mathfrak{R}^2 \times \mathfrak{R}$$

a partir de una función de covarianza puramente espacial  $\varphi(-)$ , y otra puramente temporal  $\psi(-)$ .

Teniendo en cuenta los resultados que hemos obtenido en las secciones anteriores, podríamos aplicar el resultado de Gneiting utilizando una función de Matérn para modelizar la estructura de covarianza espacial, y probando con dos funciones exponenciales diferentes,  $(\rho(u) = \exp(-d \cdot u^2))$  (Gneiting 1) y  $\rho(u) = \exp(-d \cdot u^e)$  (Gneiting 2)) para modelizar la estructura de correlación temporal.

En la tabla 4.4 podemos ver los resultados de los distintos ajustes con modelos de covarianza no separables. Los ajustes se han realizado por el

Cuadro 4.4: Resultados de ajustar varios modelos de covarianza no separables a nuestro variograma espacio temporal empírico .

	Cressie 1	Cressie 2	Gneiting 1	Gneiting 2
	$\tau^2=1.10e+09$	$\tau^2= 1.08e+09$	$\tau^2=1.579e+09$	$\tau^2= 1.147e+09$
	$\sigma^2=2.84e+09$	$\sigma^2= 2.59e+09$	$\sigma^2=1.185e+09$	$\sigma^2= 3.906e+09$
	$a=5.65e-01$	$a= 2.88e-01$	$d=1.036e+00$	$d= 3.174e-01$
	$b=8.57e-03$	$b= 8.08e-03$		$e= 2.885e-01$
			$\phi=3.539e+01$	$\phi= 4.688e+01$
			$\kappa=1$	$\kappa=2$
RSE	201100000	204600000	2.37e+08	205400000
GL	1736	1736	1736	1735

método de mínimos cuadrados. Si comparamos estos resultados con los presentados en la tabla 4.3 vemos que hemos obtenido el mejor ajuste, al suponer una estructura de correlación separable. El producto de dos funciones de Matérn nos ha proporcionado el mejor ajuste a nuestro variograma espacio temporal empírico.

#### 4.4.3. Estimación

Una vez hemos modelizado tanto la estructura de la media como la de la matriz de varianzas covarianzas, nos planteamos la estimación conjunta de todos los parámetros del modelo. Para ello volveremos a combinar el método de máxima verosimilitud con el algoritmo “backfitting”, con el objetivo de obtener estimaciones lo más precisas posible.

Tenemos en un vector todas las observaciones tomadas sobre un mismo paciente

$$Y_i = (Y_{i,1,t_1}, Y_{i,2,t_1}, \dots, Y_{i,52,t_1}, \dots, Y_{i,1,t_{n_i}}, \dots, Y_{i,52,t_{n_i}}).$$

Suponemos que los datos siguen una distribución gaussiana con media:

$$\mu_{i,j,t} = \beta_0 + \beta_1 d_j + \beta_2 r_j + \beta_3 a_{i,t} + f(x_j)$$

y dependiendo de la persona variará la longitud del vector  $Y_i$  (dependiendo del número de CV observados), la varianza para el  $i$ -ésimo paciente será:

$$V_i = \text{Var}(\epsilon) = \nu^2 J_i + \sigma^2 H_i + \tau^2 I_i$$

con  $J_i$  una matriz  $52 \cdot n_i \times 52 \cdot n_i$  de unos,  $I_i$  una matriz identidad  $52 \cdot n_i \times 52 \cdot n_i$ , y siendo  $H_i$  otra matriz  $52 \cdot n_i \times 52 \cdot n_i$ . Si definimos  $d_j^k = \text{distancia}(x_j, x_k)$ ,  $t_l^m = |t_l - t_m|$ ,  $M(h, u) = \text{matern}(h, \phi_1, \kappa_1) \cdot \text{matern}(u, \phi_2, \kappa_2)$ :

$$H_i = \begin{pmatrix} M(d_1^1, t_1^1) & \cdots & M(d_{52}^{52}, t_1^1) & \cdots & M(d_1^1, t_1^{n_i}) & \cdots & M(d_{52}^{52}, t_1^{n_i}) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ M(d_{52}^1, t_1^1) & \cdots & M(d_{52}^{52}, t_1^1) & \cdots & M(d_{52}^1, t_1^{n_i}) & \cdots & M(d_{52}^{52}, t_1^{n_i}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ M(d_1^1, t_{n_i}^1) & \cdots & M(d_{52}^{52}, t_{n_i}^1) & \cdots & M(d_1^1, t_{n_i}^{n_i}) & \cdots & M(d_{52}^{52}, t_{n_i}^{n_i}) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ M(d_{52}^1, t_{n_i}^1) & \cdots & M(d_{52}^{52}, t_{n_i}^1) & \cdots & M(d_{52}^1, t_{n_i}^{n_i}) & \cdots & M(d_{52}^{52}, t_{n_i}^{n_i}) \end{pmatrix}$$

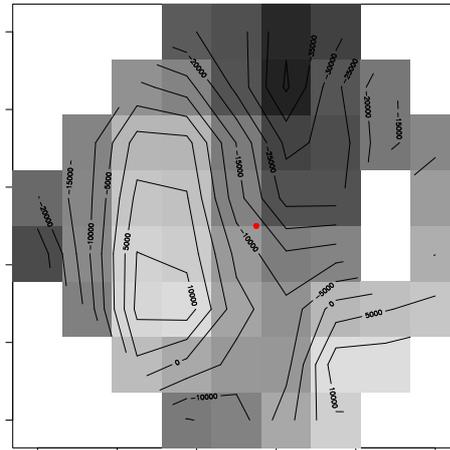
En la tabla 4.5 tenemos las estimaciones de los parámetros que hemos obtenido, y en la figura 4.8 podemos ver la estimación de la superficie suave  $f(x)$ .

#### 4.4.4. Comprobando los residuos

Como una medida para estudiar la bondad del ajuste, podemos calcular los residuos estandarizados:

$$R_i = \hat{V}^{-\frac{1}{2}} (Y_i - \hat{\mu}_i)$$

En el caso de tener un buen ajuste, la media de los residuos debería ser igual a cero y su variograma espacio temporal debería ser una superficie constante igual a 1. Para comprobar estas hipótesis, calculamos la media

Figura 4.8: Estimación de  $f(x)$ .

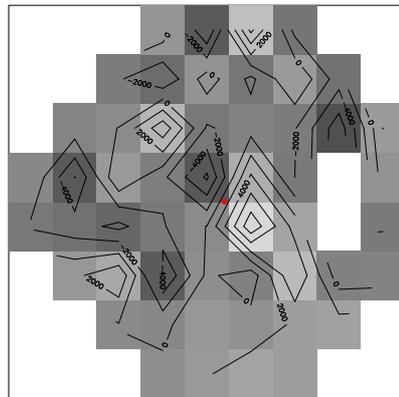
muestral de los residuos estandarizados, obteniendo el valor de  $-0.0036$ , y calculamos y dibujamos el variograma espacio temporal empírico. En la figura 4.9, dibujamos la media en cada posición de los residuos obtenidos, y el variograma espacio temporal empírico. A la vista de los resultados, podemos concluir que el ajuste es bastante bueno.

#### 4.4.5. Validación cruzada

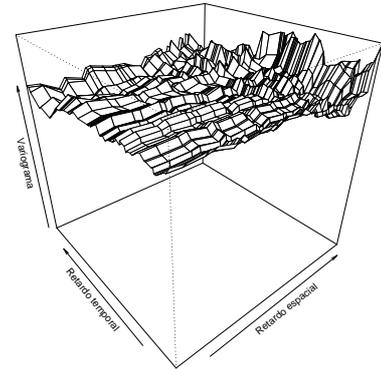
Además de estudiar los residuos del modelo como hemos hecho en la sección anterior, otra forma de comprobar la bondad de un ajuste es mediante un método de validación cruzada [199] que nos permitirá comprobar las

Cuadro 4.5: Estimaciones de los parámetros del modelo, utilizando máxima verosimilitud y el algoritmo backfitting.

Parámetro	Estimación	Parámetro	Estimación
$\beta_0$	374904.3139	$\nu^2$	1.1964e+09
$\beta_1$	-31820.6705	$\sigma^2$	1.1671e+09
$\beta_2$	1777.4231	$\tau^2$	1.0666e+09
$\beta_3$	-1307.0925		
Parámetro	Estimación	Parámetro	Estimación
$\phi_1$	0.4438	$\phi_2$	0.3256
$\kappa_1$	14	$\kappa_2$	1.25



(a)



(b)

Figura 4.9: (a) Media empírica de los residuos estandarizados en cada posición. (b) Variograma espacio temporal empírico calculado sobre los residuos estandarizados.

suposiciones que hemos ido haciendo, tanto sobre el modelo como sobre los datos.

Como comentábamos en el capítulo 3, un estudio de validación cruzada consiste en borrar de la base de datos una observación, volver a estimar los parámetros del modelo a partir de las observaciones restantes, y predecir el valor de la observación que hemos “borrado”. Una vez hecho esto se vuelve a introducir esta observación en la base de datos, se quita otra y se repiten los pasos de estimación de parámetros y de predicción de la observación borrada. Este proceso se va repitiendo iterativamente hasta haberlo hecho sobre todas las observaciones de la base de datos. Una medida del error global cometido (observaciones-predicciones) nos servirá como medida de la bondad del ajuste.

Para hacer este estudio, consideraremos únicamente a aquellos pacientes que tenemos en la base de datos con tres observaciones o más (en estas condiciones tenemos a 17 de los 55 pacientes incluidos inicialmente en el estudio), e iremos extrayendo sucesivamente de la base de datos un campo visual de cada uno de ellos.

Al proceder de esta forma no estaremos haciendo propiamente un estudio de validación cruzada, sino una simplificación del mismo. En un estudio de validación cruzada propiamente dicho, en vez de quitar cada vez todo un campo visual, deberíamos quitar una única observación (posición), es decir en vez de quitar un valor muestral cada vez, estamos quitando cada vez 52 valores muestrales. Quitar un campo visual en cada caso significa repetir los pasos de estimación de parámetros y de predicción tantas veces como campos visuales tenemos en la base de datos. Si en cada paso quitáramos un único valor muestral (una posición), este número quedaría multiplicado por 52, lo cual requeriría mucho tiempo de computación.

En este estudio de validación cruzada, no vamos a tener en cuenta a los 38

pacientes que tenemos en la base de datos con únicamente 2 observaciones ya que si los incluyéramos, al quitar uno de sus campos visuales, no tendríamos suficiente información para estimar los parámetros.

Así pues cada vez quitamos un campo visual, volvemos a estimar todos los parámetros del modelo completo, y estimamos por kriging simple ([199], [68]) los valores del campo visual “eliminado” a partir de las nuevas estimaciones de los parámetros, y de la información proporcionada por el resto de campos visuales que tenemos del paciente en cuestión (recordar que estamos suponiendo independencia entre pacientes).

Si llamamos  $Y_{i_0, \cdot, t_0}^*$  a la estimación por kriging simple del campo visual perteneciente a la  $t_0$ -ésima exploración del paciente número  $i_0$ , entonces:

$$Y_{i_0, \cdot, t_0}^* - \mu_{i_0, \cdot, t_0} = \sum_i \sum_j \sum_t w_{i,j,t} \left( Y_{i,j,t} - \mu_{i,j,t} \right) \quad (4.9)$$

y los pesos  $w_{i,j,t}$  se obtienen minimizando la varianza del error, es decir, minimizando:

$$Var[Y_{i_0, \cdot, t_0}^* - Y_{i_0, \cdot, t_0}] = E[(Y_{i_0, \cdot, t_0}^* - Y_{i_0, \cdot, t_0})^2] \quad (4.10)$$

Podemos utilizar kriging simple porque al haber estimado el valor de todos los parámetros en la sección 4.4.3, podemos considerar la media del modelo conocida.

Una vez quitado cada uno de los campos visuales, re-estimados los parámetros y aplicado el kriging simple para predecirlos, calculamos la media del error de validación cruzada (i.e. la media de las diferencias entre los campos visuales observados y los obtenidos al aplicar el kriging con los nuevos parámetros). Si esta media no está lejos de cero, diremos que no hay un sesgo aparente, e indicará un buen ajuste.

Por otro lado, comparamos la raíz cuadrada de los errores de validación cruzada con la desviación estándar del kriging en cada punto. (Ver en el

capítulo 3 de forma más detallada cómo evaluar la bondad de un modelo utilizando validación cruzada). Si la media de este cociente no dista demasiado de 1, tendremos una evidencia de la adecuación del modelo.

Como resultado de este estudio, obtenemos el valor -71414.67 para la media del error de validación cruzada, y un cociente de desviaciones estándar igual a 2.342. Aunque no hemos obtenido un valor para la media del error próximo a 0, ni un valor para el cociente de desviaciones típicas próximo a 1, creemos que los resultados obtenidos son aceptables, ya que al trabajar con una transformación de Box Cox sobre los datos, estamos trabajando con valores que tienen una magnitud alrededor de  $10^6$ , y además no hemos completado propiamente el estudio de validación cruzada. Además al trabajar con 38 de los 55 pacientes iniciales, tampoco esperábamos obtener resultados mucho mejores.

## 4.5. Posibles aplicaciones del modelo

Este modelo puede proporcionar gran cantidad de aplicaciones. En esta sección vamos a comentar aquellas que nosotros creemos que pueden ser más relevantes.

En principio parece que las aplicaciones más inmediatas del modelo pueden encaminarse hacia tres objetivos muy concretos: la detección de posiciones no sanas, utilizando intervalos de confianza, la predicción de la evolución de los campos visuales de pacientes sanos a lo largo del tiempo, y la simulación de campos visuales, también de pacientes sanos.

Sabemos que cada una de estas aplicaciones necesitaría un estudio experimental mucho más extenso, aunque nosotros nos limitaremos a exponerlas para destacar la utilidad del modelo que hemos estado construyendo a lo largo de todo el capítulo.

Sólo trataremos con detalle una de las tres aplicaciones: la simulación de

campos visuales, ya que a nuestro parecer se trata de una aplicación muy importante, y por ello le dedicaremos el capítulo 5 donde la trataremos con más profundidad.

#### 4.5.1. Construcción de intervalos de confianza para identificar posiciones “enfermas”

A partir del modelo que acabamos de construir, podemos calcular intervalos de confianza, bien para posiciones de forma aislada, o para grupos de posiciones. En particular, si construimos intervalos de confianza para posiciones de forma aislada, dada una nueva observación en dicha posición, podremos clasificarla de inmediato como “normal” o “enferma”, dependiendo de si cae dentro o fuera del intervalo.

Este procedimiento está actualmente incorporado en la mayoría de los paquetes estadísticos comerciales implementados en los campímetros, como el STATPAC [194] o el SITA [15], con la diferencia de que en estos paquetes, los intervalos de confianza para los “valores normales” se han construido empíricamente a partir de un conjunto muy extenso de CV de pacientes normales. En la figura 4.11 analizamos varios CV de nuestra base de datos utilizando los intervalos de confianza construidos a partir de nuestro modelo. Las posiciones “defectuosas” se detectan con una confianza del 95%. En la misma figura presentamos también la clasificación que ofrece el SITA para la misma confianza. Como se puede ver, hay un alto grado de concordancia entre ambas clasificaciones.

#### 4.5.2. Predicción

En este apartado queremos predecir CV de un paciente del que disponemos de observaciones, en un instante en el que no tenemos observaciones (normalmente en el futuro).

Siguiendo la metodología geoestadística clásica, si el objetivo único de nuestro estudio hubiera sido la realización de predicciones, podríamos haber utilizado kriging universal, y habernos “ahorrado” parte del trabajo que hemos ido exponiendo en las secciones anteriores. Este método nos hubiera permitido estimar la media del proceso al mismo tiempo que obteníamos las predicciones.

Sin embargo, debido a la complejidad de la media de nuestro modelo, hemos preferido alejarnos un poco de la Geoestadística clásica, ya que el kriging universal se hubiera complicado muchísimo en este caso. A partir de las estimaciones que hemos obtenido en la sección 4.4.3, podemos considerar la media conocida, y realizar las predicciones utilizando kriging simple (ecuaciones 4.9 y 4.10).

En la figura 4.10 vemos una predicción de un CV de un paciente de 70.5 años. Hemos hecho esta predicción (por kriging simple) utilizando el modelo y la información contenida en los 4 CV que teníamos de este paciente a las edades de 66.8, 68.2, 68.7 y 69.4 años. Esta predicción podría ser guardada por el oftalmólogo y comparada con un futuro CV observado al paciente a la edad de 70.5 años.

Otra aplicación muy interesante sería la predicción en posiciones espaciales no observadas. Este procedimiento posibilitaría el acortar el tiempo de realización de los tests, ya que el número de posiciones a explorar sería menor. Esta posibilidad debería estudiarse con detenimiento, y si se considerara adecuada, podría incorporarse en los distintos paquetes perimétricos comerciales.

### 4.5.3. Simulación

Dejamos esta aplicación para desarrollarla con más detenimiento en el capítulo 5.

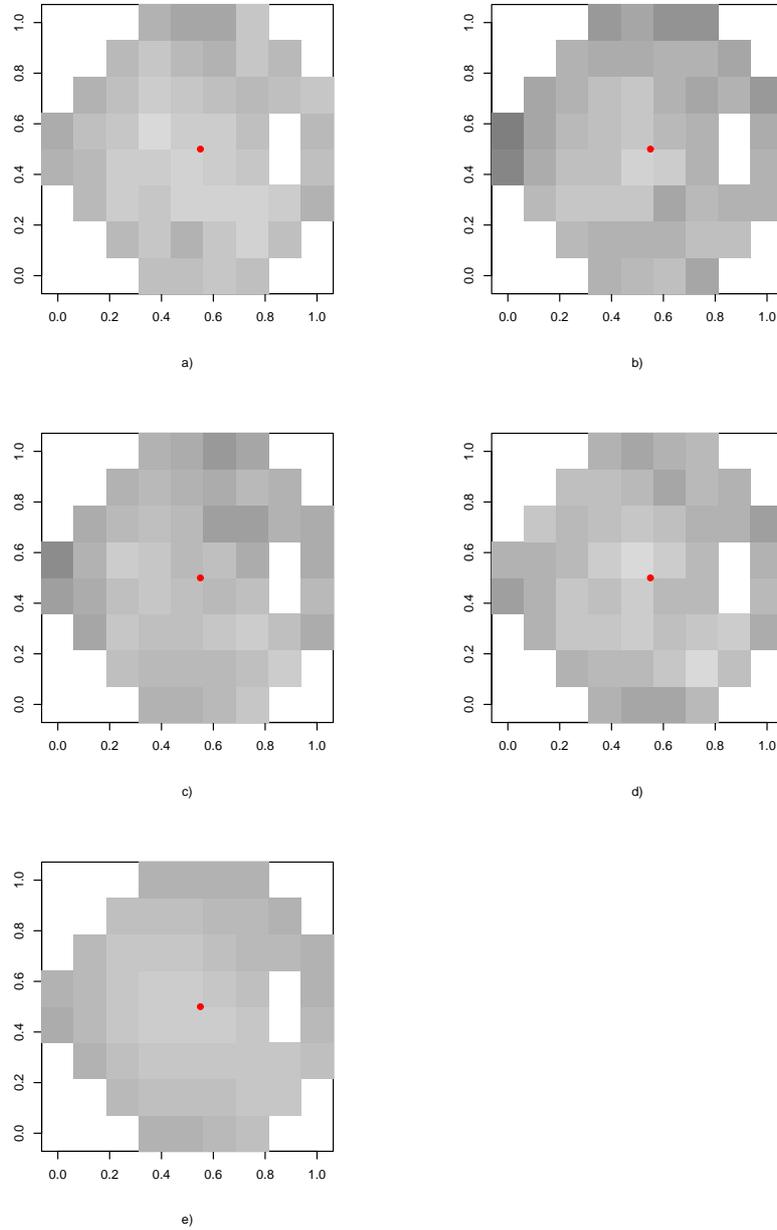


Figura 4.10: En e) tenemos la predicción del CV de un paciente sano a la edad de 70.5 años. En a), b), c) y d) vemos las observaciones que disponíamos de este paciente a las edades de 66.8, 68.2, 68.7 y 69.4 años.

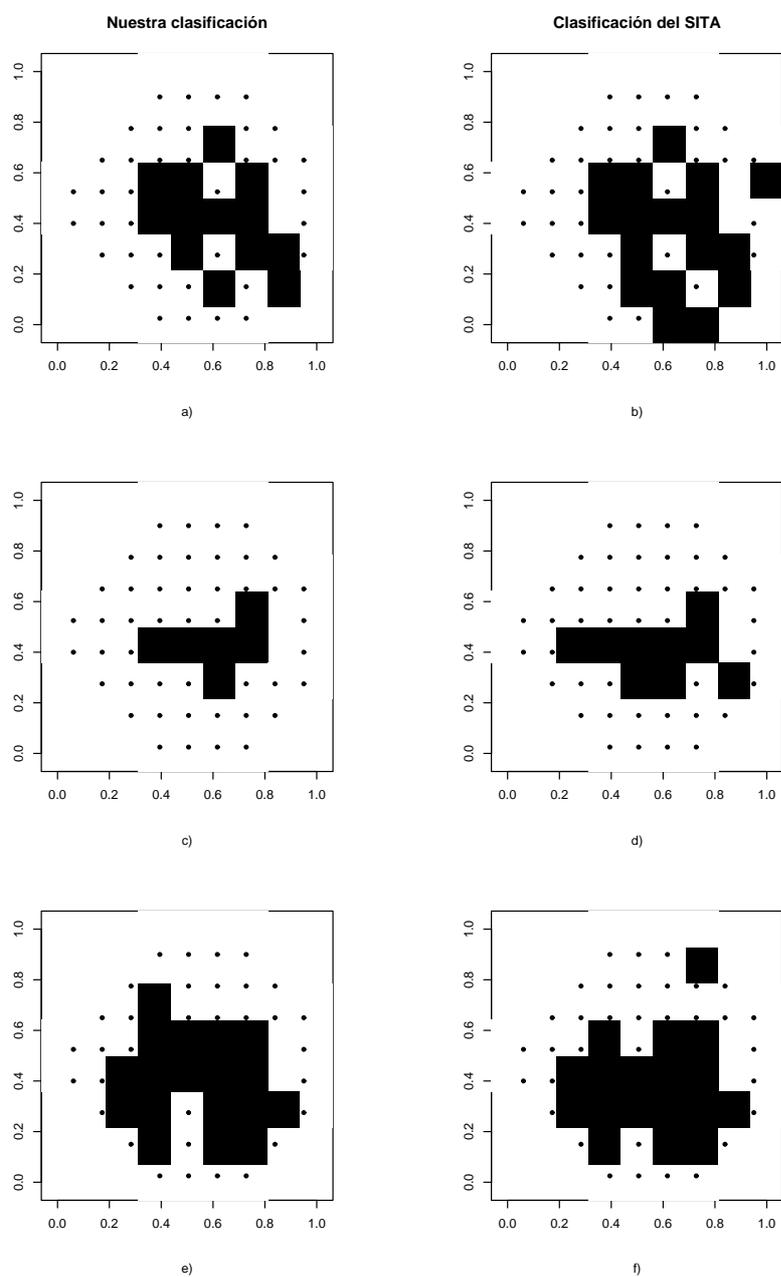


Figura 4.11: Comparación entre la clasificación de posiciones como “normales” o “enfermas”. En a), c) y e) hemos utilizado intervalos de confianza construidos a partir de nuestro modelo, y en b),d) y f) vemos la clasificación dada por el SITA para los mismos CV, ambos trabajando a una confianza del 95 %.



## Capítulo 5

# Simulación condicionada y sin condicionar de campos visuales de pacientes sanos

### Contenidos

---

<b>5.1. Introducción</b> . . . . .	<b>122</b>
5.1.1. Notación . . . . .	126
<b>5.2. Simulación no condicional</b> . . . . .	<b>126</b>
5.2.1. Simulación directa . . . . .	127
5.2.2. Simulación no condicional de campos visuales . . .	129
<b>5.3. Simulación condicional</b> . . . . .	<b>134</b>
5.3.1. Primer método.- Método secuencial gaussiano . . .	136
5.3.2. Segundo método.- Basado en Kriging Simple . . .	136
5.3.3. Simulación condicional de campos visuales . . . .	137

---

## 5.1. Introducción

En el capítulo 4, hemos modelizado la distribución espacio temporal de campos visuales de pacientes sanos. Nuestro objetivo en este capítulo es el utilizar este modelo para realizar simulaciones.

Tal y como explican Vesti et al [197], se necesitan grandes bases de datos con largas series de CV pertenecientes a un mismo paciente, tanto para analizar la eficiencia y eficacia de los distintos métodos que permiten detectar y hacer un seguimiento del glaucoma, como para evaluar y comparar la eficiencia de los nuevos métodos que van surgiendo. Sin embargo, el tiempo necesario que conlleva realizar cada test, y la lentitud con la que evoluciona la enfermedad hacen que sea difícil para el investigador el disponer de un conjunto de datos suficientemente largo como para poder sacar conclusiones a partir de él. Además no hay una “regla de oro” objetiva que nos permita decidir entre la eficacia de dos métodos sobre un conjunto de datos extenso.

En la literatura podemos encontrar dos planteamientos o dos objetivos distintos en la utilización de simulaciones en la investigación relacionada con el análisis de campos visuales.

El primero de los objetivos que encontramos es el de estudiar la eficacia y eficiencia de los distintos procedimientos de test. Cada perímetro lleva implementado un procedimiento de test con el que trabajar. Un procedimiento de test es un software que lleva implementado distintas herramientas y donde se define el funcionamiento del test. Todos los CV de nuestra base de datos han sido adquiridos con campímetros Humphrey y utilizando el test llamado SITA que ya comentamos en el capítulo 2 [15, 14, 12]. Pero continuamente se proponen modificaciones o novedades en los procedimientos de test, buscando siempre conseguir tests de menor duración, que proporcionen tanta información como sea posible. Otros procedimientos de test muy utilizados han sido, el método FASTPAC [81], el STATPAC [194, 93], y el SITA Fast

[13].

Siguiendo esta línea de investigación, en uno de sus trabajos, Turpin et al. [196] explican detalladamente cómo utilizan una de estas técnicas de simulación. Como punto de partida, toman los CV que han obtenido en un análisis exploratorio realizado sobre una muestra amplia tanto de personas normales ( $n_1 = 506$ ), como de enfermos de glaucoma ( $n_2 = 352$ ). El proceso de simulación toma uno de los CV de la base de datos, almacena estos datos como “datos de entrada”, imagina que tiene delante a un paciente y comienza el procedimiento de test. Obviamente considera cada uno de estos “datos de entrada” como el CV real del paciente, por lo que sería deseable que al finalizar el test, el campímetro devolviera estos mismos valores. Entonces la simulación del desarrollo del test empieza eligiendo los primeros puntos a testear, la intensidad con que se mostrarán inicialmente los estímulos, el procedimiento de escalera a utilizar, etc,....

En su versión más sencilla, el proceso de simulación va presentando estímulos en las distintas posiciones, obteniendo una respuesta negativa del paciente (no respuesta) cuando el estímulo presentado tiene una intensidad luminosa menor a la del “dato de entrada” correspondiente, y obteniendo una respuesta positiva cuando el estímulo presenta una intensidad superior a la del “dato de entrada”. Si las dos intensidades coinciden, la respuesta será positiva (o negativa), con una probabilidad del 50%. Al finalizar el test, para probar la precisión del método utilizado compara el CV resultante con el considerado inicialmente y que ha servido para realizar el test.

Esta es una versión muy simplificada del procedimiento de simulación, que normalmente se completa al introducir medidas de error, diferentes reglas de decisión, etc...

Procedimientos de simulación similares se han utilizado también en otros trabajos para comprobar la bondad y la eficacia de distintos procedimientos

de test. Entre ellos, podemos citar el trabajo de Johnson y Shapiro [105], quienes utilizan un programa de simulación llamado KRAKEN [183] para comparar el procedimiento del test de la escalera tradicional, con un nuevo procedimiento de escalera conocido como MOBS (búsqueda binaria modificada). Shapiro y Johnson [182] utilizan de nuevo el mismo programa de simulación para examinar la precisión y reproductibilidad de la perimetría manual kinética. Usando otro simulador de campos visuales, conocido como Barramundi (<http://www.computing.edu.au/~andrew/Barramundi/barramundi.htm>), Turpin et al [196] desarrollan estrategias de umbral para la perimetría Tecnológica de doble frecuencia (Frequency Doubling Technology Perimetry), i.e. utilizan simulación por ordenador para comparar una técnica rápida de búsqueda binaria eficiente (REBS) con un procedimiento basado en la máxima verosimilitud (ZEST).

Un objetivo completamente distinto ha sido el perseguido por Spry et al [189], quienes desarrollaron un programa de simulación de campos visuales (en C++), que simula una serie de campos visuales entre un CV inicial y un CV final que son fijados con anterioridad.

El CV inicial y el final se fijan de antemano, y se simulan nuevos CV entre ellos mediante interpolación, permitiendo un decaimiento lineal, cuadrático, etc... entre los valores de los datos. Su intención, por tanto, es la de simular campos visuales para analizar la evolución de un paciente entre dos exploraciones dadas, espaciadas en el tiempo.

Para tener en cuenta la fluctuación que ocurre a lo largo de una sesión de test, los umbrales interpolados y un término de fluctuación (short term fluctuation value) se utilizan como media y desviación típica respectivamente de una distribución normal. Un valor de esta distribución normal se genera independientemente para cada posición espacial y secuencialmente para todos los campos dentro de las series.

Nuestro objetivo está más próximo a este segundo planteamiento que al primero. En el capítulo 4 hemos modelizado la distribución espacio temporal de CV de pacientes sanos. Nuestra intención en este capítulo es la de simular varios CV en pacientes sanos para reflejar su evolución en el tiempo. ¿Cuáles son las diferencias entre nuestro método de simulación y el propuesto por Spry et al.?

El modelo propuesto por Spry et al realiza simulaciones independientemente sobre cada posición del CV. En ningún momento consideran que los valores de las distintas posiciones de un mismo CV están fuertemente relacionados.

Nosotros tenemos un modelo estadístico espacio temporal a partir del cual simular, considerando la correlación existente entre los valores observados (o simulados) tanto en el espacio como en el tiempo.

La simulación desarrollada por Spry et al. se basa en un procedimiento de interpolación entre un CV inicial y otro final, prefijados de antemano. Nuestra tarea ahora es simular varios CV para un mismo paciente a distintas edades, distinguiendo para ello entre dos procedimientos generales de simulación. En algunos casos el paciente se ha sometido a algún análisis de CV y estamos interesados en ver cómo evoluciona en el tiempo. Otra posibilidad es simular CV de un paciente a distintas edades, pero sin tener ninguna observación inicial.

Como consideramos innecesario recordar las características del modelo espacio temporal a partir del que vamos a realizar las simulaciones, (ya que es el que acabamos de construir en el capítulo 4), organizamos el resto del capítulo en dos secciones. En la sección 5.2, analizaremos los distintos métodos existentes para simular distribuciones gaussianas no condicionalmente, y aplicaremos uno de ellos para simular CV a partir de nuestro modelo. En la sección 5.3, veremos dos métodos distintos para simular campos aleato-

rios gaussianos condicionalmente, y finalmente aplicaremos uno de ellos para obtener las simulaciones que nos interesan.

### 5.1.1. Notación

Recordamos brevemente la notación que utilizamos en el capítulo 4, y que seguiremos utilizando en este capítulo.

Denotamos por  $Y_{i,j,t}$  a la observación número  $t$  del paciente número  $i$ , en la posición número  $j$ . Llamaremos  $Y_{i,\cdot,t} = (Y_{i,1,t}, Y_{i,2,t}, \dots, Y_{i,52,t})'$  al vector con los 52 valores obtenidos en el test número  $t$  para el paciente número  $i$ ,  $i \in \{1, \dots, N\}$ ,  $t \in \{t_1, \dots, t_{n_i}\}$ , y llamaremos  $Y_i$  al vector con todas las observaciones que tenemos del paciente número  $i$ :

$$Y_i = (Y_{i,1,t_1}, Y_{i,2,t_1}, \dots, Y_{i,52,t_1}, \dots, Y_{i,1,t_{n_i}}, \dots, Y_{i,52,t_{n_i}})$$

Identificamos el “tiempo”  $t$  con la edad del paciente.

Como ya tenemos el modelo probabilístico completamente determinado, a la hora de enfrentarnos a la simulación, deberemos elegir entre dos estrategias dependiendo del fin que nos propongamos. Si estamos interesados en simular CV de un mismo paciente sano a distintas edades, podemos utilizar alguno de los métodos disponibles para simular de una forma no condicional (sección 5.2). Alternativamente si hemos observado un CV en un paciente sano y queremos simular su evolución en el tiempo, podemos utilizar métodos para simular condicionalmente (sección 5.3).

## 5.2. Simulación no condicional

Nuestro objetivo en esta sección es revisar los métodos que permiten simular una función aleatoria gaussiana estacionaria, con media  $\mu$ , y función de covarianza  $C$ .

En la literatura podemos encontrar varios métodos para simular funciones aleatorias gaussianas o vectores aleatorios [38, 179, 124, 43].

Alguno de ellos, como el “método espectral”, desarrollado por Shinozuka y Jan [184] entre otros, y el “Turning Bands Method” desarrollado por Matheron [143], simulan procesos aleatorios, no gaussianos en general, con media cero y covariograma  $C(-)$ . En estos casos, se puede aplicar el teorema central del límite para obtener las simulaciones del proceso gaussiano que nos interesa. (Consultar [38, 179, 124] para más detalles.)

Si fijamos la dimensión del proceso a simular y ésta no es demasiado grande, podemos simular directamente distribuciones gaussianas utilizando un método muy intuitivo. Lo veremos en la sección 5.2.1, y en la sección 5.2.2 lo aplicaremos a nuestro problema. (Una vez más, recomendamos consultar [38, 179, 124] para encontrar explicaciones más detalladas.)

### 5.2.1. Simulación directa

Queremos simular en  $n$  localizaciones espaciales específicas  $x_1, x_2, \dots, x_n$ , un proceso gaussiano  $Y = (Y_1, \dots, Y_n)$  con una cierta media  $\mu$ , y matriz de covarianza  $V = (v_{i,j})_{i,j \in \{1, \dots, n\}}$ , (en particular queremos que  $v_{i,j} = C(Y_i, Y_j)$ ).

Es relativamente sencillo simular un vector aleatorio  $Z = (Z_1, \dots, Z_n)$ , con  $(Z_1, \dots, Z_n)$  componentes independientes generadas a partir de una distribución gaussiana con media 0 y varianza 1,  $Z_i \sim N(0, 1)$ ,  $\forall i \in \{1, \dots, n\}$  i.i.d., por tanto, si encontramos una matriz  $W$  tal que  $W \cdot W' = V$  entonces podremos definir  $Y$  como:

$$Y = \mu + WZ \tag{5.1}$$

y nuestro problema de simulación quedará resuelto.

Hay distintos métodos para obtener  $W$ . El primero sería utilizar la des-

composición de Cholesky para descomponer  $V$  como:

$$V = WW'$$

con  $W$  una matriz triangular inferior  $n \times n$ .

Pero si  $n$  es mayor que 1000 y  $V$  es una matriz dispersa, utilizar este método puede llevarnos a obtener resultados numéricos imprecisos.

El segundo de los métodos propuesto para obtener  $W$  se basa en la descomposición espectral de  $V$ .

Si llamamos  $\lambda_1, \lambda_2, \dots, \lambda_n$  a los valores propios de  $V$ , y  $Q$  es una matriz ortogonal formada por los vectores propios de  $V$  como columnas, entonces:

$$V = Q \cdot \text{diag}\{\lambda_1, \dots, \lambda_n\} \cdot Q'$$

y podemos definir:

$$W = Q \cdot \text{diag}\{\lambda_1^{1/2}, \lambda_2^{1/2}, \dots, \lambda_n^{1/2}\} \quad (5.2)$$

El tercer método para calcular  $W$  se basa en el hecho de que la ecuación (5.1) puede formularse alternativamente como una convolución de las variables aleatorias  $\{Z_i : i \in \{1, \dots, n\}\}$  generadas independientemente. Por tanto es equivalente (5.1) a la siguiente ecuación:

$$Y(r) = \mu + (w * Z)(r) = \mu + \sum_{i \in \{x_1, \dots, x_n\}} w(r-i) \cdot Z(i), \forall r \in \{x_1, \dots, x_n\}$$

Debemos imponer dos restricciones sobre  $w : \sum_h w^2(h) = 1$  y  $w(r) = w(-r), \forall r$ . Entonces, como  $Y(r)$  será gaussiana, por ser suma de variables gaussianas, y

$$E[Y(r)] = \mu + \sum_{i \in \{x_1, \dots, x_n\}} w(r-i) \cdot E[Z(i)] = \mu$$

$$\begin{aligned} \text{Cov}[Y(r), Y(r+u)] &= \sum_{i, j \in \{x_1, \dots, x_n\}} w(r-i) \cdot w(r+u-j) \cdot \text{Cov}[Z(i), Z(j)] = \\ &= \sum_{i \in \{x_1, \dots, x_n\}} w(r-i) \cdot w(r-i+u) = (w * w)(u) \end{aligned}$$

Si formulamos  $Cov[Y(r), Y(r + u)] = \rho(u)$ , siendo  $\rho(-)$  una función de covarianzas conocida, entonces:

$$(w * w)(u) = \rho(u)$$

Aplicando una transformación de Fourier (por ejemplo la transformada rápida de Fourier, “Fast Fourier transform”) a ambos lados de la igualdad, tendremos:

$$FFT((w * w)(u)) = FFT(\rho(u))$$

y como la transformada de Fourier de una convolución, equivale al producto de las transformadas de Fourier:

$$w = FFT^{-1} \sqrt{FFT(\rho(u))}$$

### 5.2.2. Simulación no condicional de campos visuales

Una vez hemos visto cómo simular campos aleatorios gaussianos, veamos cómo aplicarlos a la simulación de campos visuales.

Queremos utilizar el modelo desarrollado en el capítulo 4 para simular CV pertenecientes a un hipotético paciente sano.

Sea  $N$  el número de CV que queremos simular de esta persona hipotética, y sea  $t_i$  su edad en cada simulación. Como en cada CV hay 52 posiciones, simularemos un total de  $52 \times N$  valores, i.e. queremos simular un proceso gaussiano en las  $52 \times N$  posiciones:  $\{(1, t_1), (2, t_1), \dots, (52, t_1), (1, t_2), \dots, (52, t_2), \dots, (1, t_N), \dots, (52, t_N)\}$  con media, varianza y covarianza (ver capítulo 4):

$$\mu_{j,t_i} = \beta_0 + \beta_1 \cdot d_j + \beta_2 \cdot r_j + \beta_3 \cdot t_i + f(j), \quad (5.3)$$

$$var(Y_{j,t_i}) = \nu^2 + \sigma^2 + \tau^2, \quad (5.4)$$

$$Cov(Y_{j,t_i}, Y_{k,t_l}) = \nu^2 + \sigma^2 \rho(\text{distancia}(j, k), |t_i - t_l|), \quad (5.5)$$

Cuadro 5.1: Parámetros del modelo de CV

Parámetro	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
Valor estimado	374904.3139	-31820.6705	1777.4231	-1307.0925
Parámetro	$\nu^2$	$\sigma^2$	$\tau^2$	
Valor estimado	1.1964e+09	1.1671e+09	1.0666e+09	
Parámetro	$\phi_1$	$\kappa_1$	$\phi_2$	$\kappa_2$
Valor estimado	0.4438	14	0.3256	1.25

donde hemos eliminado el subíndice  $i$  de la notación que utilizábamos en el capítulo 4, porque vamos a simular CV de un único paciente. Hemos llamado  $d_j$  a la distancia euclídea de la posición espacial número  $j$  al centro del CV;  $r_j$  a la fila donde está la posición número  $j$ ;  $f(j)$  a la superficie suave que ya estimamos en el capítulo 4 (ver figura 4.8);  $\rho(u, v) = \text{matern}(u, \phi_1, \kappa_1) \cdot \text{matern}(v, \phi_2, \kappa_2)$  y  $(\beta_0, \beta_1, \beta_2, \beta_3, \nu^2, \sigma^2, \tau^2, \phi_1, \phi_2, \kappa_1, \kappa_2)$  al vector de parámetros que toman los valores que se detallan en la tabla 5.1.

Si utilizamos el índice  $k = 1, \dots, 52 \times N$  para enumerar las  $52 \times N$  localizaciones, podemos definir  $Y_{(i-1)*52+j} = Y_{j,t_i}$ , y utilizando la notación que hemos introducido en la sección anterior, nuestro problema se reduce a simular un vector gaussiano  $Y = (Y_1, \dots, Y_{52N})'$  con vector de medias  $\mu = (\mu_1, \dots, \mu_{52N})'$ , (siendo  $\mu_{(i-1)*52+j} = \beta_0 + \beta_1 d_j + \beta_2 r_j + \beta_3 t_i + f(j)$ ), y matriz de covarianzas  $V$  definida como:

$$V = \nu^2 \mathbf{1}_{52N \times 52N} + \sigma^2 H + \tau^2 I \quad (5.6)$$

Si desarrollamos la expresión de  $V$ , denotando por  $N$  al número total de CV a simular, por  $d_i^j = \text{distancia}(i, j)$  a la distancia euclídea entre la  $i$ -ésima y la  $j$ -ésima posición espacial, y por  $t_l^m = |t_l - t_m|$  al tiempo en valor absoluto transcurrido entre la  $l$ -ésima y la  $m$ -ésima exploración, nos

quedará la siguiente expresión:

$$V = \begin{pmatrix} \nu^2 + \sigma^2 + \tau^2 & \nu^2 + \sigma^2 \rho(d_1^2, t_1^1) & \dots & \nu^2 + \sigma^2 \rho(d_1^{52}, t_1^N) \\ \nu^2 + \sigma^2 \rho(d_1^2, t_1^1) & \nu^2 + \sigma^2 + \tau^2 & \dots & \nu^2 + \sigma^2 \rho(d_2^{52}, t_1^N) \\ \vdots & \vdots & \dots & \vdots \\ \nu^2 + \sigma^2 \rho(d_1^{52}, t_1^1) & \nu^2 + \sigma^2 \rho(d_2^{52}, t_1^1) & \dots & \nu^2 + \sigma^2 \rho(d_{52}^{52}, t_1^N) \\ \nu^2 + \sigma^2 \rho(d_1^1, t_2^1) & \nu^2 + \sigma^2 \rho(d_1^2, t_2^1) & \dots & \nu^2 + \sigma^2 \rho(d_1^{52}, t_2^N) \\ \vdots & \vdots & \vdots & \vdots \\ \nu^2 + \sigma^2 \rho(d_1^{52}, t_2^1) & \nu^2 + \sigma^2 \rho(d_2^{52}, t_2^1) & \dots & \nu^2 + \sigma^2 \rho(d_{52}^{52}, t_2^N) \\ \vdots & \vdots & \vdots & \vdots \\ \nu^2 + \sigma^2 \rho(d_1^1, t_1^N) & \nu^2 + \sigma^2 \rho(d_1^2, t_1^N) & \dots & \nu^2 + \sigma^2 \rho(d_1^{52}, t_1^N) \\ \vdots & \vdots & \vdots & \vdots \\ \nu^2 + \sigma^2 \rho(d_1^{52}, t_1^N) & \nu^2 + \sigma^2 \rho(d_2^{52}, t_1^N) & \dots & \nu^2 + \sigma^2 + \tau^2 \end{pmatrix}$$

Para elegir el algoritmo de simulación, tendremos en cuenta el número total,  $N$ , de CV que queremos simular. Para clarificar todo el proceso, veamos un ejemplo. Vamos a simular  $N = 3$  CV de un paciente a las edades de 42, 64 y 76 años. Veamos cómo quedan las primeras filas y columnas de la matriz de varianzas-covarianzas ( $V$ ), cuando particularizamos a este caso (3 simulaciones a edades  $t_1 = 42$ ,  $t_2 = 64$  y  $t_3 = 76$ ). En este caso particular:

$$V = \begin{pmatrix} 3,4 & 2,2 & 1,9 & 1,7 & 2,1 & 2,2 & 2,1 & 1,9 & \dots & 1,2 \\ 2,2 & 3,4 & 2,2 & 1,9 & 1,9 & 2,1 & 2,2 & 2,1 & \dots & 1,2 \\ 1,9 & 2,2 & 3,4 & 2,2 & 1,6 & 1,9 & 2,1 & 2,2 & \dots & 1,2 \\ 1,7 & 1,9 & 2,2 & 3,4 & 1,4 & 1,6 & 1,9 & 2,1 & \dots & 1,2 \\ 2,1 & 1,9 & 1,6 & 1,4 & 3,4 & 2,2 & 1,9 & 1,6 & \dots & 1,2 \\ 2,2 & 2,1 & 1,9 & 1,6 & 2,2 & 3,4 & 2,2 & 1,9 & \dots & 1,2 \\ 2,1 & 2,2 & 2,1 & 1,9 & 1,9 & 2,2 & 3,4 & 2,2 & \dots & 1,2 \\ \vdots & \ddots & \vdots \\ 1,2 & 1,2 & 1,2 & 1,2 & 1,2 & 1,2 & 1,2 & 1,2 & \dots & 3,4 \end{pmatrix} \times 10^9 \quad (5.7)$$

Como el número de campos visuales a simular,  $N$ , no es demasiado grande, vamos a utilizar el método de “simulación directa”. En él debemos descomponer la matriz de varianzas (ecuación 5.6) como  $V = WW'$ , y para ello utilizaremos el segundo de los métodos comentados (ecuación 5.2), i.e. utilizaremos la descomposición espectral de  $V$ , en valores y vectores propios, y calcularemos  $W$  aplicando la ecuación 5.2. Una vez calculadas  $W$  y  $\mu$ , obtenemos las simulaciones utilizando la ecuación 5.1.

En la figura 5.1, podemos ver las simulaciones de varios CV obtenidas para tres pacientes hipotéticos, todas ellas a las edades de 42, 64 y 76 años.

¿Como evaluar la bondad de estas simulaciones? En el capítulo 4, tras construir el modelo realizamos un análisis de residuos y un estudio de validación cruzada para comprobar su bondad. Por otro lado, todos los algoritmos de simulación que hemos utilizado han sido ampliamente estudiados en la literatura, por lo que si los combinamos, esperamos obtener buenos resultados. Pero de todas formas nos gustaría verificarlos.

Una de las formas más sencillas de evaluar estas simulaciones es preguntando su opinión a oftalmólogos expertos. Ellos están acostumbrados a ver gran cantidad de CV cada día, y pueden decirnos si bajo su opinión los CV que hemos obtenido podrían considerarse “realistas” o si no. Al mostrar nuestras simulaciones a dos oftalmólogos expertos, efectivamente las encontraron bastante realistas.

Además de esta evaluación “cualitativa” nos planteamos el evaluar la “credibilidad” de las simulaciones por medio de un procedimiento gráfico, similar a un test no paramétrico de Montecarlo [47, 42]. Este procedimiento se basa en las siguientes ideas. Normalmente en la estadística espacial y en el análisis de series temporales se dispone únicamente de una realización del proceso en cada posición de monitorización, y se realizan inferencias a partir de esta única observación. Siguiendo esta idea, podríamos decir que esta-

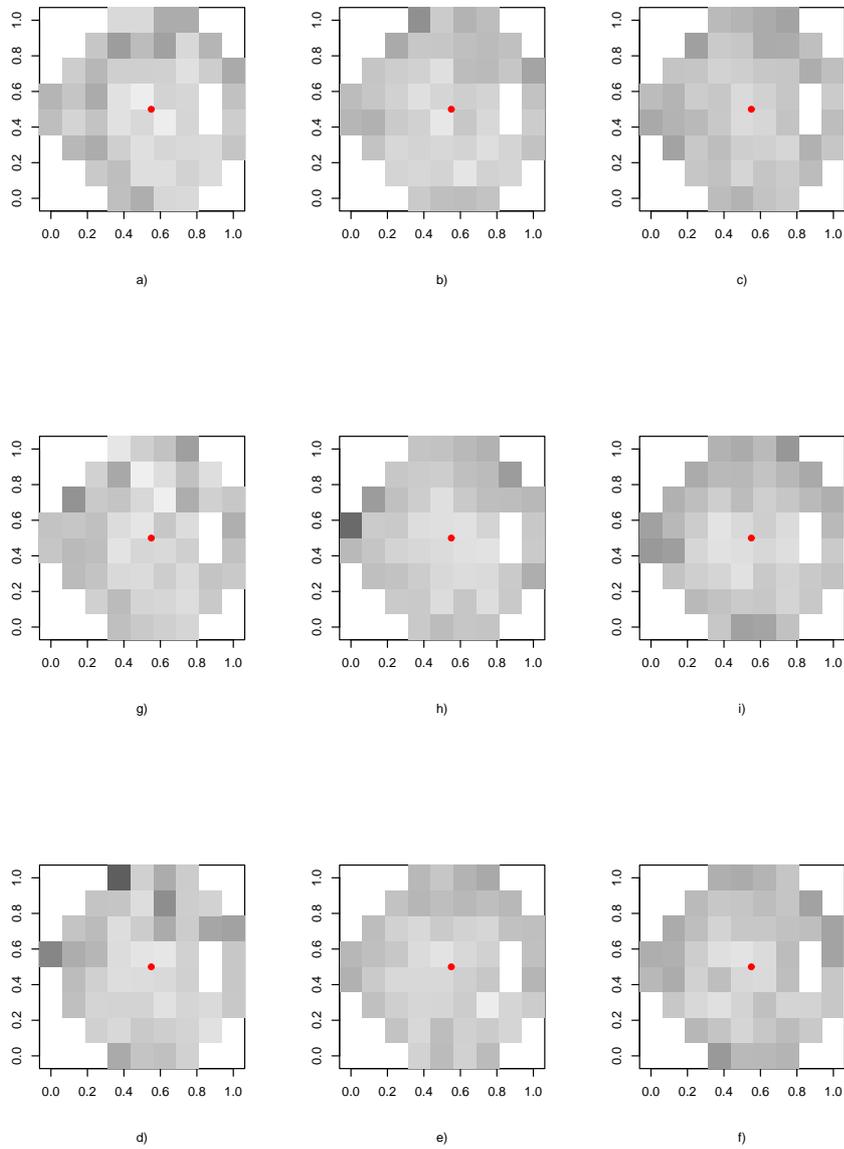


Figura 5.1: Simulaciones de tres diferentes pacientes de la misma edad. En a), g), d) los tres tienen 42 años, en b), h) y e) tienen 64, y en c), i) y f) tienen 76.

mos interesados en probar que el conjunto de  $N$  campos visuales simulados se ajustan a un modelo particular. Bajo esta interpretación, necesitaremos un test de bondad de ajuste. Si no queremos hacer ninguna suposición sobre el modelo, deberíamos basar el test sólo en el conjunto de datos original. De esta forma deberíamos considerar cada CV observado para cada paciente como una muestra simulada de este modelo desconocido, y podríamos aplicar un test gráfico de Montecarlo. Para ello, debemos elegir un estadístico cualquiera, calcularlo para cada paciente, y dibujar las envolventes superior e inferior de los valores obtenidos. Entonces, calcular el mismo estadístico sobre los datos que hemos obtenido tras la simulación y ver si el valor obtenido está entre las dos envolventes.

Para aplicar este procedimiento a nuestros datos, elegimos como estadístico el variograma espacial dividido por el cuadrado de la media, para eliminar así el efecto de la edad. Podemos ver los resultados en la figura 5.2. Podemos ver claramente que en todos los casos el estadístico (variograma) calculado sobre los datos simulados, está entre las dos envolventes.

### 5.3. Simulación condicional

Al igual que en la sección anterior, nuestra intención es simular un proceso aleatorio estacionario gaussiano  $Y$ , con media  $\mu$ , y función de covarianzas  $C$ , pero ahora tenemos una limitación adicional. Queremos simular  $Y$ , pero verificando que en un conjunto cerrado de puntos  $E$ , los valores de  $Y$  sean fijos, i.e.  $Y(x) = Y_0(x)$ ,  $\forall x \in E$  valores conocidos, o variables aleatorias previamente simuladas.

Estos procedimientos de simulación serán especialmente interesantes para nosotros, ya que si tenemos un CV de un paciente a la edad  $det_1$  años, seremos capaces de simular su evolución en el tiempo, en sucesivas exploraciones.

En la literatura, hemos encontrado dos métodos distintos para simular

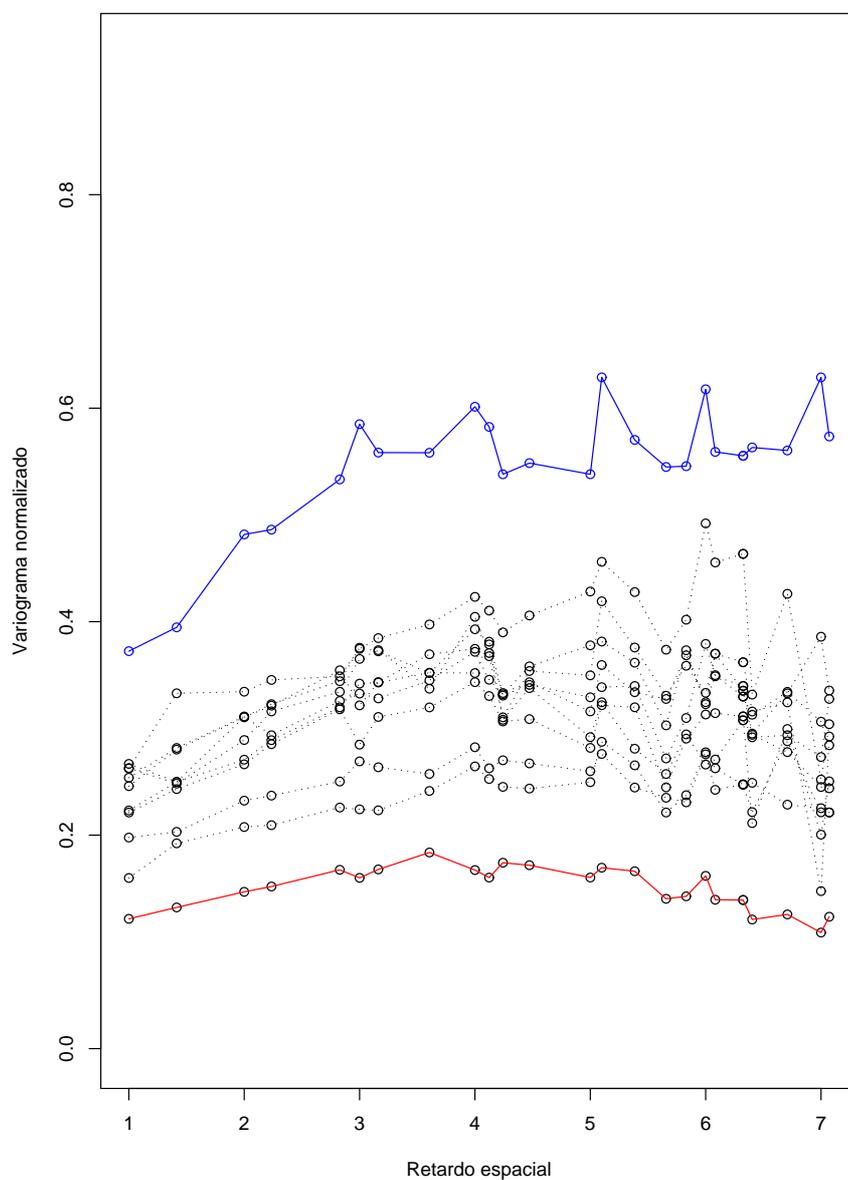


Figura 5.2: Las líneas continuas muestran las envolventes superior e inferior del variograma espacial normalizado calculado a partir de los 56 pacientes de nuestra base de datos. Con líneas discontinuas tenemos los variogramas normalizados calculados a partir de los campos visuales simulados de la figura 5.1.

campos aleatorios gaussianos condicionalmente. Vamos a describirlos brevemente y luego los aplicaremos para obtener las simulaciones que nos interesan. (Podemos encontrar más detalles sobre estos métodos en [38, 179, 124]).

### 5.3.1. Primer método.- Método secuencial gaussiano

Sea  $Y_1 = \{Y(e) : e \in E\}$  el conjunto de valores conocidos, o las variables aleatorias previamente simuladas, sea  $Y_2 = \{Y(p) : p \notin E\}$  el conjunto de variables aleatorias a simular, y sea  $Y = Y_1 \cup Y_2$ .

El método secuencial gaussiano va simulando punto a punto sucesivamente, en vez de simularlos todos a la vez, y para ello trabaja con distribuciones condicionales dados los valores ya simulados.

Siguiendo nuestra notación, queremos simular  $Y_2$  dado  $Y_1$ , y:

$$Y_2 | Y_1 \sim N(S_{21}S_{11}^{-1}Y_2, S_{22} - S_{21}S_{11}^{-1}S_{12})$$

siendo  $S_{ij} = Cov(Y_i, Y_j)$

Si definimos  $\mu := S_{21}S_{11}^{-1}Y_2$  y  $V = S_{22} - S_{21}S_{11}^{-1}S_{12}$ , podemos utilizar el método de simulación directa que hemos comentado en la sección 5.2.1 para obtener las simulaciones de  $Y_2$ .

### 5.3.2. Segundo método.- Basado en Kriging Simple

Siguiendo la misma notación, tenemos  $Y_1 = \{Y(e) : e \in E\}$  el conjunto de valores conocidos, o de variables aleatorias ya simuladas,  $D$  todos los puntos en los que queremos simular (incluyendo a  $E$ , i.e.  $E \subset D$ ), e  $Y_2 = \{Y(p) : p \in D \setminus E\}$  el conjunto de variables aleatorias a simular,  $Y = Y_1 \cup Y_2$ .

Utilizando la información contenida en  $Y_1$ , podemos calcular el predictor del kriging simple [68]:

$$Y^*(x) = \sum_{e \in E} \lambda_e(x) Y_1(e), \forall x \in D \quad (5.8)$$

y expresar  $Y$  como la suma de dos funciones gaussianas independientes,  $Y^*$  e  $Y - Y^*$  (la demostración de su independencia se puede encontrar en [38] y en [124]):

$$Y(x) = Y^*(x) + \left( Y(x) - Y^*(x) \right), \forall x \in D$$

El segundo término de esta suma, se puede reemplazar por un término basado en una simulación no condicional, para ello podemos utilizar cualquier método de simulación no condicional y simular un campo aleatorio gaussiano  $\{Z(x), x \in D\}$  con media 0 y covarianza  $C$  en todo  $D$ . Una vez simulado  $\{Z(x)\}$ , podemos distinguir entre  $Z_1 = \{Z(e) : e \in E\}$ , y  $Z_2 = \{Z(p) : p \in D \setminus E\}$ , y podemos calcular los siguientes estimadores kriging:

$$Z^*(x) = \sum_{e \in E} \lambda_e(x) Z_1(e), \forall x \in D$$

La simulación condicional [107], se define ahora como:

$$Y(x) = Y^*(x) + \left( Z(x) - Z^*(x) \right), \forall x \in D \quad (5.9)$$

Si  $x \in E$ , entonces  $Y^*(x) = Y(x)$  y  $Z^*(x) = Z(x)$ , por tanto:

$$Y(x) = Y(x) + Z(x) - Z(x) = Y(x)$$

### 5.3.3. Simulación condicional de campos visuales

Nos planteamos de nuevo la simulación de varios CV de un paciente a distintas edades, pero ahora vamos a tener un punto de partida.

Suponemos que el paciente ya se ha sometido alguna vez a tests de campo visual, y tenemos uno de sus campos visuales. Nuestro objetivo ahora será simular la evolución de éste campo visual en el tiempo. Para ello denotaremos por  $t_1$  a la edad del paciente cuando se observó este CV que tomaremos como inicio de nuestro estudio, y por  $t_2, \dots, t_N$  a las edades en las que queremos simular.

Por tanto, al igual que en la sección 5.2.2, queremos simular un proceso gaussiano sobre  $52 \times N$  posiciones,  $\{(1, t_1), (2, t_1), \dots, (52, t_1), (1, t_2), \dots, (52, t_2), \dots, (1, t_N), \dots, (52, t_N)\}$  con media, varianza y covarianza como en las ecuaciones 5.3, 5.4 y 5.5. Pero ahora siendo  $\{Y(1, t_1), Y(2, t_1), \dots, Y(52, t_1)\}$  valores conocidos, e iguales a los obtenidos en el CV del paciente.

Para ver un ejemplo, cogemos un CV de una persona de las incluidas en nuestra base de datos, es un paciente con 49 ños y vamos a suponer que estamos interesados en simular la evolución de este CV a las edades de 50, 51 y 52. Por tanto,  $N = 4$ ,  $t_1 = 49$ ,  $t_2 = 50$ ,  $t_3 = 51$  y  $t_4 = 52$ . Podemos ver el CV inicial en la figura 5.3 a). Utilizamos el segundo de los métodos propuestos para llevar a cabo las simulaciones. Como primer paso, obtenemos las estimaciones por simple kriging (ec. 5.8) siendo  $\lambda_e(x)$  el  $e$ -ésimo elemento del vector:

$$\Lambda_e = V_{52 \times 52}^{-1} V_x$$

donde  $V_{52 \times 52}$  es el primer bloque cuadrado de tamaño 52 de la matriz  $V$  dada en la tabla 5.7, y donde  $V_x$  es un vector que contiene los primeros 52 elementos de la  $x$ -ésima fila de la matriz  $V$ . En la figura 5.4 podemos ver las distintas  $Z^*(x)$  que hemos obtenido (estimaciones kriging) .

A continuación, utilizamos uno de los métodos expuestos en la sección 5.2.2 para simular un campo aleatorio gaussiano  $Z$ , con media 0, y matriz de varianzas  $V$  (como en la tabla 5.4). En la figura 5.5 podemos ver las distintas  $Z$  que hemos obtenido.

El tercer paso consiste en obtener las estimaciones kriging  $Z^*$  para  $Z$ , de nuevo con los mismos coeficientes  $\lambda_e(x)$  que utilizamos en el paso 2. Las tenemos en la figura 5.6.

Por último podemos obtener ya las simulaciones  $Y^*$  (ecuación 5.9, figura 5.3).

En las figuras 5.7 y 5.8 podemos ver más simulaciones condicionales,

obtenidas a partir de CV de otros pacientes. En ellas vemos también la evolución en el tiempo de los CV de dos pacientes distintos, con un retardo temporal de 1 año entre dos simulaciones consecutivas.

Para evaluar las simulaciones, en este caso podemos compararlas con nuestros datos reales. En la figura 5.9 podemos ver 3 series distintas, con 4 CV cada una. La primera de las series muestra los datos “reales” (observados) de un paciente, y la segunda y la tercera son simulaciones condicionadas al primer CV de ese paciente.

Hemos de tener en cuenta que estamos trabajando con simulaciones, no estamos prediciendo futuras observaciones a partir de las primera, por lo tanto necesariamente existirá una diferencia entre los CV “reales” y los “simulados”, debido a las distintas fuentes de aleatoriedad existentes.

Por lo tanto, como se puede ver, los CV simulados son similares a los reales, pero no idénticos.

Como medida numérica de la bondad de las estimaciones, vamos a calcular el error cuadrático medio normalizado entre las medias de 50 CV simulados (como los que mostramos en la figura 5.9), y los valores reales:  $ECM = \frac{1}{3 \times 52} \sum \frac{(Y_{jt} - \bar{Y}_{jt}^*)^2}{Y_{ijt}^2}$ , siendo  $Y^*$  los datos simulados. El ECM (error cuadrático medio) obtenido fue 0.53.

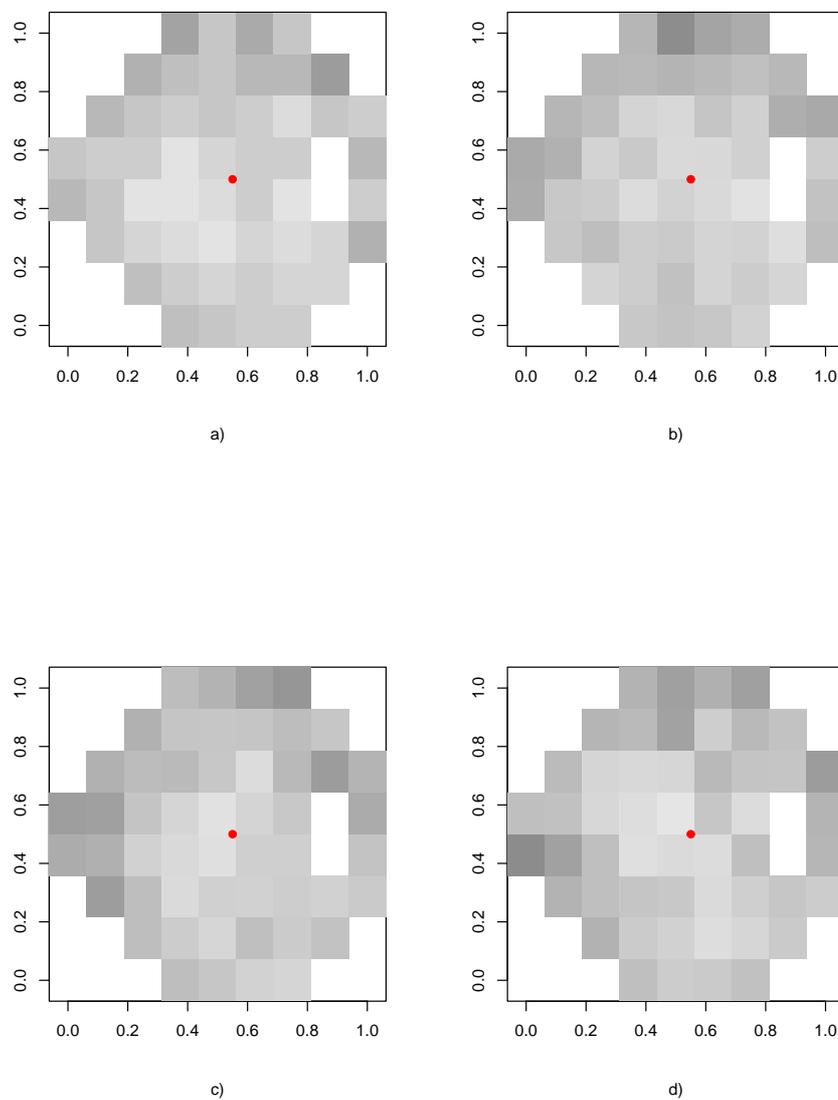


Figura 5.3: En a) tenemos el CV observado en un paciente de 49 años. En b), c) y d) mediante simulación condicionada hemos simulado CV a partir del CV que tenemos en a) con un retardo temporal de 1 año entre dos simulaciones consecutivas.

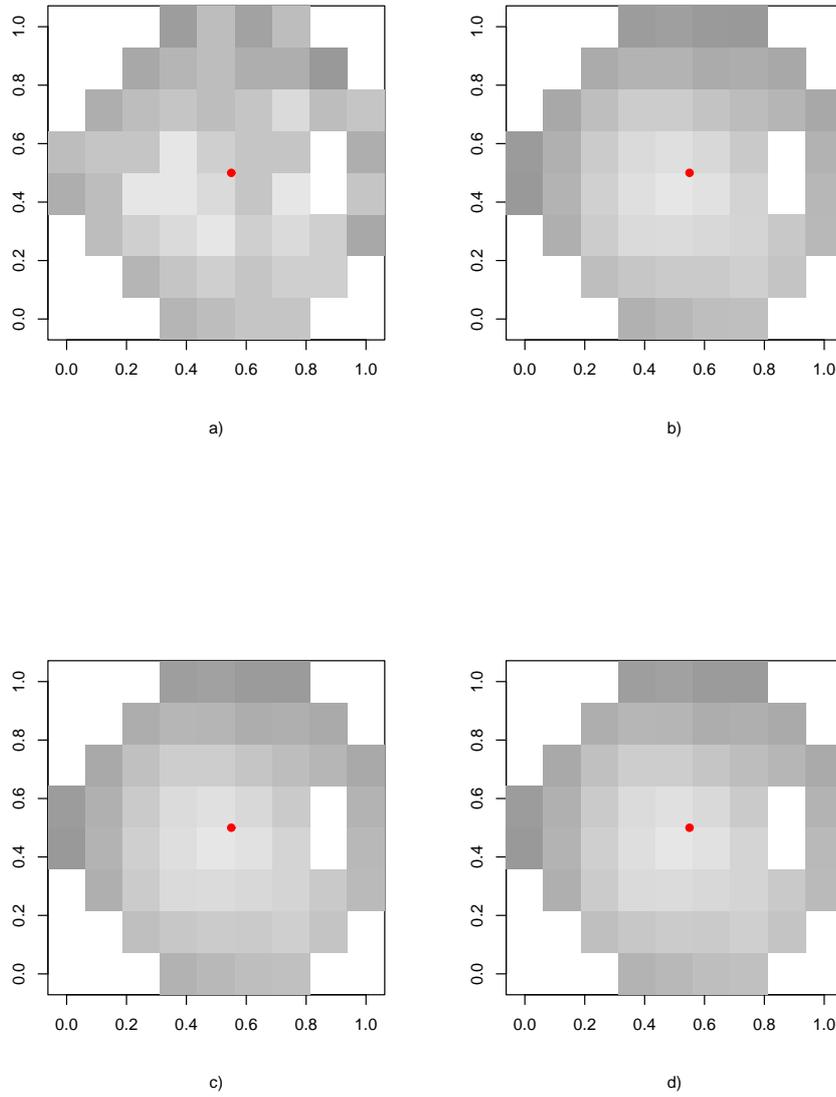


Figura 5.4: Estimación Kriging de la figura 5.3 a), a las edades de 50, 51 y 52 años.

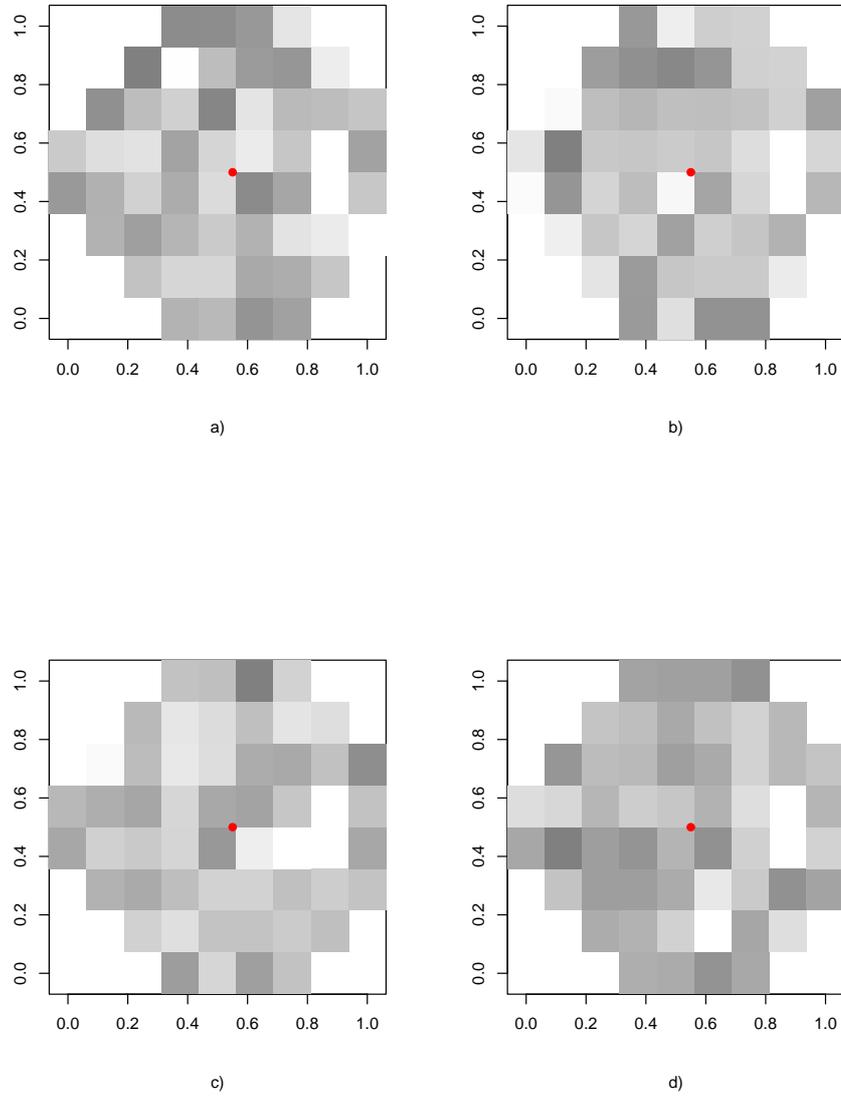


Figura 5.5: Simulaciones no condicionales de un proceso aleatorio gaussiano  $Z$  de media 0 y matriz de covarianzas  $V$  detallada en la ecuación 5.7. El retardo temporal entre dos simulaciones consecutivas es de 1 año.

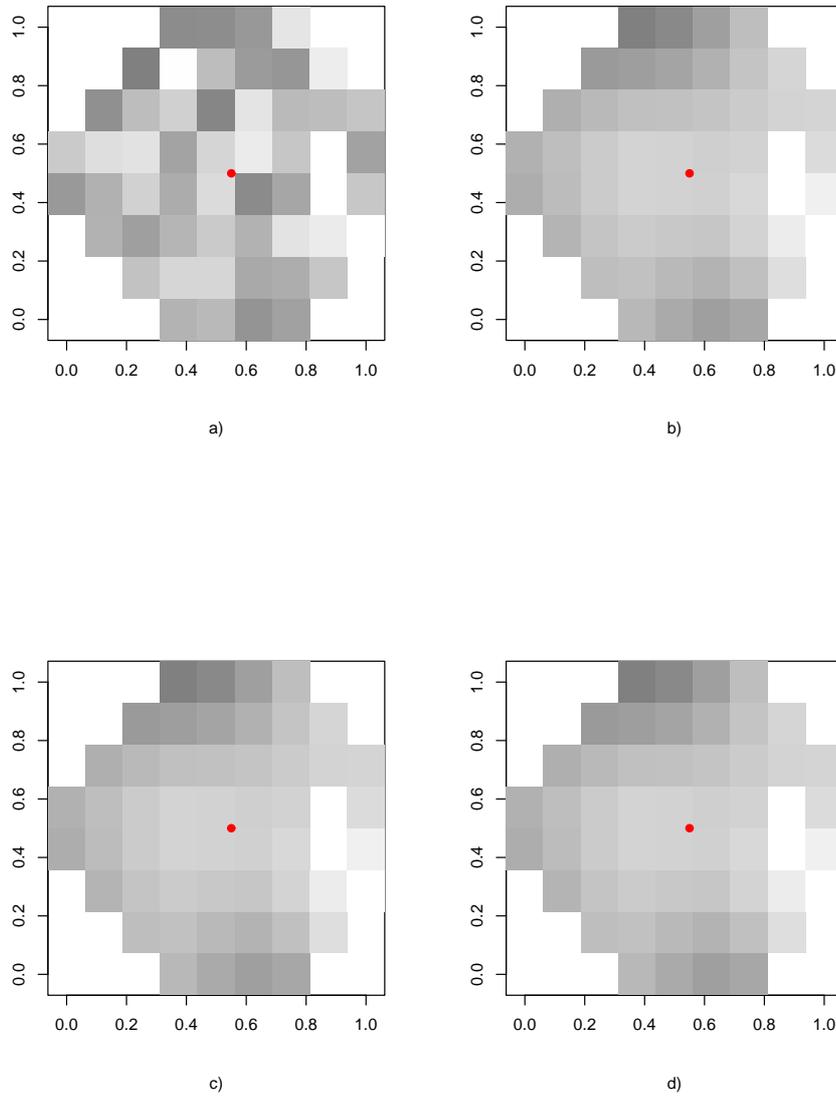


Figura 5.6: Estimaciones Kriging de la figura 5.5.

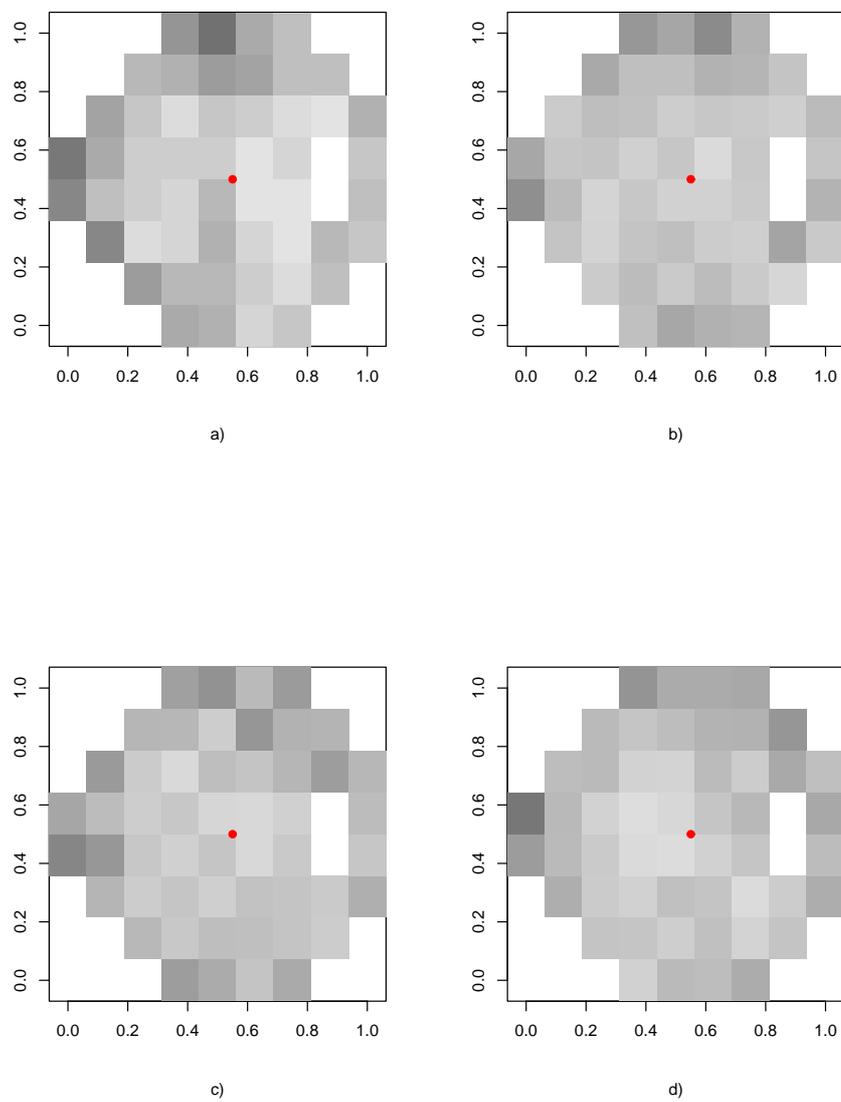


Figura 5.7: En a) vemos el CV “real” de un paciente de 63 años. En b), c) y d) hemos simulado condicionalmente CV a partir del primero, con un retardo temporal de 1 año entre dos simulaciones consecutivas.

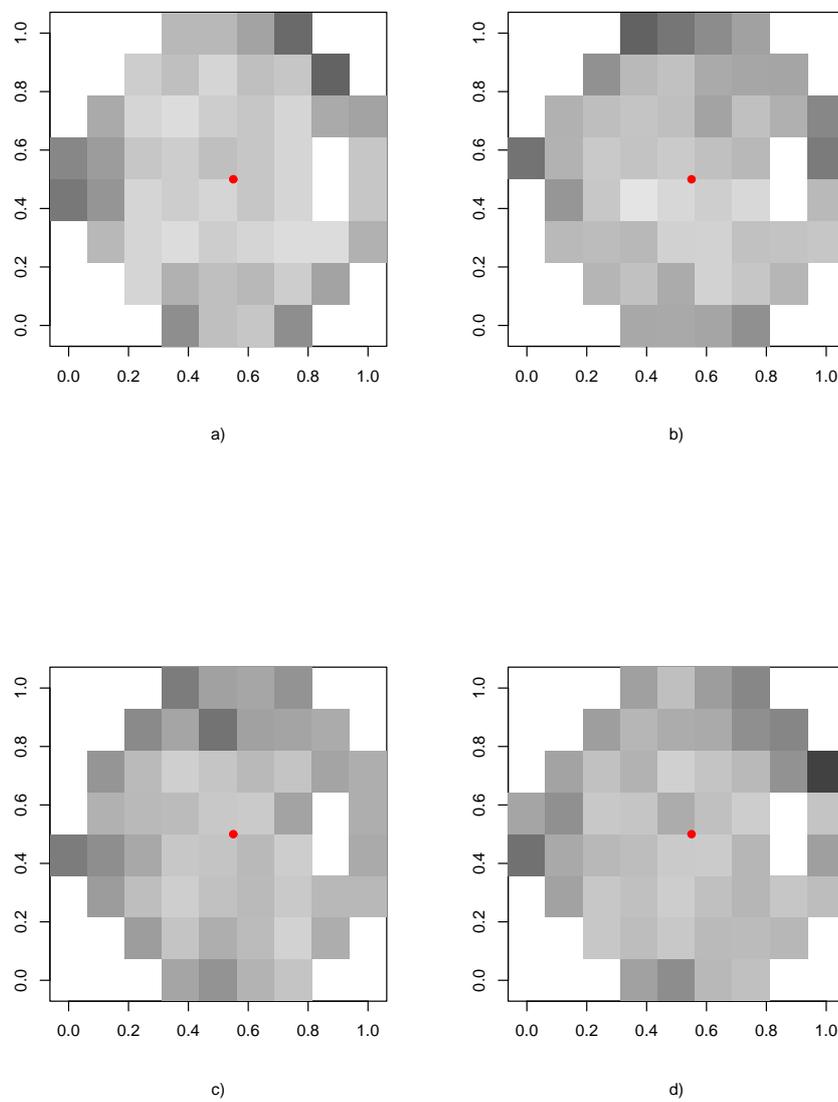


Figura 5.8: En a) vemos el CV “real” de un paciente de 58 años. En b), c) y d) hemos simulado condicionalmente CV a partir del primero, con un retardo temporal de 1 año entre dos simulaciones consecutivas.

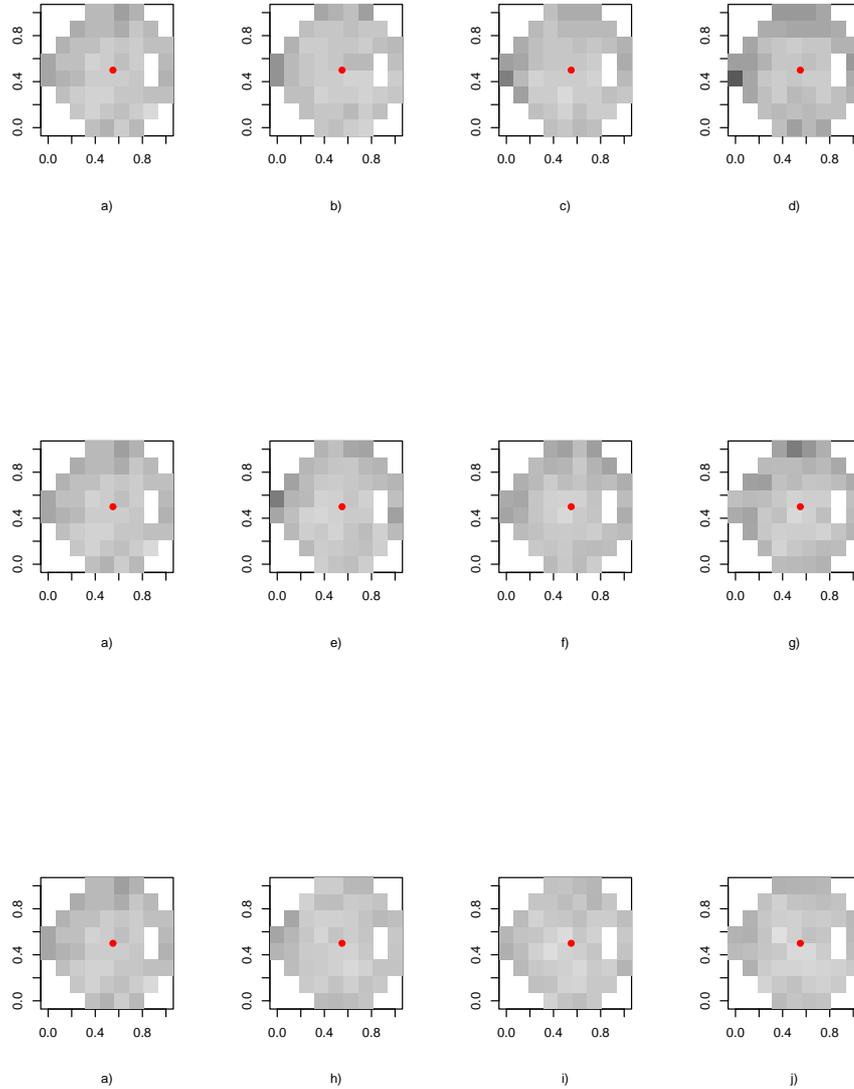


Figura 5.9: Comparación de CV “reales” y “simulados”. En a), b), c) y d) vemos 4 CV “reales”, pertenecientes al mismo paciente a distintas edades. En e), f), y g) vemos tres CV que han sido simulados condicionalmente a partir de a), al igual que en h), i) y j), a las mismas edades que tenía el paciente en, b), c) y d).

## Capítulo 6

# Modelización Geoestadística.- pacientes con Glaucoma

### Contenidos

---

6.1. Introducción . . . . .	148
6.2. Análisis espacial . . . . .	149
6.3. Buscando homogeneidad en la varianza espacial	152
6.4. Modelización espacio temporal . . . . .	156
6.5. Bondad del modelo . . . . .	162
6.5.1. Validación cruzada . . . . .	162

---

## 6.1. Introducción

Como un primer paso para el estudio espacio temporal de campos visuales, en el capítulo 4, hemos modelizado la distribución espacio temporal de CV de pacientes sanos [101]. Este modelo nos ha permitido realizar predicciones, simulaciones, y construir intervalos de confianza para clasificar posiciones de nuevas exploraciones como “normales” o “enfermas”.

En este capítulo nos planteamos ampliar este modelo al caso de pacientes con glaucoma.

Seguiremos utilizando la misma notación, llamando  $Y_{i,j,t}$  a la observación número  $t$  del paciente número  $i$ , posición número  $j$ ;  $Y_{i,\cdot,t} = (Y_{i,1,t}, Y_{i,2,t}, \dots, Y_{i,52,t})$  al vector con los 52 valores obtenidos en el test número  $t$  realizado al paciente  $i$ ,  $i \in \{1, \dots, N\}$ ,  $t \in \{t_1, \dots, t_{n_i}\}$ , y finalmente

$$Y_i = (Y_{i,1,t_1}, \dots, Y_{i,52,t_1}, \dots, Y_{i,1,t_{n_i}}, \dots, Y_{i,52,t_{n_i}})$$

al vector que recoge todas las observaciones que tenemos del paciente número  $i$ . Identificamos el “tiempo”  $t$  con la edad del paciente.

En el capítulo 4, al trabajar con CV de pacientes sanos, vimos que para lograr estacionariedad en media y varianza necesitábamos transformar los datos, aplicándoles una transformación de Box-Cox. En este capítulo vamos a seguir trabajando con los datos sometidos a la misma transformación.

La diferencia con el trabajo que realizamos en el capítulo 4, es que ahora en la base de datos tenemos CV tanto de pacientes sanos como de pacientes enfermos. Pero en vez de hablar de CV sanos y de CV enfermos, vamos a hablar de posiciones sanas y de posiciones enfermas dentro de cada CV. Cada una de las posiciones de cada uno de los CV que hemos observado puede estar “sana” o “enferma”, y dependiendo de esta condición, la intensidad de visión que pueda percibir el paciente en ella será mayor o menor.

Para poder etiquetar cada una de las posiciones como “normal” o “en-

ferma”, definimos una variable  $s_{i,,t} = (s_{i,1,t}, s_{i,2,t}, \dots, s_{i,52,t})$ , con  $s_{i,j,t} = 0$  si en el instante  $t$  la posición  $j$ -ésima del  $i$ -ésimo paciente está sana, y  $s_{i,j,t} = 1$  si está afectada por el glaucoma. Análogamente, denotaremos por  $s_i$  al vector con todas las “etiquetas” de todas las posiciones observadas en los distintos CV que tenemos del paciente número  $i$ , es decir:

$$s_i = (s_{i,1,t_1}, \dots, s_{i,52,t_1}, \dots, s_{i,1,t_{n_i}}, \dots, s_{i,52,t_{n_i}})$$

En el capítulo 4, teníamos una base de datos con 131 pacientes sanos con una única observación, 38 con 2, 9 con 3 y 8 con 4. Todos ellos satisfaciendo unos requisitos mínimos de fiabilidad. Añadimos ahora los CV que tenemos de pacientes clasificados por los oftalmólogos como “CV con glaucoma”, juntando un total de 131 pacientes con una única exploración, y otros 160 pacientes cuyo número de exploraciones oscila entre 2 y 9.

Tras caracterizar el comportamiento de CV de pacientes sanos, en el capítulo 4 construimos intervalos de confianza para las intensidades percibidas en cada posición. Estos intervalos de confianza, a un nivel del 95 %, nos van a permitir clasificar cada una de las posiciones de todos los CV que tenemos en nuestra base de datos como “sana” o “enferma”. Una vez tenemos todas las posiciones etiquetadas, vamos a continuar trabajando suponiendo que la clasificación es la correcta y que en todo momento es conocido el estado (normal/enfermo) de cada una de las posiciones.

Siguiendo las mismas pautas que en el capítulo 4, empezaremos estudiando la distribución espacial de los CV.

## 6.2. Análisis espacial

Empezamos realizando un análisis exploratorio espacial de los datos, y para tener repeticiones independientes del mismo suceso, extraemos aleatoriamente de la base de datos un CV de cada uno de los pacientes. Entre

estos CV tenemos algunos con todas sus posiciones sanas, otros con algunas posiciones afectadas, y otros pocos con casi la totalidad de sus posiciones afectadas por la enfermedad.

Al centrarnos en el análisis espacial, vamos a suprimir momentáneamente el índice  $t$  de la notación, denotando ahora por  $Y_{ij}$  al valor observado en la  $j$ -ésima posición del CV que hemos seleccionado del paciente número  $i$ .

Como ya hemos discriminado entre posiciones sanas y posiciones enfermas, calculamos la media y la desviación típica de cada grupo de observaciones en cada posición, y realizamos varios gráficos para comparar los resultados obtenidos. En la figura 6.1 (a) vemos claramente que existen diferencias en las medias dependiendo de si la posición es sana o enferma. La media empírica de las observaciones tomadas en posiciones sanas, es mayor que la media obtenida en las posiciones enfermas.

En el caso de pacientes sanos, hemos modelizado la media como:

$$\mu_{ij} = \beta_0 + \beta_1 \text{dist}(x_j, 0) + \beta_2 \text{fila}(x_j) + \beta_3 \text{edad}_i + f(j),$$

siendo  $f(\cdot)$  una función suave que depende de la posición. Para conseguir ahora que la media  $\mu_{ij}$  dependa de si la posición  $j$ -ésima es normal o defectuosa, y buscando para ello el modelo más simple posible, podemos formular:

$$\mu_{ijt} = \beta_0 + \beta_1 \text{dist}(x_j, 0) + \beta_2 \text{fila}(x_j) + \beta_3 \text{edad}_{it} + f(j) + \beta_4 s_{ijt} \quad (6.1)$$

siendo de nuevo  $f(\cdot)$  una función suave que depende de la posición.

Con respecto a la varianza, en la figura 6.1 (b), vemos que su comportamiento en principio es análogo al de la media, ya que la varianza en posiciones sanas también es mayor que la varianza en posiciones enfermas.

Para ayudarnos con la modelización, y comprender mejor las diferencias entre medias y varianzas de posiciones sanas y de posiciones enfermas, en la figura 6.2 vemos una serie de mapas que reflejan la disposición espacial de las medias y varianzas empíricas obtenidas en cada posición. Podemos ver que

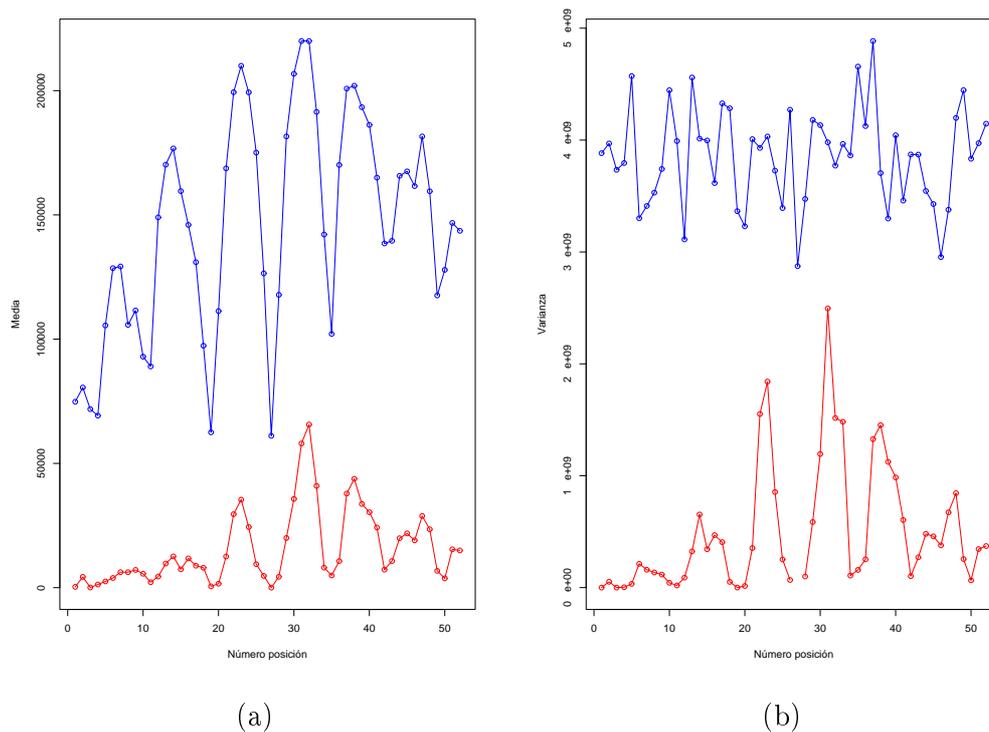


Figura 6.1: En (a) vemos la media empírica de las observaciones recogidas en cada posición, (en rojo tenemos la media para las posiciones enfermas, y en azul la de las posiciones sanas ). Las posiciones han sido enumeradas del 1 al 52 siguiendo el orden natural. En (b) tenemos un gráfico análogo para las varianzas empíricas. Las varianzas para las posiciones enfermas (en rojo) son también las que presentan valores más bajos.

las medias tanto en posiciones sanas como enfermas tienen una estructura “concéntrica” muy similar, por lo que la modelización dada en la ecuación (6.1) puede ser adecuada.

Sin embargo, mirando los mapas de la varianza de la figura 6.2, vemos que en las posiciones enfermas ya no tenemos homogeneidad en la varianza, por lo tanto deberemos buscar una nueva transformación de los datos para poder seguir trabajando bajo la hipótesis de homogeneidad, o recurrir a modelizaciones más complejas que tengan en cuenta esta heterogeneidad.

Siguiendo la idea del artículo de Cressie y Majure [40] en la que también utilizan la Geoestadística en un problema espacio-temporal con heterogeneidad en la varianza, transformaremos los datos para conseguir homogeneidad, y poder seguir, (trabajando con los datos transformados), más o menos los mismos pasos que seguimos en el capítulo 4 en la modelización de pacientes sanos.

### 6.3. Buscando homogeneidad en la varianza espacial

Nuestro objetivo en este apartado es buscar una transformación con la que obtengamos homogeneidad en la varianza. En el caso de posiciones sanas vimos en el capítulo 4 que sí que podíamos suponer la varianza homogénea (notar que estamos trabajando ya con los datos transformados por una transformación de Box-Cox). El problema lo tenemos con las posiciones enfermas. Para encontrar la transformación más adecuada para conseguir homogeneidad, continuamos trabajando con la base de datos “reducida” en la que tenemos un único CV de cada paciente.

A la vista de los gráficos de la varianza de la figura 6.2, y de estudios de regresión comprobamos que tanto la distancia al centro de cada posición

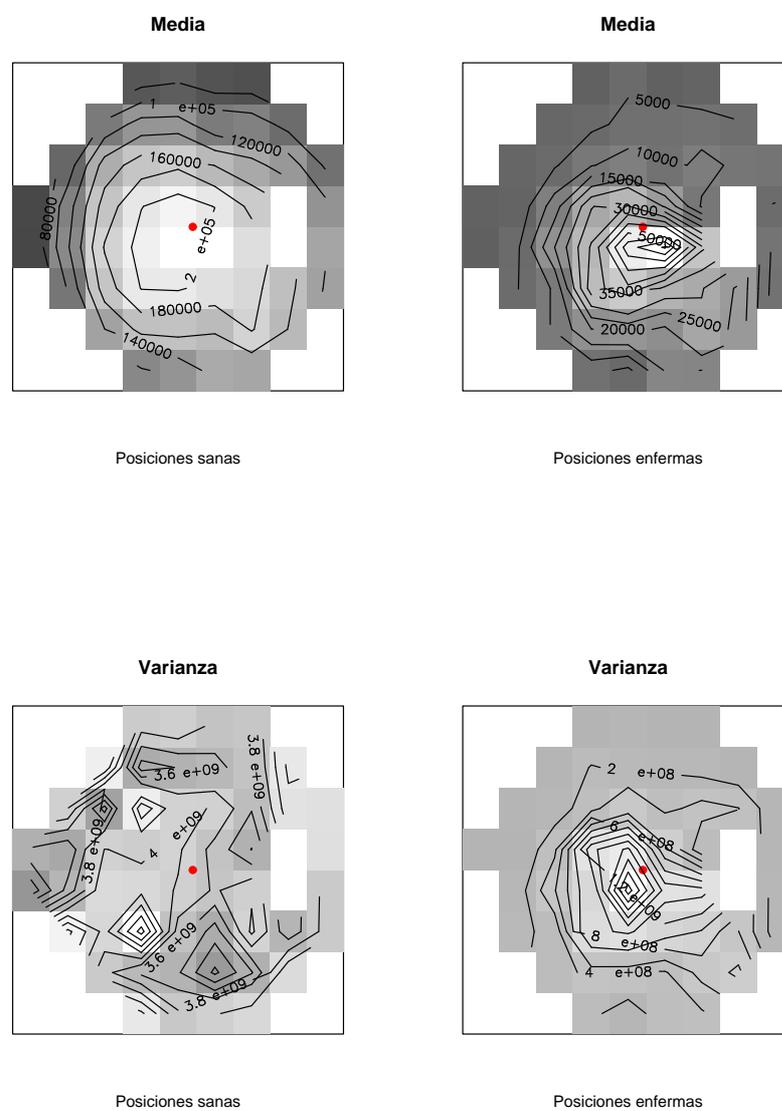


Figura 6.2: Mapas de medias y varianzas en posiciones sanas y en posiciones enfermas.

como la fila en la que cada posición se encuentra, son covariables significativas para la varianza, por lo que proponemos modelizar para cada posición  $Y_{ij}$

$$Var(Y_{ij}) = \sigma^2 \cdot \exp\{s_{ij} \cdot (k_1 dist(x_j, 0) + k_2 fila(x_j))\}$$

(cuando volvamos al contexto espacio-temporal, modelizaremos  $Y_{ijt}$  como:  $Var(Y_{ijt}) = \sigma^2 \cdot \exp\{s_{ijt} \cdot (k_1 dist(x_j, 0) + k_2 fila(x_j))\}$ ). Ya que con esta formulación tenemos una varianza constante para las posiciones sanas, y una varianza dependiendo de un conjunto de covariables, para las posiciones enfermas.

Ajustamos por mínimos cuadrados este modelo a los valores obtenidos empíricamente para la varianza en las posiciones enfermas, obteniendo para  $k_1$  y  $k_2$  los valores:  $k_1 = -1,0515$ ,  $k_2 = 0,3120$ , y construimos para cada campo visual la matriz

$$\begin{aligned} D_i = & \text{diag}(\sigma^2 \cdot \exp\{s_{i1} \cdot (-1,05 \cdot dist(x_1, 0) + 0,31 \cdot fila(x_1))\}, \\ & \sigma^2 \cdot \exp\{s_{i2} \cdot (-1,05 \cdot dist(x_2, 0) + 0,31 \cdot fila(x_2))\}, \dots, \\ & \dots, \sigma^2 \cdot \exp\{s_{i52} \cdot (-1,05 \cdot dist(x_{52}, 0) + 0,31 \cdot fila(x_{52}))\}) \end{aligned} \quad (6.2)$$

Para comprobar si este modelo que acabamos de ajustar para las varianzas es adecuado, vamos a suponer por un momento que cada uno de los campos visuales observados  $Y_i \sim N(\mu_i, \sigma^2 \cdot D_i)$ , siendo  $D_i$  como en la ecuación 6.2, y con

$$\mu_i = \underbrace{\begin{pmatrix} 1 & dist(x_1, 0) & fila(x_1) & edad_i & s_{i1} \\ 1 & dist(x_2, 0) & fila(x_2) & edad_i & s_{i2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & dist(x_{52}, 0) & fila(x_{52}) & edad_i & s_{i52} \end{pmatrix}}_{X_i} \cdot \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}}_{\beta}$$

es decir, no consideramos dependencia espacial entre las posiciones.

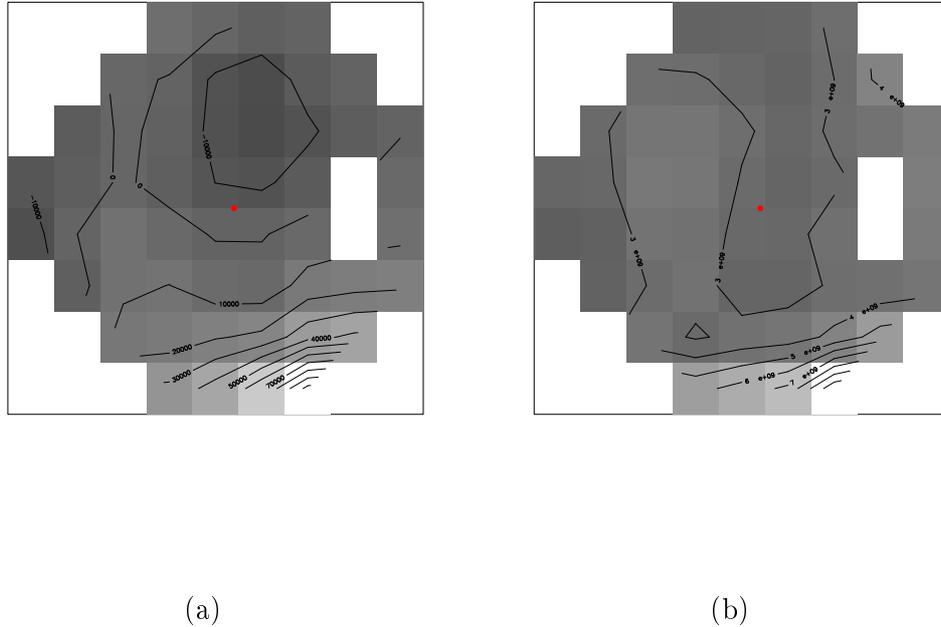


Figura 6.3: En (a) vemos la media empírica de los residuos y en (b) tenemos la varianza empírica, también de los residuos.

Si estimamos el parámetro  $\beta$  por mínimos cuadrados,  $\hat{\beta} = \left( \sum_i X_i' D_i X_i \right)^{-1} \left( \sum_i X_i' D_i Y_i \right)$ , podremos construir los residuos estandarizados  $R_i = D_i^{-1/2} (Y_i - X_i \cdot \hat{\beta})$ , y comprobar si ya podemos suponer homogeneidad en la varianza. En la figura 6.3 vemos los mapas para la media y para la varianza empírica de los residuos. Tras observar la figura 6.3 (b), vemos que con los datos transformados podemos seguir trabajando asumiendo homogeneidad en la varianza.

Transformamos por tanto los datos, y abusando de notación volvemos a denotar por  $Y_i$  al CV resultante al transformar el original  $Y_i$ . La transforma-

ción es la siguiente:

$$Y_i = D_i^{-1/2} \tilde{Y}_i \quad (6.3)$$

En adelante, y al menos que digamos lo contrario, al hablar de  $Y_{i,j}$ , nos referiremos siempre a los datos transformados. Antes de continuar, modificamos todos los datos de nuestra base de datos (i.e. aplicamos la ecuación 6.3 a todos los CV de todos los pacientes de nuestra base de datos “completa”).

## 6.4. Modelización espacio temporal

En este capítulo nos proponemos generalizar el modelo encontrado en el capítulo 4 para que abarque la modelización de posiciones o CV tanto sanos como enfermos.

Siguiendo las pautas de la Geoestadística, buscamos un modelo en el que podamos expresar

$$Y_i = \mu_i + \epsilon_i$$

siendo  $Y_i = (Y_{i,1,t_1}, \dots, Y_{i,52,t_1}, \dots, Y_{i,1,t_{n_i}}, \dots, Y_{i,52,t_{n_i}})'$ , y  $\epsilon_i = (\epsilon_{i,1,t_1}, \dots, \epsilon_{i,52,t_1}, \dots, \epsilon_{i,1,t_{n_i}}, \dots, \epsilon_{i,52,t_{n_i}})'$ .

Tal y como vimos en la modelización de los pacientes sanos, para formular la media planteamos una modelización semiparamétrica, utilizando los modelos aditivos de Hastie y Tibshirani [92], que nos permiten modelizarla como:

$$\mu_{ijt} = \beta_0 + \beta_1 \text{dist}(x_j, 0) + \beta_2 \text{fila}(x_j) + \beta_3 \text{edad}_{it} + \beta_4 s_{ijt} + f(j) \quad (6.4)$$

siendo  $f(\cdot)$  una función suave, de la posición espacial en la que se toma cada observación.

Combinando el método de mínimos cuadrados con el algoritmo backfitting (algoritmo 2.1 capítulo 4), podemos estimar todos los parámetros de

la media. En la tabla 6.1 podemos ver las estimaciones obtenidas para los parámetros de la parte paramétrica, y en la figura 6.4 podemos ver la estimación de la superficie suave.

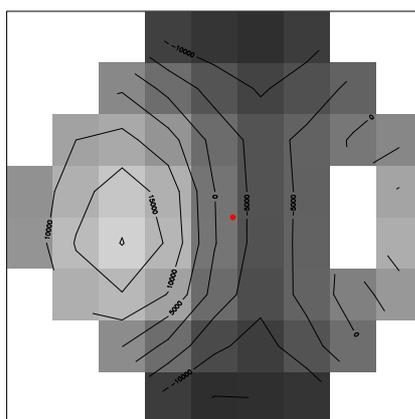


Figura 6.4: Estimación de la superficie suave  $f(x)$  que aparece en la formulación de la media (ecuación 6.4).

Cuadro 6.1: Estimaciones de los parámetros de la media modelizada en la ecuación 6.4

$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
326237.47	-24994.37	4624.46	-2193.94	-87561.70

Una vez estimada la media, la restamos de todas las observaciones, y nos quedamos con los residuos. Para modelizar la estructura de covarianza, y detectar las diferentes fuentes de aleatoriedad presentes en los datos, dibujamos

el variograma espacio temporal empírico de estos residuos. Recordemos que el variograma espacio temporal empírico se calcula como:

$$\hat{\gamma}(h, u) = \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{2 |N(h, u)|} \sum_{(j,k,t_1,t_2) \in N(h,u)} (Y_{i,j,t_1} - Y_{i,k,t_2})^2 \right]$$

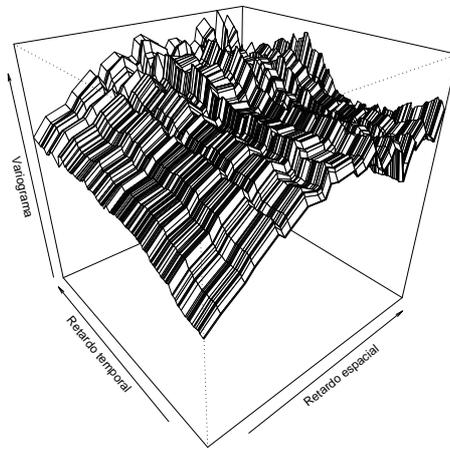


Figura 6.5: Variograma espacio temporal empírico.

De nuevo, al analizar el gráfico del variograma (figura 6.5), detectamos la presencia de las tres fuentes de aleatoriedad: un efecto aleatorio reflejando la variabilidad existente entre distintos pacientes, la correlación espacio-temporal, y un término de error de medida.

Ahora debemos buscar un modelo paramétrico adecuado para ajustar

esta superficie. Como ya comentamos en los capítulos 3 y 4, el principal problema que existe a la hora de modelizar una función de covarianzas, es que debemos garantizar que la función resultante sea definida positiva.

Como nosotros queremos generalizar un modelo ya construido, vamos a suponer que se mantiene la estructura de la matriz de varianzas covarianzas. En el capítulo 4 utilizamos un modelo separable para modelizar la correlación espacio temporal. Vamos a seguir utilizando el mismo modelo, y tras el ajuste comprobaremos su bondad. Nuestro modelo quedará:

$$Y_i = (Y_{i,1,t_1}, \dots, Y_{i,52,t_1}, \dots, Y_{i,1,t_{n_i}}, \dots, Y_{i,52,t_{n_i}})$$

$$Y_i = \mu_i + \epsilon_i$$

$$\mu_{ijt} = \beta_0 + \beta_1 \cdot \text{dist}(x_j, 0) + \beta_2 \cdot \text{fila}(x_j) + \beta_3 \cdot \text{edad}_{it} + \beta_4 \cdot s_{ijt} + f(x_j)$$

Si llamamos:

$$X_i = \begin{pmatrix} 1 & \text{dist}(x_1, 0) & \text{fila}(x_1) & \text{edad}_{it_1} & s_{i1t_1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \text{dist}(x_{52}, 0) & \text{fila}(x_{52}) & \text{edad}_{it_1} & s_{i52t_1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \text{dist}(x_1, 0) & \text{fila}(x_1) & \text{edad}_{it_{n_i}} & s_{i1t_{n_i}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \text{dist}(x_{52}, 0) & \text{fila}(x_{52}) & \text{edad}_{it_{n_i}} & s_{i52t_{n_i}} \end{pmatrix}, \text{ y } \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}$$

podemos formular:

$$\mu_i = X_i \beta + f(x)$$

$$\epsilon_i = (\epsilon_{i,1,t_1}, \dots, \epsilon_{i,52,t_1}, \dots, \epsilon_{i,1,t_{n_i}}, \dots, \epsilon_{i,52,t_{n_i}})$$

$$\epsilon_i \sim N(0, V_i)$$

$$V_i = \text{Var}(\epsilon) = \nu^2 J_i + \sigma^2 H_i + \tau^2 I_i$$

Cuadro 6.2: Estimaciones de los parámetros del modelo de la distribución espacio temporal de todos los pacientes

$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
5.2546e+05	-2.5824e+04	3.7241e+02	-2.3787e+03	-9.5882e+03
$\nu^2$	$\sigma^2$	$\tau^2$		
1.548006e+09	1.911588e+09	1.247544e+09		
$\phi_1$	$\kappa_1$	$\phi_2$	$\kappa_2$	
3.665576e-01	14	4.128466e-01	1.25	

siendo  $J_i$  una matriz  $52n_i \times 52n_i$  de unos,  $I_i$  una matriz identidad  $52n_i \times 52n_i$ , y siendo  $H_i$  otra matriz  $52n_i \times 52n_i$ . Si definimos  $d_j^k = \text{distancia}(x_j, x_k)$ ,  $t_l^m = |t_l - t_m|$ ,  $M(h, u) = \text{matern}(h, \phi_1, \kappa_1) \cdot \text{matern}(u, \phi_2, \kappa_2)$ , la matriz  $H_i$  quedará de nuevo como:

$$H_i = \begin{pmatrix} M(d_1^1, t_1^1) & \cdots & M(d_{52}^{52}, t_1^1) & \cdots & M(d_1^1, t_{n_i}^1) & \cdots & M(d_{52}^{52}, t_{n_i}^1) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ M(d_{52}^1, t_1^1) & \cdots & M(d_{52}^{52}, t_1^1) & \cdots & M(d_{52}^1, t_{n_i}^1) & \cdots & M(d_{52}^{52}, t_{n_i}^1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ M(d_1^1, t_{n_i}^1) & \cdots & M(d_{52}^{52}, t_{n_i}^1) & \cdots & M(d_1^1, t_{n_i}^1) & \cdots & M(d_{52}^{52}, t_{n_i}^1) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ M(d_{52}^1, t_{n_i}^1) & \cdots & M(d_{52}^{52}, t_{n_i}^1) & \cdots & M(d_{52}^1, t_{n_i}^1) & \cdots & M(d_{52}^{52}, t_{n_i}^1) \end{pmatrix}$$

Una vez construido el modelo, estimamos los parámetros combinando el algoritmo de máxima verosimilitud con el algoritmo backfitting (algoritmo 2.2 capítulo 4). En la tabla 6.2 podemos ver las estimaciones obtenidas para los parámetros del modelo, y en el gráfico 6.6 la superficie suave estimada.

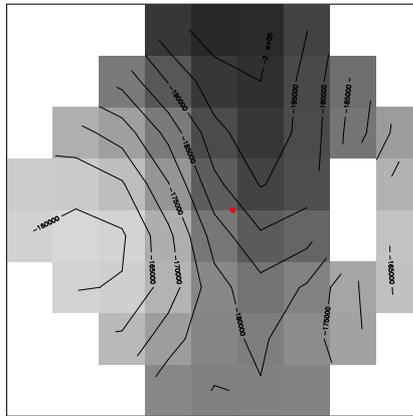


Figura 6.6: Estimación de  $f(x)$  de la media de la distribución espacio temporal conjunta.

## 6.5. Bondad del modelo

Como una medida para estudiar la bondad del ajuste, podemos calcular los residuos estandarizados:

$$R_i = \hat{V}_i^{-\frac{1}{2}} (Y_i - \hat{\mu}_i)$$

Si hemos conseguido un buen ajuste, la media de los residuos debe ser igual a cero y su variograma espacio temporal debe ser una superficie constante igual a 1. En nuestro caso, al calcular la media muestral de los residuos estandarizados, obtenemos el valor de  $-0,0353$ .

Calculamos y dibujamos también el variograma espacio temporal empírico de los residuos. En la figura 6.7 podemos ver un mapa con la media empírica de los residuos estandarizados en cada posición, y el variograma espacio temporal empírico. A la vista de los resultados, podemos concluir que hemos conseguido un ajuste bastante bueno.

### 6.5.1. Validación cruzada

Al igual que hicimos en el capítulo 4, además de estudiar los residuos del modelo, comprobaremos la bondad del ajuste mediante un método de validación cruzada ([199] o capítulo 3).

Al igual que hicimos allí, consideraremos únicamente a aquellos pacientes que tenemos en la base de datos con tres observaciones o más, e iremos extrayendo sucesivamente de la base de datos un campo visual de cada uno de ellos. De esta forma estamos simplificando el estudio de validación cruzada, pero reducimos considerablemente el tiempo computacional.

Una vez “quitado” cada uno de los campos visuales, volvemos a estimar todos los parámetros, aplicamos kriging simple para predecir las observaciones eliminadas, y calculamos la media de las diferencias entre los campos visuales observados y los obtenidos al aplicar el kriging con los nuevos

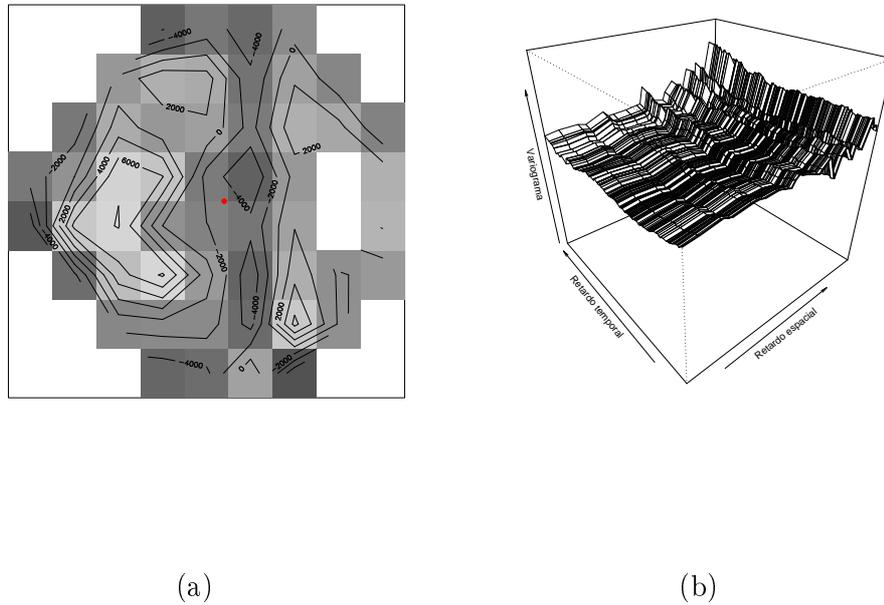


Figura 6.7: (a)Media empírica de los residuos estandarizados en cada posición. (b) Variograma espacio temporal empírico calculado sobre los residuos estandarizados.

parámetros. Si el valor obtenido (que llamamos media del error de validación cruzada), está próximo a cero, nos indicará ausencia de sesgo, y la bondad del ajuste. Por otro lado, la media del cociente entre la raíz cuadrada de los errores de validación cruzada con la desviación estándar del kriging en cada punto, no debe distar demasiado de 1, para tener una evidencia de la adecuación del modelo.

Con nuestro modelo y nuestros datos, obtenemos el valor 5094.48 para la media del error de validación cruzada, y un cociente de desviaciones estándar igual a 0.8669.



## Parte III

# Series temporales y Campos

## Aleatorios de Markov



# Capítulo 7

## Introducción a las series temporales multivariantes y a los campos aleatorios de Markov

### Contenidos

---

<b>7.1. Introducción.</b>	<b>169</b>
<b>7.2. Series temporales multivariantes estacionarias.</b>	<b>172</b>
7.2.1. Procesos autoregresivos multivariantes (Procesos <i>VAR</i> ).	173
7.2.2. Procesos de medias móviles multivariantes (Procesos <i>MA</i> ).	195
7.2.3. Procesos <i>ARMA</i> multivariantes.	198
<b>7.3. Series multivariantes no estacionarias.</b>	<b>199</b>
7.3.1. Introducción.	199
7.3.2. Procesos estacionarios en tendencia.	200
7.3.3. Procesos con raíces unitarias.	202

7.3.4. Comparación: Procesos estacionarios en tendencia y procesos integrados . . . . .	210
7.3.5. Procesos con cambios de régimen (o con cambios estructurales) . . . . .	213
<b>7.4. CAM y algoritmos de estimación de parámetros</b>	<b>219</b>
7.4.1. Introducción . . . . .	219
7.4.2. Campos aleatorios de Markov. Conceptos básicos .	222
7.4.3. Verosimilitud de Montecarlo . . . . .	227
7.4.4. Algoritmos basados en la pseudoverosimilitud . . .	229
7.4.5. Algoritmos basados en el gradiente estocástico . .	230
7.4.6. Estudio comparativo . . . . .	232

---

## 7.1. Introducción.

En este capítulo hacemos una revisión de los conceptos básicos de series temporales multivariantes desde el punto de vista clásico y de la metodología de Box y Jenkins [21]. Nos centraremos en el estudio de los modelos, técnicas y metodología general, que necesitaremos en los capítulos 8 y 9 para construir un modelo que caracterice la evolución espacio temporal de CV de pacientes con glaucoma. No pretendemos profundizar en todos los modelos y métodos desarrollados en la literatura, sino centrarnos en aquellos que puedan sernos útiles en nuestra modelización.

La mayor parte del capítulo se ha escrito tras consultar los siguientes textos [132, 85, 149, 167, 88], que se recomiendan como referencia para aquel que quiera profundizar más en cualquiera de los puntos que aquí se traten.

Al hablar de una serie temporal multivariante, normalmente nos referimos a un conjunto de observaciones de una variable  $k$ -dimensional,  $y$ , tomadas en puntos igualmente espaciados en el tiempo, y es habitual plantearnos dos objetivos: su análisis y la predicción de futuras observaciones. El objetivo del análisis es resumir las propiedades de la serie y destacar sus rasgos característicos. Para ello se puede trabajar tanto en el dominio del tiempo como en el dominio de las frecuencias. En el dominio del tiempo centramos nuestra atención en las relaciones existentes entre las observaciones recogidas en tiempos distintos, mientras que en el dominio de frecuencias nos centramos más en estudiar sus movimientos cíclicos. Las dos formas de análisis no son competitivas, sino complementarias, ya que procesan la misma información bajo dos perspectivas distintas. En este trabajo utilizaremos únicamente el análisis en el dominio del tiempo.

La principal razón para modelizar una serie temporal suele ser la de poder utilizar el modelo para predecir futuras observaciones, y este será uno de los “objetivos” que plantearemos en el capítulo 9.

El primer paso en el análisis de una serie temporal es la selección de un modelo matemático factible, y el modelo matemático que genera una serie temporal es el proceso estocástico. Una serie temporal multivariante  $\{y_t\}_{t=0, \pm 1, \pm 2, \dots}$  es una realización (o una parte finita de una realización) de un proceso estocástico multivariante en el espacio  $Z$  de los números enteros, i.e. para  $t \in Z$  fijo, una función  $y_t: \Omega \rightarrow \mathfrak{R}^k$  con  $(\Omega, A, P)$  un espacio de probabilidad, o, para un  $w \in \Omega$  fijo, una secuencia de vectores  $y_t(w)$ ,  $t \in Z$ . El proceso estocástico subyacente se denomina proceso generador de la serie temporal.

En adelante, al hablar de procesos estocásticos nos referiremos a procesos estocásticos multivariantes, a no ser que en algún caso explicitemos que trabajamos en el caso univariante.

Para caracterizar y/o describir un proceso estocástico deberemos conocer la función de distribución conjunta de cualquier subconjunto finito de  $\{y_t : t \in S \subset Z\}$ , aunque en la práctica a menudo será suficiente con estudiar y determinar los momentos de primer y segundo orden de la distribución. Estos momentos se definen como  $E(y_t) := \mu_t$ ,  $Var(y_t) := \Sigma_t$ ,  $Cov(y_t, y_s) = E[(y_t - \mu_t)(y_s - \mu_s)]$ .

Veamos a continuación, en este contexto, una serie de definiciones básicas que ya vimos en el capítulo 3.

**Definición 1.1** *Un proceso estocástico (serie temporal) es estacionario en sentido débil (o estacionario en la covarianza), si existen los momentos de primer y segundo orden, y no dependen de  $t$ , es decir si:*

1.  $E(y_t) = \mu_y = cte \quad \forall t \in T.$
2.  $Var(y_t) = \Sigma_y = cte \quad \forall t \in T.$
3.  $Cov(y_t, y_{t+k}) = \Gamma(k) := \Gamma_k.$

**Definición 1.2** Una serie temporal  $\{y_t\}$ ,  $t = 0, \pm 1, \pm 2, \dots$  se dice estacionaria en sentido estricto, si la distribución de probabilidad de los vectores aleatorios  $(y_{1,t_1}, y_{2,t_2}, \dots, y_{k,t_k})$  y  $(y_{1,t_1+l}, y_{2,t_2+l}, \dots, y_{k,t_k+l})$  es la misma, para tiempos  $t_1, \dots, t_k$  arbitrarios y  $\forall l = 0, \pm 1, \pm 2, \dots$ .

Aunque en el caso gaussiano ambas son equivalentes, en general la estacionariedad en sentido estricta es más exigente que la estacionariedad débil. En adelante, al hablar de estacionariedad nos referiremos a la estacionariedad débil, a menos que digamos lo contrario.

**Definición 1.3** Dada  $\{y_t\}$  una serie temporal estacionaria, la covarianza entre  $y_t$  e  $y_{t+l}$  dependerá sólo de la diferencia entre los tiempos de observación  $l$ . En este caso se define la matriz de covarianzas cruzadas para el retardo  $l$  como:

$$\Gamma(l) = E[(y_t - \mu)(y_{t+l} - \mu)'] = \begin{bmatrix} \gamma_{11}(l) & \gamma_{12}(l) & \cdots & \gamma_{1k}(l) \\ \gamma_{21}(l) & \gamma_{22}(l) & \cdots & \gamma_{2k}(l) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{k1}(l) & \gamma_{k2}(l) & \cdots & \gamma_{kk}(l) \end{bmatrix}$$

con

$$\gamma_{ij}(l) = \text{Cov}(y_{it}, y_{j,t+l}) = E[(y_{it} - \mu_i)(y_{j,t+l} - \mu_j)]$$

**Definición 1.4** Dada  $y_t$  una serie temporal (multivariante) estacionaria, se define la matriz de correlaciones cruzadas (temporal) para un retardo  $l$  como:

$$\rho(l) = V^{-\frac{1}{2}} \Gamma(l) V^{-\frac{1}{2}} = \begin{bmatrix} \rho_{11}(l) & \rho_{12}(l) & \cdots & \rho_{1k}(l) \\ \rho_{21}(l) & \rho_{22}(l) & \cdots & \rho_{2k}(l) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1}(l) & \rho_{k2}(l) & \cdots & \rho_{kk}(l) \end{bmatrix}$$

para  $l = 0, \pm 1, \pm 2, \dots$ , con  $V^{-1/2} = \text{Diag}\{\gamma_{11}(0)^{-1/2}, \dots, \gamma_{kk}(0)^{-1/2}\}$ , y  $\gamma_{ii}(0) = \text{Var}(y_{it})$ . Donde:  $\rho_{ij}(l) = \text{Corr}(y_{it}, y_{j,t+l}) = \gamma_{ij}(l) / [\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}$ .

Notar que al estar trabajando con procesos estacionarios se cumple que  $\gamma_{ij}(l) = \gamma_{ij}(-l)$ , y por tanto  $\Gamma(l)' = \Gamma(-l)$  y  $\rho(l)' = \rho(-l)$ . Además las matrices  $\Gamma(l)$  y  $\rho(l)$  son definidas no negativas, y por tanto verifican que  $\sum_{i=1}^n \sum_{j=1}^n b_i' \Gamma(i-j) b_j \geq 0$  para todo entero positivo  $n$ , y para cualesquiera vectores  $b_1, \dots, b_n$  de dimensión  $k$ .

**Definición 1.5** *Dada una serie temporal  $\{y_t\}$  se define el operador retardo,  $L$ , como aquel operador tal que:*

$$Ly_t = y_{t-1}$$

*Análogamente  $L^2 y_t := L(Ly_t) = y_{t-2}$ , y así sucesivamente.*

Vamos a dedicar este capítulo al estudio de los modelos clásicos para series temporales multivariantes. Empezaremos en la sección 7.2 estudiando las series temporales multivariantes estacionarias, y dejaremos la sección 7.3 para estudiar las no estacionarias.

En los capítulos 8 y 9 utilizaremos la metodología que vamos a ir describiendo en este capítulo, para modelizar y analizar la distribución espacio temporal de campos visuales. Las observaciones serán series temporales multivariantes, y cada observación en el tiempo, así como una variable no observada que asociaremos en cada posición, podrán ser consideradas como un campo aleatorio de Markov. Por eso en este capítulo vamos a añadir una tercera sección (sección 7.4) en la que veremos el concepto de campo aleatorio de Markov y en la que haremos un estudio sobre algunos algoritmos de estimación de parámetros en estos modelos.

## 7.2. Series temporales multivariantes estacionarias.

En esta sección introducimos una clase muy importante de series temporales definidas en términos de ecuaciones lineales en diferencias con coeficien-

tes constantes. La imposición de esta estructura define una familia paramétrica de procesos estacionarios multivariantes. Vamos a estudiar en la sección 7.2.1 los procesos autoregresivos multivariantes (procesos  $VAR$ ), los procesos de medias móviles multivariantes (procesos  $VMA$ ) en la sección 7.2.2, y por último en 7.2.3 veremos los procesos autoregresivos de medias móviles multivariantes (procesos  $VARMA$ ). De todos estos modelos, los que veremos con más detalle serán los procesos  $VAR$  ya que serán los que utilizaremos en los capítulos 8 y 9 en el análisis de campos visuales de pacientes afectados por glaucoma. La estructura lineal de todos estos procesos conduce a una teoría muy simple y hace que tengan una interpretación muy intuitiva.

Estos modelos se vieron muy popularizados a partir del trabajo de Box y Jenkins [21]. La disponibilidad de códigos informáticos bien documentados, ha ayudado mucho a su difusión y aplicabilidad.

### 7.2.1. Procesos autoregresivos multivariantes (Procesos $VAR$ ).

Decimos una serie temporal  $\{y_t\}$  es un proceso autoregresivo multivariante, o que es un proceso  $VAR$  de orden  $p$  ( $VAR(p)$ ), si podemos expresar:

$$y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \epsilon_t \quad t = 0, \pm 1, \pm 2, \cdots \quad (7.1)$$

siendo  $y_t = (y_{1t}, y_{2t}, \cdots, y_{kt})$  un vector aleatorio de dimensión  $k$ ;  $A_i$  matrices  $k \times k$  de coeficientes (normalmente desconocidos);  $\nu = (\nu_1, \nu_2, \cdots, \nu_k)$  vector de ordenadas (también desconocidas), y  $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, \cdots, \epsilon_{kt})$  un proceso aleatorio (ruido blanco gaussiano), tal que  $E(\epsilon_t) = 0$ ,  $Var(\epsilon_t) = \Sigma_\epsilon$ ,  $Cov(\epsilon_t, \epsilon_s) = 0, \forall t \neq s$ .

Utilizando la notación de retardos, podemos expresar la ecuación (7.1) como:

$$\begin{aligned} [I_k - A_1 L - A_2 L^2 - \cdots - A_p L^p] y_t &= \nu + \epsilon_t \quad \text{o:} \\ A^p(L) y_t &= \nu + \epsilon_t \end{aligned}$$

siendo  $A^p(L)$  una matriz de polinomios en  $L$ .

Para estudiar bajo qué condiciones un proceso  $VAR(p)$  es estacionario, y para encontrar de una forma sencilla las expresiones de la media marginal y de la matriz de varianzas-covarianzas marginal, vamos a utilizar el hecho de que cualquier proceso  $VAR(p)$  con  $p > 1$ , puede expresarse como un proceso  $VAR(1)$ , sin más que definir:

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}_{(kp \times 1)} \quad v = \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(kp \times 1)} \quad \xi_t = \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(kp \times 1)}$$

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_k & 0 & \cdots & 0 & 0 \\ 0 & I_k & \cdots & 0 & 0 \\ \vdots & & \ddots & & 0 \\ 0 & 0 & \cdots & I_k & 0 \end{bmatrix}_{(kp \times kp)}$$

Entonces la ecuación (7.1) es equivalente a:

$$Y_t = v + AY_{t-1} + \xi_t, \quad \xi_t \sim N(0, W) \quad (7.2)$$

con:

$$E(\xi_t \xi'_\tau) = \begin{cases} \Sigma_\epsilon & \text{si } t = \tau \\ 0 & \text{en otro caso} \end{cases} \quad W = \begin{bmatrix} \Sigma_\epsilon & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{(kp \times kp)}$$

Calculamos los momentos de primer y segundo orden de este proceso:

$$\begin{aligned}
 Y_1 &= v + AY_0 + \xi_1 \\
 Y_2 &= v + AY_1 + \xi_2 = (I + A)v + A^2Y_0 + \xi_2 + A\xi_1 \\
 Y_3 &= v + AY_2 + \xi_3 = (I + A + A^2)v + A^3Y_0 + \xi_3 + A\xi_2 + A^2\xi_1 \\
 &\vdots \\
 Y_t &= \sum_{i=0}^{t-1} A^i v + A^t Y_0 + \sum_{i=0}^{t-1} A^i \xi_{t-i} \\
 &\vdots
 \end{aligned} \tag{7.3}$$

Entonces:

$$E(Y_t) = \sum_{i=0}^{t-1} A^i v + A^t E(Y_0) + \sum_{i=0}^{t-1} A^i E(\xi_{t-i}) = \sum_{i=0}^{t-1} A^i v + A^t E(Y_0). \tag{7.4}$$

Tomando varianzas en (7.2), suponiendo estacionariedad, y definiendo  $\Xi = \text{Var}(Y(t)) \forall t$ , debe cumplirse que:

$$\Xi = A\Xi A' + W. \tag{7.5}$$

Para resolver esta ecuación, definimos el operador  $vec$ , como un operador que al aplicarlo sobre una matriz  $A$  de dimensión  $m \times n$ , la convierte en un vector columna  $mn \times 1$  al “apilar” las columnas de  $A$  una a continuación de otra, es decir:

$$\text{Si } A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}_{(3 \times 2)} \quad \text{entonces } \quad vec(A) = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}_{(6 \times 1)}$$

Si  $A, B$  y  $C$  son matrices cuyas dimensiones permiten que el producto  $ABC$  exista, entonces:

$$\text{vec}(ABC) = (C' \otimes A) \cdot \text{vec}(B)$$

con  $\otimes$  representando al producto de Kronecker.

Si aplicamos el operador  $\text{vec}$  a los dos lados de la ecuación (7.5) tendremos:

$$\begin{aligned} \text{vec}(\Xi) &= (A' \otimes A) \cdot \text{vec}(\Xi) + \text{vec}(W). \quad \text{Definiendo} \quad F = A' \otimes A \\ \text{vec}(\Xi) &= F \cdot \text{vec}(\Xi) + \text{vec}(W) \\ \text{vec}(\Xi) &= (I - F)^{-1} \text{vec}(W) \end{aligned} \tag{7.6}$$

**Proposición 7.2.1** *Una condición necesaria para que un proceso VAR(p) como (7.2) sea estacionario es que todos los valores propios de la matriz  $A$  sean en módulo menor que la unidad.*

En la demostración vamos a ver que al hablar de estacionariedad nos estamos refiriendo a estacionariedad suponiendo que el proceso empezó en un pasado infinito, llamaremos a este tipo de estacionariedad, estacionariedad asintótica. En adelante siempre que digamos que un proceso VAR es estacionario, implícitamente estaremos suponiendo que esta estacionariedad es asintótica.

**Demostración** Vamos a suponer que todos los valores propios de  $A$  son en módulo menor que la unidad, y comprobaremos a partir de aquí la estacionariedad del modelo VAR, para ello veremos qué pasa con la media y qué pasa con la varianza.

**Media.** En (7.4) hemos visto que

$$E(Y_t) = \sum_{i=0}^{t-1} A^i v + A^t E(Y_0)$$

Si todos los valores propios de  $A$  tienen módulo menor que la unidad, la secuencia  $A^i, i = 0, 1, \dots$  es absolutamente sumable, y:

$$(I + A + A^2 + \dots + A^i) \cdot v \xrightarrow{i \rightarrow \infty} (I - A)^{-1} \cdot v$$

Además  $A^i$  tiende a 0 rápidamente cuando  $i \rightarrow \infty$  por tanto, en el límite podemos ignorar el término  $A^t \cdot E(Y_0)$  (aunque en la práctica el proceso ha empezado en un instante específico, es a menudo conveniente suponer que empezó en el pasado infinito). Por tanto:

$$E(Y_t) := \mu = (I - A)^{-1}v \quad \text{constante } \forall t \quad (7.7)$$

**Varianza.** Sabemos por (7.6) que

$$vec(\Xi) = (I - F)^{-1}vec(W).$$

Para que  $vec(\Xi)$  exista, la matriz  $(I - F)$  debe ser no singular. Si existiera algún valor propio de  $A$  igual a la unidad esta condición ya no se cumpliría, pero no es el caso.  $F = A' \otimes A$ , y los valores propios de  $A \otimes A$  son de la forma  $\lambda_i \cdot \lambda_j$ , (con  $\lambda_i, \lambda_j$  valores propios de  $A$ ). Como  $|\lambda_i| < 1 \forall i$ ,  $(I - F)$  es no singular, y la varianza existe y no depende de  $t$ .

**Definición 2.1** Diremos que un proceso  $\{y_t\} \sim VAR(p)$  con  $A^p(L)y_t = \nu + \epsilon_t$  es estable, si todos los valores propios de  $A$  son en módulo menor que la unidad.

La proposición 7.2.1 puede reformularse ahora diciendo que todo proceso  $VAR(p)$  estable es estacionario.

Además, como los valores propios de la matriz  $A$  son aquellos valores  $\lambda$  que cumplen:

$$|I_k \lambda^p - A_1 \lambda^{p-1} - A_2 \lambda^{p-2} - \dots - A_p| = 0 \quad (7.8)$$

hemos visto que un proceso  $VAR(p)$  como (7.2) es estacionario en sentido débil si todos los valores propios de  $A$  son en módulo menor que la unidad, es decir si todos los  $\lambda$  que cumplen (7.8) son en módulo menor que la unidad. Equivalentemente, el proceso  $VAR(p)$  es estacionario en sentido débil si todos los valores de  $z$  que cumplen que

$$|I_k - A_1z - A_2z^2 - \dots - A_pz^p| = 0$$

son en módulo mayor que la unidad.

### Expresión de la media y de la covarianza de un proceso VAR(p).

Si tenemos un proceso estacionario  $y_t$ , tomado esperanzas en (7.1) obtendremos la expresión del vector de medias marginal:

$$\begin{aligned}\mu_y &= \nu + A_1\mu_y + A_2\mu_y + \dots + A_p\mu_y, \\ \mu_y &= (I - A_1 - A_2 - \dots - A_p)^{-1}\nu.\end{aligned}\tag{7.9}$$

La expresión de las autocovarianzas  $\Gamma(h)$ , para un proceso  $VAR(p)$  estable vendrá dada por un conjunto de ecuaciones que permitirán su cálculo recursivamente. Estas ecuaciones, que deduciremos a continuación, reciben el nombre de “ecuaciones de Yule-Walker”.

Dado un proceso  $VAR(p)$ ,  $Y_t = v + AY_{t-1} + \xi_t$ , la varianza

$$Var(Y_t) := \Xi = E((Y_t - \mu) \cdot (Y_t - \mu)') =$$

$$= E \left\{ \begin{bmatrix} y_t - \mu \\ y_{t-1} - \mu \\ \dots \\ y_{t-p+1} - \mu \end{bmatrix} \times \begin{bmatrix} (y_t - \mu)' & (y_{t-1} - \mu)' & \dots & (y_{t-p+1} - \mu)' \end{bmatrix} \right\} =$$

$$= \begin{pmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{p-1} \\ \Gamma'_1 & \Gamma_0 & \cdots & \Gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma'_{p-1} & \Gamma'_{p-2} & \cdots & \Gamma_0 \end{pmatrix}$$

Hemos visto que  $\Xi = A\Xi A' + W$ , y que  $vec(\Xi) = (I - F)^{-1}vec(W)$  siendo  $F = A' \otimes A$ .

Si llamamos  $\Xi_h = E((Y_t - \mu)(Y_{t-h} - \mu)')$ , se cumplirá que  $\Xi_h = A \Xi_{h-1}$  para  $h = 1, 2, \dots$ , o análogamente que  $\Xi_h = A^h \Xi$  para  $h = 1, 2, \dots$ , y como:

$$\Xi_h = \begin{pmatrix} \Gamma_h & \Gamma_{h+1} & \cdots & \Gamma_{h+p-1} \\ \Gamma_{h-1} & \Gamma_h & \cdots & \Gamma_{h+p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma'_{h-p+1} & \Gamma'_{h-p+2} & \cdots & \Gamma_h \end{pmatrix}$$

tendremos que:

$$\Gamma_h = Cov(y_t, y_{t-h}) = A_1 \Gamma_{h-1} + A_2 \Gamma_{h-2} + \cdots + A_p \Gamma_{h-p}. \quad (7.10)$$

### Estimación de parámetros.

Existen varios métodos que nos permiten estimar los parámetros de un proceso autoregresivo multivariante.

Podríamos, por ejemplo, utilizar las ecuaciones de Yule-Walker que acabamos de deducir (consultar por ejemplo [132] o [85] para ver el desarrollo de estos métodos de estimación), sin embargo, en este apartado vamos a estudiar la estimación de los parámetros del modelo por máxima verosimilitud, ya que será éste el método que utilizaremos para estimar los parámetros en los capítulos 8 y 9.

Supongamos por tanto que tenemos un proceso  $VAR(p)$   $\{y_t\}$ ,

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \epsilon_t \quad \text{con } \epsilon_t \sim i.i.d.N(0, \Sigma_\epsilon)$$

que hemos observado en  $T+p$  instantes en el tiempo, y que queremos estimar sus parámetros por máxima verosimilitud.

La expresión de la verosimilitud completa en este tipo de modelos resulta muy complicada, por ello lo habitual en este caso es trabajar con la función de verosimilitud condicionada a las  $p$  primeras observaciones, y basar la estimación en las  $T$  observaciones restantes.

La expresión de la verosimilitud condicionada es:

$$f(y_T, y_{T-1}, \dots, y_1 \mid y_0, y_{-1}, \dots, y_{-p+1}; \Theta, \Sigma_\epsilon) \quad \text{con} \quad \Theta = (c, A_1, A_2, \dots, A_p)'$$

Si llamamos  $X_t = (1 \ y_{t-1} \ \dots \ y_{t-p+1} \ y_{t-p})$ , tendremos que:

$$y_t \mid y_{t-1}, y_{t-2}, \dots, y_{t-p} \sim N(X_t \Theta, \Sigma_\epsilon)$$

y a partir de aquí ya podemos obtener la log-verosimilitud (condicionada):

$$\begin{aligned} L(\Theta, \Sigma_\epsilon) &= \sum_{t=1}^T \log f(y_t \mid y_{t-1}, \dots, y_{t-p}; \Theta) = \\ &= \frac{-Tk}{2} \log(2\pi) + \frac{T}{2} \log |\Sigma_\epsilon^{-1}| - \frac{1}{2} \sum_{t=1}^T \left[ (y_t - X_t \Theta)' \Sigma_\epsilon^{-1} (y_t - X_t \Theta) \right] \end{aligned}$$

Para simplificar un poco más la notación, expresamos

$$\underbrace{\begin{pmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_T \end{pmatrix}}_Y = \underbrace{\begin{pmatrix} 1 & y'_0 & y'_{-1} & \dots & y'_{-p+1} \\ 1 & y'_1 & y'_0 & \dots & y'_{-p+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & y'_{T-1} & y'_{T-2} & \dots & y'_{-p+T} \end{pmatrix}}_X \cdot \underbrace{\begin{pmatrix} c \\ A'_1 \\ A'_2 \\ \vdots \\ A'_p \end{pmatrix}}_\Theta + \underbrace{\begin{pmatrix} \epsilon'_1 \\ \epsilon'_2 \\ \vdots \\ \epsilon'_T \end{pmatrix}}_\epsilon$$

$$\text{siendo ahora } \epsilon = \begin{pmatrix} \epsilon'_1 \\ \epsilon'_2 \\ \vdots \\ \epsilon'_T \end{pmatrix} \sim N(0, I_T \otimes \Sigma_\epsilon)$$

Si definimos  $\tilde{\Sigma} = I_T \otimes \Sigma_\epsilon$ , la log-verosimilitud (condicionada) quedará [180]:

$$L(\Theta, \tilde{\Sigma}) = \frac{-Tk}{2} \log(2\pi) + \frac{T}{2} \log |\tilde{\Sigma}^{-1}| - \frac{1}{2} \text{tr} \left[ (Y - X\Theta)'(Y - X\Theta)\tilde{\Sigma}^{-1} \right] \quad (7.11)$$

Derivando e igualando a 0 obtendremos los estimadores máximo verosímiles de los parámetros del modelo:

$$\begin{aligned} \frac{dL(\Theta, \tilde{\Sigma})}{d\Theta} &= \frac{d}{d\Theta} \left( -\frac{1}{2} \text{tr} \left[ (Y - X\Theta)'(Y - X\Theta)\tilde{\Sigma}^{-1} \right] \right) = \\ &= -\frac{1}{2} \frac{d}{d\Theta} \left( \text{tr} \left[ (Y'Y\tilde{\Sigma}^{-1} - Y'X\Theta\tilde{\Sigma}^{-1} - \Theta'X'Y\tilde{\Sigma}^{-1} + \Theta'X'X\Theta\tilde{\Sigma}^{-1}) \right] \right) = \\ &= -\frac{1}{2} (-X'Y\tilde{\Sigma}^{-1} - X'Y\tilde{\Sigma}^{-1} + 2X'X\Theta\tilde{\Sigma}^{-1}) \end{aligned}$$

$$\text{Por lo que:} \quad \hat{\Theta} = (X'X)^{-1}(X'Y) \quad (7.12)$$

Como dada una matriz cuadrada  $M$  cualquiera, se cumple que:

$$\frac{d \log |M|}{dM} = (M')^{-1}$$

entonces:

$$\begin{aligned} \frac{dL(\Theta, \tilde{\Sigma})}{d\tilde{\Sigma}^{-1}} &= \frac{d}{d\tilde{\Sigma}^{-1}} \left( \frac{T}{2} \log |\tilde{\Sigma}^{-1}| - \frac{1}{2} \text{tr} \left[ (Y - X\Theta)'(Y - X\Theta)\tilde{\Sigma}^{-1} \right] \right) = \\ &= \frac{T}{2} \tilde{\Sigma} - \frac{1}{2} (Y - X\Theta)'(Y - X\Theta) \end{aligned}$$

Por tanto:

$$\hat{\tilde{\Sigma}} = \frac{1}{T} (Y - X\hat{\Theta})'(Y - X\hat{\Theta}) \quad (7.13)$$

Una vez estimados los parámetros nos planteamos la distribución asin-

tónica de los estimadores. Para ello definimos:

$$x'_t := [1 \ y'_{t-1} \ y'_{t-2} \ \cdots \ y'_{t-p}]_{(1 \times (kp+1))}$$

$$\mathbf{Q} := E(x_t x'_t)$$

$$\hat{\theta} := \text{vec}(\hat{\Theta}) \quad (\text{expresamos los parámetros estimados en forma vectorial}), \text{ con: } \hat{\theta}_i = \left[ \sum_{t=1}^T x_t x'_t \right]^{-1} \left[ \sum_{t=1}^T x_t y_{it} \right].$$

$$\theta := \text{vec}(\Theta) \quad \text{expresión en forma vectorial de los parámetros poblacionales.}$$

$\text{vech}()$  operador que apila verticalmente los elementos de una matriz simétrica que están sobre su diagonal o, equivalentemente, por debajo de ella, i.e.

$$\text{vech} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} = \begin{pmatrix} \sigma_{11} \\ \sigma_{21} \\ \sigma_{31} \\ \sigma_{22} \\ \sigma_{32} \\ \sigma_{33} \end{pmatrix}$$

A partir de aquí se obtienen los siguientes resultados sobre la distribución asintótica de los estimadores (las demostraciones, se pueden consultar por ejemplo en [85, 132]):

$$\text{a) } \hat{\theta} \xrightarrow{p} \theta;$$

$$\text{b) } \hat{\tilde{\Sigma}} \xrightarrow{p} \tilde{\Sigma};$$

$$\text{c) } \sqrt{T}(\hat{\theta} - \theta) \xrightarrow{L} N(0, (\Sigma \otimes \mathbf{Q}^{-1})), \text{ siendo } \otimes \text{ el producto de Kronecker .}$$

$$\text{d) } \begin{pmatrix} \sqrt{T}[\hat{\theta} - \theta] \\ \sqrt{T}[\text{vech}(\hat{\tilde{\Sigma}}) - \text{vech}(\tilde{\Sigma})] \end{pmatrix} \xrightarrow{L} N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} (\tilde{\Sigma} \otimes \mathbf{Q}^{-1}) & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \right)$$

Siendo  $\Sigma_{22}$  la matriz que en cada coordenada recoge la covarianza entre pares de elementos de  $\hat{\tilde{\Sigma}} - \tilde{\Sigma}$ . En particular el elemento de  $\Sigma_{22}$  correspondiente

a la covarianza entre  $\hat{\sigma}_{ij} - \tilde{\sigma}_{ij}$  y  $\hat{\sigma}_{lm} - \tilde{\sigma}_{lm}$  viene dado por  $\sigma_{ij}\sigma_{jm} + \sigma_{im}\sigma_{jl}$ ,  $\forall i, j, l, m = 1, 2, \dots, k$ .

### Comprobación de la bondad del ajuste.

Hasta este momento, hemos estado suponiendo que nuestros datos se podían modelizar utilizando un modelo  $VAR(p)$ , y hemos visto cómo estimar los parámetros del modelo. Dependiendo de cual sea nuestro objetivo final tras la modelización, los métodos que vamos a ver en este apartado serán más o menos “decisivos” a la hora de hablar de la “adecuación” del modelo.

Los modelos  $VAR$  pueden verse, en general, como métodos que proporcionan un filtro para transformar los datos en series multivariantes de ruido blanco. Si es éste el objetivo final del estudio, será muy importante disponer de criterios para ver si realmente los residuos del modelo pueden considerarse generados por un proceso de ruido blanco multivariante, o si no. Por otro lado, el objetivo del estudio puede estar enfocado también en la realización de predicciones. En este caso ya no será tan importante saber si los residuos son realmente ruido blanco gaussiano, siempre y cuando el modelo proporcione buenas predicciones.

En esta sección vamos a repasar brevemente herramientas para comprobar la bondad del ajuste. Así veremos dos procedimientos para decidir si el orden ( $p$ ) del modelo  $VAR$  que hemos elegido es adecuado, y veremos también tests que nos permitan contrastar si los residuos del modelo pueden considerarse generados por un proceso de ruido blanco, y si siguen una distribución gaussiana.

De todas formas, queremos destacar que son muchas las técnicas desarrolladas en la literatura para abarcar cada uno de los objetivos anteriores. En [167, 132] o [85], pueden consultarse estos métodos mucho más detallados, y encontrar otros que no comentamos aquí.

**Minimizando el error cuadrático medio de predicción.**

Si la predicción es el objetivo final del análisis, tiene sentido elegir el orden del modelo de forma que se minimice el error cuadrático medio de predicción.

Más adelante en esta sección, en el subapartado en que hablemos de cómo hacer predicciones, veremos que dado un proceso  $VAR(p)$ ,  $\{y_t\}$ ,

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \epsilon_t \quad \text{con} \quad \epsilon_t \sim i.i.d.N(0, \Sigma_\epsilon)$$

que hemos observado hasta un instante  $t$ , podemos plantearnos predecir su valor en  $t+1$ . A este valor lo llamaremos  $\hat{y}_{t+1|t}$ , y veremos (ecuación (7.15)) que el error cuadrático medio de predicción en este caso es:

$$ECM(\hat{y}_{t+1|t}) = \Sigma_\epsilon + \frac{kp+1}{T} \Sigma_\epsilon = \frac{T+kp+1}{T} \Sigma_\epsilon$$

siendo  $T$  el tamaño muestral,  $k$  la dimensión de la serie temporal, y  $p$  el orden del proceso  $VAR$  ajustado. Para poder calcularlo (y minimizarlo), deberemos reemplazar la matriz de covarianzas de los residuos ( $\Sigma_\epsilon$ ), por una estimación, e idear alguna forma de resumir en un escalar el error cuadrático medio, que resultará ser una matriz.

Supongamos que llamamos  $\hat{\Sigma}_\epsilon$  al estimador máximo verosímil de  $\Sigma_\epsilon$ . Akaike [2, 3] sugiere aproximar  $\Sigma_\epsilon$  por esta estimación corregida por sus grados de libertad, y calcular el determinante de la matriz resultante. Este criterio se llama criterio de error de predicción final (FPE) y se define como:

$$\begin{aligned} FPE(p) &= \det \left[ \frac{T+kp+1}{T} \cdot \frac{T}{T-kp-1} \cdot \hat{\Sigma}_\epsilon \right] = \\ &= \left[ \frac{T+kp+1}{T-kp-1} \right]^k \cdot \det(\hat{\Sigma}_\epsilon) \end{aligned}$$

Por tanto este método propone plantear y estimar los parámetros de modelos VAR de varios órdenes, y quedarnos con el modelo de aquel orden que minimice el valor de  $FPE(p)$ .

Basándose en otro razonamiento distinto, Akaike [4] derivó otro criterio muy similar al que acabamos de exponer, que se conoce con el nombre de *AIC* (Criterio de Información de Akaike). Para un proceso  $VAR(p)$  este criterio se define como:

$$AIC(p) = \ln |\hat{\Sigma}_\epsilon| + \frac{2pk^2}{T}$$

y también nos quedaremos con el orden  $p$  del proceso que minimice esta cantidad.

Lutkepohl [132] compara la eficacia de estos estimadores, y hace hincapié en destacar el buen comportamiento de los dos criterios al trabajar con muestras pequeñas. Además, destaca que aunque trabajemos con muestras pequeñas puede ser que estos criterios no nos proporcionen el orden más adecuado para el modelo, pero sí que proporcionan el orden que minimiza el error cuadrático medio de predicción. Por último, comenta también que estos dos criterios difieren esencialmente en un término de orden  $O(T^{-2})$ , por lo que serán prácticamente equivalentes para valores de  $T$  moderados o altos.

### **Tests de Cociente de Verosimilitud para comprobar el orden del proceso $VAR$ ajustado.**

Al estar nuestro modelo basado en una distribución gaussiana, uno de los métodos más sencillos para realizar contrastes de hipótesis sobre la adecuación del modelo, será el de realizar tests de cocientes de verosimilitudes.

En esta sección, en el subapartado en el que hemos visto la estimación de parámetros, hemos obtenido la expresión de la verosimilitud y hemos calculado los estimadores máximo verosímil de los parámetros del modelo. Sustituyendo la expresión de los estimadores en la verosimilitud, se obtiene fácilmente que:

$$L(\hat{\Theta}, \hat{\Sigma}) = \frac{-Tk}{2} \log(2\pi) + \frac{T}{2} \log |\hat{\Sigma}^{-1}| - \frac{Tk}{2}.$$

Esto simplifica mucho la realización de tests de cocientes de verosimilitudes.

Para comprobar la adecuación del orden del proceso *VAR* ajustado, podemos contrastar que el orden sea  $p = p_0$  (orden que hemos ajustado), frente a la alternativa de que sea otro valor  $p = p_1 > p_0$ . El doble del cociente de verosimilitudes, en este caso, será:

$$2(L_1^* - L_0^*) = T \{ \log | \hat{\Sigma}_1^{-1} | - \log | \hat{\Sigma}_0^{-1} | \} \quad (7.14)$$

que bajo la hipótesis nula tendrá asintóticamente una distribución  $\chi^2$  con  $k^2(p_1 - p_0)$  grados de libertad.

Para el caso de muestras pequeñas, Sims ([187], pag 17) sugiere una modificación al test de cociente de verosimilitudes, y recomienda sustituir (7.14) por

$$(T - (kp + 1)) \{ \log | \hat{\Sigma}_1^{-1} | - \log | \hat{\Sigma}_0^{-1} | \}$$

que es más conservador.

### **Distribuciones asintóticas de la autocovarianza y autocorrelación de un proceso de ruido blanco.**

Si el modelo ajustado es adecuado a los datos, podremos modelizar los residuos como un proceso de ruido blanco. Fuller [57] y Hannan [87] entre otros, estudiaron las distribuciones asintóticas de las autocovarianzas y autocorrelaciones de un proceso de ruido blanco multivariante.

Sea  $u_t$  un proceso de ruido blanco de dimensión  $k$ , con una matriz de varianzas-covarianzas no singular,  $\Sigma_u$ , (y una matriz de correlaciones  $\rho_u$ ). Estimamos las matrices de autocovarianza de  $u_t$  con:

$$C_i = \frac{1}{T} \sum_{t=i+1}^T u_t u'_{t-i} \quad i = 0, 1, \dots, h < T$$

**Proposición 7.2.2** *Sea  $u_t$  un proceso estandarizado de ruido blanco, independiente e idénticamente distribuido, es decir  $u_t$  y  $u_s$  tienen la misma distribución multivariante con matriz de covarianzas no singular  $\Sigma_u$ , y matriz de correlación  $\rho_u$ . Entonces,  $\forall h \geq 1$ ,*

$$\begin{aligned}\sqrt{T}\mathbf{c}_h &\xrightarrow{d} N(0, I_h \otimes \Sigma_u \otimes \Sigma_u) \quad y \\ \sqrt{T}\mathbf{r}_h &\xrightarrow{d} N(0, I_h \otimes \rho_u \otimes \rho_u)\end{aligned}$$

siendo  $\mathbf{c}_h := \text{vec}(\mathbf{C}_h)$ ,  $\mathbf{C}_h = (C_1, C_2, \dots, C_h)$  las matrices de autocovarianza estimadas hasta un retardo  $h$ ,  $\mathbf{r}_h := \text{vec}(\mathbf{R}_h)$  y  $\mathbf{R}_h = (R_1, R_2, \dots, R_h)$  las matrices de autocorrelación estimadas hasta un retardo  $h$ .

(La demostración puede encontrarse, entre otros, en [57, 87]).

Notar que el segundo resultado implica que las varianzas de la distribución asintótica de los elementos de  $\sqrt{T}\mathbf{r}_h$  son todas igual a la unidad, por lo tanto, para muestras grandes, para  $i > 0$ ,  $\sqrt{T}r_{mni}$  (siendo  $r_{mni}$  el elemento  $(m, n)$ -ésimo de  $R_i$ ), tendrá aproximadamente una distribución normal estándar. Si asumimos una significación aproximadamente del 5% y llamamos  $\rho_{mni}$  a los verdaderos coeficientes de correlación correspondientes a  $r_{mni}$ , el contraste:

$$\begin{cases} H_0 : \rho_{mni} = 0 \\ H_1 : \rho_{mni} \neq 0 \end{cases}$$

rechazará la hipótesis nula si:

$$|\sqrt{T}r_{mni}| > 2 \text{ o, equivalentemente: } |r_{mni}| > 2/\sqrt{T}.$$

Por tanto, una forma sencilla de cotrastrar la hipótesis de que una serie temporal múltiple es ruido blanco, será calcular las correlaciones y comparar sus valores absolutos con  $2/\sqrt{T}$ . La hipótesis de ruido blanco es rechazada si alguno de los coeficientes de correlación estimados excede esta cota.

Este método goza de gran popularidad debido a su sencillez, pero puede ser conveniente no basarse exclusivamente en él, ya que realiza un contraste de hipótesis sobre cada una de las autocorrelaciones de los residuos del modelo, basándose en que estos son asintóticamente independientes.

Notar también que estamos trabajando con distribuciones asintóticas, por lo tanto, este resultado no será demasiado bueno cuando el tamaño de la muestra sea pequeño.

### Test de Pormanteau.

Los “Tests de Pormanteau” son tests globales para estudiar las autocorrelaciones de los residuos hasta un retardo  $h$ . La versión multivariante de estos tests ha sido estudiada entre otros por [99, 162, 127].

Este test, realiza un contraste de hipótesis sobre un conjunto de coeficientes de correlación residual, calculados hasta un retardo  $h$ . El test propuesto es el siguiente:

$$\begin{cases} H_0 : \rho_{\mathbf{h}} = (\rho_1, \rho_2, \dots, \rho_h) = 0 \\ H_1 : \rho_{\mathbf{h}} \neq 0 \end{cases}$$

Utilizando la misma notación que hemos utilizado en el apartado anterior, el estadístico propuesto es:

$$P_h = T \sum_{i=1}^h \text{tr}(R_i' R_0^{-1} R_i R_0^{-1})$$

Obviamente este estadístico es muy sencillo de calcular a partir de los residuos. En el libro de Lutkepohl [132], puede encontrarse la demostración de que este estimador sigue una distribución  $\chi^2$  con  $k^2(h-p)$  grados de libertad.

Notar que volvemos a estar trabajando con distribuciones asintóticas. Para muestras pequeñas se ha sugerido modificar el estadístico de test ([41,

130] o [99]), y utilizar

$$\bar{P}_h = T^2 \sum_{i=1}^h (T-i)^{-1} \text{tr}(R_i' R_0^{-1} R_i R_0^{-1}).$$

### Test de Normalidad.

Como es de sobra conocido, una forma muy sencilla de contrastar normalidad, se basa en los momentos centrales de tercer y cuarto orden de la distribución normal. La generalización de estas medidas de asimetría y curtosis al caso multivariante han sido propuestas por Mardia [137] y Seber [181], quienes para una v.a. multivariante,  $x$ , de dimensión  $k$ , que suponemos que sigue una distribución de media  $\mu$  y matriz de varianzas  $\Sigma$ , proponen definir:

$$\begin{aligned} \beta_{1k} &= E[\{(y - \mu)' \Sigma^{-1} (x - \mu)\}^3] \\ \beta_{2k} &= E[\{(y - \mu)' \Sigma^{-1} (x - \mu)\}^2] \end{aligned}$$

siendo  $y$  una v.a. independiente de  $x$  pero con la misma distribución. Cuando  $x \sim N_k(\mu, \Sigma)$ ,  $\beta_{1k} = 0$  y  $\beta_{2k} = k(k+2)$  [181].

Otros autores [132], en vez de utilizar estos índices multivariantes, prefieren seguir trabajando con índices univariantes. Si  $u_t$  es un proceso gaussiano multivariante, de dimensión  $k$ , de forma que  $u_t \sim N(\mu_u, \Sigma_u)$ , y  $P$  es una matriz triangular inferior, con elementos en la diagonal positivos, tal que  $P \cdot P' = \Sigma_u$ , entonces  $P^{-1} \cdot (u_t - \mu_u) \sim N(0, I_k)$ .

Si definimos  $w_t = P^{-1}(u_t - \mu_u) = (w_{1t}, w_{2t}, \dots, w_{kt})' \sim N(0, I_k)$ , las componentes de  $w_t$  son independientes, entonces podemos utilizar tests univariantes para ver si todas las  $w_t \sim N(0, 1)$ . Notar que con este método no estamos comprobando que  $w_t$  sigue una distribución normal multivariante, sino que cada una de sus componentes  $w_{it}$  es normal univariante.

### Cómo hacer predicciones.

Uno de los principales objetivos del análisis de series temporales es la predicción de futuras observaciones, a partir de la información recogida hasta

el instante  $t$ .

Normalmente al hacer predicciones, se define la “predicción óptima” como aquella que minimiza una función de coste escogida con antelación. Por lo tanto, a la hora de hacer una predicción se busca minimizar el coste o la pérdida esperada. En general las predicciones dependerán por tanto de la función de pérdida a minimizar.

En el contexto de modelos  $VAR$ , los predictores que minimizan el  $ECM$  (error cuadrático medio) de predicción son los más utilizados, es decir, si queremos predecir una observación  $y_{t+h}$ , el mejor predictor según este método será aquel  $\hat{y}_{t+h}$  que minimize

$$ECM(\hat{y}_{t+h}) := E[(y_{t+h} - \hat{y}_{t+h})^2]$$

[70, 72] entre otros, probaron que la predicción que minimiza el  $ECM$ , también minimiza un amplio rango de funciones de pérdida diferentes al  $ECM$ .

Supongamos que hemos modelizado un proceso como un  $VAR(p)$ , que tenemos observaciones hasta un instante  $t$ , y que queremos predecir futuras ocurrencias del proceso.

La predicción del valor  $y_{t+h}$  se basará en el hecho, conocido en la literatura estadística, de que la predicción óptima en el sentido de minimizar el  $ECM$  de predicción coincide con la esperanza de la distribución de probabilidad condicionada  $E(y_{t+h} | \{y_s : s \leq t\})$  (la demostración puede encontrarse por ejemplo en [132]).

Como estamos trabajando con un proceso  $VAR(p)$ ,  $\{y_t\}$ ,

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \epsilon_t \quad \text{con} \quad \epsilon_t \sim i.i.d.N(0, \Sigma_\epsilon)$$

tomando esperanzas condicionadas:

$$\begin{aligned}
\tilde{y}_{t+1|t} &:= E(y_{t+1} \mid y_1, \dots, y_t) = \\
&= c + A_1 E(y_t \mid y_1, \dots, y_t) + A_2 E(y_{t-1} \mid y_1, \dots, y_t) + \dots + \\
&+ A_p E(y_{t-p+1} \mid y_1, \dots, y_t) = c + A_1 y_t + A_2 y_{t-1} + \dots + A_p y_{t-p+1} \\
\tilde{y}_{t+2|t} &:= E(y_{t+2} \mid y_1, \dots, y_t) = \\
&= c + A_1 E(y_{t+1} \mid y_1, \dots, y_t) + A_2 E(y_t \mid y_1, \dots, y_t) + \dots + \\
&+ A_p E(y_{t-p+2} \mid y_1, \dots, y_t) = c + A_1 \tilde{y}_{t+1} + A_2 y_t + \dots + A_p y_{t-p+2} \\
&\vdots
\end{aligned}$$

obtendremos las predicciones de forma recursiva.

Estas expresiones serán válidas en caso de que los parámetros del modelo *VAR* sean conocidos (no se hayan estimado). Si las predicciones se realizan a partir de estimaciones de los parámetros, tendremos que cambiar en las ecuaciones anteriores los coeficientes  $c, A_1, A_2, \dots, A_p$  por sus estimaciones  $\hat{c}, \hat{A}_1, \hat{A}_2, \dots, \hat{A}_p$ .

Vamos a separar en dos pasos el cálculo de la varianza de las predicciones. Primero supondremos que los coeficientes del modelo son conocidos, y luego calcularemos la varianza de las predicciones cuando se han estimado los parámetros.

### 1. Parámetros del modelo *VAR* conocidos.

La predicción en  $t + 1$  condicionada a la información hasta el instante  $t$  se obtiene:

$$\left. \begin{aligned}
y_{t+1} &= c + A_1 y_t + A_2 y_{t-1} + \dots + A_p y_{t-p+1} + \epsilon_{t+1} \\
\tilde{y}_{t+1|t} &= c + A_1 y_t + A_2 y_{t-1} + \dots + A_p y_{t-p+1}
\end{aligned} \right\} \rightarrow$$

$$\rightarrow y_{t+1} - \tilde{y}_{t+1|t} = \epsilon_{t+1}$$

Si definimos  $B_0 = I$  tenemos:  $y_{t+1} - y_{t+1|t} = B_0\epsilon_{t+1}$ .

La predicción en  $t+2$  condicionada a la información hasta el instante  $t$  es:

$$\left. \begin{aligned} y_{t+2} &= c + A_1 y_{t+1} + A_2 y_t + \cdots + A_p y_{t-p+2} + \epsilon_{t+2} \\ \tilde{y}_{t+2|t} &= c + A_1 \tilde{y}_{t+1|t} + A_2 y_t + \cdots + A_p y_{t-p+2} \end{aligned} \right\} \rightarrow$$

$$\rightarrow y_{t+2} - \tilde{y}_{t+2|t} = \epsilon_{t+2} + A_1 \epsilon_{t+1}$$

Si llamamos  $B_1 = A_1$  tendremos que  $y_{t+2} - \tilde{y}_{t+2|t} = B_0\epsilon_{t+2} + B_1\epsilon_{t+1}$ .

Si definimos  $B_0 = I$ ;  $B_1 = A_1$ ; y  $B_s = A_1 B_{s-1} + A_2 B_{s-2} + \cdots + A_p B_{s-p}$  para  $s = 1, 2, \dots$ , y  $B_s = 0 \forall s < 0$ , es facil ver por inducción que:

$$y_{t+s} - \tilde{y}_{t+s|t} = \epsilon_{t+s} + B_1\epsilon_{t+s-1} + \cdots + B_{s-1}\epsilon_{t+1}$$

Entonces:

$$\begin{aligned} ECM(\tilde{y}_{t+s|t}) &:= E[(y_{t+s} - \tilde{y}_{t+s|t})(y_{t+s} - \tilde{y}_{t+s|t})'] = \\ &= Var[\epsilon_{t+s} + B_1\epsilon_{t+s-1} + B_2\epsilon_{t+s-2} + \cdots + B_{s-1}\epsilon_{t+1}] = \\ &= \Sigma_\epsilon + B_1\Sigma_\epsilon B_1' + B_2\Sigma_\epsilon B_2' + \cdots + B_{s-1}\Sigma_\epsilon B_{s-1}' \end{aligned}$$

$$E[(y_{t+s} - \tilde{y}_{t+s|t})(y_{t+s} - \tilde{y}_{t+s|t})'] = \sum_{i=0}^{s-1} B_i \Sigma_\epsilon B_i'$$

Notar que:  $ECM(\tilde{y}_{t+s|t}) = ECM(\tilde{y}_{t+s-1|t}) + B_{s-1}\Sigma_\epsilon B_{s-1}'$ , por lo tanto el error de predicción será mayor cuanto a más largo plazo estemos haciendo las predicciones.

En la sección 7.2.2 veremos que todo proceso  $VAR(p)$  puede expresarse como:  $y_t = c + \sum_{i=0}^{\infty} B_i \epsilon_{t-i}$  con  $B_i$  como los que hemos definido en este apartado. Por lo tanto cuando  $s \rightarrow \infty$ ,  $ECM(\tilde{y}_{t+s|t}) = \Gamma_0 = Var(y_t) = \sum_{i=0}^{\infty} B_i \Sigma_\epsilon B_i'$

2. **Parámetros del modelo VAR estimados.**

De una forma muy parecida a la anterior, se demuestra que en este caso:

$$ECM(\hat{y}_{t+h|t}) = \sum_{i=0}^{h-1} B_i \Sigma_\epsilon B_i' + E[(\tilde{y}_{t+h|t} - \hat{y}_{t+h|t})^2]$$

Para calcular  $E[(\tilde{y}_{t+h|t} - \hat{y}_{t+h|t})^2]$  usamos la distribución asintótica de los estimadores  $\hat{\Theta} = (\hat{c}, \hat{A}_1, \dots, \hat{A}_p)$ , y llegamos a que:

$$ECM(\hat{y}_{t+h|t}) = \sum_{i=0}^{h-1} B_i \Sigma_\epsilon B_i' + \frac{1}{T} \Omega(h) \quad \text{con:}$$

$$\Omega(h) = \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} tr \left[ (P')^{h-1-i} \Gamma^{-1} P^{h-1-j} \Gamma \right] B_i \Sigma_\epsilon B_j'$$

siendo:

$$P := \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ c & A_1 & A_2 & \cdots & A_{p-1} & A_p \\ 0 & I_k & 0 & \cdots & 0 & 0 \\ 0 & 0 & I_k & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I_k & 0 \end{pmatrix}_{(kp+1) \times (kp+1)} ;$$

$$Z_t := \begin{pmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}_{(kp+1) \times 1} \quad \text{y} \quad \Gamma := E(Z_t Z_t')$$

La demostración de este resultado puede encontrarse en [132], páginas 87-88.

A partir de este resultado, es conveniente fijarse en la varianza del error de predicción cuando predecimos en  $t+1$ , condicionada a la información

recogida hasta  $t$ . En este caso:

$$ECM(\hat{y}_{t+1|t}) = B_0 \Sigma B_0' + \frac{1}{T} \Omega(1) \quad \text{con:}$$

$$\Omega(1) = tr \left[ (P')^{1-1} \Gamma^{-1} P^{1-1} \Gamma \right] B_0 \Sigma_\epsilon B_0'$$

Por lo tanto:

$$ECM(\hat{y}_{t+1|t}) = \Sigma_\epsilon + \frac{kp+1}{T} \Sigma_\epsilon = \frac{T+kp+1}{T} \Sigma_\epsilon \quad (7.15)$$

Este resultado destaca que el error cuadrático medio de predicción a un paso, cuando los coeficientes del proceso  $VAR$  se han estimado, depende tanto de la dimensión  $k$  del proceso, como del orden  $p$ , como del tamaño muestral utilizado para la estimación.

### Intervalos de confianza para las predicciones.

Estamos trabajando con:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + \epsilon_t \quad \text{con} \quad \epsilon_t \sim i.i.d.N(0, \Sigma_\epsilon),$$

y acabamos de ver que bajo estas condiciones,

$$y_{t+s} - \hat{y}_{t+s|t} \sim N(0, \Sigma_{Pred}) \quad (7.16)$$

donde  $\Sigma_{Pred} = \sum_{i=0}^{s-1} B_i \Sigma_\epsilon B_i'$  si los parámetros del modelo  $VAR$  eran conocidos, o  $\Sigma_{Pred} = \sum_{i=0}^{s-1} B_i \Sigma_\epsilon B_i' + \frac{1}{T} \Omega(s)$  si los parámetros del modelo eran desconocidos y se han estimado con anterioridad. Por lo tanto el error de predicción de cada componente individual es normal, y si llamamos  $\hat{y}_{i,t+s|t}$  a la  $i$ -ésima componente de  $\hat{y}_{t+s|t}$ , y llamamos  $\sigma_{i,s}$  a la raíz cuadrada del  $i$ -ésimo elemento de la diagonal de  $\Sigma_{Pred}$ , tendremos que:

$$\frac{y_{i,t+s} - \hat{y}_{i,t+s|t}}{\sigma_{i,s}} \sim N(0, 1)$$

Por lo tanto un intervalo de confianza al  $(1 - \alpha)\%$  para  $y_{i,t+s}$  será:

$$[\hat{y}_{i,t+s|t} - z_{\alpha/2}\sigma_{i,s}, \hat{y}_{i,t+s|t} + z_{\alpha/2}\sigma_{i,s}]$$

A partir de (7.16) también podemos construir intervalos de confianza conjuntos para dos o más variables. Por ejemplo, supongamos que queremos construir un intervalo de confianza para las  $N$  primeras componentes. Entonces si definimos la matriz  $F = [I_N \ 0]_{N \times k}$ , entonces:

$$[y_{t+s} - \hat{y}_{t+s|t}]' F' (F \Sigma_{Pred} F')^{-1} F [y_{t+s} - \hat{y}_{t+s|t}] \sim \chi_N^2$$

Para simplificar notación, en la ecuación anterior hemos denotado por  $\Sigma_{Pred} = \sum_{i=0}^{s-1} B_i \Sigma_\epsilon B_i'$ .

Ahora se puede utilizar la distribución  $\chi_N^2$  para construir el “elipsoide de confianza” para las primeras  $N$  componentes del proceso, con la significatividad deseada.

En la práctica, la construcción de este elipsoide puede resultar engorrosa si el valor de  $N$  es mayor de 2 o 3, y otros métodos para construir intervalos de confianza pueden ser más adecuados en este caso. Entre ellos destacar el “método de Bonferroni”, que podemos encontrar detallado en la página 35 de [132].

### 7.2.2. Procesos de medias móviles multivariantes (Procesos MA).

En el apartado anterior hemos estudiado los procesos autoregresivos multivariantes. En la modelización clásica de series temporales estacionarias se trabaja normalmente con tres tipos de modelos: los modelos autoregresivos (que ya hemos visto), los modelos de medias móviles (a los que dedicamos este apartado), y los modelos autoregresivos de medias móviles (que veremos en el apartado siguiente).

Nuestra motivación para esta revisión teórica, es la de revisar los resultados que aplicaremos en los capítulos 8 y 9. Como allí trabajaremos únicamente con procesos autoregresivos, veremos los procesos de medias móviles y los autoregresivos de medias móviles muy por encima. Para saber más acerca de estos procesos recomendamos consultar [85] y [132].

Un proceso de medias móviles de orden  $q$  (un proceso  $MA(q)$ ) toma la forma:

$$y_t = \mu + \epsilon_t + B_1\epsilon_{t-1} + B_2\epsilon_{t-2} + \cdots + B_q\epsilon_{t-q}$$

siendo  $B_j$  matrices de coeficientes, y  $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{kt})$  un proceso de ruido, i.e.  $E(\epsilon_t) = 0$ ,  $Var(\epsilon_t) = \Sigma_\epsilon$ ,  $Cov(\epsilon_t, \epsilon_s) = 0, \forall t \neq s$ . Entonces:

$$\begin{aligned} E(y_t) &= \mu \\ Var(y_t) &= \Gamma_0 = E[(y_t - \mu)(y_t - \mu)'] = \\ &= E[\epsilon_t\epsilon_t'] + B_1E[\epsilon_{t-1}\epsilon_{t-1}]B_1' + \cdots + B_qE[\epsilon_{t-q}\epsilon_{t-q}]B_q' = \\ &= \Sigma_\epsilon + B_1\Sigma_\epsilon B_1' + \cdots + B_q\Sigma_\epsilon B_q' \\ Cov(y_t, y_{t-h}) &= \begin{cases} B_h\Sigma_\epsilon + B_{h+1}\Sigma_\epsilon B_{h+1}' + \cdots + B_q\Sigma_\epsilon B_q' & 1 \leq h \leq q \\ \Sigma_\epsilon B_{-h}' + B_1\Sigma_\epsilon B_{-h+1}' + \cdots + B_{q+h}\Sigma_\epsilon B_{q+h}' & -q \leq h \leq -1 \\ 0 & |h| > q \end{cases} \end{aligned}$$

con  $B_0 = I$ .

Por tanto cualquier proceso  $MA(q)$  multivariante es estacionario en sentido débil.

**Proposición 7.2.3** *Cualquier proceso  $VAR(p)$  estacionario, es equivalente a un proceso  $MA(\infty)$ .*

**Demostración** Al empezar a hablar de los procesos  $VAR(p)$ , en la ecuación (7.3) vimos que si consideramos que el proceso  $VAR(p)$  se ha iniciado en un

pasado infinito, podemos llegar a que

$$Y_t = \mu + \sum_{i=0}^{\infty} A^i \xi_{t-i} \quad (7.17)$$

Con

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_k & 0 & \cdots & 0 & 0 \\ 0 & I_k & \cdots & 0 & 0 \\ \vdots & & \ddots & & 0 \\ 0 & 0 & \cdots & I_k & 0 \end{bmatrix}_{(kp \times kp)}$$

Como  $Y_t = (y_t, y_{t-1}, \dots, y_{t-p+1})'$  si multiplicamos (7.17) por la matriz  $J = [I_k \ 0 \ \cdots \ 0]$  tendremos

$$y_t = JY_t = J\mu + \sum_{i=0}^{\infty} JA^i J' J \xi_{t-i} = J\mu + \sum_{i=0}^{\infty} JA^i J' \epsilon_{t-i} \quad (7.18)$$

Si el proceso  $y_t$  es estable, todos los valores propios de la matriz  $A$  serán en módulo menor que 1, por tanto la serie  $A^i$  es absolutamente sumable, y la serie  $JA^i J'$  también lo es, luego (7.18) está bien definida.

A partir de esta expresión, es fácil ver la relación existente entre las matrices  $A_i$  del modelo expresado en forma autorregresiva, y las matrices  $B_i$  del modelo expresado en forma de medias móviles.

$$B_i = JA^i J \rightarrow \begin{cases} B_0 = I \\ B_1 = A_1 \\ B_2 = A_1^2 + A_2 = A_1 B_1 + A_2 \\ \vdots \\ B_i = A_1 B_{i-1} + \cdots + A_p B_{i-p}, \forall i \geq 1, \text{ con } B_j = 0 \forall j < 0 \end{cases} \quad (7.19)$$

### Cómo hacer predicciones.

Por último, veamos cómo hacer predicciones cuando tenemos modelos de medias móviles, a partir de la información recogida hasta el instante  $t$ .

La filosofía a la hora de hacer las predicciones, es la que ya comentamos en el apartado correspondiente para los procesos autoregresivos: la predicción óptima es aquella que minimiza una función de coste elegida de antemano. Por lo que la finalidad es buscar el valor que minimice el coste o la pérdida esperada.

La función de pérdida habitual sigue siendo el *ECM*, y por lo tanto, la predicción óptima coincide con la esperanza de la distribución de probabilidad condicionada  $E(y_{t+h} | \{y_s : s \leq t\})$ , y se puede demostrar que si  $y_t$  sigue un proceso  $MA(q)$  y tenemos observaciones hasta el instante  $t$ , podemos obtener predicciones del proceso en el futuro como sigue:

$$\tilde{y}_{t+h|t} = \begin{cases} \mu + \sum_{i=h}^q B_i \epsilon_{t+h-i} & \text{si } h < q \\ \mu & \text{si } h = q \\ 0 & \text{si } h > q \end{cases} \quad (7.20)$$

y la varianza de estas predicciones será:

$$\text{Var}(\tilde{y}_{t+h|t}) = \begin{cases} \sum_{i=h}^q B_i' \Sigma_\epsilon B_i & \text{si } h < q \\ 0 & \text{si } h \geq q \end{cases} \quad (7.21)$$

### 7.2.3. Procesos ARMA multivariantes.

La expresión general de un modelo  $VARMA(p, q)$ , donde  $p$  es el orden de la parte autoregresiva, y  $q$  el orden de la parte  $MA$  es:

$$y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \epsilon_t + B_1 \epsilon_{t-1} + \cdots + B_q \epsilon_{t-q}$$

con  $\epsilon_i \sim N(0, \Sigma_\epsilon)$  i.i.d.  $\forall i$ .

Definiendo  $A(L) = I - A_1 L - \cdots - A_p L^p$  y  $B(L) = I + B_1 L + \cdots + B_q L^q$  podemos expresar el modelo  $VARMA(p, q)$  como:

$$A(L)y_t = \nu + B(L)\epsilon_t \quad (7.22)$$

Decir que las propiedades de un proceso  $VARMA$  dependen de las propiedades que verifiquen la parte  $VAR$  y la parte  $MA$ .

## 7.3. Series multivariantes no estacionarias.

### 7.3.1. Introducción.

En el apartado 7.2, hemos estudiado modelos para series temporales multivariantes estacionarias, y hemos visto que estas series se pueden modelizar como:

$$y_t = \nu + B(L)\epsilon_t$$

siendo  $B(L)$  un polinomio de la forma  $B(L) = I + B_1L + B_2L^2 + \dots + B_qL^q$ ,  $q$  el orden del proceso (puede ser  $q = \infty$  o cualquier número natural finito); donde las  $B_j$  son matrices de coeficientes, y donde  $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{kt})$  es un proceso de ruido, i.e.  $E(\epsilon_t) = 0$ ,  $Var(\epsilon_t) = \Sigma$ ,  $Cov(\epsilon_t, \epsilon_s) = 0, \forall t \neq s$ .

Estas series se caracterizan, por definición, por tener invariantes sus momentos de primer y segundo orden. Sin embargo, muchas de las series temporales que aparecen en la práctica son no estacionarias, ya que presentan tendencias, ciclos, movimientos estacionales, etc ... Por ello vamos a dedicar este apartado a estudiar cómo trabajar con series no estacionarias.

Como las causas posibles de no estacionariedad son muy numerosas, nos centraremos sólo en el estudio de tres tipos de no estacionariedad, que son las que después nos aparecerán en los casos prácticos que estudiaremos en los capítulos 8 y 9. Para un estudio más a fondo de procesos temporales no estacionarios recomendamos consultar [85, 136, 120, 167, 163] o [132].

La causa más frecuente de no estacionariedad, es la debida a la aparición de tendencias en las series temporales. En la literatura se proponen mayoritariamente dos aproximaciones o modelizaciones para trabajar en estos casos. La primera de las aproximaciones, consiste en incluir en el modelo una tendencia temporal determinística y expresar:

$$y_t = \alpha + \delta t + B(L)\epsilon_t \tag{7.23}$$

Procesos de este tipo reciben el nombre de “estacionarios en tendencia”, ya que si pudiéramos restar la tendencia  $\alpha + \delta t$  de (7.23) el proceso resultante sería estacionario.

La segunda aproximación, viene dada por los procesos con raíces unitarias.

Vamos a estructurar este apartado como sigue: en la sección 7.3.2 estudiaremos los procesos estacionarios en tendencia, y en la sección 7.3.3 los procesos con raíces unitarias. En la sección 7.3.4 compararemos los dos tipos de procesos, y por último en 7.3.5 hablaremos de otro tipo de procesos no estacionarios como son los procesos con cambios en régimen.

### 7.3.2. Procesos estacionarios en tendencia

Los procesos estacionarios en tendencia son de la forma:

$$y_t = \alpha + \delta t + B(L)\epsilon_t \quad (7.24)$$

con  $y_t = (y_{1t}, y_{2t}, \dots, y_{kt})' \forall t$ ,  $\alpha = (\alpha_1, \dots, \alpha_k)'$ ,  $\delta = (\delta_1, \dots, \delta_k)'$ , y siendo  $B(L) = I + B_1L + B_2L^2 + \dots + B_qL^q$ , con  $B_j$  las matrices de coeficientes con  $q = \infty$  o igual a cualquier número natural finito y cumpliendo las condiciones de estacionariedad de la proposición 7.2.1 y ecuación 7.19. Además  $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{kt})$  con  $E(\epsilon_t) = 0$ ,  $Var(\epsilon_t) = \Sigma$ , y  $Cov(\epsilon_t, \epsilon_s) = 0, \forall t \neq s$ . Si los parámetros  $\alpha$  y  $\delta$  fueran conocidos, podríamos convertir el proceso en estacionario simplemente restando la media, ya que la esperanza de un proceso estacionario en tendencia resulta ser:

$$E(y_t) := \mu_t = \alpha + \delta t$$

La estimación de los parámetros de un proceso estacionario en tendencia suele hacerse por mínimos cuadrados ordinarios, pero las distribuciones asintóticas de las estimaciones de estos coeficientes no pueden calcularse siguiendo las pautas utilizadas para los procesos estacionarios. Estas distribuciones

serán ahora más complejas, ya que por ejemplo las estimaciones de los distintos parámetros tendrán distintas velocidades de convergencia asintótica. La aproximación general que se utiliza para obtener estas distribuciones asintóticas fue sugerida por Sims, Sotck y Watson [187], y puede consultarse por ejemplo en el capítulo 16 de [85].

Nosotros, no vamos a detenernos en la exposición de la teoría asintótica, ni en obtener las expresiones de los estimadores de los parámetros, sino simplemente en exponer la forma de trabajar y destacar aquellas propiedades que distinguirán este tipo de procesos, de los procesos de raíces unitarias que veremos en la sección 7.3.3. Veremos también que este tipo de modelos no será adecuado para la modelización que nos plantearemos en los capítulos 8 y 9.

Para predecir a partir de estos modelos, nos fijamos en la ecuación (7.20). Sea cual sea el valor  $q$  del orden del proceso  $MA$ , si disponemos de observaciones del proceso hasta un instante  $t$ , y queremos predecir futuras observaciones, las ecuaciones de predicción serán:

$$\begin{aligned}
\tilde{y}_{t+1|t} &:= E(y_{t+1} \mid y_1, \dots, y_t) = \\
&= \alpha + \delta(t+1) + B_1\epsilon_t + B_2\epsilon_{t-1} + \dots + B_q\epsilon_{t-q+1} \\
\tilde{y}_{t+2|t} &:= E(y_{t+2} \mid y_1, \dots, y_t) = \\
&= \alpha + \delta_{t+2} + B_2\epsilon_t + \dots + B_q\epsilon_{t-q+2} \\
&\vdots \\
\tilde{y}_{t+q|t} &:= E(y_{t+q} \mid y_1, \dots, y_t) = \mu \\
\tilde{y}_{t+q+k|t} &= \alpha + \delta(t+q+k) \quad \forall k > 0
\end{aligned}$$

de donde podemos concluir que

$$\tilde{y}_{t+h|t} = \begin{cases} \alpha + \delta(t+h) + \sum_{i=h}^q B_i\epsilon_{t+h-i} & \text{si } h < q \\ \alpha + \delta(t+h) & \text{si } h \geq q \end{cases}$$

Por tanto:

$$\tilde{y}_{t+h|t} - y_{t+h} = \begin{cases} \sum_{i=0}^{h-1} B_i \epsilon_{t+h-i} & \text{si } h < q \\ 0 & \text{si } h \geq q \end{cases}$$

y el error de predicción será:

$$E\left(y_{t+h} - \tilde{y}_{t+h|t}\right)^2 = \begin{cases} \sum_{i=0}^{h-1} B_i' \Sigma B_i & \text{si } h < q \\ 0 & \text{si } h \geq q \end{cases}$$

Dejamos aquí estos resultados de momento. Los retomaremos en la sección 7.3.4, cuando veamos las diferencias entre estos procesos y los procesos con raíces unitarias.

### 7.3.3. Procesos con raíces unitarias

En la sección 7.2.1 vimos que un proceso  $VAR(p)$ ,

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + \epsilon_t$$

es estacionario si todos los valores propios de  $A$  (siendo la matriz  $A$  como en la sección 7.2.1) son en módulo menor que la unidad, i.e. si todos los  $\lambda$  que cumplen:  $|I_k \lambda^p - A_1 \lambda^{p-1} - A_2 \lambda^{p-2} - \dots - A_p| = 0$  son en módulo menor que la unidad. Si existe algún valor propio de  $A$  que sea en módulo mayor o igual que 1, el proceso será no estacionario.

La existencia de valores propios de  $A$  mayores que 1 da lugar a procesos explosivos (i.e. procesos en los que la varianza tiende a infinito, [132, 85]), que resultan poco aplicables a casos prácticos. Por ello vamos a centrarnos en el caso de que alguno de los valores propios de  $A$  sea igual a 1, y el resto estén dentro del círculo unidad.

Si tenemos un proceso  $VAR(p)$ , y  $A$  tiene  $m$  valores propios iguales a 1, decimos que tenemos un proceso  $VAR(p)$  con  $m$  raíces unitarias.

Vamos ahora a distinguir entre varios tipos de procesos VAR con raíces unitarias. Todo proceso  $VAR(p)$  puede expresarse:

$$(I_k - A_1L - A_2L^2 - A_3L^3 - \dots - A_pL^p)y_t = \nu + \epsilon_t \quad (7.25)$$

pero teniendo en cuenta la siguiente igualdad de polinomios:

$$\begin{aligned} (I_k - A_1L - A_2L^2 - A_3L^3 - \dots - A_pL^p) &= \\ &= (I_k - \rho L) - (\zeta_1L + \zeta_2L^2 + \dots + \zeta_{p-1}L^{p-1})(1 - L) \end{aligned}$$

con

$$\begin{aligned} \rho &= A_1 + A_2 + \dots + A_p, \\ \zeta_s &= -[A_{s+1} + A_{s+2} + \dots + A_p], \quad \forall s = 1, 2, \dots, p-1, \end{aligned}$$

podremos reformularlo como:

$$(I_k - \rho L)y_t - (\zeta_1L + \zeta_2L^2 + \dots + \zeta_{p-1}L^{p-1})(1 - L)y_t = \nu + \epsilon_t \quad (7.26)$$

Si  $\rho = I_k$ , es decir

$$A_1 + A_2 + \dots + A_p = I_k, \quad (7.27)$$

entonces

$$|I_k - A_1 - A_2 - \dots - A_p| = 0 \quad (7.28)$$

por lo que el proceso será no estacionario. En este caso, la ecuación (7.26) quedará:

$$(I_k - \zeta_1L - \zeta_2L^2 - \dots - \zeta_{p-1}L^{p-1})\Delta y_t = \nu + \epsilon_t$$

Si las raíces de  $|I_k\lambda^{p-1} - \zeta_1\lambda^{p-2} - \zeta_2\lambda^{p-3} - \dots - \zeta_{p-1}| = 0$  son ahora en módulo menor que la unidad,  $\Delta y_t$  será un proceso  $VAR(p-1)$ . Si  $\zeta$  vuelve

a tener un valor propio igual a 1, podremos repetir el todo el razonamiento anterior y concluir que  $\Delta^2 y_t$  es un proceso  $VAR(p-2)$  y así sucesivamente.

Luego dado un proceso  $VAR(p)$  no estacionario con una raíz unitaria, se cumple que:

$$|I_k - A_1 - \dots - A_p| = 0$$

y en este caso puede ocurrir que  $I_k = A_1 + A_2 + \dots + A_p$  o que  $I_k \neq A_1 + A_2 + \dots + A_p$ , ya que, aunque (7.27) implica (7.28), (7.28) no implica necesariamente (7.27).

A continuación vamos a estudiar la clase particular de modelos no estacionarios para los que se verifica (7.27). Estos modelos se llaman modelos integrados, y hemos visto que al diferenciarlos darán lugar a procesos estacionarios. Después veremos una clase más general de procesos no estacionarios, llamados procesos cointegrados. Estos procesos verifican (7.28) pero no cumplen (7.27). Como en los capítulos posteriores no vamos a necesitarlos, veremos su definición, y daremos referencias de dónde poder estudiarlos con más profundidad.

### Procesos integrados

Sea  $y_t$  un proceso  $VAR(p)$  de dimensión  $k$ ,

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + \epsilon_t$$

y supongamos que

$$I_k = A_1 + A_2 + \dots + A_p,$$

o dicho de otra forma, que el polinomio  $I_k - A_1 z - A_2 z^2 - \dots - A_p z^p$  tiene  $d < p$  raíces unitarias, y el resto de raíces fuera del círculo unidad. Entonces

podemos escribir

$$I_k - A_1L - A_2L^2 - \dots - A_pL^p = (1 - L)^d\alpha(L),$$

siendo  $\alpha(L)$  un polinomio en  $L$  de grado  $p - d$ , que ya satisface la condición de estacionariedad, por lo tanto podemos expresar el modelo  $VAR(p)$  como

$$\begin{aligned}\alpha(L)(1 - L)^d y_t &= \nu + \epsilon_t && \text{es decir,} \\ \alpha(L)\Delta^d y_t &= \nu + \epsilon_t\end{aligned}$$

y concluir que  $\Delta^d y_t$  es un proceso  $VAR(p - d)$  estacionario. A partir de aquí los pasos de estimación de parámetros, predicción, estudio de la bondad del modelo, etc, ... pueden hacerse tal y como hemos visto para el caso de los modelos  $VAR$  estacionarios.

Este tipo de modelos, que se transforman en estacionarios tras diferenciarlos, se llaman modelos integrados, y se define el orden de integración como el número de veces que hay que diferenciar el proceso para lograr la estacionariedad.

**Definición 3.1** *Un proceso integrado  $\{y_t\}$  se denomina proceso  $VARIMA(p, d, q)$ , si tomando diferencias de orden  $d$ , el proceso resultante es un  $VARMA(p, q)$ .*

En el capítulo 8, utilizaremos un modelo  $VARI(1, 1)$  para modelizar la evolución temporal de los campos visuales. Diferenciaremos la serie para lograr estacionariedad, y a partir de ahí trabajaremos como si tuviéramos un proceso  $VAR(1)$  estacionario.

### Procesos cointegrados

Los procesos cointegrados, surgen en el ámbito de la economía, donde en ocasiones se trabaja con variables económicas multivariantes, en las que se puede observar que existe un equilibrio a largo plazo entre sus componentes.

Existen situaciones, en que cada componente de la variable multivariante por sí misma es un proceso no estacionario, pero donde una combinación de las distintas componentes da lugar a un proceso estacionario.

Dado un proceso (una serie temporal)  $\{y_t\}$  multivariante (de dimensión  $k$ ), se dice que es un proceso cointegrado de orden  $(d, b)$  (abreviadamente, que  $y_t \sim CI(d, b)$ ), si todas las componentes de  $y_t$  son integradas de orden  $d$ , y además existe una combinación lineal de las componentes de  $y_t$ ,  $cy_t$ , con  $c = (c_1, c_2, \dots, c_k) \neq 0$ , que es integrado de orden  $d - b$ . Así por ejemplo, si todas las componentes de  $y_t$  son  $I(1)$ , y  $cy_t$  es estacionario ( $I(0)$ ), entonces se dice que  $y_t \sim CI(1, 1)$ . Al vector  $c$  se le llama “vector de cointegración”. Estos procesos fueron desarrollados en el contexto de modelización de series económicas, e introducidos por [71, 53]. Desde entonces su popularidad en el trabajo tanto teórico, como aplicado a modelizaciones econométricas, ha ido en aumento.

A partir de la definición, podemos concluir de inmediato que un vector de co-integración no es único.

A pesar de que la definición de sistema cointegrado es la que acabamos de dar, la mayor parte de la literatura referente a procesos cointegrados, se centra en el caso particular de procesos  $CI(1, 1)$  (cada componente  $y_t \sim I(1)$ , y existe una matriz  $C$  tal que  $Cy_t \sim I(0)$ ). La particularización es tal, que a menudo al hablar de procesos cointegrados ya se tiene en cuenta en la definición que todas las componentes del proceso  $y_t$  son  $I(1)$ , y se dice que el proceso es cointegrado de orden  $r$ , si existe una matriz  $C$  de rango  $r$  tal que  $y_t \sim I(1)$  y  $Cy_t \sim I(0)$ . Esta será la definición de cointegración que nosotros adoptaremos también en adelante.

**Proposición 7.3.1** *Dado un proceso VAR(p)*

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + \epsilon_t \quad (7.29)$$

de dimensión  $k$ , si la matriz  $\Pi = I_k - A_1 - \dots - A_p$  es de rango  $r$ , entonces el proceso  $\{y_t\}$  es cointegrado de rango  $r < k$ .

La demostración es muy simple si re-escribimos (7.29) como:

$$\begin{aligned} y_t - y_{t-1} = & c - (I_k - A_1)y_{t-1} + (I_k - A_1)y_{t-2} - \\ & -(I - A_1 - A_2)y_{t-2} + (I - A_1 - A_2)y_{t-3} - \\ & -(I - A_1 - A_2 - A_3)y_{t-3} + (I - A_1 - A_2 - A_3)y_{t-4} - \\ & -(I - A_1 - A_2 - A_3 - A_4)y_{t-4} + \dots - \\ & -(I - A_1 - A_2 - \dots - A_p)y_{t-p} + \epsilon_t \end{aligned}$$

ya que si llamamos  $D_1 = -(I - A_1)$ ,  $D_2 = -(I - A_1 - A_2)$ ,  $\dots$ ,  $D_i = -(I - A_1 - A_2 - \dots - A_i)$ , la expresión anterior nos queda:

$$\Delta y_t = c + D_1 \Delta y_{t-1} + D_2 \Delta y_{t-2} + \dots + D_{p-1} \Delta y_{t-p+1} - \Pi y_{t-p} + \epsilon_t$$

Si  $\Pi$  es una matriz de rango  $r < k$ , se puede factorizar como  $\Pi = HC$  siendo  $H$  una matriz  $k \times r$ , y  $C$  una matriz  $r \times k$ , ambas de rango  $r$ . En este caso, despejando el término que contiene  $y_{t-p}$  en el segundo término de la ecuación anterior, vemos que podemos expresar  $\Pi y_{t-p}$  como combinación lineal de procesos estacionarios, por lo que  $\Pi y_{t-p} = HC y_{t-p}$  debe ser estacionaria, y debe seguir siendo estacionaria si lo multiplicamos por  $(H'H)^{-1}H'$ , por lo tanto  $Cy_t$  es estacionaria, y cada fila de  $Cy_t$  representa una relación de cointegración.

Un desarrollo teórico sobre propiedades de estos modelos (en el caso multivariante), así como todo lo referente a estimación de parámetros y comprobación de la bondad del ajuste, puede encontrarse en [132, 85, 167].

### Tests de integración y de cointegración

Tanto en [85] como en [132], podemos encontrar varios de los tests de integración y de cointegración propuestos en la literatura. Entre ellos podemos

citar el test de Johansen [104] que construye un test de cociente de verosimilitudes para contrastar un rango de cointegración específico  $r = r_0$ , frente a un rango de cointegración mayor  $r = r_1$ . Y los tests de Engle y Granger [53] y Engle y Yoo [54], quienes proponen una batería de tests que pueden utilizarse para contrastar si existe una relación de cointegración entre un conjunto de variables. Algunos de estos tests, necesitan testear en algún momento si existen raíces unitarias en algún proceso univariante. Los test de raíces unitarias en el caso univariante, fueron desarrollados en gran parte por Dickey y por Fuller en [57, 45]. Un test de raíces unitarias para procesos  $VAR(p)$  se discute en [55], y podemos encontrar otros tests de cointegración en [159, 192].

### **Predicción a partir de procesos integrados**

En este apartado vamos a extraer sólo las ecuaciones de predicción para procesos integrados de orden 1. El no deducir las ecuaciones para un modelo integrado o cointegrado de cualquier orden (en general), es debido a que en los capítulos 8 y 9 sólo vamos a trabajar con modelos autoregresivos integrados de primer orden. El sacar las ecuaciones en particular para este caso, nos ayudará a comprender mejor la situación en el capítulo 9. De todas formas, las ecuaciones de predicción para modelos integrados o cointegrados de cualquier orden pueden encontrarse en la sección 11.3 de [132].

Nos centramos por tanto en un modelo autoregresivo integrado de orden 1. Para poder comparar con las predicciones en los modelos estacionarios en tendencia, y al mismo tiempo calcular más fácilmente el error de predicción, usaremos la representación  $MA(\infty)$  (la relación entre ambas representaciones la podemos encontrar en la ecuación 7.19). En la práctica, será mucho más sencillo el hacer la predicciones de forma recursiva como vimos en la sección

7.2.1 para un proceso VAR(p).

$$(1 - L)y_t = \delta + B(L)\epsilon_t \quad \text{i.e.} \quad \Delta y_t = \delta + B(L)\epsilon_t$$

siendo  $B(L)$  un proceso  $MA$  de orden infinito.

Como hemos supuesto que la serie  $\{y_t\}$  tenía únicamente una raíz unitaria, la serie  $\Delta y_t$  será estacionaria. En particular,  $\Delta y_t$  es un proceso estacionario  $MA(\infty)$ , por lo tanto podemos aplicar las ecuaciones de predicción de estos procesos (ec 7.20) y tendremos que:

$$\Delta \tilde{y}_{t+h|t} = \tilde{y}_{t+h|t} - \tilde{y}_{t+h-1|t} = \delta + \sum_{i=h}^{\infty} B_i \epsilon_{t+h-i}$$

y como

$$\begin{aligned} \tilde{y}_{t+h|t} &= (\tilde{y}_{t+h|t} - \tilde{y}_{t+h-1|t}) + (\tilde{y}_{t+h-1|t} - \tilde{y}_{t+h-2|t}) + \cdots + (\tilde{y}_{t+1|t} - y_t) + y_t = \\ &= \Delta \tilde{y}_{t+h|t} + \Delta \tilde{y}_{t+h-1|t} + \cdots + \Delta \tilde{y}_{t+1|t} + y_t = \\ &= h\delta + y_t + (B_h + \cdots + B_1)\epsilon_t + (B_{h+1} + \cdots + B_2)\epsilon_{t-1} + \cdots \end{aligned}$$

Luego

$$\tilde{y}_{t+h|t} = h\delta + y_t + (B_h + \cdots + B_1)\epsilon_t + (B_{h+1} + \cdots + B_2)\epsilon_{t-1} + \cdots$$

Veamos ahora cuanto vale el error de predicción. Es este caso:

$$\begin{aligned} y_{t+h} - \tilde{y}_{t+h|t} &= (\Delta y_{t+h} + \Delta y_{t+h-1} + \cdots + \Delta y_{t+1} + y_t) - \\ &\quad - (\Delta \tilde{y}_{t+h|t} + \Delta \tilde{y}_{t+h-1|t} + \cdots + \Delta \tilde{y}_{t+1|t} + y_t) \end{aligned}$$

y como  $\forall j \in \{1, \dots, h\}$  se cumple que

$$\begin{aligned} \Delta y_{t+j} - \Delta \tilde{y}_{t+j|t} &= \left( \delta + \epsilon_{t+j} + \sum_{i=1}^{\infty} B_i \epsilon_{t+j-i} \right) - \left( \delta + \sum_{i=j}^{\infty} B_i \epsilon_{t+j-i} \right) = \\ &= \epsilon_{t+j} + \sum_{i=1}^{j-1} B_i \epsilon_{t+j-i} \end{aligned}$$

tendremos que:

$$\begin{aligned}
 y_{t+h} - \tilde{y}_{t+h|t} &= \epsilon_{t+h} + \sum_{i=1}^{h-1} B_i \epsilon_{t+h-i} + \epsilon_{t+h-1} + \sum_{i=1}^{h-2} B_i \epsilon_{t+h-i} + \cdots + \epsilon_{t+1} = \\
 &= \epsilon_{t+h} + (I + B_1) \epsilon_{t+h-1} + (I + B_1 + B_2) \epsilon_{t+h-2} + \cdots + \\
 &\quad + (I + \sum_{i=1}^{h-1} B_i) \epsilon_{t+1}
 \end{aligned}$$

por lo que la varianza del error de predicción será:

$$\begin{aligned}
 E[y_{t+h} - \tilde{y}_{t+h|t}]^2 &= \\
 &= \Sigma + (I + B_1)' \Sigma (I + B_1) + (I + B_1 + B_2)' \Sigma (I + B_1 + B_2) + \\
 &\quad + \cdots + (I + B_1 + B_2 + \cdots + B_{h-1})' \Sigma (I + B_1 + B_2 + \cdots + B_{h-1}).
 \end{aligned}$$

### 7.3.4. Comparación: Procesos estacionarios en tendencia y procesos integrados

Para comparar los modelos estacionarios en tendencia y los modelos integrados, nos fijaremos principalmente en las diferencias existentes cuando queremos predecir a partir de ellos, en la varianza del error de predicción, y en las transformaciones que se requieren en cada caso para lograr la estacionariedad. Para más detalles consultar [85].

Recordamos las expresiones de los modelos que queremos comparar: los modelos estacionarios en tendencia  $y_t = \alpha + \delta t + B(L)\epsilon_t$ , como (7.24), y los modelos con raíces unitarias (integrados)  $(1 - L)y_t = \delta + B(L)\epsilon_t$ .

Para poder comparar las dos modelizaciones, vamos a particularizar al caso en que tanto en el modelo estacionario en tendencia como en el modelo con raíces unitarias,  $B(L)$ , es un proceso  $MA(\infty)$ .

#### Comparación de predicciones

Supongamos que tenemos observaciones hasta un instante  $t$  y que queremos predecir la variable en el instante  $t + h$ . Hemos visto que si trabajamos

con modelos estacionarios en tendencia, la ecuación de predicción será:

$$\tilde{y}_{t+h|t} = \alpha + \delta(t+h) + \sum_{i=h}^{\infty} B_i \epsilon_{t+h-i}$$

y al trabajar con modelos integrados de primer orden, la ecuación de predicción es:

$$\tilde{y}_{t+h|t} = h\delta + y_t + (B_h + \dots + B_1)\epsilon_t + (B_{h+1} + \dots + B_2)\epsilon_{t-1} + \dots$$

Por lo tanto, el parámetro  $\delta$  juega un papel muy parecido en las dos modelizaciones. Con cualquiera de los dos modelos, la predicción convergerá a una función lineal del horizonte de predicción  $h$ , con pendiente  $\delta$ . La principal diferencia está en la ordenada de esta función lineal, ya que para procesos estacionarios en tendencia la predicción converge a una recta que es independiente del valor de  $y_t$ , pero la ordenada en el origen de esta recta en el caso de modelos con raíces unitarias, va cambiando continuamente con cada nueva observación sobre  $y$ .

### Comparación de los errores de predicción

Veamos que también hay diferencias en los errores de predicción al trabajar con unos procesos u otros.

En el caso de procesos estacionarios en tendencia, el error de predicción es:

$$E\left(y_{t+h} - \tilde{y}_{t+h|t}\right)^2 = \sum_{i=0}^{h-1} B_i' \Sigma B_i$$

Podemos ver, que este error de predicción crece con el horizonte de predicción, es decir, predicciones a más largo plazo implicarán un mayor error de predicción.

Sin embargo, notar que en el límite este error de predicción coincide con la varianza no condicionada de la componente estacionaria del modelo, i.e. con la varianza de  $B(L)\epsilon$ .

En el caso de procesos integrados:

$$\begin{aligned} E[y_{t+h} - \tilde{y}_{t+h|t}]^2 &= \\ &= \Sigma + (I + B_1)' \Sigma (I + B_1) + (I + B_1 + B_2)' \Sigma (I + B_1 + B_2) + \\ &+ \cdots + (I + B_1 + B_2 + \cdots + B_{h-1})' \Sigma (I + B_1 + B_2 + \cdots + B_{h-1}). \end{aligned}$$

Al igual que en el caso anterior, la varianza también crece con el horizonte de predicción, pero ahora no converge cuando  $h \rightarrow \infty$ . En el caso univariante, se ha demostrado que la variación del error de predicción se aproxima linealmente a una función lineal de  $h$ , con pendiente  $(1 + B_1 + B_2 + \cdots)^2 \sigma^2$  (siendo ahora los  $B_i$  escalares en lugar de matrices). En el caso multivariante, nos quedamos con la idea de que este error de predicción no está acotado [85].

### Diferentes transformaciones para lograr estacionariedad

Una última diferencia entre procesos estacionarios en tendencia y procesos con raíces unitarias, es la forma de transformarlos para lograr series estacionarias.

Si realmente tenemos un proceso estacionario en tendencia, el tratamiento adecuado es restar  $y_t - \delta t$  para obtener una serie estacionaria. Sin embargo, si tuviéramos una serie con raíces unitarias, al restar  $y_t - \delta t$  lograríamos quitar la dependencia temporal de la estructura de la media, pero no de la estructura de la varianza.

Veámoslo con un ejemplo muy sencillo. Cogemos una serie temporal univariante con raíces unitarias (en particular un proceso de paso aleatorio con pendiente), y veamos que pasa al restarle  $\delta t$ .

Supongamos que tenemos el proceso con raíces unitarias univariante más simple

$$y_t = y_{t-1} + \delta + \epsilon_t$$

o lo que es lo mismo:

$$y_t = y_0 + \delta t + (\epsilon_1 + \epsilon_2 + \dots + \epsilon_t)$$

si restamos  $y_t - \delta t$  tendremos:

$$y_t - \delta t = y_0 + (\epsilon_1 + \dots + \epsilon_t).$$

Por tanto

$$E(y_t - \delta t) = 0 \quad \text{pero} \quad \text{Var}(y_t - \delta t) = t\sigma^2$$

Como podemos ver, no hemos conseguido que la serie sea estacionaria.

En caso de series con raíces unitarias, la forma de lograr series estacionarias es diferenciándolas. Sin embargo, si intentamos diferenciar un proceso estacionario en tendencia:

$$\begin{aligned} y_t &= \alpha + \delta t + \epsilon_t + B_1\epsilon_{t-1} + B_2\epsilon_{t-2} + \dots \\ y_{t-1} &= \alpha + \delta(t-1) + \epsilon_{t-1} + B_1\epsilon_{t-2} + B_2\epsilon_{t-3} + \dots \\ \Delta y_t &= \delta + (\epsilon_t - \epsilon_{t-1}) + B_1(\epsilon_{t-1} - \epsilon_{t-2}) + B_2(\epsilon_{t-2} - \epsilon_{t-3}) - \dots \\ \Delta y_t &= \delta + (1-L)B(L)\epsilon_t \end{aligned}$$

Obtenemos una serie estacionaria, pero hemos introducido una raíz unitaria en la parte  $MA$ , hecho que hará el trabajo con este modelo mucho más complicado.

### 7.3.5. Procesos con cambios de régimen (o con cambios estructurales)

Como hemos visto en los capítulos anteriores, la suposición de estacionariedad es sumamente importante, y el que esta condición no se verifique puede deberse a diversas causas, siendo las más frecuentes, la aparición de tendencias, componentes estacionales o varianzas que varían en el tiempo. En

las secciones 7.3.2 y 7.3.3 hemos estudiado la no estacionariedad debida a la existencia de tendencias y varianzas que dependen del tiempo, y hemos visto que en este caso podemos recobrar la estacionariedad fácilmente (restando la tendencia o diferenciando, dependiendo del modelo utilizado).

En este apartado, vamos a hablar de un tipo particular de series temporales no estacionarias que no podremos transformar en estacionarias mediante ninguna transformación. Estas series son las llamadas series con cambios de régimen o con cambios estructurales. Estos cambios estructurales, surgen habitualmente cuando se produce algún acontecimiento (social, económico, el inicio o el final de una guerra, la entrada de una nueva legislación, una caída de la bolsa, el contraer una enfermedad,...) que causa cambios drásticos en alguna variable, y que hace que a partir de aquí el comportamiento de la serie temporal sea distinto a su comportamiento hasta este momento.

La solución a este problema, ha sido desarrollar varias modelizaciones que permitan tener en cuenta los cambios estructurales. Entre las modelizaciones desarrolladas, las dos que han tenido mayor difusión y aplicabilidad han sido los modelos con cambios en régimen, que comentaremos a continuación, y los modelos en espacio de estados, en los que los parámetros van cambiando continuamente, que no comentaremos demasiado ya que no los utilizaremos después. Junto a estos modelos, se han desarrollado también gran cantidad de trabajos y tests que permiten “detectar” en estas series temporales la existencia de cambios estructurales, y estudiar su efecto. Así por ejemplo encontramos en la literatura tests que contrastan la hipótesis de cero cambios frente a la alternativa de un único cambio [33], o frente a la alternativa de múltiples cambios [52]. Encontramos también procedimientos que intentan estimar el número de cambios estructurales [117] en una serie temporal. Este es realmente un problema de selección de modelo, que abordan entre otros Kim y Maddala [118], P. Phillips [158], y Bai y Perron [6, 7], quienes

discuten además las distribuciones asintóticas de los estimadores. También se han desarrollado tests que además de decir si existe o no algún cambio estructural en la serie temporal, en el caso de concluir que exista, identifican su localización. Este es el caso de los tests CUSUM (plots of the cumulative sum) [23, 161, 111]; de los tests propuestos por Banerjee, Lumsdaine y Stock [8], Zivot y Andrews [209], o Chu y White [37]; o los tests que desde una aproximación bayesiana son discutidos por Kim y Maddala [118].

La literatura sobre la materia, y los tests y estudios derivados de las referencias que acabamos de dar son extensísimos, ya que es un área en la que actualmente existen muchos investigadores trabajando. Para un tratamiento más extenso y ver en qué consiste exactamente cada test, además de las referencias que hemos dado, recomendamos consultar [136], y las referencias que allí se citan.

La existencia de cambios estructurales en las series temporales, puede ir acompañada también de la existencia de raíces unitarias en el proceso, o de cointegración. Los estudios de Rappoport y Reichelin [166], Hendry y Neale [96] y Perron [157] entre otros, han demostrado que las inferencias sobre raíces unitarias se ven afectadas por la existencia de cambios estructurales (los test de raíces unitarias en presencia de cambios estructurales tienden a rechazar menos veces la hipótesis nula de que no hayan raíces unitarias). Lo mismo pasa con los tests de cointegración, como han demostrado Gregory, Nason y Watt [75], y Campos, Ericsson y Hendry [26]. Volvemos a referir a [136] y a las referencias que acabamos de citar, a todo aquel que quiera saber más sobre cómo quedan modificados estos tests (de raíces unitarias y de cointegración), en presencia de cambios estructurales.

Antes de centrarnos en los modelos con cambios en régimen, comentar brevemente otra aproximación para modelizar series no estacionarias, que permite que los cambios sean graduales (ver [136]). Se trata de los modelos

en espacios de estados [89], que permiten que los cambios se realicen de forma continua en el tiempo.

Dada una serie temporal multivariante  $y_t$  de dimensión  $N$ , un modelo en espacio de estados formula

$$y_t = Z_t \alpha_t + d_t + \epsilon_t, \quad t = 1, \dots, T \quad (7.30)$$

donde  $Z_t$  es una matriz  $N \times m$  de covariables,  $d_t$  es un vector  $N \times 1$ ; y  $\epsilon_t$  es un vector  $N \times 1$  de errores, tales que  $E(\epsilon_t) = 0$  y  $Var(\epsilon_t) = H_t$

En general  $Z_t$  y  $d_t$  son conocidos, y aunque pueden variar con el tiempo, normalmente esta variación es determinista. Los elementos de  $\alpha_t$  no son observables, y se supone que están generados por un proceso de Markov de primer orden, i.e.:

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t, \quad t = 1, \dots, T \quad (7.31)$$

siendo  $T_t$  una matriz  $m \times m$ ,  $c_t$  un vector  $m \times 1$ ,  $R_t$  una matriz  $m \times g$  y  $\eta_t$  un vector  $g \times 1$  de errores, con  $E(\eta_t) = 0$  y  $Var(\eta_t) = Q_t$ .

A las matrices  $Z_t, d_t, H_t, T_t, c_t, R_t$  y  $Q_t$  se las llama matrices del sistema, y se supone que no son estocásticas, i.e. que aunque pueden cambiar con el tiempo, lo hacen de una forma predeterminada.  $\alpha_t$  recibe el nombre de vector de estados, la ecuación (7.30) se llama ecuación de observación, y (7.31) es la ecuación de transición.

Para completar la especificación del espacio de estados se requieren varias suposiciones más [89]:

1.  $E(\alpha_0) = a_0$  y  $Var(\alpha_0) = P_0$  (aunque esta suposición puede relajarse en algunas situaciones [89]),
2. los distintos errores están incorrelados entre sí en todos los instantes, y están incorrelados también con  $\alpha_0$ , i.e.  $E(\epsilon_t \eta'_s) = 0 \quad \forall s, t = 1, \dots, T$  y  $E(\epsilon_t \alpha'_0) = 0$  y  $E(\eta_t \alpha'_0) = 0$  para  $t = 1, \dots, T$

En ocasiones, podemos encontrarnos con que las matrices  $Z_t, H_t, T_t, R_t$  y  $Q_t$  pueden depender a su vez de un conjunto de parámetros desconocidos (hiperparámetros) que deberán estimarse también.

Una vez tenemos un modelo expresado en forma de espacio de estados, la estimación de los parámetros del modelo, y la predicción de futuras observaciones realizadas a partir de la información disponible hasta el instante  $t$  se realizará utilizando el Filtro de Kalman [109, 110, 144].

Aunque los modelos de estados junto con el filtro de Kalman son utilizados en multitud de casos prácticos, no vamos a centrarnos en ellos, y pasamos a la sección siguiente en la que nos centramos en los modelos con cambios en régimen que son los que necesitaremos en nuestra modelización en los capítulos 8 y 9.

### **Modelos con cambios en régimen**

Muchas veces, al observar el comportamiento de una serie temporal, vemos que cambia abruptamente a partir de un instante, debido a que se ha producido algún suceso que ha influido en su comportamiento. Es por ejemplo el caso del comportamiento de variables económicas ante anuncios de acontecimientos bélicos, cambios en las políticas económicas, ... o el de variables biológicas al contraer una enfermedad, o cuando se aplica un tratamiento médico.

Una idea muy sencilla para modelizar series temporales que se ven sometidas a estos cambios abruptos, es la de modelizar cada período por separado (un modelo para el comportamiento de la serie antes del cambio, y otra para su comportamiento a partir del cambio). Pero si un proceso ha cambiado en el pasado, puede ser que vuelva a cambiar en el futuro, y esta posibilidad debería tenerse en cuenta si nos planteamos realizar predicciones.

Por lo tanto, ante procesos en los que observamos cambios abruptos, va-

mos a suponer que el proceso está influenciado por una variable no observada,  $s_t$ , que llamaremos estado o régimen. Entonces, si inicialmente hubiéramos modelizado la serie como un  $VAR(p)$ ,  $y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t$  con  $u_t \sim iid(0, \Sigma)$ , e  $y_0, \dots, y_{1-p}$  fijos, al tener un proceso con cambios en régimen, si suponemos que  $s_t \in \{1, 2, \dots, M\}$ , el modelo a plantear sería:

$$y_t = \begin{cases} \nu_1 + A_{11}y_{t-1} + \dots + A_{p1}y_{t-p} + \Sigma_1^{1/2}u_t & \text{si } s_t = 1 \\ \nu_2 + A_{12}y_{t-1} + \dots + A_{p2}y_{t-p} + \Sigma_2^{1/2}u_t & \text{si } s_t = 2 \\ \vdots & \\ \nu_M + A_{1M}y_{t-1} + \dots + A_{pM}y_{t-p} + \Sigma_M^{1/2}u_t & \text{si } s_t = M \end{cases}$$

o equivalentemente:

$$y_t = \nu_{s_t} + A_{1s_t}y_{t-1} + \dots + A_{ps_t}y_{t-p} + \Sigma_{s_t}^{1/2}u_t. \quad (7.32)$$

Aunque éste es el modelo general, habitualmente sólo alguno de los parámetros del modelo dependerá del estado, siendo el resto de parámetros invariantes respecto al régimen.

Además, los cambios en régimen son, a su vez, procesos aleatorios (muy pocas veces responden a procesos deterministas), por lo tanto para tener un modelo completo deberemos especificar una distribución de probabilidad para el paso de un régimen a otro.

Los modelos con cambios en régimen fueron introducidos en la literatura económica por Quandt [164], pero los modelos con cambios de régimen más difundidos, son los llamados modelos con cambios en régimen de Markov (Markov-switching model), introducidos en la literatura estadística por Hamilton [82, 83]. En estos modelos se supone que el proceso no observado, el de los regímenes, está gobernado por un proceso de Markov discreto (estados discretos y tiempos discretos), que viene definido por determinadas probabilidades de transición:

$$p_{ij} = Pr(s_{t+1} = j \mid s_t = i), \sum_j p_{ij} = 1, \forall i, j \in \{1, \dots, M\}. \quad (7.33)$$

Una de las características de los modelos de cambios en régimen estudiados en la literatura, es que aunque la serie  $y_t$  sea una serie multivariante, todas sus componentes se encuentran en el mismo estado  $s_t$  en el instante  $t$ . Los cambios en régimen se producen en el mismo instante para todas las componentes de la serie. No hemos encontrado ninguna referencia a trabajos en los que los estados no sean univariantes. Por supuesto que, como se comenta en [85], podríamos “redefinir” los estados, y definir para cada  $t$ ,

$$\begin{aligned} \tilde{s}_t = 1 & \text{ si } & (s_{1,t} = 1, s_{2,t} = 0, s_{3,t} = 0, \dots, s_{k,t} = 0) \\ \tilde{s}_t = 2 & \text{ si } & (s_{1,t} = 0, s_{2,t} = 1, s_{3,t} = 0, \dots, s_{k,t} = 0) \\ & & \vdots \\ \tilde{s}_t = M & \text{ si } & (s_{1,t} = 1, s_{2,t} = 1, s_{3,t} = 1, \dots, s_{k,t} = 1) \end{aligned}$$

con  $M = 2^k$ . En el caso en que  $k$  sea grande será bastante complicado trabajar con estos  $\tilde{s}$  y además, como veremos en los capítulos 8 y 9 donde  $k = 52$ , el cálculo de las probabilidades de paso  $p_{ij}$  son difíciles de calcular. En ambos capítulos buscaremos otra forma de modelizar el proceso con cambios en régimen. Aquellos lectores interesados en saber cómo estimar los parámetros, y cómo predecir cuando se trabaja con este tipo de modelos pueden consultar [84, 121, 120]. En los capítulos 8 y 9 propondremos una generalización de estos métodos para trabajar con cambios de régimen multivariantes con  $k$  grande.

## 7.4. Campos aleatorios de Markov y algoritmos de estimación de parámetros

### 7.4.1. Introducción

En este apartado queremos revisar brevemente el concepto de “campos aleatorios de Markov” en un cierto contexto que comentaremos a continua-

ción, y principalmente estudiar algoritmos que permitan la estimación de parámetros en estos modelos. Estos métodos nos serán de utilidad en los capítulos 8 y 9, cuando nos planteemos la modelización de la distribución espacio temporal de CV de pacientes que sufren glaucoma por dos motivos:

- Trabajaremos entonces con series temporales multivariantes, y podremos considerar cada una de las componentes de la serie (fijado  $t$ ) como un campo aleatorio de Markov de tipo gaussiano.
- En cada posición del campo visual consideraremos una variable latente indicando si la posición es sana o enferma. Dado  $t$  también se supondrá que sigue un Campo Aleatorio de Markov en este caso de tipo binario.

Notar que estaremos en un contexto muy similar al del análisis bayesiano de imágenes (segmentación, restauración, ...) pues toda imagen digitalizada se identifica con un retículo regular y cada píxel de la imagen se identifica con una posición del retículo. Asociada al retículo tenemos un conjunto de variables aleatorias. En este marco, el problema (segmentación, restauración,...) se reformula como un problema de estimación de parámetros, ya que se considera la imagen observada  $Y$  como una versión con ruido de la verdadera imagen  $S$ . i.e.  $S$ , la verdadera imagen, se considera como un parámetro más de la distribución de probabilidad de  $Y$ . Distribuciones a priori para  $S$  pueden construirse para incorporar en el modelo la información que a priori se tenga de la verdadera imagen. El modelo de campo aleatorio de Markov permite incorporar en el modelo relaciones de dependencia entre píxeles vecinos y dar probabilidades altas a imágenes cuyos rasgos sean similares a los deseados.

Por este motivo, para hacer este estudio, nos centraremos en este contexto, el del análisis de imágenes, ya que pensamos que es más intuitivo y en él podremos comprobar y comparar la eficacia de los distintos métodos más fácilmente. Además, un CV puede considerarse como una imagen a niveles

de gris, por lo que la aplicación de lo que aquí veamos a la modelización que planteemos en los capítulos 8 y 9, será inmediata.

Los modelos de campos aleatorios de Markov (CAM) han sido muy utilizados en el ámbito del análisis de imágenes [31, 126]. Los primeros trabajos en la materia, son los artículos de Besag [18] y de Geman y Geman [61]. A partir de entonces su aplicabilidad ha ido en aumento, convirtiéndose en una herramienta muy utilizada en el análisis de imágenes médicas, como por ejemplo para la reconstrucción de datos de tomografías computerizadas (single photon emission computed tomography data) [62, 5, 74, 202, 201, 178], en la detección de contornos en ventrículos [193], y en la segmentación y alineamiento de angiografías fluoresceínicas para la detección y evaluación de enfermedades como la retinopatía diabética [51, 100, 186].

Otras aplicaciones de los CAM incluyen la restauración de películas [108], la restauración y descompresión de imágenes de satélite [102, 139], el reconocimiento de caracteres escritos [25], reconocimiento del habla [131, 73, 147] e imágenes de sonar [206, 145], etc...

En muchas de las técnicas de procesamiento de imágenes, la estimación de parámetros tiene un papel importantísimo, pero la mayoría de los procedimientos clásicos de estimación no pueden aplicarse en situaciones en las que existen datos faltantes o cuando la constante normalizadora es desconocida.

En este apartado vamos a revisar varios métodos de estimación de parámetros. Tras su revisión haremos un estudio de simulación para comparar los que hemos considerado más relevantes.

Seguiremos la siguiente organización: en la sección 7.4.2 introduciremos los campos aleatorios de Markov y derivaremos el problema de estimación de parámetros en este contexto. En la sección 7.4.3 introduciremos el método de verosimilitud de Montecarlo. En la sección 7.4.4 revisaremos los métodos de estimación de parámetros basados en la pseudoverosimilitud y en la

sección 7.4.5 los basados en el gradiente estocástico. Haremos un análisis comparativo con los tres algoritmos expuestos más relevantes en la sección 7.4.6.

### 7.4.2. Campos aleatorios de Markov. Conceptos básicos

Como ya hemos comentado, para introducir la metodología general de los campos aleatorios de Markov (CAM) y revisar brevemente algunos algoritmos de estimación de parámetros, vamos a centrarnos en el contexto del análisis de imágenes.

Sea por tanto  $Y = \{Y(x)\}_{x \in X}$  la imagen observada, y sea  $S = \{S(x)\}_{x \in X}$  la verdadera imagen.

Deberemos asumir un modelo para los datos observados  $Y$  dado  $S$ , dependiendo de un vector de parámetros  $\beta$ :

$$p_{\beta}(Y/S) = c(\beta/S)^{-1} \exp(-h_{\beta}(Y/S)). \quad (7.34)$$

y podemos formular también una distribución a priori de la verdadera imagen, utilizando un modelo probabilístico de Markov con vector de parámetros  $\alpha$ :

$$\pi_{\alpha}(S) = c(\alpha)^{-1} \exp(-h_{\alpha}(S))$$

En estos modelos, a la función  $h_{\alpha}(S)$  se le llama “función de energía” y es la suma de funciones definidas sobre cliques. Un clique es un conjunto de píxeles con la particularidad de que todos son mutuamente vecinos respecto a una vecindad dada.  $c(\alpha)$  es la constante normalizadora, que en la mayoría de modelos tiene una expresión analítica desconocida.

Al vector de parámetros  $\alpha$  se le llama vector de hiperparámetros.

Como hemos comentado, vamos a centrarnos principalmente en la estimación conjunta del vector de parámetros  $\theta = (\alpha, \beta)$ .

Estos parámetros se estiman a menudo utilizando técnicas “ad-hoc” [112], o una aproximación completamente bayesiana, asumiendo un modelo a priori para  $\theta$  [146, 139]. Nosotros vamos a suponer que no disponemos de ningún conocimiento a priori de  $\theta$ , y vamos a centrarnos en el método de máxima verosimilitud (ML).

Si tuviéramos acceso a una realización de  $S$  (i.e. si tuviéramos “training data”), para encontrar el estimador máximo verosímil deberíamos maximizar el logaritmo de la verosimilitud, donde:

$$l(\theta) = \log p_{\theta}(S, Y) = \log \{p_{\beta}(Y/S)\pi_{\alpha}(S)\}. \quad (7.35)$$

Pero en la gran mayoría de métodos basados en CAM, el objetivo final es la estimación de la verdadera imagen ( $S$ ) a partir de los datos observados ( $Y$ ), y de la distribución a priori ( $\pi_{\alpha}(S)$ ). Por lo tanto, en adelante vamos a suponer que no disponemos de ninguna observación de la verdadera imagen ( $S$ ) y basaremos la estimación de  $\theta$  sólo en la imagen observada ( $Y$ ).

Si la verdadera imagen ( $S$ ) no ha sido observada, tenemos un problema de datos faltantes, y para obtener el estimador máximo verosímil de los parámetros, en vez de maximizar la log-verosimilitud (7.35), deberemos trabajar con la distribución marginal:

$$p_{\theta}(Y) = \int p_{\beta}(Y/S)\pi_{\alpha}(S)dS \quad (7.36)$$

y maximizar la log-verosimilitud de esta distribución marginal:

$$\log p_{\theta}(Y) = \log \int p_{\beta}(Y/S)\pi_{\alpha}(S)dS = E_{\theta}[\log p_{\theta}(S, Y) | Y], \quad (7.37)$$

expresión normalmente muy difícil de maximizar.

Aún en el caso de disponer de una observación completa  $Z = (S, Y)$ , el cálculo del estimador máximo verosímil es una tarea complicada debido a la intratabilidad de la constante normalizadora de  $p_{\theta}(S, Y)$  (de hecho, sólo

en el caso de que la distribución a priori  $\pi_\alpha(S)$  sea gaussiana, la constante normalizadora tiene una expresión analítica).

En la literatura podemos encontrar varias aproximaciones para resolver estas dificultades (la existencia de datos faltantes y la intratabilidad de la constante normalizadora). Revisemos los distintos métodos desarrollados para tratar cada problema por separado, y luego ya veremos sus generalizaciones para aquellos casos en que se presentan los dos “problemas”.

### Estimación de parámetros con datos faltantes

El algoritmo más popular para trabajar en problemas de datos faltantes ( $S$  desconocida), es el algoritmo EM [44].

El algoritmo EM es un algoritmo iterativo, que partiendo de una inicialización de los parámetros  $\theta = \theta_0$ , construye  $\theta_{n+1}$  a partir de  $\theta_n$  maximizando:

$$g_n(\theta) = E_{\theta_n}[\log p_\theta(S, Y) | Y]. \quad (7.38)$$

La comparación de las ecuaciones (7.37) y (7.38) nos da una idea intuitiva del algoritmo EM. Este algoritmo intenta rellenar en cada paso los datos faltantes, utilizando para ello el modelo y los valores que en cada momento tengan los parámetros.

Cada iteración del algoritmo consta de dos pasos. El primero es el “Paso E” (esperanza), en el que se calcula la media de la función de log-verosimilitud condicionada sobre los datos observados y el valor de los parámetros en esta iteración  $\theta = \theta_n$ . En el segundo paso, llamado “Paso M” (maximización), maximizamos esta media con respecto a  $\theta$ .

Cuando no podemos calcular la esperanza en (7.38), podemos aproximarla utilizando métodos de Montecarlo. Esta es la base de gran cantidad de algoritmos, llamados algoritmos EM Estocásticos [46, 150, 172, 165, 65], cuya idea original se debe a Celeux y Diebolt. Aunque este es el método más conocido y utilizado, existen otras aproximaciones y métodos derivados, como el

método propuesto por Noda et al. [152] que combina el algoritmo EM con la descomposición del campo medio, o el algoritmo de máxima verosimilitud generalizado (GLM), consistente en maximizar alternativamente (7.35) con respecto a  $S$  y a  $\theta$ . Este último método es subóptimo, pero es rápido y sencillo. Fue introducido en el contexto de análisis de imágenes por Lakshamanan et al [123], siendo utilizado también más recientemente en [30].

Otro procedimiento general para trabajar en problemas con datos faltantes, es el “Iterated Conditional Estimation” (ICE) propuesto por Pieczynski [160]. Este método es similar al algoritmo EM pero converge más rápidamente. Se basa en el hecho de que si  $\hat{\theta} = \hat{\theta}(S, Y)$  es un buen estimador de  $\theta$  en un contexto en el que no hayan observaciones faltantes (por ejemplo el estimador máximo verosímil),  $E_{\theta}(\hat{\theta}(S, Y))$  será la mejor aproximación a él (mejor en sentido de mínimos cuadrados) cuando  $S$  sea desconocida (no observada). Desafortunadamente, esta esperanza condicional depende de  $\theta$ . Para superar este obstáculo, proponen un algoritmo iterativo que en cada paso aproxime la esperanza por la media de una muestra de  $n$  realizaciones de  $S$ , dado el anterior valor de  $\theta$ .

### Estimación de parámetros cuando desconocemos la constante normalizadora

Suponemos ahora que desconocemos la constante normalizadora, pero que la verdadera imagen  $S$  es conocida. En este caso el método tradicionalmente usado ha sido el sustituir la función de verosimilitud (7.35) por otra función más manejable, como por ejemplo la función de pseudoverosimilitud [17, 18, 186], dada por:

$$PL_{\theta}(S, Y) = \prod_{x \in X} p_{\beta}(Y(x)/S(x))\pi_{\alpha}(S(x)/S(z)_{x \neq z}, \alpha).$$

También pueden utilizarse ante este problema algoritmos basados en el

gradiente estocástico, de entre los que destaca el algoritmo de Younes [207]. Para explicar brevemente en qué consiste este algoritmo simplificamos notación, llamando  $p_\theta(S, Y) = c(\theta)^{-1} \exp(-h_\theta(S, Y))$ . Es éste un algoritmo iterativo, que partiendo de  $S$  (conocida) y de un valor inicial del vector de parámetros  $\theta = \theta_0$ , actualiza las estimaciones de los parámetros como sigue:

$$\theta_{n+1} = \theta_n + \frac{c}{n+1} \left\{ h'(S_{n+1}, Y_{n+1}) - h'(S, Y) \right\}$$

donde  $Y$  es la imagen observada,  $h'(-)$  es el gradiente de  $h(-)$  con respecto a  $\theta$ , y  $(S_{n+1}, Y_{n+1})$  es una realización de  $(S, Y)$  simulada mediante el muestreador de Gibbs de acuerdo con su distribución conjunta para  $\theta = \theta_n$ .

Los métodos que acabamos de exponer, permiten estimar todos los parámetros del modelo conjuntamente, aunque normalmente al trabajar con un problema de estimación donde  $S$  es conocida y la constante normalizadora es intratable, se suele partir el problema de estimación en dos pasos, y estimar por separado los parámetros de la distribución de  $S$  y los de la distribución de  $Y/S$ .

Veamos ahora métodos que nos permitan estimar parámetros cuando tenemos datos faltantes, y la constante normalizadora es desconocida. Veremos cómo la mayoría de métodos son aproximaciones o generalizaciones de los métodos ya comentados [78].

Son muchos los métodos propuestos en la literatura para estimar parámetros en problemas de datos faltantes (u observaciones imperfectas), y constante normalizadora intratable. La mayoría de ellos se enfocan a resolver aplicaciones muy específicas, como por ejemplo la segmentación o restauración de imágenes, y se basan en la idea de ir alternando pasos de estimación-restauración: si suponemos que hemos completado las observaciones perdidas,  $\hat{S}$ , y tenemos  $\theta = g(S, Y)$  un buen estimador de  $\theta$  (en un contexto de no datos faltantes), entonces  $\hat{\theta} = g(\hat{S}, Y)$  será un buen estimador del vector de parámetros.

Como nos interesan los métodos más generales, vamos a explicar en detalle únicamente el método de verosimilitud de Montecarlo, el método de Chalmond (Chalmond's Gibbsian EM algorithm), y el algoritmo de Younes. Algoritmos más particulares, similares o muy relacionados con éstos, son también comentados en los apartados respectivos. Sólo recordar que estamos interesados en la estimación máximo verosímil, en modelos de campos aleatorios de Markov, y bajo una perspectiva no bayesiana.

### 7.4.3. Verosimilitud de Montecarlo

En esta sección vamos a introducir y aplicar al contexto del análisis de imágenes, el algoritmo de verosimilitud de Montecarlo (MCL). Este algoritmo ha sido ampliamente utilizado en el contexto de procesos puntuales, y permite obtener una buena aproximación del estimador máximo verosímil, en presencia de datos faltantes y siendo la constante normalizadora desconocida.

El método general de verosimilitud de Montecarlo (MCL) fue introducido en la literatura estadística por Geyer et al [64], pero su generalización para trabajar con datos faltantes y constante normalizadora desconocida fue desarrollada por Gelfand et al [59].

El método MCL se basa en aproximar el logaritmo de la función de verosimilitud (7.37) utilizando métodos MCMC [65] (ver [59] y [63] para una revisión de este método más detallada que la dada a continuación), y su fundamento teórico reside en dos conocidos resultados probabilísticos.

Sea  $p(S) = c(\theta) \exp(-h_\theta(S))$  cualquier distribución de probabilidad general con vector de parámetros  $\theta$ . Sea  $\theta = \theta_0$  un valor inicial para este vector. Teniendo en cuenta que  $c(\theta) = 1 / \int \exp(-h_\theta(S)) dS$ , es fácil comprobar que:

$$\frac{c(\theta)}{c(\theta_0)} = E_{\theta_0} \left( \frac{\exp(-h_\theta(S))}{\exp(-h_{\theta_0}(S))} \right).$$

Por otro lado, si volvemos a nuestro caso particular, usando la ecuación

(7.36) es también fácil comprobar que:

$$p_{\theta}(Y) = \frac{c(\theta/Y)}{c(\theta)}, \quad (7.39)$$

siendo  $c(\theta/Y)$  la constante normalizadora de la distribución condicional de  $S$  dado  $Y$  y siendo  $c(\theta)$  la constante normalizadora de la distribución conjunta (7.35), i.e.  $p_{\theta}(S, Y) = c(\theta)^{-1} \exp(-h_{\theta}(S, Y))$ .

Si combinamos los dos resultados anteriores, el logaritmo de la función de verosimilitud verifica:

$$l(\theta) \propto \log E_{\theta_0} \left\{ \frac{\exp(-h_{\theta}(S, Y))}{\exp(-h_{\theta_0}(S, Y))} / Y \right\} - \log E_{\theta_0} \left\{ \frac{\exp(-h_{\theta}(S, Y))}{\exp(-h_{\theta_0}(S, Y))} \right\} \quad (7.40)$$

Finalmente, estas esperanzas pueden ser aproximadas utilizando métodos de Montecarlo. Para ello necesitamos muestras de la distribución conjunta de  $S$  e  $Y$ , y muestras de la distribución condicional dada  $Y$ , ambas para  $\theta = \theta_0$ , que podremos obtener utilizando métodos MCMC como por ejemplo el muestreador de Gibbs [61].

Si  $(S_1, Y_1), \dots, (S_k, Y_k)$  son muestras de la distribución conjunta,  $S_1^*, \dots, S_k^*$  son muestras de la distribución condicional, e  $y$  es el valor observado de  $Y$ , podemos aproximar  $l(\theta)$  por:

$$\hat{l}(\theta)_k \propto \log \frac{1}{k} \sum_{i=1}^k \left\{ \frac{\exp(-h_{\theta}(S_i^*, y))}{\exp(-h_{\theta_0}(S_i^*, y))} \right\} - \log \frac{1}{k} \sum_{i=1}^k \left\{ \frac{\exp(-h_{\theta}(S_i, Y_i))}{\exp(-h_{\theta_0}(S_i, Y_i))} \right\} \quad (7.41)$$

Maximizando esta aproximación obtendremos una aproximación de Montecarlo,  $\hat{\theta}_k$ , al MLE. Cuanto mayor sea el valor de  $k$ , mejor será la aproximación. De hecho si  $k$  es fijo, sólo si  $\theta$  está en una vecindad de  $\theta_0$ ,  $\hat{l}(\theta)_k$  será una buena aproximación de  $l(\theta)$ . El tamaño de la vecindad depende del tamaño muestral,  $k$ . Cuando el valor de  $\theta_0$  no esté próximo al MLE  $\hat{\theta}$ , la aproximación dada por este método podría llegar a ser muy mala, o podría también no encontrar ningún máximo. Por ello, varios autores [64] proponen un algoritmo recursivo que repita el procedimiento anterior realizando

únicamente maximizaciones locales, partiendo en cada iteración del máximo obtenido en la iteración anterior.

Este método permite construir tests de cocientes de verosimilitud e intervalos de confianza basados en verosimilitudes. Además no se trata de un algoritmo iterativo, por lo que no deberemos imponer ningún criterio de parada. La precisión de la aproximación dependerá del tamaño de la muestra de Montecarlo, y esta puede decidirse a priori, dependiendo de la relación que exijamos entre tiempo computacional y precisión de los resultados. Debemos enfatizar el hecho de que este método aproxima la verosimilitud, por lo tanto no es ninguna simplificación de la estimación máximo verosímil.

#### 7.4.4. Algoritmos basados en la pseudoverosimilitud

1. **Algoritmo EM de Gibbs de Chalmond** Chalmond [29] describe una versión del algoritmo EM que consiste fundamentalmente en sustituir la función de log-verosimilitud en (7.38) por la pseudoverosimilitud, y en utilizar muestreador de Gibbs para calcular la esperanza.

Si reemplazamos  $\pi_\alpha(S)$  por la pseudoverosimilitud:

$$PL_\alpha(S) = \prod_{x \in X} \pi_\alpha(S(x)/S(z)z \neq x, \alpha), \quad (7.42)$$

y análogamente  $p_\theta(S, Y)$  por la pseudoverosimilitud:

$$PL_\theta(S, Y) = \prod_{x \in X} p_\beta(Y(x)/S(x))\pi_\alpha(S(x)/S(z)x \neq z, \alpha), \quad (7.43)$$

el algoritmo EM (7.38) actualizaría la estimación actual  $\theta_n$  como:

$$\theta_{n+1} = \text{Arg Max}_\theta E[\log PL_\theta(S, Y)/\theta_n]. \quad (7.44)$$

Podemos utilizar el muestreador de Gibbs para calcular la esperanza por métodos de Montecarlo.

Al aproximar la verosimilitud por la pseudoverosimilitud, estamos suponiendo independencia condicional, por lo tanto pese a sus buenas propiedades asintóticas (sección 7.1 de [38]), en la práctica este método puede no funcionar bien.

2. **Algoritmo “Reversible Jump MCMC”** Un algoritmo muy relacionado con el algoritmo de Chalmond es el llamado “Reversible Jump MCMC” [9]. Este algoritmo ha sido utilizado para segmentar imágenes, e intenta estimar conjuntamente  $S$  y  $\theta$  bajo una perspectiva bayesiana. Asume una distribución de probabilidad a priori para todos los parámetros del modelo, y al igual que en el algoritmo de Chalmond, sustituye la log-verosimilitud por la función de pseudoverosimilitud. Utiliza un algoritmo MCMC (muestreador de Gibbs o Metropolis-Hasting) para muestrear, en una búsqueda secuencial, de la distribución a posteriori completa. Por último utiliza Simulated Annealing para estimar la solución máximo a posteriori (MAP).

#### 7.4.5. Algoritmos basados en el gradiente estocástico

1. **Algoritmo de Younes.** Younes [208] propone un algoritmo de gradiente estocástico para maximizar la expresión de la log-verosimilitud marginal (ecuación 7.37) para modelos exponenciales.

En el contexto de análisis de imágenes, es muy habitual que la distribución conjunta  $p_\theta(S, Y)$  pertenezca a la familia exponencial. En este caso, esta distribución conjunta puede formularse como:

$$p_\theta(S, Y) = c(\theta)^{-1} \exp(-\langle \theta, H(S, Y) \rangle)$$

y en este caso es fácil comprobar que maximizar (7.37) es equivalente a encontrar una solución a la ecuación:

$$E_\theta(H) = E_\theta(H/Y) \tag{7.45}$$

El esquema del algoritmo es el siguiente: llamamos  $\theta_n$  al valor actual de los parámetros, y llamamos  $(S_1^{n+1}, Y_1^{n+1})$  y  $S_2^{n+1}$  a los valores obtenidos tras actualizar el valor de una posición utilizando el muestreador de Gibbs para la distribución conjunta y la distribución condicional respectivamente, con parámetros  $\theta_n$ .

Este método actualiza el valor de los parámetros como:

$$\theta_{n+1} = \theta_n + \frac{c}{n+1} \{H(S_i^{n+1}, Y_i^{n+1}) - H(S_i^{n+1}, y)\} \quad (7.46)$$

siendo  $c$  una constante dada.

Younes prueba que si el valor de  $c$  es pequeño y  $\theta$  está en un compacto, el algoritmo converge a una solución de (7.45). Pero la implementación de este algoritmo requiere ciertas precauciones ya que se necesita elegir adecuadamente el valor de  $c$  y el valor inicial de los parámetros. Dependiendo de si partimos de los valores iniciales adecuados o si no, podemos encontrarnos con diferencias significativas en el tiempo computacional, hasta el extremo de que una mala elección puede llevar a que el algoritmo no converja. Estos valores deben en general ser encontrados por el investigador, experimentando el algoritmo sobre un modelo dado.

Esta aproximación se ha estudiado en [178] para aplicaciones bayesianas a trabajos con tomografías.

2. **ICE+ Algoritmo del gradiente estocástico.** Un método muy relacionado, es el propuesto por Salzenstein et al. [173], quienes proponen combinar el algoritmo "Iterated Conditional Estimation" (ICE) con el algoritmo del gradiente estocástico. Como se ha explicado en la sección 7.4.2, para usar el método ICE necesitamos:

- (i) Un buen estimador  $\hat{\theta} = \hat{\theta}(S, Y)$  de  $\theta$ , en un contexto en el que no hayan observaciones faltantes.

- (ii) Simulaciones de  $S$  de acuerdo con su distribución a posteriori para cada  $\theta$ .

Si la constante normalizadora es desconocida, tendremos problemas tanto para encontrar el estimador de  $\theta$ ,  $\hat{\theta}$ , como para obtener las simulaciones de  $S$ . Para resolver el problema de la constante normalizadora al hacer la simulación, utilizaremos el muestreador de Gibbs. En cuanto a la estimación de parámetros, para estimar los hiperparámetros ( $\alpha$ ) de  $\pi_\alpha(S)$ , se propone utilizar un algoritmo de gradiente estocástico [207]. Los parámetros de la distribución de  $Y$  se toman igual a sus correspondientes medias muestrales.

Un algoritmo similar, pero en el contexto de segmentación de imágenes se plantea por Kato et al. en [113, 114]. Este algoritmo va combinando pasos de estimación y restauración. Ellos proponen estimar  $S$  y  $\theta$  conjuntamente maximizando de una forma iterativa ([60]):

$$(i) \quad \hat{S} = \operatorname{argmax}_S p_{\hat{\theta}}(S, Y) \quad (7.47)$$

$$(ii) \quad \hat{\theta} = \operatorname{argmax}_\theta p_{\hat{\theta}}(\hat{S}, Y) \quad (7.48)$$

Los problemas derivados de desconocer la constante normalizadora, aparecerán sólo en el paso 2 cuando se plantee la estimación de los hiperparámetros del modelo. En este paso, proponen la aproximación del gradiente utilizando relajación estocástica.

#### 7.4.6. Estudio comparativo

En este apartado hacemos un estudio comparativo sobre tres de los algoritmos que acabamos de exponer: el algoritmo MCL, el algoritmo de Chalmoud y el de Younes, y vamos a trabajar con imágenes sintéticas porque sólo en este caso podemos conocer exactamente los valores de los parámetros (va-

lores que nos han servido para simular las imágenes) y podremos comparar fácilmente el “valor real” de los parámetros con las estimaciones.

### Modelo

Para comparar los tres métodos vamos a considerar un problema muy sencillo: la segmentación de una imagen con  $L$  colores que se ha degradado con un ruido aditivo gaussiano con media 0 y varianza  $\sigma^2$ .

Al igual que hicimos en la sección 7.4.1, denotamos por  $S$  a la imagen original, no degradada, y por  $Y$  a la imagen degradada. Llamamos  $X$  al grid de dimensión  $p \times q$ , y  $x \in X$  a las posiciones del grid.

Entonces:

$$p_{\sigma^2}(Y/X) = c(\sigma^2/X)^{-1} \exp\left(\frac{-1}{2\sigma^2} \sum_x (Y(x) - S(x))^2\right). \quad (7.49)$$

Asumimos una distribución a priori muy simple para  $x$ :

$$\pi_\alpha(S) = c(\alpha)^{-1} \exp\left(-\alpha \sum_x \sum_{z:z \sim x} H(S(z), S(x))\right) \quad (7.50)$$

$$\text{con } H(S(z), S(x)) = \begin{cases} 1 & \text{if } S(z) \neq S(x) \\ 0 & \text{otherwise} \end{cases}$$

Para simplificar la notación definimos:  $n(x) = \sum_{z:z \sim x} H(S(x), S(z))$  y  $n(x, l) = \sum_{z:z \sim x} H(S(z), l)$  para  $l = 1, \dots, L$ .

La expresión de la log verosimilitud es:

$$l(\theta) \propto \frac{-1}{2\sigma^2} \sum_x (Y(x) - S(x))^2 - \alpha \sum_x n(x), \quad (7.51)$$

y estamos interesados en la estimación de los parámetros  $\theta = (\alpha, \sigma^2)$ .

Veamos cómo se pueden aplicar los distintos métodos para obtener las estimaciones de los parámetros.

1. **Verosimilitud de Montecarlo.** Siguiendo con la notación que hemos utilizado en la sección 7.4.3, y teniendo en cuenta la ecuación 7.51, para

aplicar el método de verosimilitud de Montecarlo necesitamos fijar unos valores iniciales para el vector de parámetros  $\theta_0 = (\sigma_0^2, \alpha_0)$ . Entonces la ecuación 7.41 será ahora:

$$\begin{aligned} \hat{l}(\theta)_k &\propto \\ &\propto \log \left( \frac{1}{k} \sum_{i=1}^k \left\{ \exp \left( \frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma^2} \right) \sum_x (Y_i^*(x) - S_i^*(x))^2 + (\alpha_0 - \alpha) n_i^*(x) \right) \right\} \right) - \\ &- \log \left( \frac{1}{k} \sum_{i=1}^k \left\{ \exp \left( \frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma^2} \right) \sum_x (Y_i(x) - S_i(x))^2 + (\alpha_0 - \alpha) n_i(x) \right) \right\} \right) \end{aligned}$$

y esta será la expresión a maximizar.

2. **Algoritmo de Chalmond.** Al aplicar el método de Chalmond, en cada paso del algoritmo tendremos que maximizar:

$$\begin{aligned} E[\log PL_\theta(S, Y)/\theta_n] &= \\ &= -p \cdot q \log(\sqrt{2\pi\sigma_n^2}) - \frac{1}{k} \left( \sum_{i=1}^k [p \cdot q \cdot \log \left( \sum_{l=1}^L \exp\{-\alpha_n n_i(x, l)\} \right)] - \right. \\ &\left. - \frac{1}{2\sigma_n^2} \sum_x (Y_i(x) - S_i(x))^2 - \alpha_n \sum_x n_i(x) \right] \end{aligned}$$

3. **Algoritmo de Younes.** Por último, si aplicamos el algoritmo de Younes, las actualizaciones vendrán dadas por:

$$\begin{aligned} H(S_i^{n+1}, Y_i^{n+1}) - H(S_i^{n+1}, y) &= \\ &= \left( \begin{array}{c} \frac{1}{2} \sum_x [(Y_i^*(x) - S_i^*(x))^2 - (Y_i(x) - S_i(x))^2] \\ \sum_x n_i^*(x) - \sum_x n_i(x) \end{array} \right) \end{aligned}$$

## Resultados

Hemos simulado un conjunto de imágenes a partir de la distribución (7.50), tomando  $L = 5$  y dos valores de  $\alpha$  distintos (0.3 y 0.6).

Simulamos un centenar de imágenes para cada valor de  $\alpha$ , sobre un retículo regular de dimensión  $64 \times 64$  (ver figuras 7.1 y 7.4). Como podemos ver,

las imágenes generadas con  $\alpha = 0.3$  presentan mucho más ruido que las generadas con  $\alpha = 0.6$ .

Añadimos ruido blanco con varianza 0.8, (figuras 7.2 y 7.5), y ruido blanco con varianza 1 (figuras 7.3 y 7.6) a todas las imágenes, y aplicamos los tres métodos de estimación sobre las imágenes resultantes.

En primer lugar tomamos cuatro puntos iniciales diferentes  $\theta_0 = (\alpha_0, \sigma_0^2)$  para iniciar los tres métodos de estimación. Podemos ver que la convergencia del método de verosimilitud de Montecarlo depende en gran medida del valor inicial que asignemos a la varianza. Valores de  $\sigma_0^2$  próximos al valor de  $\sigma^2$  que se usó en la simulación de la imagen, proporcionan buenas estimaciones, y valores de  $\sigma_0^2$  alejados de  $\sigma^2$  pueden llevar a que no se consigan estabilizar las estimaciones a lo largo del procedimiento de estimación. Por ello decidimos utilizar dos valores iniciales distintos para el parámetro  $\alpha$  para el método de verosimilitud de Montecarlo, y utilizaremos la varianza muestral de la imagen observada como valor inicial del parámetro  $\sigma^2$  (en las tablas 7.1 y 7.2 denotaremos la varianza muestral por  $s_y^2$ ). La convergencia de todos los procedimientos, se aseguró observando la convergencia de las estimaciones de los parámetros.

Las tablas 7.1 y 7.2 muestran la media muestral y la desviación típica de las estimaciones, y el tiempo computacional aproximado que se necesita para alcanzar la convergencia de las estimaciones (cociente contra el tiempo que ha utilizado el método de verosimilitud de Montecarlo) para cada procedimiento de estimación para cada punto inicial. Cuando no se alcanza la convergencia de alguna de las estimaciones, dibujamos una recta en lugar de un valor numérico, para indicar esta situación.

Al analizar los resultados obtenidos vemos que el método de Chalmond converge más rápidamente que el método de verosimilitud de Montecarlo y que el método de Younes; a pesar de todo proporciona las estimaciones menos

precisas. Este método tiene la ventaja de no depender del valor inicial. La convergencia del estimador de  $\alpha$  con el método de Younes es ligeramente más rápido que con el método MCL, pero el método de Younes tiene una dificultad adicional, y es que depende de la constante,  $U$ , y no hay forma de elegir este valor excepto mediante experimentación. Finalmente, podemos ver que el método MCL proporciona la mejor estimación si, como ya hemos comentado, el valor inicial para el parámetro de la varianza está muy próximo al valor verdadero de este parámetro. Además, los valores más pequeños de  $\alpha$  son los que proporcionan las peores estimaciones. Esto ocurre porque este parámetro controla la suavidad de la verdadera imagen. Por tanto, si  $\alpha$  es muy pequeño, el algoritmo no puede distinguir entre el ruido debido al efecto de este parámetro en la generación de la imagen y el efecto del ruido gaussiano que hemos añadido. De hecho, a través de este estudio raras veces hemos conseguido la convergencia de las estimaciones utilizando el método de Younes, y las estimaciones obtenidas con el método de Chalmond quedan muy lejos de los verdaderos valores. Debemos destacar también que la convergencia falla más a menudo cuando los valores iniciales del parámetro  $\alpha$  es menor que el verdadero valor.

Como conclusión general podemos decir que un algoritmo de estimación muy eficiente podría resultar al mezclar el método de Chalmond o el de Younes con el método de verosimilitud de Montecarlo. Podríamos así utilizar el método de Chalmond o el de Younes para acercarnos al estimador máximo verosímil, y aplicar entonces el algoritmo de verosimilitud de Montecarlo para completar el procedimiento de estimación.

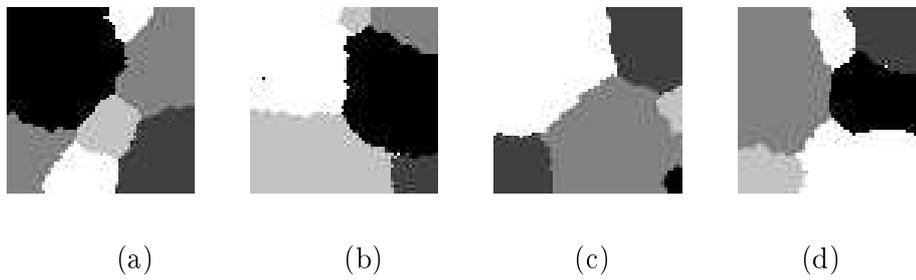


Figura 7.1: Ejemplos de las imágenes utilizadas en nuestros experimentos, generadas a partir de la distribución (7.50) con  $\alpha = 0,6$

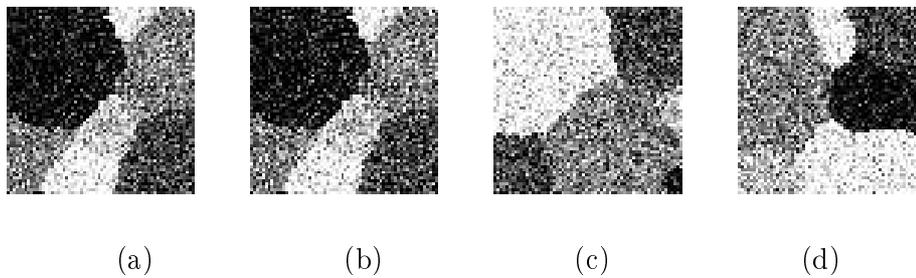


Figura 7.2: Añadimos un proceso de ruido blanco con varianzax0,8 a las imágenes de la figura 7.1

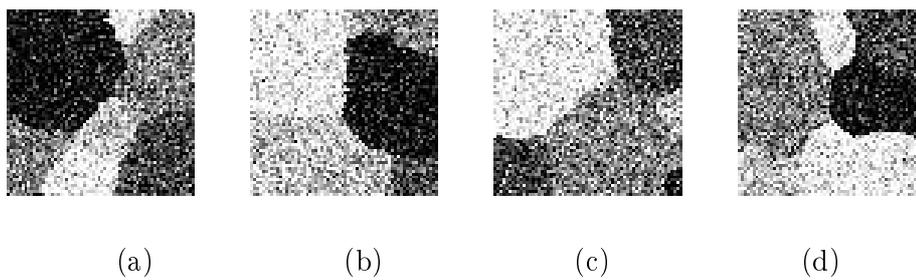


Figura 7.3: Añadimos un proceso de ruido blanco con varianzax1 a las imágenes de la figura 7.1

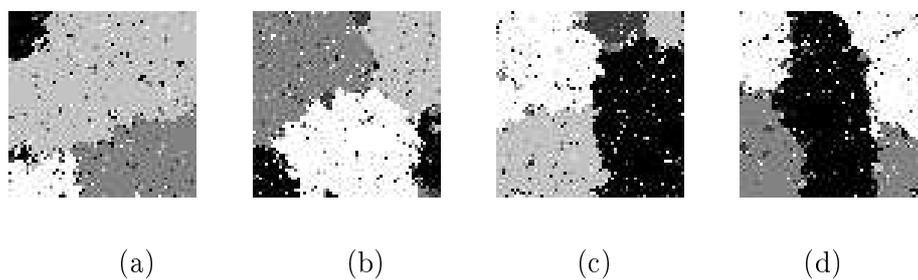


Figura 7.4: Ejemplos de las imágenes utilizadas en nuestros experimentos, generadas a partir de la distribución (7.50) con  $\alpha = 0,3$

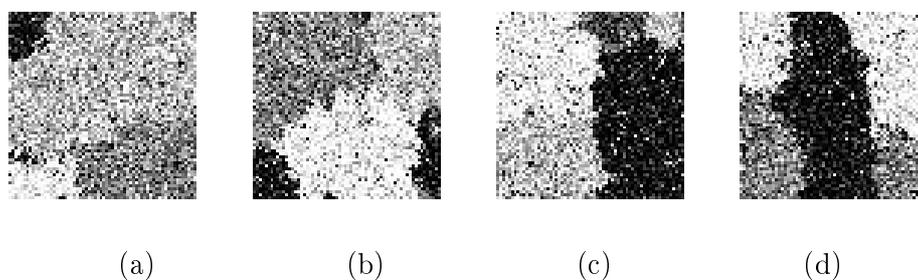


Figura 7.5: Añadimos un proceso de ruido blanco con varianza 0,8 a las imágenes de la figura 7.4

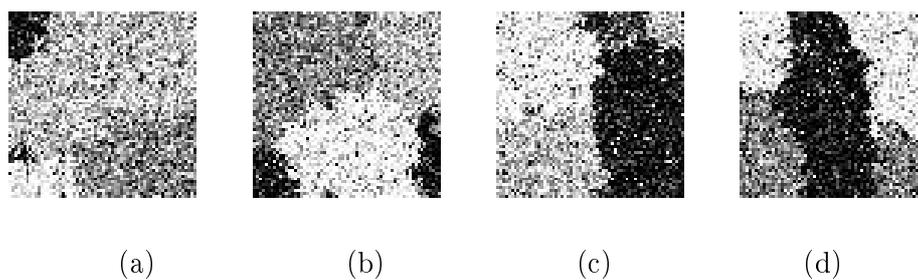


Figura 7.6: Añadimos un proceso de ruido blanco con varianza 1 a las imágenes de la figura 7.4

$(\alpha, \sigma^2) = (0.3, 0.8)$				
Algor.	$\theta_0$	media	desv. típ.	tiempo
MCL	$(0.4, s_y^2)$	(0.3, 0.920)	(0, 0.120)	1
	$(0.2, s_y^2)$	(0.3, -)	(0, -)	1
Younes	(0.4, 0.96)	(-, 1.363)	(, 0.016)	$\frac{3}{4}$
	(0.2, 0.96)	(1.4, 0.790)	(0.32, 0.223)	$\frac{3}{4}$
Chalm.	(0.4, 0.96)	(0.080, 0.88)	(0, 0.002)	$\frac{1}{2}$
	(0.2, 0.96)	(0.080, 0.770)	(0, 0.001)	$\frac{1}{2}$

$(\alpha, \sigma^2) = (0.6, 0.8)$				
Algor.	$\theta_0$	media	desv. típ.	tiempo
MCL	$(0.7, s_y^2)$	(0.648, 0.813)	(0.204, 0.127)	1
	$(0.5, s_y^2)$	(0.717, 0.800)	(0.156, 0.035)	1
Younes	(0.7, 0.96)	(0.453, 0.802)	(0.113, 0.004)	$\frac{3}{4}$
	(0.5, 0.96)	(1.439, 0.790)	(0.270, 0.156)	$\frac{3}{4}$
Chalm.	(0.7, 0.96)	(0.450, 0.806)	(0.001, 6e-9)	$\frac{1}{2}$
	(0.7, 0.64)	(0.571, 0.808)	(0.013, 0.001)	$\frac{1}{2}$

Cuadro 7.1: Resultados del estudio experimental. En la tabla tenemos la media muestral y las desviaciones típicas de las estimaciones y el tiempo computacional que hemos necesitado para alcanzar la convergencia de las estimaciones (tiempo=cociente respecto al tiempo tardado por el método de verosimilitud de Montecarlo) para cada procedimiento de estimación, y para cada punto inicial. Las líneas indican que no se ha alcanzado la convergencia.

$(\alpha, \sigma^2) = (0.3, 1)$				
Algor.	$\theta_0$	media	desv. típ.	tiempo
MCL	$(0.4, s_y^2)$	(0.390, 1.100)	(0.180, 0.184)	1
	$(0.2, s_y^2)$	(0.345, -)	(0.127, -)	1
Younes	(0,4, 1,16)	(-,1.654)	(-,0.138)	$\frac{3}{4}$
	(0,2, 1,16)	(-,1.059)	(-,0.122)	$\frac{3}{4}$
Chalm.	(0,4, 1,16)	(0.067, 0.706)	(0.000, 1.8e-7)	$\frac{1}{2}$
	(0,2, 1,16)	(0.070, 0.982)	(0.000, 4.06e-7)	$\frac{1}{2}$

$(\alpha, \sigma^2) = (0.6, 1)$				
Algor.	$\theta_0$	media	desv. típ.	tiempo
MCL	$(0.7, s_y^2)$	(0.758, 1.080)	(0.045, 0.000)	1
	$(0.5, s_y^2)$	(0.686, 1.077)	(0.196, 0.020)	1
Younes	(0.7, 1.16)	(0.345, 1.069)	(0.079, 0.011)	$\frac{3}{4}$
	(0.5, 1.16)	(-, 1.071)	(-, 0.160)	$\frac{3}{4}$
Chalm.	(0.7, 1.16)	(0.193, 0.900)	(0.010, 0.022)	$\frac{1}{2}$
	(0.5, 1.16)	(-, 1.026)	(-, 0.025)	$\frac{1}{2}$

Cuadro 7.2: Resultados del estudio experimental. En la tabla tenemos la media muestral y las desviaciones típicas de las estimaciones y el tiempo computacional que hemos necesitado para alcanzar la convergencia de las estimaciones (tiempo=cociente respecto al tiempo tardado por el método de verosimilitud de Montecarlo) para cada procedimiento de estimación, y para cada punto inicial. Las líneas indican que no se ha alcanzado la convergencia.

## Capítulo 8

# Estimación bayesiana del estatus en campos visuales de pacientes con glaucoma

### Contenidos

---

<b>8.1. Introducción . . . . .</b>	<b>243</b>
8.1.1. Conjunto de datos y notación . . . . .	244
8.1.2. Objetivos . . . . .	245
<b>8.2. Rasgos característicos de los campos visuales . .</b>	<b>247</b>
<b>8.3. Modelización de los umbrales condicionados a los estatus . . . . .</b>	<b>248</b>
<b>8.4. Modelización a priori de los estatus . . . . .</b>	<b>253</b>
<b>8.5. Distribución conjunta . . . . .</b>	<b>258</b>
<b>8.6. Estimación de parámetros . . . . .</b>	<b>260</b>
8.6.1. Paso 2. Estimación de los parámetros de la distribución a priori de los estatus . . . . .	262
8.6.2. Paso 3. Estimación de los parámetros de la distribución de los umbrales condicionados a los estatus	262
8.6.3. Paso 4. Estimación de los estatus . . . . .	267

8.6.4. Simulación de los estatus . . . . .	268
<b>8.7. Resultados y conclusiones . . . . .</b>	<b>271</b>

---

## 8.1. Introducción

En este capítulo y el siguiente vamos a aplicar las metodologías introducidas en el capítulo anterior en dos tipos de estudios relacionados con campos visuales de pacientes que sufren glaucoma.

En los capítulos 4 y 6 de este trabajo, estudiamos la evolución espacio temporal de CV, tanto de pacientes sanos (capítulo 4), como de pacientes con glaucoma (capítulo 6), utilizando en ambos casos la metodología propia de la Geoestadística. Nuestro objetivo ahora, es abordar problemas similares utilizando la metodología clásica de series temporales multivariantes y los campos aleatorios de Markov, que hemos estado viendo en el capítulo 7. Es decir, una serie de campos visuales de un paciente será considerada una serie temporal multivariante. Dado  $t$ , las componentes de  $y_t$  se asumirán un campo aleatorio de Markov. Utilizaremos un modelo VARIMA para modelizar cada serie.

En este capítulo, nos planteamos el problema de “etiquetar” cada una de las posiciones de los CV observados como “sana” o “enferma”, usando métodos estadísticos bayesianos. Para ello además de una variable que recoja las observaciones, definiremos una segunda variable que nos permita identificar cada posición como sana o enferma. El objetivo de este capítulo será estimar los valores de esta nueva variable en cada posición.

En el siguiente capítulo (capítulo 9), nos proponemos un objetivo distinto, mucho más próximo al que nos planteábamos en el capítulo 6, ya que buscaremos una modelización para la distribución espacio temporal de CV de pacientes con glaucoma, con el objetivo de poder realizar predicciones y simulaciones.

Como antecedente directo del estudio que vamos a realizar, debemos citar el trabajo de Olsson y Rootzen de 1.994 [154]. En este trabajo los autores proponen un modelo para la distribución espacial de CV de pacientes enfermos

de glaucoma. Aunque ellos tenían acceso a datos internos del campímetro, y a todas las respuestas (tanto afirmativas como negativas) de los pacientes, (i.e. tenían información muy distinta a la nuestra), nos basaremos en sus conclusiones, e intentaremos generalizar alguno de sus modelos para incluir en ellos la componente temporal.

### 8.1.1. Conjunto de datos y notación

En este capítulo y el siguiente, vamos a trabajar con una base de datos formada por 10 pacientes con un número de observaciones cada uno que oscila entre 4 y 9. La razón de que la base de datos sea tan pequeña se ha debido a varios factores. Primero a que en los hospitales públicos españoles estas pruebas se realizan con una periodicidad de un año, por lo que es difícil tener largas series de observaciones de un mismo paciente. Además, aunque teóricamente se realiza una revisión al año, este período puede acortarse o alargarse dependiendo de la mayor o menor gravedad del estado del paciente. Al trabajar con series temporales, vamos a exigir que las observaciones estén igualmente espaciadas, con un retardo de un año entre dos observaciones consecutivas, por lo que no hemos podido incluir en nuestra base de datos muchas de las exploraciones que nos han proporcionado los oftalmólogos. De todas formas continuamos trabajando para obtener una base de datos mucho más extensa sobre la que poder repetir y validar todos los métodos que vamos a desarrollar a continuación.

Notar que el tipo de datos con los que trabajamos son un poco diferentes a los habituales en series temporales multivariantes y en general en estadística espacio-temporal, puesto que tenemos repeticiones independientes de la serie (10 pacientes). Por tanto deberemos adaptar muchos de los resultados del capítulo anterior a esta situación, la mayoría de ocasiones esta

adaptación es inmediata. El hecho de tener repeticiones independientes de la misma variable podría en cierta forma “compensar” el hecho de que las series de observaciones sobre cada paciente sean tan cortas.

Siguiendo con la misma notación que ya introdujimos en el capítulo 4, sea  $P$  el número de pacientes que tenemos en nuestra base de datos, y sea  $p$ , con  $p = 1, \dots, P$  cada uno de ellos. Llamamos  $N_p$  al número de observaciones que tenemos del paciente número  $p$ , y utilizaremos el índice  $t$ ,  $t = 1, \dots, N_p$  para diferenciarlas.

Vimos en el capítulo 2, que cada campo visual contiene información sobre la intensidad de visión del paciente en 52 puntos de su retina. Por eso, continuando con la misma notación definimos una variable continua,  $z$ , que recogerá el umbral de visión de cada paciente en cada uno de los puntos. Definimos  $z_{itp}$  como el umbral observado en la  $i$ -ésima posición, de la  $t$ -ésima observación del paciente número  $p$ , y llamamos  $Z_p = (Z_{1p}, Z_{2p}, \dots, Z_{N_p p})$  al vector con todos los umbrales observados en todas las exploraciones de un paciente, con  $Z_{tp} = (z_{1tp}, z_{2tp}, \dots, z_{52tp})$ .

Asociada a cada una de las posiciones del campo visual, definimos una nueva variable, que llamamos “estatus”. Será una variable binaria en cada posición, latente, no observada, que nos permitirá clasificar cada posición como sana ( $s = 1$ ), o enferma ( $s = -1$ ). De forma análoga a la notación utilizada para los umbrales, llamaremos  $s_{itp}$  al valor de la variable estatus en la  $i$ -ésima posición del  $t$ -ésimo campo visual del paciente número  $p$ , y definiremos  $S_p = (S_{1p}, S_{2p}, \dots, S_{N_p p})$  con  $S_{tp} = (s_{1tp}, s_{2tp}, \dots, s_{52tp})$ .

### 8.1.2. Objetivos

Recordemos brevemente el problema médico que hemos estado siguiendo durante todo este trabajo. Como hemos explicado de forma bastante extensa

en el capítulo 2, el glaucoma es una enfermedad muy grave en la que el ojo pierde sin dolor y gradualmente la visión, empezando por la periferia del campo visual y extendiéndose paulatinamente hacia zonas más centrales, pudiendo llegar a la ceguera. Una información muy útil para diagnosticar glaucoma es la medición del campo visual por medio de un test perimétrico. Llamaremos “campo visual” al resultado de esta medición, y consiste en un mapa numérico que informa al oftalmólogo sobre el estado de visión del paciente. Un campo visual (CV) puede considerarse como un vector con 52 valores numéricos, uno por cada uno de los puntos testeados en la retina, y cada uno de estos valores numéricos es una estimación de la intensidad de visión del paciente en el punto correspondiente. Los puntos se encuentran colocados sobre el grid regular de la tabla 2.1 (tabla del capítulo 2). El paciente se somete a este tipo de mediciones habitualmente una vez al año, pero esta frecuencia puede variar dependiendo del estado de la enfermedad.

El objetivo que nos planteamos en este capítulo, es la estimación del estatus de cada uno de los puntos testeados en la retina, es decir, decidir qué posiciones del CV son “normales” y cuales han sido afectadas por el glaucoma, en base a la información proporcionada por todos los campos visuales del paciente.

Todos los paquetes implementados en los campímetros (SITA, SITA FAST, FASTPAC,...) proporcionan esta información al oftalmólogo, ya que presentan unos mapas en los que destacan las posiciones que consideran “dañadas”. Como comentámos en el capítulo 2, para hacer esta clasificación se basan en estudios empíricos realizados sobre bases de datos bastante grandes. Nuestro objetivo ahora es desarrollar un modelo bayesiano que permita estimar a partir de él, el valor de la variable estatus en cada posición.

Estamos por tanto planteando un problema muy parecido al problema de la segmentación de una imagen, (en el contexto del análisis de imágenes),

bajo una perspectiva bayesiana (ver [18, 61, 38, 78, 205] o [100] para más detalles). Los umbrales,  $Z$ , pueden considerarse como la “imagen observada”, y los estatus,  $S$ , como la “imagen con las etiquetas”, donde cada valor representa el verdadero estado (normal/enfermo) de cada posición, y del que no disponemos de observaciones.

En estos problemas, se construye la distribución de probabilidad de la imagen observada,  $f(Z/S)$ , y se formula una distribución a priori sobre  $S$ ,  $\pi(S)$ . A partir de aquí se calcula la distribución a posteriori de  $S$  dado  $Z$ :

$$f(S/Z) \propto f(Z/S)\pi(S)$$

y la estimación de  $S$  se obtiene habitualmente tomando  $\hat{S}$  como la moda de esta distribución a posteriori. Este estimador se conoce como estimador máximo a posteriori (*MAP*), y será el que utilizaremos nosotros también.

## 8.2. Rasgos característicos de los campos visuales

Como hemos comentado en el capítulo 2, los campos visuales han sido extensivamente estudiados por los oftalmólogos, por lo tanto muchas de sus características fisiológicas son bien conocidas y pueden encontrarse documentadas en la literatura médica. Recordemos brevemente aquellas características de los campos visuales que tendremos en cuenta a la hora de formular nuestros modelos.

Así por ejemplo respecto a campos visuales sanos, recordemos que es conocido que el valor esperado de los umbrales decrece linealmente con la edad, y que este decrecimiento no es homogéneo en todo el campo visual, sino que es más rápido en las posiciones periféricas que en las centrales. Además, las posiciones más periféricas son aquellas que presentan niveles más bajos en las intensidades de los umbrales, y valores más altos para su varianza.

Por otro lado, se sabe también que no hay una influencia significativa de la edad sobre la variabilidad de los umbrales.

Con respecto a las posiciones afectadas por glaucoma, se sabe que los defectos aparecen agrupados formando patrones típicos, y que la enfermedad se va extendiendo siguiendo el trazado de las fibras nerviosas de la retina, por lo tanto podremos utilizar este hecho para definir un tipo de “distancia fisiológica” que nos permita definir una relación de vecindad para la variable “estatus”.

Por último, recordar que aunque el umbral medio disminuye con la edad, esta disminución es más pronunciada en las posiciones “enfermas” que en las “sanas”.

El resto de capítulo se ha organizado como sigue: en la sección 8.3 se propone un modelo para la variable umbral condicionada a la variable estatus, a continuación en la sección 8.4 se propone un modelo a priori para el estatus y se formula la verosimilitud en la sección 8.5. El problema de estimación de parámetros se aborda en la sección 8.6 y por último en la sección 8.7 se presentan los resultados empíricos y las conclusiones del trabajo.

### **8.3. Modelización de los umbrales condicionados a los estatus**

Sabemos que en general, el valor esperado de los umbrales decrece linealmente con la edad; que este decrecimiento no es homogéneo en todo el campo visual, siendo más rápido en las posiciones más periféricas, y que el decrecimiento es más pronunciado en las posiciones enfermas que en las sanas.

Así pues, para cada paciente y en cada posición tenemos una serie temporal no estacionaria.

En el capítulo 7.3 hemos visto que en este caso podemos plantearnos dos tipos de modelos, y trabajar con procesos estacionarios en tendencia (sección 7.3.2), o con modelos con raíces unitarias (sección 7.3.3). Además, independientemente del tipo de modelo que finalmente cojamos, deberemos tener en cuenta que también estamos en una situación de cambios en régimen (sección 7.3.5), ya que el comportamiento de la serie cambia cuando una posición pasa de ser normal a ser enferma.

Si intentamos modelizar los umbrales condicionados a los estatus con un modelo estacionario en tendencia, al querer reflejar que el decrecimiento de los umbrales con la edad es más rápido en las posiciones enfermas que en las sanas, nos encontramos con que los distintos regímenes (sano/enfermo), afectan a la pendiente de la recta, y deberíamos plantear un modelo multivariante donde cada componente sea del tipo:

$$z_{itp} = \begin{cases} a_i + b_i t + \epsilon_{itp} & \text{si } s_{itp} = 1 \\ a_i + (b_i + c_i)t + \epsilon_{itp} & \text{si } s_{itp} = -1 \end{cases}$$

Este modelo tiene la peculiaridad de no ser estacionario para cada uno de los regímenes. Este hecho es bastante inusual en los modelos con cambios en régimen, donde se asume habitualmente que, dado el régimen, las dos series son estacionarias, de hecho no hemos podido encontrar en la literatura ninguna referencia en la que se de esta situación. Una forma para lograr que la serie sea estacionaria en cada “trozo” sería diferenciarla, y tendríamos:

$$\Delta z_{itp} := z_{itp} - z_{i-1tp} = \begin{cases} b_i + \epsilon_{itp} & \text{si } s_{itp} = s_{i-1tp} = 1 \\ (b_i + c_i) + \epsilon_{itp} & \text{si } s_{itp} = s_{i-1tp} = -1 \\ \vdots & \vdots \end{cases}$$

pero este modelo resultaría muy complicado de expresar en forma multivariante.

Dejamos abierto este camino para seguir trabajando en el futuro, y nos decantamos por plantear un modelo utilizando procesos con raíces unitarias. Dentro de estos procesos y debido a la precariedad de nuestra base de datos, optamos por elegir el modelo más parsimonioso y así proponemos modelizar la variable umbral con un modelo VARI(1,1), con cambios de régimen (volveremos a proponer el mismo modelo en el capítulo 9, y allí comprobaremos la bondad del ajuste).

En principio podríamos haber planteado un modelo cointegrado, pero no lo hemos hecho por dos motivos. El primero es que los modelos cointegrados se basan en el conocimiento a nivel práctico de que el comportamiento de las series temporales de alguna forma se “compensan” o “tienen un equilibrio” a largo plazo. Esta hipótesis es bastante natural en Economía pero en nuestro caso parece no tener demasiado sentido a nivel práctico. El segundo motivo para no trabajar con modelos cointegrados ha sido, como decíamos antes, el buscar el modelo más parsimonioso. Por otro lado, la “precariedad” de nuestra base de datos nos impide poder aplicar los tests de cointegración con cambios estructurales propuestos en la sección 7.3.5. Los dejamos también como trabajo futuro para cuando podamos contar con una base de datos más extensa en el tiempo.

Si denotamos por  $\Delta$  al operador “primera diferencia”, es decir:  $\Delta Z_{tp} = Z_{tp} - Z_{t-1p}$ , el modelo que proponemos para los umbrales condicionados a los estatus es:

$$\Delta Z_{tp} = C_0 + c_1 \Delta S_{tp} + \Psi \left( \Delta Z_{t-1p} - c_1 \Delta S_{t-1p} \right) + \epsilon_{tp} \quad (8.1)$$

con  $\epsilon_{tp} \mid \Delta Z_{t-1p}, S_p \sim N(0, M(I - C)^{-1})$

Como hemos dicho se trata de un modelo VARI(1,1) con cambios en régimen que afectan sólo al termino constante y que vienen dados por la variable  $\Delta S_{tp}$ , es decir por cambios en el estatus.

Veamos a continuación cuáles son y qué significado tienen los parámetros del modelo:

- $C_0$  es un vector con 52 componentes,  $C_0 = (c_{01}, c_{02}, \dots, c_{052})$ . Si tenemos en mente el modelo de predicción con procesos integrados, que vimos en la sección 7.3.3,  $C_0$  podría interpretarse como las 52 pendientes resultantes al hacer la predicción de la intensidad en cada posición.
- El parámetro  $c_1$  representa el cambio en la pendiente debido al cambio en el estatus.
- La matriz  $\Psi$  es la matriz de los coeficientes autoregresivos. Vamos a suponer que estos coeficientes no dependen de la posición, y que sobre cada umbral en un instante  $t$  influye únicamente el valor del umbral en la misma posición en el instante anterior, y su valor en las posiciones vecinas (también en el instante anterior), es decir que:

$$\Psi = (\psi_{ij})_{i,j=1,\dots,n} \quad \text{con} \quad \psi_{ij} = \begin{cases} \alpha_1 & \text{si } j \sim i \\ \alpha_0 & \text{si } i = j \\ 0 & \text{en otro caso} \end{cases} \quad (8.2)$$

- $M(I-C)^{-1}$  es la matriz de varianzas covarianzas de los residuos, no vamos a distinguir distintas varianzas dependiendo del valor del estatus, pero sí que vamos a considerar diferencias debido a la excentricidad. Supondremos que existe una relación lineal entre la varianza en cada posición y su distancia al centro del CV. Para ello definimos:

$$M = \sigma_0^2 I + \sigma^2 D = \sigma^2 \left( \sigma_1^2 I + D \right) \quad (8.3)$$

siendo  $\sigma^2$ ,  $\sigma_1^2$  parámetros a estimar,  $I$  la matriz identidad y  $D = \text{diag}(d_1, \dots, d_{52})$ , con  $d_i$  la distancia euclídea de la posición  $i$  al centro del CV.

La dependencia espacial, que será Markoviana, vendrá dada por la matriz  $C$  y vamos a suponerla constante entre posiciones vecinas, es decir:

$$C = (\gamma_{ij})_{i,j=1,\dots,52} \quad \text{con:} \quad \gamma_{ij} = \begin{cases} c & \text{si } j \sim i \\ 0 & \text{en otro caso} \end{cases} \quad (8.4)$$

siendo  $c$  un nuevo parámetro a estimar.

Para simplificar notación, definimos una matriz de vecindades  $H$  como:

$$H = (\delta_{ij})_{i,j=1,\dots,52} \quad \text{con:} \quad \delta_{ij} = \begin{cases} 1 & \text{si } j \sim i \\ 0 & \text{en otro caso} \end{cases}$$

Con esta nueva notación, podemos cambiar las expresiones de las matrices  $\Psi$  y  $C$ . La expresión de  $\Psi$  en la ecuación (8.2) quedará:

$$\Psi = \alpha_0 I + \alpha_1 H$$

y la definición de  $C$  en la ecuación (8.4) quedará:

$$C = cH$$

Sustituyendo la expresión de  $\Psi$  y de  $C$  en (8.1), tendremos que con la nueva notación el modelo propuesto para los umbrales condicionados a los estatus es:

$$\begin{aligned} \Delta Z_{tp} &= C_0 + c_1 \Delta S_{tp} + \left( \alpha_0 I + \alpha_1 H \right) \left( \Delta Z_{t-1p} - c_1 \Delta S_{t-1p} \right) + \epsilon_{tp} \\ \text{con } \epsilon_{tp} \mid \Delta Z_{t-1p}, S_p &\sim N(0, \sigma^2 (\sigma_1^2 I + D)(I - cH)^{-1}) \end{aligned} \quad (8.5)$$

Para poder hallar las expresiones de los estimadores de los parámetros, necesitamos modificar un poco más la notación y expresar (8.5) en forma

vectorial. Para ello definimos:

$$\begin{aligned} \Delta Z_p &= \begin{pmatrix} \Delta Z'_{3p} \\ \dots \\ \Delta Z'_{N_p p} \end{pmatrix}; \Delta Z_{-1,p} = \begin{pmatrix} \Delta Z'_{2p} \\ \dots \\ \Delta Z'_{N_p-1p} \end{pmatrix}; \\ \Delta S_p &= \begin{pmatrix} \Delta S'_{3p} \\ \dots \\ \Delta S'_{N_p p} \end{pmatrix}; \Delta S_{-1,p} = \begin{pmatrix} \Delta S'_{2p} \\ \dots \\ \Delta S'_{N_p-1p} \end{pmatrix}; \\ H\Delta Z_{-1,p} &= \begin{pmatrix} H\Delta Z'_{2p} \\ \dots \\ H\Delta Z'_{N_p-1p} \end{pmatrix}; H\Delta S_{-1,p} = \begin{pmatrix} H\Delta S'_{2p} \\ \dots \\ H\Delta S'_{N_p-1p} \end{pmatrix} \end{aligned} \quad (8.6)$$

Con esta nueva notación, el modelo definido en (8.5) quedará:

$$\begin{aligned} \Delta Z_p &= 1' \cdot C_0 + c_1 \Delta S_p + \alpha_0 \Delta Z_{-1,p} + \alpha_1 H \Delta S_{-1,p} - \alpha_0 c_1 \Delta S_{-1,p} - \\ &\quad - \alpha_1 c_1 H \Delta S_{-1,p} + \tilde{\epsilon}_p \\ \text{con} \quad \epsilon_{tp} \mid \Delta Z_{t-1p}, S_p &\sim N(0, \sigma^2 (\sigma_1^2 I + D)(I - cH)^{-1}) \end{aligned} \quad (8.7)$$

## 8.4. Modelización a priori de los estatus

Recordemos que los estatus son variables binarias, no observadas y definidas sobre cada posición como  $s_{itp} \in \{\pm 1\}$ , siendo  $s_{itp} = -1$  cuando la posición correspondiente está enferma, y siendo  $s_{itp} = 1$  cuando la posición está normal.

En la sección 7.3.5 hemos estado hablando de los modelos con cambios de régimen, y en particular de los “Markov-switching model” [82, 83, 84, 85], que suponen que el proceso no observado (el de los regímenes) está gobernado por un proceso de Markov discreto (estados discretos y tiempos discretos), que viene definido por determinadas probabilidades de transición:

$$p_{ij} = Pr(s_{t+1} = j \mid s_t = i), \quad \sum_j p_{ij} = 1, \forall i, j \in \{1, \dots, M\}. \quad (8.8)$$

Como ya comentamos entonces, una de las características de estos modelos, es que aunque la serie  $z_t$  sea multivariante, todas sus componentes se encuentran en el mismo estado  $s_t$  en el instante  $t$ . Pero en nuestro caso no ocurre esto, como comentábamos al principio de este trabajo, el glaucoma es una enfermedad que se va extendiendo con el tiempo. Al principio sólo unas pocas posiciones están enfermas y con el tiempo este número va en aumento. Por tanto cada posición tiene un determinado estatus. Como también comentábamos en la sección 7.3.5, una solución a este problema sería “redefinir” los estados, y definir para cada paciente  $p$  y cada instante  $t$ ,

$$s_{tp} = \begin{cases} 1 & \text{si } (s_{1tp} = 1, \quad s_{2tp} = 1, \quad s_{3tp} = 1, \quad \dots, \quad s_{52tp} = 1) \\ 2 & \text{si } (s_{1tp} = -1, \quad s_{2tp} = 1, \quad s_{3tp} = 1, \quad \dots, \quad s_{52tp} = 1) \\ \vdots & \vdots \\ M & \text{si } (s_{1tp} = -1, \quad s_{2tp} = -1, \quad s_{3tp} = -1, \quad \dots, \quad s_{52tp} = -1) \end{cases}$$

con  $M = 2^{52}$ . Sin embargo, el cálculo de las probabilidades de transición  $p_{ij} = Pr(s_{tp} = j / s_{t-1p} = i)$  resulta muy complicado.

Olsson y Rootzen [154] propusieron modelizar la distribución espacial de los estatus utilizando un modelo autológico:

$$\pi(S) = \frac{1}{Z_s} \exp\{-U(s)\} \quad \text{con} \quad U(s) = \frac{-1}{2} \sum_i \alpha_i s_i - \frac{1}{4} \sum_i \sum_{j \sim i} \beta_{ij} s_i s_j$$

y definieron  $\beta_{ij}$  de forma que reflejara el hecho de que la enfermedad evoluciona siguiendo el trazado de las fibras nerviosas.

En la literatura médica pueden encontrarse mapas con el trazado de las fibras nerviosas en el ojo [1]. A partir de estos mapas, Olson y Rootzen definieron lo que llamaron “dirección preferida entre dos puntos”, que denotaron por  $e_{ij}$ . Para definirla, llamaron  $(x_i, y_i)$  a las coordenadas (en grados) de la posición  $i$ -ésima del CV, y definieron:  $\bar{x}_{i,j} = \frac{x_i + x_j}{2}$ ;  $\bar{y}_{i,j} = \frac{y_i + y_j}{2}$ ; y

$\vec{e}_{i,j} = \frac{\vec{e}_{i,j}^0}{\|\vec{e}_{i,j}^0\|}$ , siendo

$$\vec{e}_{i,j}^0 = (e_{i,j}^{0,x}, e_{i,j}^{0,y}) \quad \text{con} \quad e_{i,j}^{0,x} = \begin{cases} \bar{x}_{i,j} - 15 & \text{si } 0 < \bar{x}_{i,j} \\ -3 & \text{en otro caso} \end{cases}$$

$$e_{i,j}^{0,y} = \begin{cases} \bar{y}_{i,j} & \text{si } 0 < \bar{x}_{i,j} \\ -1 & \text{si } \bar{x}_{i,j} < 0 \text{ y } 0 < \bar{y}_{i,j} \\ 1 & \text{si } \bar{x}_{i,j} < 0 \text{ y } \bar{y}_{i,j} < 0 \\ 0 & \text{en otro caso} \end{cases}$$

Una vez definida esta “distancia preferida” entre cada par de posiciones, para cada par de puntos  $i, j$ , construyen el vector  $\vec{d}_{ij}$  trazado del punto  $i$ -ésimo al  $j$ -ésimo, y lo descomponen en una componente  $d_{\parallel}$  paralela a  $\vec{e}_{i,j}$  y otra parte  $d_{\perp}$ , perpendicular a  $\vec{e}_{i,j}$ . Abusando de notación y llamando de nuevo  $d_{\parallel}$  y  $d_{\perp}$  a los módulos de estos vectores, definen:

$$\beta_{ij} = \frac{b}{\rho_{ij}} \quad \text{si } i \neq j, j \in N_i \quad \text{con: } \rho_{ij} = \sqrt{d_{\parallel}^2 + k^2 d_{\perp}^2} \quad (8.9)$$

siendo  $N_i$  la vecindad de segundo orden de  $i$ .

El objetivo que nos planteamos es modificar este modelo autológico, para incluir en él la componente temporal.

En la bibliografía hemos encontrado dos generalizaciones del modelo de Ising al contexto espacio-temporal.

La primera de ellas, es la utilizada por Hainsworth y Mardia [80] quienes en el contexto de la restauración de secuencias de imágenes, utilizan un modelo de Ising espacio-temporal definido a partir de la distribución conjunta, definiendo:

$$\pi(X_1, X_2, \dots, X_T) = \exp\{-U(X_1, \dots, X_T)\}/Z \quad (8.10)$$

siendo  $Z$  la constante normalizadora apropiada, y  $U(X_1, \dots, X_T)$  una función de energía que descompondrán como suma de unas determinadas funciones potenciales.

La otra aproximación es la utilizada entre otros, por Järpe [103] y por Guyon y Hardouin [79], que se basa en generalizar el modelo de Ising al caso espacio temporal, definiendo las siguientes probabilidades condicionadas:

$$\begin{aligned} Pr\left(X(t) = x(t)/X(t-1) = x(t-1)\right) &= \\ &= Z^{-1} \exp\left\{\phi \sum_i x_i(t) \sum_{j \sim i} x_j(t) + \psi \sum_i x_i(t)x_i(t-1)\right\} \end{aligned} \quad (8.11)$$

Este modelo está basado en el concepto de “Markov Chain of Markov Field”, introducido por Guyon [78] y por Bayomog [11].

**Definición 4.1** Sea  $A = \{1, \dots, k\}$  un conjunto de posiciones formando un retículo en un espacio medible en el que se ha definido una relación de vecindad  $\sim$ . Denotemos por  $\delta_i$  para  $i \in A$  el conjunto de posiciones en  $A$  vecinas de  $i$  y por  $A - i$  todas las posiciones de  $A$  excepto la  $i$ -ésima. Un proceso estocástico  $\{X(t)\}$ ,  $t \in Z$  con  $X(t) = \{X_i(t) \mid i \in A\}$ , es un MCMF si cumple que:

$$X_i(t) \mid \{X_{A-i}(t), \{X(s) : s = 1, 2, \dots, t-1\}\} \stackrel{d}{=} X_i(t) \mid \{X_{\delta_i}(t), X(t-1)\}$$

Notar que  $X = \{X(t)\}$  es una cadena de Markov y  $X(t)$  condicionado a  $X(t-1)$  es un campo aleatorio de Markov en  $A$ .

¿Cuales son las ventajas y desventajas de cada una de las aproximaciones? Ambas aproximaciones sólo difieren en el hecho de formular el modelo conjuntamente o a partir de las distribuciones condicionales en  $t$  dado  $t-1$ . Obviamente el modelo definido en (8.11) permite obtener fácilmente probabilidades de transición, i.e. permite calcular fácilmente  $Pr(X(t) = x_t / X(< t))$ . Por otra parte, el modelo definido en la ecuación (8.10) permite simular fácilmente a partir de él, pero no nos permite hacer simulaciones ni predicciones de lo que ocurrirá en el instante  $T+1$ , ya que el modelo está explícitamente definido para los instantes  $1, \dots, T$ , y no podemos extraer información fuera de este intervalo temporal.

Cuadro 8.1: Penalizamos el que una posición enferma en el instante  $t - 1$  pase a sana en  $t$

$s_{it-1}$	$s_{it}$	$(1 + s_{itp})(1 - s_{it-1p})$
-1	-1	0
-1	1	2
1	-1	0
1	-1	0

Como lo que nos interesa en este apartado es la estimación bayesiana de la variable estatus en los puntos observados, y por tanto su expresión conjunta, trabajaremos con la formulación (8.10). En el siguiente capítulo en el que nos plantearemos realizar predicciones trabajaremos con la formulación (8.11).

Proponemos por tanto la siguiente distribución a priori para la variable estatus:

$$\begin{aligned} \pi\left(S_{2p}, \dots, S_{Tp} \mid S_{1p}\right) &\propto \\ &\propto \exp\left\{\sum_{t=2}^T \sum_{i=1}^{52} \left(\sum_{j \sim i} \beta_{ij} s_{itp} s_{jtp} - \phi(1 + s_{itp})(1 - s_{it-1p})\right)\right\} \end{aligned} \quad (8.12)$$

siendo  $\beta_{i,j}$  como en la ecuación (8.9). Con el término  $\phi(1 + s_{itp})(1 - s_{it-1p})$ , intentamos penalizar evoluciones no deseables de la enfermedad, ya que fisiológicamente se sabe que si una posición está enferma en el instante  $t$ , no puede evolucionar a un estado sano. El parámetro  $\phi$  controlará estas evoluciones no deseadas (ver tabla 8.1), y cuanto mayor sea su valor, menor probabilidad habrá de que una posición enferma evolucione a normal. Como una “corrección de bordes” vamos a sustituir en la ecuación (8.12)  $\phi$  por  $\phi_t$ , donde:

$$\phi_t = \begin{cases} \phi & \text{si } t < T \\ 2\phi & \text{si } t = T \end{cases}$$

Una situación ideal se daría al considerar  $\phi \rightarrow \infty$ , pero en la realidad hemos visto que factores como el efecto aprendizaje presente en las primeras exploraciones que realiza un paciente, las pérdidas de fijación, y las respuestas falsas positivas y falsas negativas, hacen que nos encontremos con series de tests en los que hay posiciones inicialmente clasificadas como enfermas que evolucionan a sanas, cosa fisiológicamente imposible. Por todo ello, consideraremos  $\phi$  como un parámetro más a estimar.

Por tanto, en la distribución de los estatus tenemos tres parámetros a estimar:  $(\phi, b, k)$ .

Para trabajar desde una perspectiva completamente bayesiana, ahora deberíamos definir distribuciones a priori para los hiperparámetros del modelo, pero no vamos a hacerlo de momento. Dejamos la modelización completamente bayesiana para un trabajo futuro.

## 8.5. Distribución conjunta

Con la distribución a priori de los estatus

$$\begin{aligned} & \pi\left(S_{2p}, \dots, S_{Tp} \mid S_{1p}\right) \propto \\ & \propto \exp\left\{\sum_{t=2}^T \sum_{i=1}^{52} \left(\sum_{j \sim i} \beta_{ij} s_{itp} s_{jtp} - \phi_t(1 + s_{itp})(1 - s_{it-1p})\right)\right\}, \end{aligned}$$

y la distribución de los umbrales condicionados a los estatus

$$\begin{aligned} \Delta Z_p = & \quad 1' \cdot C_0 + c_1 \Delta S_p + \alpha_0 \Delta Z_{-1,p} + \alpha_1 H \Delta S_{-1,p} - \alpha_0 c_1 \Delta S_{-1,p} - \\ & - \alpha_1 c_1 H \Delta S_{-1,p} + \tilde{\epsilon}_p \end{aligned}$$

con  $\epsilon_{tp} \mid \Delta Z_{t-1p}, S_p \sim N(0, \sigma^2(\sigma_1^2 I + D)(I - cH)^{-1})$ , la distribución

conjunta quedará:

$$\begin{aligned}
f\left(\Delta Z, S_3, \dots, S_T \mid Z_1, Z_2, S_1, S_2\right) &\propto \tag{8.13} \\
&\propto \prod_{p=1}^P (2\pi\sigma^2)^{-\frac{n \cdot (N_p - 2)}{2}} \mid (\sigma_1^2 I + D)^{-1} (I - cH) \mid^{\frac{N_p - 2}{2}} \cdot \\
&\cdot \exp\left\{\frac{-1}{2\sigma^2} \left( \Delta Z_p - 1 \cdot 1' C_0 - c_1 \cdot \Delta S_p - \alpha_0 \Delta Z_{-1,p} - \alpha_1 H \Delta S_{-1,p} + \right. \right. \\
&\left. \left. + \alpha_0 c_1 \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} \right)' \cdot (\sigma_1^2 I + D)^{-1} \cdot (I - cH) \cdot \right. \\
&\left. \left( \Delta Z_p - 1 \cdot 1' C_0 - c_1 \cdot \Delta S_p - \alpha_0 \Delta Z_{-1,p} - \alpha_1 H \Delta S_{-1,p} + \alpha_0 c_1 \Delta S_{-1,p} + \right. \right. \\
&\left. \left. + \alpha_1 c_1 H \Delta S_{-1,p} \right) + \sum_{t=3}^{N_p} \sum_{i=1}^{52} \left( \sum_{j \sim i} \beta_{ij} s_{itp} s_{jtp} - \phi_t (1 + s_{itp})(1 - s_{it-1p}) \right)\right).
\end{aligned}$$

Notar que hemos calculado la distribución conjunta, condicionada a las dos primeras observaciones. Estamos trabajando con un modelo equivalente a un  $VAR(2)$  y, como comentamos en la sección 7.2.1, la estimación resulta mucho más sencilla si maximizamos la distribución condicional, pues la distribución conjunta suele dar lugar a ecuaciones no lineales. Por este motivo, a partir de ahora tomaremos las dos primeras observaciones como fijas y conocidas.

La matriz de vecindad  $H$ , es una matriz cuadrada y simétrica. Si llamamos  $h_1 \leq h_2 \leq \dots \leq h_n$  a sus valores propios ordenados, entonces  $\mid I - cH \mid = \prod_{i=1}^n (1 - ch_i)$ , y la logverosimilitud quedará:

$$\begin{aligned}
\log(L) &\propto \sum_{p=1}^P \left[ \frac{-n \cdot (N_p - 2)}{2} \log(2\pi\sigma^2) - \right. \tag{8.14} \\
&\left. - \frac{N_p - 2}{2} \sum_{i=1}^n \log(\sigma_1^2 + d_i) + \frac{N_p - 2}{2} \sum_{i=1}^n \log(1 - ch_i) - \right. \\
&\left. - \frac{1}{2\sigma^2} \left( \Delta Z_p - 1 \cdot 1' C_0 - c_1 \cdot \Delta S_p - \alpha_0 \Delta Z_{-1,p} - \alpha_1 H \Delta S_{-1,p} + \right. \right.
\end{aligned}$$

$$\begin{aligned}
& +\alpha_0 c_1 \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} \Big)' \cdot \text{diag} \left( \frac{1}{\sigma_1^2 + d_i} \right) (I - cH) \cdot \\
& \left( \Delta Z_p - 1 \cdot 1' C_0 - c_1 \cdot \Delta S_p - \alpha_0 \Delta Z_{-1,p} - \alpha_1 H \Delta S_{-1,p} + \right. \\
& \left. + \alpha_0 c_1 \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} \right) + \\
& \left. + \sum_{t=3}^{N_p} \sum_{i=1}^n \left( \sum_{j=1}^n \beta_{ij} s_{itp} s_{jtp} - \phi_t (1 + s_{itp})(1 - s_{it-1p}) \right) \right]
\end{aligned}$$

## 8.6. Estimación de parámetros

A partir de la distribución de  $f(Z/S)$  y de la distribución a priori de  $S$ ,  $\pi(S)$ , hemos construido en el apartado anterior la distribución conjunta  $f(Z, S)$ . Nuestro objetivo es estimar los parámetros (de  $\pi(S)$  y de  $f(Z/S)$ ) y los valores de la variable estatus, conjuntamente. Como comentábamos en la sección 7.4, para estimar  $S$  desde el punto de vista bayesiano debemos obtener la distribución a posteriori  $f(S/Z) \propto f(S, Z)$  y tomaremos como estimador de  $S$ , la moda de esta distribución a posteriori. Lógicamente, debido a la complejidad de todas las distribuciones propuestas y la dimensión del espacio en que nos encontramos, ninguna de estas tareas pueden obtenerse de forma analítica y tendremos que recurrir a aproximaciones y/o métodos de Montecarlo.

De los métodos revisados en la sección 7.4 tenemos que fijarnos en aquellos que intentan estimar a la vez los parámetros y maximizar la distribución a posteriori, que como hemos comentado son los métodos usados habitualmente en los problemas de restauración y segmentación de imágenes.

Si nos fijamos en la expresión de la distribución conjunta 8.14, vemos además, que si sólo nos interesara la estimación de los parámetros del modelo, podríamos estimar por separado los de la distribución a priori de los estatus, de los de la distribución de los umbrales condicionados a los estatus. Además,

la distribución de los umbrales condicionada a los estatus es de tipo gaussiano y sólo en este tipo de modelos de Markov la constante normalizadora es conocida y el estimador máximo verosímil de los parámetros es relativamente sencillo de obtener.

Por todos estos motivos, de los algoritmos estudiados en el apartado 7.4 vamos a utilizar un algoritmo de la sección 7.4.5, similar al introducido en el contexto de segmentación de imágenes por Kato et al. en [113, 114].

El proceso iterativo es el siguiente:

1. Elegir valores iniciales para todos los parámetros (incluidos los estatus) e hiperparámetros.
2. Estimar los hiperparámetros de la distribución de los estatus.
3. Estimar por máxima verosimilitud los parámetros de la distribución de los umbrales.
4. Estimar los estatus, utilizando el “Simulated Annealing”.
5. Repetir los pasos 2-4 hasta obtener la convergencia de las estimaciones.

Notar que en el paso de estimación de parámetros usamos un algoritmo de estimación similar al EM estocástico [46, 150, 172, 165], pero en el paso  $E$  del algoritmo, calculamos la media muestral sólo a partir de una única simulación del proceso “faltante”. Este paso es combinado con Simulated Annealing para obtener el máximo a posteriori (MAP) de la distribución de los estatus.

Para ver otros trabajos en los que se han utilizado procedimientos iterativos similares para la estimación de parámetros consultar [60, 114].

Veamos a continuación estos pasos poco a poco:

### 8.6.1. Paso 2. Estimación de los parámetros de la distribución a priori de los estatus

Como la constante normalizadora de la función de distribución de los estatus (ec. 8.12) es desconocida, vamos a utilizar el algoritmo de Younes [207] para estimar los parámetros. Notar que ahora no estamos en un problema de datos faltantes, ya que conocemos los valores de la variable “estatus” (bien porque la hemos inicializado, o porque la tomamos igual a la estimada en el paso 4 del algoritmo).

Partiendo de la ecuación (8.12), podemos escribir la distribución de los estatus, condicionada a la primera observación de todos los paciente como:

$$\pi_{(b,k,\phi)}(S/S_1) \propto \exp \left\{ \sum_{p=1}^P \sum_{t=2}^{N_p} \sum_{i=1}^{52} \left( \sum_{j=1}^{52} \frac{b}{\sqrt{d_{\parallel i,j}^2 + k \cdot d_{\perp i,j}^2}} \cdot s_{i,t,p} \cdot s_{j,t,p} - \phi_t(1 + s_{i,t,p})(1 - s_{i,t-1,p}) \right) \right\}$$

Si llamamos  $\theta = (b, k, \phi)$  al vector de parámetros, el método de Younes actualiza

$$\theta_{n+1} = \theta_n + \frac{c}{n+1} \left\{ h'(s^*) - h'(s) \right\}$$

con  $s^*$  denotando a una realización de la variable  $S$  simulada con el muestreador de Gibbs de acuerdo con su distribución para  $\theta = \theta_n$ . En nuestro caso:

$$h'(s_{itp}) = \begin{pmatrix} -\sum_j \frac{1}{\sqrt{d_{\parallel i,j}^2 + k \cdot d_{\perp i,j}^2}} \cdot s_{itp} \cdot s_{jtp} \\ \sum_j \frac{b \cdot d_{\perp i,j}^2}{2 \cdot \sqrt{d_{\parallel i,j}^2 + k \cdot d_{\perp i,j}^2}} s_{itp} \cdot s_{jtp} \\ (1 + s_{itp})(1 - s_{it-1p}) \end{pmatrix}$$

### 8.6.2. Paso 3. Estimación de los parámetros de la distribución de los umbrales condicionados a los estatus

Como la distribución de los umbrales condicionados a los estatus es gaussiana, vamos a estimar sus parámetros por máxima verosimilitud.

Para simplificar notación y que las expresiones resultantes sean más manejables definimos:

$$\begin{aligned}
V^{-1} &= \text{diag}\left(\frac{1}{\sigma_1^2 + d_i}\right)(I - cH) \\
A_p := A_p(c_1, \alpha_0, \alpha_1) &= c_1 \cdot \Delta S_p + \alpha_0 \Delta Z_{-1,p} + \alpha_1 H \Delta S_{-1,p} - \\
&\quad - \alpha_0 c_1 \Delta S_{-1,p} - \alpha_1 c_1 H \Delta S_{-1,p} \\
B_p := B_p(C_0, \alpha_0, \alpha_1) &= 1' C_0 + \alpha_0 \Delta Z_{-1,p} + \alpha_1 H \Delta S_{-1,p} \\
D_p := D_p(C_0, c_1, \alpha_1) &= 1' C_0 + c_1 \cdot \Delta S_p + \alpha_1 H \Delta S_{-1,p} - \alpha_1 c_1 H \Delta S_{-1,p} \\
E_p := E_p(C_0, c_1, \alpha_0) &= 1' C_0 + c_1 \cdot \Delta S_p + \alpha_0 \Delta Z_{-1,p} - \alpha_0 c_1 \Delta S_{-1,p}
\end{aligned}$$

Entonces el **estimador máximo verosímil de  $C_0$**  será:

$$\begin{aligned}
\log(L) &\propto \\
&\propto \sum_{p=1}^P \left( \Delta Z_p - 1 \cdot C_0' - A_p \right)' \cdot V^{-1} \cdot \left( \Delta Z_p - 1 \cdot C_0' - A_p \right) = \\
&= \sum_{p=1}^P \text{tr} \left[ \left( \Delta Z_p - 1 \cdot C_0' - A_p \right)' \cdot \left( \Delta Z_p - 1 \cdot C_0' - A_p \right) \cdot V^{-1} \right]
\end{aligned}$$

al derivar e igualar a 0 tendremos que:

$$\hat{C}_0 = \sum_{p=1}^P \left( A_p' 1 - \Delta Z_p' 1 \right) \left( 1' 1 \right)' \quad (8.15)$$

**2. Estimador máximo verosímil de  $c_1$ :**

$$\begin{aligned}
\log(L) &\propto \sum_{p=1}^P \left( \Delta Z_p - c_1 \cdot \Delta S_p + c_1 \cdot \alpha_0 \cdot \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} - B_p \right)' \cdot \\
&V^{-1} \cdot \left( \Delta Z_p - c_1 \cdot \Delta S_p + c_1 \cdot \alpha_0 \cdot \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} - B_p \right) =
\end{aligned}$$

$$= \sum_{p=1}^P \text{tr} \left[ \left( \Delta Z_p - c_1 \cdot \Delta S_p + c_1 \cdot \alpha_0 \cdot \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} - B_p \right)' \cdot \left( \Delta Z_p - c_1 \cdot \Delta S_p + c_1 \cdot \alpha_0 \cdot \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} - B_p \right) \cdot V^{-1} \right]$$

al derivar e igualar a 0, si llamamos:

$$P1 := \text{tr} \left[ \left( -\Delta Z'_p \cdot \Delta S_p + \alpha_0 \Delta Z'_p \Delta S_{-1,p} + \alpha_1 \cdot \Delta Z'_p \cdot H \Delta S_{-1,p} + B'_p \cdot \Delta S_p - \alpha_0 B'_p \cdot \Delta S_{-1,p} - \alpha_1 B'_p \cdot H \Delta S_{-1,p} \right) V^{-1} \right]$$

$$P2 := \text{tr} \left[ \left( \Delta S'_p \cdot \Delta S_p - 2\alpha_0 \Delta S'_p \cdot \Delta S_{-1,p} - 2\alpha_1 \Delta S'_p \cdot H \Delta S_{-1,p} + \alpha_0^2 \Delta S_{-1,p} \cdot \Delta S_{-1,p} + 2\alpha_0 \alpha_1 \Delta S'_{-1,p} \cdot H \Delta S_{-1,p} + \alpha_1^2 H \Delta S'_{-1,p} \cdot H \Delta S_{-1,p} \right) V^{-1} \right]$$

$$\text{tendremos que } \hat{c}_1 = -\frac{P1}{P2} \quad (8.16)$$

### 3. Estimador máximo verosímil de $\alpha_0$

$$\begin{aligned} \log(L) &\propto \sum_{p=1}^P \left( \Delta Z_p + c_1 \cdot \alpha_0 \Delta S_{-1,p} - \alpha_0 \Delta Z_{-1,p} - D_p \right)' \cdot V^{-1} \cdot \\ &\quad \cdot \left( \Delta Z_p + c_1 \cdot \alpha_0 \Delta S_{-1,p} - \alpha_0 \Delta Z_{-1,p} - D_p \right) = \\ &= \sum_{p=1}^P \text{tr} \left[ \left( \Delta Z_p + c_1 \cdot \alpha_0 \Delta S_{-1,p} - \alpha_0 \Delta Z_{-1,p} - D_p \right)' \cdot \right. \\ &\quad \left. \cdot \left( \Delta Z_p + c_1 \cdot \alpha_0 \Delta S_{-1,p} - \alpha_0 \Delta Z_{-1,p} - D_p \right) \cdot V^{-1} \right] \end{aligned}$$

al derivar e igualar a 0, si al igual que antes llamamos:

$$P3 := \text{tr} \left[ \left( c_1 \Delta Z'_p \cdot \Delta S_{-1,p} - \Delta Z'_p \cdot \Delta Z_{-1,p} - c_1 D'_p \cdot \Delta S_{-1,p} + D'_p \Delta Z_{-1,p} \right) V^{-1} \right]$$

$$P4 := \operatorname{tr} \left[ \left( c_1^2 \Delta S'_{-1,p} \cdot \Delta S_{-1,p} - 2c_1 \Delta S'_{-1,p} \cdot \Delta Z_{-1,p} + \Delta Z'_{-1,p} \cdot \Delta Z_{-1,p} \right) V^{-1} \right]$$

tendremos que:

$$\hat{\alpha}_0 = -\frac{P3}{P4} \quad (8.17)$$

#### 4. Estimador máximo verosímil de $\alpha_1$ :

$$\begin{aligned} \log(L) &\propto \sum_{p=1}^P \left( \Delta Z_p + \alpha_1 c_1 H \Delta S_{-1,p} - \alpha_1 H \Delta Z_{-1,p} - E_p \right)' \\ &\quad \cdot V^{-1} \cdot \left( \Delta Z_p + \alpha_1 c_1 H \Delta S_{-1,p} - \alpha_1 H \Delta Z_{-1,p} - E_p \right) = \\ &= \sum_{p=1}^P \operatorname{tr} \left[ \left( \Delta Z_p + \alpha_1 c_1 H \Delta S_{-1,p} - \alpha_1 H \Delta Z_{-1,p} - E_p \right)' \cdot \right. \\ &\quad \left. \cdot \left( \Delta Z_p + \alpha_1 c_1 H \Delta S_{-1,p} - \alpha_1 H \Delta Z_{-1,p} - E_p \right) \cdot V^{-1} \right] \end{aligned}$$

al derivar e igualar a 0 llamaremos ahora:

$$P5 := \operatorname{tr} \left[ \left( c_1 \cdot \Delta Z'_p \cdot H \Delta S_{-1,p} - \Delta Z'_p \cdot H \Delta Z_{-1,p} - c_1 E'_p \cdot H \Delta S_{-1,p} + E'_p H \Delta Z_{-1,p} \right) \cdot V^{-1} \right]$$

$$P6 := \operatorname{tr} \left[ \left( c_1^2 H \Delta S'_{-1,p} \cdot H \Delta S_{-1,p} - 2 \cdot c_1 \cdot H \Delta S_{-1,p} \cdot H \Delta Z_{-1,p} + H \Delta Z'_{-1,p} \cdot H \Delta Z_{-1,p} \right) \cdot V^{-1} \right]$$

$$\text{y podremos expresar } \hat{\alpha}_1 = \frac{P5}{P6} \quad (8.18)$$

#### 5. Estimador máximo verosímil de $\sigma^2$

Para simplificar notación definimos:

$$\begin{aligned} \hat{\Upsilon}_p &= 1' \cdot \hat{C}_0 + \hat{c}_1 \cdot \Delta S_p + \hat{\alpha}_0 \Delta Z_{-1,p} + \hat{\alpha}_1 H \Delta Z_{-1,p} - \hat{\alpha}_0 \hat{c}_1 \Delta S_{-1,p} - \\ &\quad - \hat{\alpha}_1 \hat{c}_1 H \Delta S_{-1,p} \end{aligned} \quad (8.19)$$

Entonces la expresión de la logverosimilitud (ec. 8.14) queda:

$$\begin{aligned} \log(L) \propto & \sum_{p=1}^P \left[ \frac{-n \cdot (N_p - 2)}{2} \log(2\pi) - \frac{n \cdot (N_p - 2)}{2} \log(\sigma^2) - \right. \\ & - \frac{N_p - 2}{2} \sum_{i=1}^n \log(\sigma_1^2 + d_i) + \frac{N_p - 2}{2} \sum_{i=1}^n \log(1 - ch_i) - \\ & - \frac{1}{2\sigma^2} \left( \Delta Z_p - \hat{Y}_p \right)' \text{diag} \left( \frac{1}{\sigma_1^2 + d_i} \right) (I - cH) \cdot \left( \Delta Z_p - \hat{Y}_p \right) + \\ & \left. + \sum_{t=3}^{N_p} \sum_{i=1}^n \left( \sum_{j=1}^n \beta_{ij} s_{itp} s_{jtp} - \phi(1 + s_{itp})(1 - s_{it-1p}) \right) \right] \end{aligned}$$

Al derivar con respecto a  $\sigma^2$  e igualar a 0 tendremos:

$$\begin{aligned} & \sum_{p=1}^P \left[ -n(N_p - 2) + \right. \\ & \left. + \frac{1}{\sigma^2} \text{tr} \left[ \left( \Delta Z_p - \hat{Y}_p \right)' \cdot \left( \Delta Z_p - \hat{Y}_p \right) \cdot \text{diag} \left( \frac{1}{\sigma_1^2 + d_i} \right) (I - cH) \right] \right] = 0 \\ \hat{\sigma}^2 = & \frac{\sum_{p=1}^P \text{tr} \left[ \left( \Delta Z_p - \hat{Y}_p \right)' \cdot \left( \Delta Z_p - \hat{Y}_p \right) \cdot \text{diag} \left( \frac{1}{\sigma_1^2 + d_i} \right) (I - cH) \right]}{\sum_{p=1}^P n \cdot (N_p - 2)} \quad (8.20) \end{aligned}$$

## 6. Estimador máximo verosímil de $c$ y de $\sigma_1^2$ .

Si sustituimos (8.19) y (8.20) en el logaritmo de la verosimilitud, tendremos:

$$\begin{aligned} \log(L) \propto & \sum_{p=1}^P \left[ -n \cdot (N_p - 2) \cdot \log(\hat{\sigma}^2) - (N_p - 2) \sum_{i=1}^n \log(\sigma_1^2 + d_i) + \right. \\ & \left. + (N_p - 2) \sum_{i=1}^n \log(1 - ch_i) \right] \end{aligned}$$

$$\log(L) \propto \left( -n \cdot \log(\hat{\sigma}^2) - \sum_{i=1}^n \log(\sigma_1^2 + d_i) + \sum_{i=1}^n \log(1 - ch_i) \right) \quad (8.21)$$

siendo  $h_i$  los valores propios de la matriz  $H$  que es una matriz simétrica  $H = (h_{i,j})$ , tal que  $h_{i,j} = 1$  si las dos posiciones son vecinas, y 0 en otro caso.

No podemos derivar esta expresión respecto a  $c$  ni respecto a  $\sigma_1^2$ , ya que el EMV de  $\sigma^2$  depende de  $\hat{\Upsilon}$  y para estimar  $\Upsilon$  necesitamos conocer  $\sigma_1^2$  y  $c$ . Por tanto vamos a estimarlos mediante un grid search.

Iremos dando valores a  $(\sigma_1^2, c)$  en un grid, y a partir de ellos calcularemos  $\hat{\Upsilon}_{t,p,(\forall t,p)}$ , y  $\hat{\sigma}^2$ . Cuando tengamos todos los parámetros del modelo estimados, evaluaremos la verosimilitud (8.21) y nos quedaremos con aquellos valores de  $\sigma_1^2$  y de  $c$  que la maximicen.

### 8.6.3. Paso 4. Estimación de los estatus

Una vez estimados todos los parámetros e hiperparámetros del modelo, podemos construir la distribución a posteriori de  $S$  condicionado a  $Z$ :

$$f(S/Z) \propto f(Z/S) \cdot \pi(S)$$

y podemos estimar  $S$  como la moda de esta distribución a posteriori. Para obtener este máximo, utilizaremos, como hemos comentado anteriormente, el algoritmo de “Simulated Annealing” con el muestreador de Gibbs [61]. Es bien conocido [78] que este método funciona bien cuando la función a maximizar es de Markov. El esquema del algoritmo es:

- Paso 4.1: *Inicialización*:  $k = 1$ . Definimos un parámetro  $T(k)$  llamado temperatura, y lo inicializamos,  $T(1)$ .
- Paso 4.2: *Actualización completa*. Se actualizan todos los valores de  $S$ .

El nuevo valor para  $s_{itp}$  se escoge aleatoriamente con probabilidad

$$\exp\left\{\frac{1}{T(k)} \ln\left(f(s_{itp}/S_{-itp}, \Delta Z, )\right)\right\}$$

(a continuación, añadimos un apartado más para desarrollar las expresiones de estas probabilidades condicionales, ec. 8.22- 8.24).

- Paso 4.3: *Actualizar los parámetros de control* Poner  $k = k + 1$ ,  $T(k)$ .
- Paso 4.4: Repetir los pasos 4.2-4.3 hasta obtener la convergencia de las estimaciones.

#### 8.6.4. Simulación de los estatus

Vamos a desarrollar las expresiones de las probabilidades condicionales que necesitamos en el paso 4.2 del algoritmo anterior.

A partir de la expresión de la log-verosimilitud:

$$\begin{aligned} \log(L) \propto & \sum_{p=1}^P \left[ \frac{-n \cdot (N_p - 2)}{2} \log(2\pi\sigma^2) - \frac{N_p - 2}{2} \sum_{i=1}^n \log(\sigma_1^2 + d_i) + \right. \\ & + \frac{N_p - 2}{2} \sum_{i=1}^n \log(1 - ch_i) - \frac{1}{2\sigma^2} \left( \Delta Z_p - 1 \cdot 1' C_0 - c_1 \cdot \Delta S_p - \alpha_0 \Delta Z_{-1,p} - \right. \\ & - \alpha_1 H \Delta S_{-1,p} + \alpha_0 c_1 \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} \left. \right)' \text{diag} \left( \frac{1}{\sigma_1^2 + d_i} \right) \left( I - cH \right) \cdot \\ & \left. \left( \Delta Z_p - 1 \cdot 1' C_0 - c_1 \cdot \Delta S_p - \alpha_0 \Delta Z_{-1,p} - \alpha_1 H \Delta S_{-1,p} + \alpha_0 c_1 \Delta S_{-1,p} + \right. \right. \\ & \left. \left. + \alpha_1 c_1 H \Delta S_{-1,p} \right) + \sum_{t=3}^{N_p} \sum_{i=1}^n \left( \sum_{j=1}^n \beta_{ij} s_{itp} s_{jtp} - \phi_t (1 + s_{itp}) (1 - s_{it-1p}) \right) \right] \end{aligned}$$

podemos calcular la expresión de los estatus condicionada a los umbrales.

Para simplificar la notación llamaremos:

$$\begin{aligned} R_{itp} &= \Delta Z_{itp} - C_{0i} - \alpha_0 \Delta Z_{it-1p} - \frac{\alpha_1}{r} \sum_{j \in N_i} \Delta Z_{jt-1p} \\ E_{itp} &= c_1 (s_{itp} - s_{it-1p}) + c_1 \alpha_0 (s_{it-1p} - s_{it-2p}) + c_1 \alpha_1 \frac{1}{r} \sum_{j \in N_i} (s_{jt-1p} - s_{jt-2p}) \end{aligned}$$

Con la nueva notación,

$$\begin{aligned}
\log(L) &\propto \sum_{p=1}^P \sum_{t=3}^{N_p} \left[ \frac{-n(N_p-2)}{2} \log(2\pi) - \frac{n(N_p-2)}{2} \log(\sigma^2) - \right. \\
&- \frac{N_p-2}{2} \sum_{i=1}^n \log(\sigma_1^2 + d_i) + \frac{N_p-2}{2} \sum_{i=1}^n \log(1 - ch_i) - \\
&- \frac{1}{2\sigma^2} \left( R_{tp} - E_{tp} \right)' \text{diag} \left( \frac{1}{\sigma_1^2 + d_i} \right) (I - cH) \cdot \left( R_{tp} - E_{tp} \right) + \\
&+ \log(k(\phi)) + \sum_{i=1}^n \left( \sum_{j=1}^n \beta_{ij} s_{itp} s_{jtp} - \phi_t(1 + s_{itp})(1 - s_{it-1p}) \right) \left. \right] = \\
&= \sum_{p=1}^P \sum_{t=3}^{N_p} \left[ \frac{-n(N_p-2)}{2} \log(2\pi) - \frac{n(N_p-2)}{2} \log(\sigma^2) - \right. \\
&- \frac{N_p-2}{2} \sum_{i=1}^n \log(\sigma_1^2 + d_i) + \frac{N_p-2}{2} \sum_{i=1}^n \log(1 - ch_i) - \\
&- \frac{1}{2\sigma^2} \left( \sum_{l=1}^{52} \frac{1}{\sigma_1^2 + d_l} (R_{ltp} - E_{ltp})^2 - c \cdot (R_{ltp} - E_{ltp}) \cdot \sum_{j \in N(l)} (R_j - E_j) \right) + \\
&+ \sum_{i=1}^n \left( \sum_{j=1}^n \beta_{ij} s_{itp} s_{jtp} - \phi_t(1 + s_{itp})(1 - s_{it-1p}) \right) \left. \right]
\end{aligned}$$

Para que las ecuaciones no resulten tan largas ni engorrosas, seguimos simplificando notación. Definimos  $B_{i,t,p} = R_{i,t,p} - E_{i,t,p}, \forall i, t$ .

- Si  $3 \leq t < N(p) - 1$

$$\begin{aligned}
&f(s_{itp}/Z, S_{t-1p}, S_{t+1p}, S_{-itp}) \propto \tag{8.22} \\
&\propto \exp \left\{ \frac{-1}{2\sigma^2} \cdot \sum_{l=t}^{t+2} \left[ \frac{1}{\sigma_1^2 + d_i} \left( B_{i,l,p} \right)^2 - c \left( B_{ilp} \right) \cdot \sum_{k \in N_i} \left( B_{klp} \right) \right] - \right. \\
&- \frac{1}{2\sigma^2} \sum_{l=t+1}^{t+2} \sum_{m \in N_i} \left[ \frac{1}{\sigma_1^2 + d_m} \left( B_{mlp} \right)^2 - c \left( B_{mlp} \right) \cdot \sum_{k \in N_m} \left( B_{klp} \right) \right] \left. \right\} \cdot \\
&\cdot \exp \left\{ 2 \sum_{j \in N_i^s} \beta_{ij} s_{itp} s_{jtp} - \phi(1 + s_{itp})(1 - s_{it-1p}) - \phi(1 + s_{it+1p})(1 - s_{itp}) \right\}
\end{aligned}$$

- Si  $t = N(p) - 1$

$$\begin{aligned}
 & f(s_{itp}/Z, S_{t-1p}, S_{t+1p}, S_{-itp}) \propto \tag{8.23} \\
 & \propto \exp\left\{ \frac{-1}{2\sigma^2} \cdot \sum_{l=N(p)-1}^{N(p)} \left[ \frac{1}{\sigma_1^2 + d_i} \left( B_{i,l,p} \right)^2 - c \left( B_{ilp} \right) \cdot \sum_{k \in N_i} \left( B_{klp} \right) \right] - \right. \\
 & \quad - \frac{1}{2\sigma^2} \sum_{m \in N_i} \left[ \frac{1}{\sigma_1^2 + d_m} \left( B_{mN(p)p} \right)^2 - c \cdot \left( B_{mN(p)p} \right) \cdot \right. \\
 & \quad \left. \left. \cdot \sum_{k \in N_m} \left( B_{kN(p)p} \right) \right] \right\} \cdot \exp\left\{ 2 \sum_{j \in N_i^s} \beta_{ij} s_{i,N(p)-1,p} s_{j,N(p)-1,p} - \right. \\
 & \quad \left. - \phi(1 + s_{i,N(p)-1,p})(1 - s_{i,N(p)-2,p}) - 2\phi(1 + s_{i,N(p),p})(1 - s_{i,N(p)-1,p}) \right\}
 \end{aligned}$$

- Si  $t = N(p)$

$$\begin{aligned}
 & f(s_{itp}/Z, S_{t-1p}, S_{t+1p}, S_{-itp}) \propto \tag{8.24} \\
 & \propto \exp\left\{ \frac{-1}{2\sigma^2} \cdot \left[ \frac{1}{\sigma_1^2 + d_i} \left( B_{i,N(p),p} \right)^2 - c \left( B_{iN(p)p} \right) \cdot \sum_{k \in N_i} \left( B_{kN(p)p} \right) \right] \right\} \cdot \\
 & \quad \cdot \exp\left\{ 2 \sum_{j \in N_i^s} \beta_{ij} s_{i,N(p),p} s_{j,N(p),p} - 2\phi(1 + s_{i,N(p),p})(1 - s_{i,N(p)-1,p}) \right\}.
 \end{aligned}$$

Para finalizar esta sección, queremos comentar que como criterio de convergencia del algoritmo hemos considerado simplemente la diferencia entre los valores obtenidos de los estimadores de los parámetros en dos iteraciones consecutivas, de manera, que paramos el algoritmo si el mínimo de estas diferencias es menor que un determinado nivel de tolerancia. En cuanto al método de Younes, como se encuentra dentro de una iteración del algoritmo general, hemos tenido que fijar el número de iteraciones a un valor fijo. Este valor se obtuvo probando con una única iteración. El mismo comentario lo extendemos al muestreador de Gibbs usado en el paso de simulación de los estatus.

Cuadro 8.2: Estimaciones de los parámetros de la distribución de los estatus.

$b$	$k$	$\phi$
0.75	10.25	0.45

## 8.7. Resultados y conclusiones

La estimación de los parámetros de la distribución de los estatus puede verse en la tabla 8.2. Como valores iniciales para los parámetros  $b$  y  $k$ , hemos utilizado los obtenidos por Olsson et al [154]. El valor inicial de  $\phi$  fue elegido “ad-hoc” como 0.5. En la tabla 8.3 encontramos las estimaciones de los parámetros de la distribución de los umbrales.

En cuanto a la estimación de los estatus, como ilustración de las estimaciones obtenidas, en la figura 8.1 podemos ver varios de los CV que hemos observado sobre un paciente en el tiempo, y la estimación de la variable estatus en cada posición. En los cuatro primeros gráficos, podemos ver la representación de los umbrales. Los colores más oscuros indican menor intensidad de visión en esos puntos. Los otros cuatro gráficos muestran las estimaciones de los estatus. Hemos representado en color negro las posiciones que han sido estimadas como enfermas, y en color blanco las posiciones normales.

Si comparamos estas estimaciones con la clasificación obtenida con la base empírica del SITA, nuestro método clasifica más posiciones como enfermas. Este resultado es lógico puesto que la clasificación de una posición como enferma tiene en cuenta el estatus de las posiciones vecinas y de las anteriores, lo cual no es tenido en cuenta en el SITA.

Cuadro 8.3: Estimaciones de los parámetros de la distribución de los umbrales.

Parámetro	Estimación
$C_0$	(0.065, -0.184, 0.109, -0.061, 0.102, -0.153, 0.096, -0.071, -0.187, 0.341, 0.095, 0.081, -0.110, -0.286, 0.232, 0.393, 0.471, 0.458, 0.311, 0.122, 0.052, -0.039, 0.589, 0.135, 0.183, 0.412, -0.316, 0.177, -0.170, 0.210, 0.493, 0.234, -0.102, -0.010, -0.134, -0.211, -0.033, 0.022, 0.074, 0.065, 0.103, -0.102, -0.319, 0.289, 0.012, 0.280, 0.105, 0.030, 0.058, -0.122, -0.041, -0.195)
$c_1$	1.465
$\alpha_0$	-0.440
$\alpha_1$	0.093
$\sigma^2$	1.145
$\sigma_1^2$	0.574
$c$	0.104

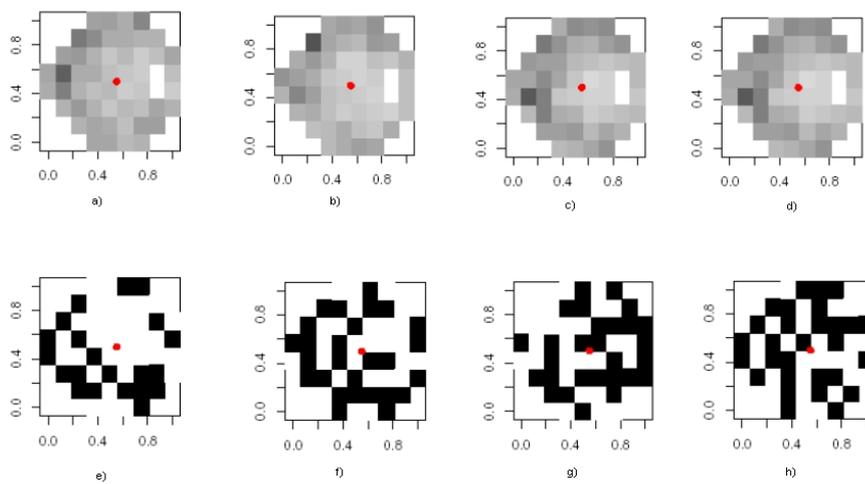


Figura 8.1: Ilustración de los resultados obtenidos. En a), b), c) y d) tenemos 4 CV observados en un mismo paciente con diferencia de un año entre observaciones consecutivas. En e), f), g) y h) vemos las estimaciones obtenidas para la variable “estatus”. En negro representamos las posiciones estimadas como “enfermas”.



## Capítulo 9

# Modelización de campos visuales a partir de la metodología clásica de series temporales multivariantes

### Contenidos

---

<b>9.1. Introducción</b>	<b>277</b>
<b>9.2. Modelización de los umbrales condicionados a los estatus</b>	<b>278</b>
<b>9.3. Modelización de los estatus</b>	<b>278</b>
<b>9.4. Verosimilitud</b>	<b>280</b>
<b>9.5. Estimación de parámetros. Búsqueda del método apropiado</b>	<b>282</b>
9.5.1. Paso 1. Simulación de los estatus	283
9.5.2. Paso 2. Estimación de los parámetros de la distribución de los estatus	284
9.5.3. Paso 3. Estimación de los parámetros de la distribución de los umbrales condicionados a los estatus	285

**9.6. Predicción. . . . . 287**

**9.7. Bondad del ajuste . . . . . 288**

---

## 9.1. Introducción

En este capítulo nos planteamos la construcción de un modelo para caracterizar la distribución espacio temporal de los CV de pacientes con glaucoma. Será por tanto un trabajo similar al que planteamos en el capítulo 6, pero ahora utilizando la metodología de las series temporales multivariantes.

El objetivo de la modelización, además de la descripción del proceso, será ver si es posible predecir el comportamiento futuro del CV de un paciente, a partir de las observaciones que tengamos de él. Para construir este modelo, en la sección 9.2 modelizaremos la distribución de los umbrales condicionados a los estatus  $f(Z/S)$ ; en la sección 9.3 modelizaremos la distribución de los estatus, y en la sección 9.4 formularemos la función de distribución conjunta  $f(Z, S)$ .

Los modelos y métodos utilizados van a ser prácticamente los mismos del capítulo anterior, por este motivo no nos extenderemos demasiado en ellos y sólo nos detendremos en aquellos que representen alguna novedad respecto al capítulo anterior. También seguimos manteniendo la misma notación introducida en la sección 8.1.1 y utilizamos la misma base de datos. Las mismas consideraciones sobre los campos visuales comentadas en la sección 8.2 deberán tenerse en cuenta.

El resto de capítulo se ha organizado como sigue: en la sección 9.2 se propone un modelo para la variable umbral condicionada a la variable estatus, a continuación en la sección 9.3 se propone un modelo para la variable estatus, y en la sección 9.4 se formula la verosimilitud conjunta. El problema de estimación de parámetros se aborda en la sección 9.5. En la sección 9.6 se estudia cómo hacer predicciones con el modelo propuesto y por último se comprueba la bondad del ajuste en la sección 9.7.

## 9.2. Modelización de los umbrales condicionados a los estatus

En la sección 8.3 del capítulo anterior construimos un modelo para la distribución de los umbrales (condicionados a los estatus), teniendo en cuenta las características de los CV. Como seguimos trabajando con el mismo conjunto de datos, continuaremos trabajando con el mismo modelo:

$$\Delta Z_{tp} = C_0 + c_1 \Delta S_{tp} + \Psi \left( \Delta Z_{t-1p} - c_1 \Delta S_{t-1p} \right) + \epsilon_{tp}$$

$$\epsilon_{tp} \mid \Delta Z_{t-1p}, \Delta S_{tp} \sim N(0, \sigma^2 (\sigma_1^2 I + D) \cdot (I - C)^{-1}),$$

con  $C_0, c_1, \Psi, \sigma^2, \sigma_1^2, D$  y  $C$  como en la sección 8.3

## 9.3. Modelización de los estatus

Siguiendo el trabajo de Olsson y Rootzen [154], quienes utilizaron un modelo de Ising para modelizar la variable estatus, en la sección 8.4 vimos dos aproximaciones que permiten generalizar el modelo de Ising al caso espacio temporal: la utilizada por Hainsworth y Mardia [80] y el modelo propuesto entre otros por Järpe [103], y por Guyon y Hardouin [79]

El objetivo final que nos planteamos en este capítulo es el poder predecir CV en tiempos no observados (en el futuro). Como ya comentamos en la sección 8.4, en este caso es conveniente modelizar los estatus como proponen Guyon et al., ya que la aproximación propuesta por Hainsworth et al. no nos permite calcular probabilidades de sucesos en tiempos futuros, y por tanto no permite realizar predicciones.

Por tanto, vamos a formular:

$$\begin{aligned} \pi\left(S_{tp}/S_{t-1p}, S_{t-2p}, \dots, S_{1p}\right) &= \pi\left(S_{tp}/S_{t-1p}\right) = \\ &= \varsigma_{tp}^{-1} \exp\left\{\sum_i \sum_{j \sim i} \beta_{ij} s_{itp} s_{jtp} - \phi \sum_i (1 + s_{itp})(1 - s_{it-1p})\right\}, \end{aligned}$$

siendo  $j \sim i$  el conjunto de posiciones vecinas de la localización  $i$ -ésima del grid, (considerando simplemente una vecindad de primer orden),  $\beta_{ij}$  como en la ecuación (8.9), y  $\varsigma_{tp} = \varsigma(S_{t-1p}, b, k, \phi) = \sum_{s_{itp}} \exp\{\sum_i \sum_{j \sim i} \beta_{ij} s_{itp} s_{jtp} - \phi \sum_i (1 + s_{itp})(1 - s_{it-1p})\}$  la constante normalizadora de la distribución. (Hemos utilizado la letra  $\varsigma$  para denotar la constante normalizadora en lugar de la letra  $Z$ , para no confundir ésta constante con la variable “umbral” a la que también habíamos denotado por  $Z$ . Notar que en la constante normalizadora sumamos sobre todas las posibles configuraciones  $s_{tp}$ , por lo que en total tendremos  $2^{52}$  sumatorios).

A partir de esta distribución, se deduce que:

$$\begin{aligned} \pi\left(S_{itp} = s_{itp}/S_{jtp} = s_{jtp} \forall j \sim i, S_{lt-1p} = s_{lt-1p} \forall l \in \{1, \dots, 52\}\right) &= \\ &= \varsigma_{itp}^{-1} \exp\left\{s_{itp} \cdot \left[\sum_{j \sim i} \beta_{ij} s_{jtp} - \phi(1 - s_{it-1p})\right]\right\} \end{aligned} \quad (9.1)$$

y como  $s_{itp} \in \{\pm 1\}$ :

$$\begin{aligned} \varsigma_{itp} &= \exp\left\{\sum_{j \sim i} \beta_{ij} s_{jtp} - \phi(1 - s_{it-1p})\right\} + \exp\left\{-\sum_{j \sim i} \beta_{ij} s_{jtp} + \phi(1 - s_{it-1p})\right\} = \\ &= 2 \cosh\left(\sum_{j \sim i} \beta_{ij} s_{jtp} - \phi(1 - s_{it-1p})\right). \end{aligned}$$

Veamos a partir de aquí cómo queda la función de verosimilitud y cómo podemos estimar los parámetros de las dos distribuciones,  $f(Z/S)$  y  $\pi(S)$ .

## 9.4. Verosimilitud

Una vez modelizada la distribución de los estatus

$$\pi\left(S_{2p}, \dots, S_{Tp} \mid S_{1p}\right) = \prod_{t=2}^T \pi(S_{tp} \mid S_{t-1p})$$

y la distribución de los umbrales condicionados a los estatus, si utilizamos la notación vectorial que introdujimos en la ecuación 8.6, tendremos:

$$\begin{aligned} \Delta Z_p &= 1' \cdot C_0 + c_1 \Delta S_p + \alpha_0 \Delta Z_{-1,p} + \alpha_1 H \Delta S_{-1,p} - \alpha_0 c_1 \Delta S_{-1,p} - \\ &\quad - \alpha_1 c_1 H \Delta S_{-1,p} + \tilde{\epsilon}_p \\ \text{con} \quad \epsilon_{tp} \mid \Delta Z_{t-1p}, S_p &\sim N(0, \sigma^2(\sigma_1^2 I + D)(I - cH)^{-1}). \end{aligned}$$

La distribución conjunta quedará:

$$\begin{aligned} f\left(\Delta Z, S_3, \dots, S_T \mid Z_1, Z_2, S_1, S_2\right) &= \\ &= \prod_p f\left(\Delta Z_p, S_{3p}, \dots, S_{N_p} \mid Z_{1p}, Z_{2p}, S_{1p}, S_{2p}\right) = \\ &= \prod_p f(\Delta Z_p \mid Z_{1p}, Z_{2p}, S_{1p}, \dots, S_{N_p}) \cdot \pi\left(S_{3p}, \dots, S_{N_p} \mid S_{1p}, S_{2p}\right) = \\ &= \prod_p \left( f(\Delta Z_p \mid Z_{1p}, Z_{2p}, S_{1p}, \dots, S_{N_p}) \cdot \prod_{t=3}^{N_p} \pi\left(S_{tp} \mid S_{t-1p}\right) \right). \end{aligned} \quad (9.2)$$

Por tanto:

$$\begin{aligned} f\left(\Delta Z, S_3, \dots, S_T \mid Z_1, Z_2, S_1, S_2\right) &= \quad (9.3) \\ &= \prod_{p=1}^P \left[ (2\pi\sigma^2)^{-\frac{n \cdot (N_p - 2)}{2}} \mid (\sigma_1^2 I + D)^{-1} (I - cH) \mid^{\frac{N_p - 2}{2}} \cdot \right. \\ &\quad \left. \cdot \exp\left\{ \frac{-1}{2\sigma^2} \left( \Delta Z_p - 1 \cdot 1' C_0 - c_1 \cdot \Delta S_p - \alpha_0 \Delta Z_{-1,p} - \alpha_1 H \Delta S_{-1,p} + \right. \right. \right. \end{aligned}$$

$$\begin{aligned}
& + \alpha_0 c_1 \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} \Big)' \cdot (\sigma_1^2 I + D)^{-1} \cdot (I - cH) \cdot \\
& \left( \Delta Z_p - 1 \cdot 1' C_0 - c_1 \cdot \Delta S_p - \alpha_0 \Delta Z_{-1,p} - \alpha_1 H \Delta S_{-1,p} + \alpha_0 c_1 \Delta S_{-1,p} + \right. \\
& \left. + \alpha_1 c_1 H \Delta S_{-1,p} \right) \cdot \\
& \cdot \prod_{t=3}^{N_p} \varsigma_{tp}^{-1} \exp \left\{ \sum_i \sum_{j \sim i} \beta_{ij} s_{itp} s_{jtp} - \phi \sum_i (1 + s_{itp})(1 - s_{it-1p}) \right\} \Big] .
\end{aligned}$$

Hemos calculado la distribución conjunta condicionada a las dos primeras observaciones, ya que de nuevo estamos trabajando con un modelo equivalente a un  $VAR(2)$ .

Si al igual que en la sección 8.5, suponemos que la matriz de vecindad  $H$ , es una matriz cuadrada y simétrica, y llamamos  $h_1 \leq h_2 \leq \dots \leq h_n$  a sus valores propios ordenados, entonces  $|I - cH| = \prod_{i=1}^n (1 - ch_i)$ , y la logverosimilitud quedará:

$$\begin{aligned}
\log(L) &= \sum_{p=1}^P \left[ \frac{-n \cdot (N_p - 2)}{2} \log(2\pi\sigma^2) - \right. & (9.4) \\
& - \frac{N_p - 2}{2} \sum_{i=1}^n \log(\sigma_1^2 + d_i) + \frac{N_p - 2}{2} \sum_{i=1}^n \log(1 - ch_i) - \\
& - \frac{1}{2\sigma^2} \left( \Delta Z_p - 1 \cdot 1' C_0 - c_1 \cdot \Delta S_p - \alpha_0 \Delta Z_{-1,p} - \alpha_1 H \Delta S_{-1,p} + \right. \\
& \left. + \alpha_0 c_1 \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} \right)' \cdot \text{diag} \left( \frac{1}{\sigma_1^2 + d_i} \right) (I - cH) \cdot \\
& \left( \Delta Z_p - 1 \cdot 1' C_0 - c_1 \cdot \Delta S_p - \alpha_0 \Delta Z_{-1,p} - \alpha_1 H \Delta S_{-1,p} + \right. \\
& \left. + \alpha_0 c_1 \Delta S_{-1,p} + \alpha_1 c_1 H \Delta S_{-1,p} \right) - \sum_{t=3}^{N_p} \log(\varsigma_{tp}) + \\
& \left. + \sum_{t=3}^{N_p} \sum_{i=1}^n \sum_{j=1}^n \left( \sum_{j=1}^n \beta_{ij} s_{itp} s_{jtp} - \phi (1 + s_{itp})(1 - s_{it-1p}) \right) \right]
\end{aligned}$$

## 9.5. Estimación de parámetros. Búsqueda del método apropiado

La estimación de los parámetros del modelo,  $(C_0, c_1, \alpha_0, \alpha_1, \sigma^2, \sigma_1^2, c, \phi, b, k)$  sigue siendo, como en el capítulo anterior, un problema complicado por el hecho de tener datos faltantes (no tenemos observaciones de la variable estatus), y porque la constante normalizadora de la distribución de probabilidad de la variable “estatus” es intratable.

De nuevo vemos que debido a la expresión de nuestra verosimilitud (ec. 9.2), si no tuviéramos datos faltantes podríamos estimar por separado los parámetros de la distribución a priori de los estatus, y los parámetros de la distribución de los umbrales condicionados a los estatus. También ocurre, como en el capítulo anterior, que la distribución de los umbrales es gaussiana y por tanto los estimadores máximo verosímiles pueden ser obtenidos de una forma relativamente sencilla.

A diferencia del capítulo anterior, en este caso no queremos estimar los estatus, y nos encontramos con el hecho de que en la verosimilitud la distribución de los estatus aparece expresada en términos de condicionales. Teniendo en cuenta todo lo expuesto, si volvemos a repasar los métodos de estimación que vimos en el apartado 7.4 vemos que el algoritmo más adecuado en este caso es un algoritmo muy similar al método de Chalmond, “Reversible Jump MCMC” [9]. Este algoritmo, como comentábamos en la sección 7.4.4, es similar al EM estocástico pero sustituyendo la verosimilitud por la pseudoverosimilitud. Nosotros utilizaremos la pseudoverosimilitud para la distribución del estatus y la verosimilitud para los parámetros de la distribución de los umbrales, y en el paso  $E$  del algoritmo, calcularemos la media muestral sólo a partir de una única simulación del proceso “faltante”.

El proceso iterativo consistirá pues es lo siguiente. Empezaremos dando

valores iniciales a todos los parámetros y simulando los valores de los estatus. Una vez concluida la simulación, consideraremos que los valores de la variable “estatus” son conocidos e iguales a los obtenidos, y estimaremos por separado los parámetros de la distribución de los estatus, y los de la distribución de los umbrales condicionados a los estatus. Para estimar los parámetros de la distribución de los estatus vamos a utilizar la pseudoverosimilitud. Por último, la estimación de los parámetros de la distribución de los umbrales condicionados a los estatus seguirá exactamente los mismos pasos que ya detallamos en las ecuaciones 8.15-8.21.

Una vez estimados todos los parámetros, volvemos a simular los estatus a partir de los nuevos valores estimados de los parámetros; volvemos a estimar todos los parámetros del modelo, y así sucesivamente hasta lograr la convergencia de los estimadores.

Veamos un poco más detalladamente cómo simular los estatus y cómo estimar todos los parámetros del modelo.

### 9.5.1. Paso 1. Simulación de los estatus

Vamos a utilizar de nuevo el muestreador de Gibbs para simular los estatus, y para ello debemos simular de la distribución:

$f(s_{itp} | \Delta Z_p, S_{-itp}, S_{-tp})$ , siendo  $S_{-itp} = (s_{1tp}, \dots, s_{i-1tp}, s_{i+1tp}, \dots, s_{52tp})$ , y  $S_{-tp}$  todos los CV que tenemos del paciente número  $p$  excepto el número  $t$ .

Como

$$f(s_{itp} | \Delta Z_p, S_{-itp}, S_{-tp}) \propto f(\Delta Z_p | s_{itp}, S_{-itp}, S_{-tp}) f(s_{itp} | S_{-itp}, S_{-tp})$$

necesitamos conocer las expresiones de estas dos distribuciones condicionales.

Fácilmente se demuestra [79] que

$$\begin{aligned} f(s_{itp} | S_{-itp}, S_{t-1p}, S_{t+1p}) &\propto \\ &\propto s_{it+1p}^{-1} \exp\left(\sum_{j \sim i} \beta_{ij} s_{it+1p} s_{jt+1p} - \phi(1 + s_{it+1p})(1 - s_{itp})\right) \cdot \\ &\cdot \exp\left(\sum_{j \sim i} \beta_{ij} s_{itp} s_{jtp} - \phi(1 + s_{itp})(1 - s_{it-1p})\right) \end{aligned}$$

Además:

$$\begin{aligned} f(\Delta Z_{tp} | s_{itp}, S_{-itp}, S_{-tp}) &\propto \\ &\propto \exp\left\{\left(\Delta Z_{tp} - C_0 - c_1 \Delta S_{tp} - \Psi\left(\Delta Z_{t-1p} - c_1 \Delta S_{t-1p}\right)\right)'\right. \\ &\cdot \left(\sigma^2(\sigma_1^2 I + D) \cdot (I - C)^{-1}\right)^{-1/2} \cdot \\ &\left.\left(\Delta Z_{tp} - C_0 - c_1 \Delta S_{tp} - \Psi\left(\Delta Z_{t-1p} - c_1 \Delta S_{t-1p}\right)\right)\right\} \end{aligned}$$

con  $C_0, c_1, \Psi, \sigma^2, \sigma_1^2, D$  y  $C$  como en la sección 8.3, y por tanto podemos ya obtener la expresión de

$$f(s_{itp} | \Delta Z_p, S_{-itp}, S_{-tp}) \quad \forall(i, t, p)$$

### 9.5.2. Paso 2. Estimación de los parámetros de la distribución de los estatus

Como hemos comentado, debido a la intratabilidad de la constante normalizadora de la distribución conjunta de los estatus, para estimar los parámetros de esta distribución vamos a utilizar métodos basados en la pseudoverosimilitud, estimando por tanto los parámetros a partir de probabilidades condicionales, que son las únicas de las que conocemos su expresión en forma cerrada.

La pseudoverosimilitud [17] para la distribución de los estatus será:

$$\begin{aligned}
 PL_{b,k,\phi}(S) &= \\
 &= \prod_{p=1}^P \prod_{t=2}^{N_p} \prod_{i=1}^{52} \pi \left( S_{itp} = s_{itp} / S_{jtp} = s_{jtp} \forall j \sim i, S_{lt-1p} = s_{lt-1p} \forall l \right)
 \end{aligned}$$

Teniendo en cuenta la ecuación (9.1), y la expresión de la constante normalizadora  $\varsigma_{itp}$  tendremos que:

$$\begin{aligned}
 PL_{b,k,\phi}(S) &= \left( \prod_{p=1}^P \prod_{t=2}^{N_p} \prod_{i=1}^{52} 2 \cosh \left( \sum_{j \sim i} \beta_{ij} s_{jtp} - \phi(1 - s_{it-1p}) \right) \right) \cdot \\
 &\cdot \exp \left\{ \sum_{p=1}^P \sum_{t=2}^{N_p} \sum_{i=1}^{52} \left( s_{itp} \cdot \left[ \sum_{j \sim i} \beta_{ij} s_{jtp} - \phi(1 - s_{it-1p}) \right] \right) \right\}
 \end{aligned}$$

Para maximizar la pseudoverosimilitud, utilizaremos uno de los algoritmos numéricos que vienen implementados en el toolbox de optimización del MATLAB.

### 9.5.3. Paso 3. Estimación de los parámetros de la distribución de los umbrales condicionados a los estatus

Volveremos a utilizar el procedimiento iterativo que ya detallamos en el Paso 3 del apartado 8.6 (ec. 8.15- 8.21).

Por último, comentar que los criterios usados para controlar la convergencia son los mismos que comentamos en la sección 8.6. En las tabla 9.2 y 9.1 tenemos las estimaciones de los parámetros que hemos obtenido para la distribución de los umbrales y de los estatus respectivamente. Notar que en este caso el significado de los parámetros de la distribución del estatus no es la misma, el parámetro  $\beta_{i,j}$  debería ser ahora aproximadamente el doble al obtenido en la formulación bayesiana.

Cuadro 9.1: Estimaciones de los parámetros de la distribución de los estatus utilizando el algoritmo iterativo propuesto en la sección 9.5

$b$	$k$	$\phi$
2.43	0.73	0.47

Cuadro 9.2: Estimaciones de los parámetros de la distribución de los umbrales utilizando el algoritmo iterativo propuesto en la sección 9.5

Parámetro	Estimación
$C_0$	(-0.105, -0.129, 0.168, -0.036, 0.158, -0.147, -0.009, -0.002, -0.194, 0.155, 0.079, 0.080, -0.120, -0.355, 0.353, 0.305, 0.514, 0.416, 0.357, -0.028, 0.073, 0.041, 0.390, -0.042, 0.222, 0.434, -0.339, -0.131, -0.324, 0.211, 0.354, 0.109, -0.170, -0.204, -0.247, -0.088, -0.289, -0.067, -0.091, -0.076, 0.157, -0.239, -0.455, 0.409, -0.244, 0.261, 0.106, 0.035, -0.229, -0.281, -0.135, -0.248)
$c_1$	0.30749
$\alpha_0$	-0.34382
$\alpha_1$	0.038115
$\sigma^2$	1.383
$\sigma_1^2$	0.574
$c$	0.112

## 9.6. Predicción.

En la sección 7.2.1 estudiamos los procesos autoregresivos multivariantes, y obtuvimos las expresiones para predecir el valor de la variable en el instante  $t+h$ , a partir de la información disponible hasta el instante  $t$ ,  $\hat{y}_{t+h|t}$  tomando aquel valor que minimiza el error cuadrático medio de predicción.

En este apartado vamos a adaptar estas expresiones a nuestro modelo (como la expresión es la misma para todos los pacientes eliminamos el subíndice  $p$ ):

$$\begin{aligned}\Delta Z_t &= C_0 + c_1 \Delta S_t + \Psi \left( \Delta Z_{t-1} - c_1 \Delta S_{t-1} \right) + \epsilon_t \\ Z_t - Z_{t-1} &= C_0 + c_1 (S_t - S_{t-1}) + \Psi \left( Z_{t-1} - Z_{t-2} - c_1 (S_{t-1} - S_{t-2}) \right) + \epsilon_t \\ Z_t &= (I + \Psi) Z_{t-1} - \Psi Z_{t-2} + C_0 + c_1 (S_t - (I + \Psi) S_{t-1} + \Psi S_{t-2}) + \epsilon_t.\end{aligned}$$

Como la forma más sencilla de hacer predicciones en un modelo VAR es de forma recursiva, veremos únicamente la predicción del valor de la variable en  $t+1$  a partir de la información que tenemos hasta el instante  $t$ .

Hemos de tener en cuenta que estamos ante un modelo con cambios en régimen. Si el régimen, en nuestro caso la variable estatus,  $S_1, \dots, S_t$ , fuera conocido la predicción en  $t+1$  sería simplemente:

$$\begin{aligned}E[Z_{t+1} | Z_l, S_l : l \leq t] &= (I + \Psi) Z_t - \Psi Z_{t-1} + C_0 + \\ &+ c_1 (S_{t+1} - (I + \Psi) S_t + \Psi S_{t-1})\end{aligned}$$

La predicción incondicional se obtiene de una forma directa utilizando propiedades básicas del operador esperanza:

$$\begin{aligned}
\hat{Z}_{t+1|t} &= E[E[Z_{t+1} | Z_l, S_l : l \leq t]] = \\
&= (I + \Psi)Z_t - \Psi Z_{t-1} + C_0 + c_1(E[S_{t+1} | Z_{l \leq t}] - \\
&\quad -(I + \Psi)E[S_t | Z_{l \leq t}] + \Psi E[S_{t-1} | Z_{l \leq t}])
\end{aligned}$$

Como las esperanzas respecto a la distribución de los estatus no pueden obtenerse de forma analítica, análogamente a lo que propone Hamilton [84] lo que proponemos es aproximarlas mediante simulaciones.

Una vez tengamos la predicción en  $Z$  para el instante  $t + 1$  podremos predecir en  $t + 2$ , y así recursivamente hasta el instante que nos interese.

En la figura 9.1, vemos como ilustración un ejemplo de predicción. A partir de 5 CVs observados a un paciente, predecimos su comportamiento un año, y dos años después de la última exploración.

## 9.7. Bondad del ajuste

En el capítulo 7, vimos que al trabajar con modelos VAR, la bondad del ajuste se realizaba a partir del análisis de los residuos, sobre los que se podía realizar una serie de tests para comprobar si son ruido blanco gaussiano.

Queremos comentar que este apartado lo presentamos simplemente como ilustración de la metodología a seguir para comprobar la bondad del modelo. Debido a la limitación de observaciones en el tiempo que tenemos en nuestra base de datos, somos conscientes de que los resultados de este apartado deberían repetirse sobre una base de datos mayor, y que por tanto no pueden tomarse como concluyentes.

Calculamos por tanto los residuos:

$$R_{tp} := Z_{tp} - \hat{Z}_{tp|t-1p}$$

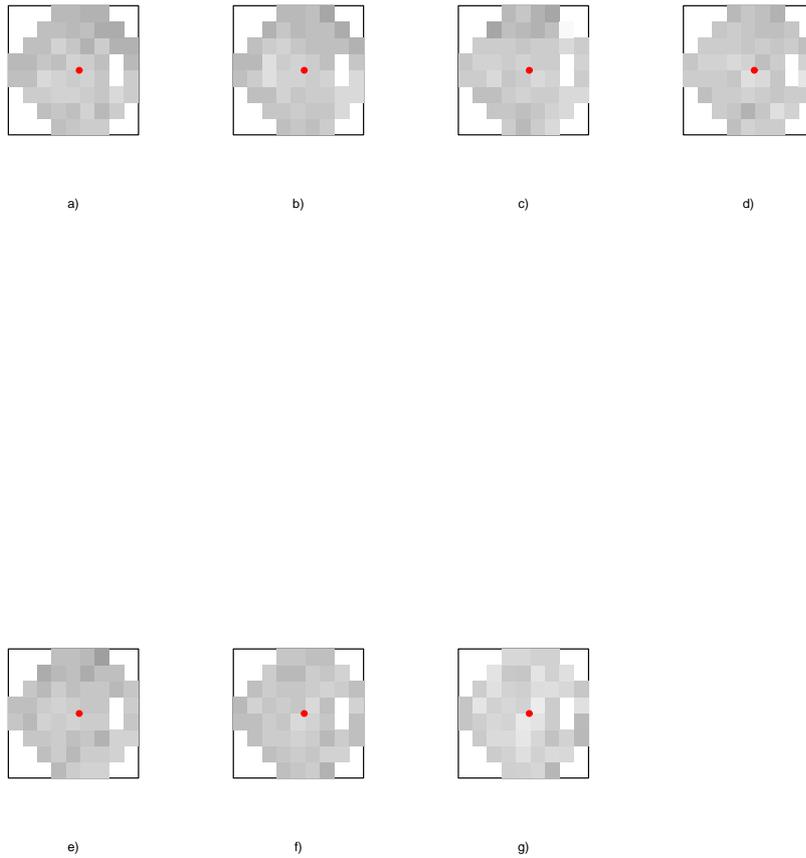


Figura 9.1: En a), b), c), d) y e) vemos 5 campos visuales observados en un paciente, en cinco exploraciones consecutivas. A partir de ellas predecimos CVs de este paciente. En f) vemos la predicción un año después de la última observación, y en g) dos años después.

Sobre estos residuos, como indicamos en el apartado 7.2.1, podemos aplicar tests para ver si las autocorrelaciones de los residuos se corresponden con las autocorrelaciones de un proceso de ruido blanco, y estudiar sus momentos de tercer y cuarto orden para ver si proceden de una distribución gaussiana.

Empezamos calculando las correlaciones para diferentes retardos temporales, y comparando los valores obtenidos (en valor absoluto) con  $2/\sqrt{T}$ . Rechazaremos la hipótesis de ruido blanco, si alguno de los coeficientes de correlación estimados excede esta cota.

Debido al reducido número de observaciones en el tiempo, las bandas de confianza son muy anchas, y resulta un contraste muy conservador, en el que va a resultar muy difícil rechazar la hipótesis de ruido blanco. Comprobamos que, efectivamente, todos los valores que obtenemos para las correlaciones caen dentro de estas cotas. En la figura 9.2, vemos un ejemplo de los coeficientes de correlación obtenidos para un paciente del que disponíamos de 9 observaciones en el tiempo. (Recordar que las dos primeras observaciones las suponemos fijas y conocidas, por lo que no las consideramos al analizar los residuos. Aceptamos por tanto la hipótesis de que los residuos de nuestro modelo siguen una distribución de ruido blanco, teniendo en cuenta las limitaciones comentadas anteriormente.

Para ver que los residuos siguen una distribución gaussiana, simplemente como ilustración, calculamos los índices univariantes de asimetría y curtosis. Para ello, trabajamos con los residuos estandarizados, sobre los que calculamos los índices para los residuos obtenidos en cada una de las 52 posiciones del CV de cada paciente. Como estamos siempre suponiendo independencia entre distintos pacientes, podemos juntar los residuos que tenemos en la misma posición de todos ellos, para tener una muestra más grande sobre la que calcular los índices de asimetría y curtosis.

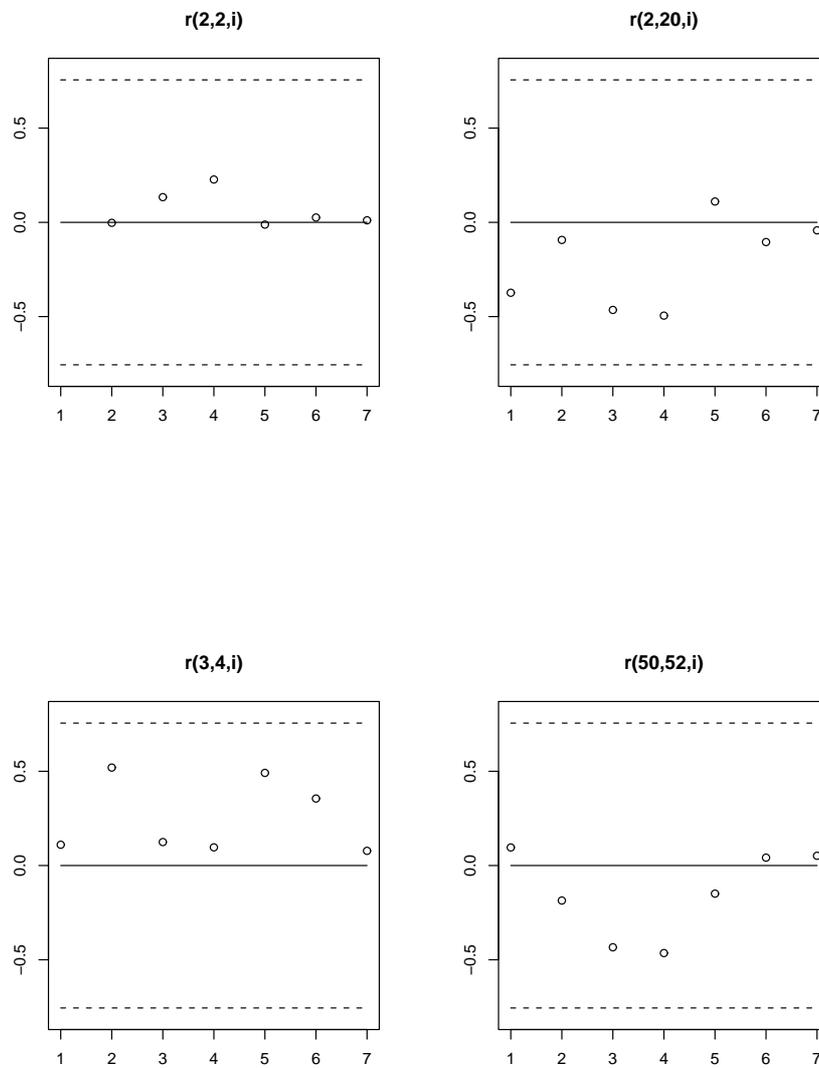


Figura 9.2: Algunas de las correlaciones estimadas para los residuos de un paciente del que tenemos 9 exploraciones.  $r(k, l, i)$  indica el  $(k, l)$ -ésimo elemento de la matriz de correlación de retardo  $i$  (ver apartado 7.2.1).

Obtenemos los siguientes resultados:

Pos.	1	2	3	4	5	6	7	8	9	10	11
Asim.	0.2	0.3	-0.6	1.1	0.1	-0.1	-0.8	-0.2	-0.7	-1.4	0.1
Curt.	2.7	2.8	2.9	3.4	3.2	5.0	3.2	2.4	3.4	5.9	1.8
Pos.	12	13	14	15	16	17	18	19	20	21	22
Asim.	0.1	0.1	0.1	0.1	0.1	-0.9	-0.4	0.3	-0.2	0.1	0.2
Curt.	2.6	2.5	3.0	2.9	2.6	5.1	3.2	3.2	2.7	2.9	3.0
Pos.	23	24	25	26	27	28	29	30	31	32	33
Asim.	-0.1	0.1	-1.8	-0.6	2.1	-1.7	0.6	-0.8	0.1	2.1	-0.1
Curt.	2.2	4.2	7.9	4.7	9.9	7.1	3.4	3.2	3.2	2.8	2.7
Pos.	34	35	36	37	38	39	40	41	42	43	44
Asim.	0.1	-0.2	0.7	0.19	-1.2	-0.9	-0.8	-0.1	0.2	0.1	0.2
Curt.	4.6	2.9	10.8	2.9	3.0	3.9	2.9	4.2	3.3	4.4	2.4
Pos.	45	46	47	48	49	50	51	52			
Asim.	-0.6	-0.5	-0.9	-0.3	0.6	0.4	-0.3	-0.28			
Curt.	2.6	3.0	4.4	2.5	2.9	2.1	2.8	4.1			

Podemos observar que en algunas posiciones, la forma de la distribución sí se asemeja a la de la distribución normal, mientras que en otras la normalidad es un poco más dudosa. Estas medidas podrían complementarse con otros tests de normalidad como por ejemplo el test de Kolmogorov-Smirnov o el test de Newton [149].

Además de estos tests, como comentábamos en la sec. 7.2.1, si el objetivo final de la modelización es la predicción, el ECM de predicción nos dará una medida de la bondad del modelo. En la sec. 7.2.1, hemos visto tests que se basan en este estadístico para elegir el orden del modelo más adecuado. Como hemos comentado, y debido de nuevo a la limitación de nuestra base de datos, calculamos este valor simplemente como una medida descriptiva más de la bondad del ajuste. El valor obtenido para el ECM es 8.24.

## Parte IV

# Conclusiones, comentarios y continuación del trabajo



## Capítulo 10

# Comentarios, conclusiones y problemas abiertos

Como ya comentábamos en el capítulo 1, el objetivo de este trabajo ha sido el estudio y la aplicación de modelos y técnicas estadísticas espacio temporales para la resolución de una serie de problemas que se plantean en el estudio del glaucoma. Hemos aprovechado también para hacer una revisión de la teoría de los modelos espacio temporales en geoestadística, y de los modelos de series temporales multivariantes.

Los problemas que nos hemos planteado han sido la modelización de la distribución espacio temporal de campos visuales de pacientes sanos, y de pacientes que sufren glaucoma (con el objetivo de realizar predicciones y simulaciones) y el poder “etiquetar” cada una de las posiciones de un CV como “sana” o “enferma”. Para ello, hemos trabajado utilizando dos metodologías muy distintas: la metodología geoestadística, y la metodología clásica de las series temporales multivariantes de Box & Jenkins.

Para llevar a cabo estas modelizaciones, hemos trabajado con una base de datos de CVs de pacientes sanos y de pacientes con glaucoma. En esta base de datos tenemos 131 pacientes sanos con una única observación, 38 pacientes sanos con 2 observaciones, 9 con 3 observaciones, y 8 con 4. La

base de datos se completa con CVs de otros 105 pacientes a los que se ha diagnosticado glaucoma, y de los que tenemos entre 2 y 9 observaciones por persona.

Debido a las características de la enfermedad, y al poco tiempo que hace que el SITA se ha introducido como estrategia estándar en las consultas oftalmológicas, no nos ha sido posible conseguir una base de datos más extensa. Nuestro objetivo es continuar trabajando en colaboración con los oftalmólogos para conseguir una mejor base de datos, en la que tengamos un conjunto de pacientes, con un número bastante mayor de observaciones sobre cada uno, y repetir todos los análisis que hemos presentado en este trabajo con esta nueva base de datos, especialmente la parte en la que hemos usado la metodología de series temporales multivariantes. En este sentido, presentamos nuestro trabajo como una ilustración de la metodología a utilizar, y no presentamos los resultados como definitivos desde el punto de vista clínico.

En la parte de geoestadística, gran parte del trabajo se ha desarrollado basándose en un modelo puramente espacial construido, en este caso sí, sobre un número bastante grande de CVs pertenecientes a distintos pacientes.

Con las dos metodologías hemos estado trabajando suponiendo que los datos siguen distribuciones normales. En el caso de la geoestadística transformamos los datos para conseguir homogeneidad, y en el caso de la metodología de series temporales trabajamos con modelos autoregresivos.

Los modelos que hemos construido bajo las dos metodologías pueden verse resumidos del siguiente modo:

1. Geoestadística

$$y_t = \beta_0 + f(x) + \beta_1 dist + \beta_2 fila + \beta_3 edad_{t152} + \beta_4 s_t + \epsilon_t$$

con  $f(x) = (f(x_1), \dots, f(x_{52}))$ , una función suave dependiente de la

posición,  $E(\epsilon_t) = \mathbf{0}_{52}$ ,  $Var(y_{it}) = \sigma^2 \forall (i, t)$ , y  $Cov(y_{i,t}, y_{j,m}) = \sigma^2 \cdot \rho(\|i - j\|, |t - m|)$ .

## 2. Modelo $VARI(1, 1)$

$$\Delta y_t = C_0 + c_1 \Delta s_t + \Phi(\Delta y_{t-1} - c_1 \cdot \Delta s_{t-1}) + \zeta_t$$

con  $\zeta_t \sim N(0, Q) i.i.d$

siendo:

$y_t = (y_{1t}, y_{2t}, \dots, y_{52t})'_{52 \times 1}$ , vector con los 52 valores observados en la  $t$ -ésima exploración del paciente (omitimos el índice  $p$ ).

$x_i$   $i$ -ésimo punto del campo visual.

$dist = (distancia(x_1, 0), distancia(x_2, 0), \dots, distancia(x_{52}, 0))'_{52 \times 1}$ ;

$fila = (fila(x_1), fila(x_2), \dots, fila(x_{52}))'_{52 \times 1}$ ;

$\beta_0 = (\beta_{0,1}, \beta_{0,2}, \dots, \beta_{0,52})'_{52 \times 1}$ ;

$s_t = (s_{1t}, s_{2t}, \dots, s_{52t})'_{52 \times 1}$ ;

$\mathbf{1}_{52} = (1, 1, \dots, 1)'_{52 \times 1}$ ;

$\beta_1, \beta_2, \beta_3, \beta_4 \in \Re$

Para empezar el análisis geoestadístico, nos planteamos el estudio de CVs de pacientes sanos. Modelizamos la media utilizando un modelo semi-paramétrico, y utilizamos un modelo separable para modelizar la función de covarianza. La complejidad de la formulación de la media ha hecho que no podamos seguir la metodología geoestadística “clásica”, al no poder estimar la estructura de la media utilizando métodos como el “kriging universal”. Estimamos los parámetros del modelo combinando el método de máxima verosimilitud con el algoritmo backfitting.

A partir de la modelización de CVs de pacientes sanos, construimos intervalos de confianza que nos permiten clasificar cada una de las posiciones

de un CV observado, como “normal” o “enferma”. A partir de esta clasificación, en el capítulo 6 generalizamos el modelo para incluir también CVs de pacientes con glaucoma.

Queremos comentar que esta línea de trabajo no está cerrada. Hemos obtenido la modelización para los umbrales, condicionando a que conocemos el valor de los estatus en cada posición, y estimamos los estatus a partir de intervalos de confianza construidos a partir del modelo para pacientes sanos. Sin embargo, no podemos simular ni predecir el comportamiento futuro de CVs de pacientes con glaucoma, puesto que para ello necesitaríamos introducir un modelo para los estatus, y simular/predecir teniendo en cuenta los dos modelos. Este problema no ha sido estudiado hasta ahora en el ámbito de la geoestadística. Notar que se trataría de una aproximación parecida a la propuesta por Diggle et al. cuando hablan de “Geoestadística basada en el modelo” [50], pero en nuestro caso, los datos observados son gaussianos, y los no observados, no gaussianos.

Dentro de este campo, también como trabajo futuro, nos planteamos continuar trabajando en la siguientes líneas:

En nuestro trabajo la decisión de utilizar un modelo de covarianza separable (frente al no separable) la basamos en el ajuste del variograma espacio temporal por medio de mínimos cuadrados, de forma que elegimos el modelo que proporcionaba mejor ajuste. Recientemente en [56] se proponen contrastes formales que podríamos aplicar a nuestros datos.

La no estacionariedad de los datos la resolvimos mediante una transformación. También recientemente, se han estudiado modelos de covarianza espacio temporal no estacionarios [191] que podríamos intentar ajustar.

El primer problema con que nos encontramos al trabajar con la metodología de series temporales, es que las observaciones deben ser igualmente espaciadas. Al trabajar con esta metodología, no hemos podido empezar construyen-

do un modelo para los pacientes sanos, tal y como hicimos en el estudio geoestadístico, por no tener suficientes datos nuestra base de datos. Notar que tenemos muchos pacientes sanos con una única observación, pero sólo 8 pacientes con 4 observaciones, y ninguno con más. Por todo ello, siguiendo el trabajo de Olsson y Rootzen [154], planteamos un modelo para los umbrales condicionados a los estatus, y otro modelo para explicar la evolución de los estatus.

Para modelizar los estatus, buscamos en la literatura generalizaciones del modelo de Ising al contexto espacio temporal, y trabajamos con los dos modelos que encontramos, utilizando cada uno en el contexto en el que resultaba más adecuado. La modelización de los umbrales se vio bastante complicada al tratarse de un modelo con cambios en régimen. Además, cambios en régimen multivariantes, problema que no ha sido estudiado en la literatura hasta ahora. Para estimar los parámetros, hemos tenido que recurrir a métodos de Montecarlo, y a métodos basados en la pseudoverosimilitud.

Al trabajar en series temporales ya desde el primer momento con el modelo “completo” y resultar éste muy complejo por ser un modelo con cambios en régimen, no hemos podido comprobar algunas hipótesis que asumíamos en el modelo como el orden, o la no cointegración. Nuestra intención es también comprobar estas hipótesis cuando tengamos una base de datos más extensa.

En cuanto a la comparación de ambas metodologías, hemos visto que la Geostatística es mucho más flexible a la hora de modelizar tanto la media del proceso (podemos usar, por ejemplo, modelos aditivos), como la covarianza (notar que por ejemplo si usamos un modelo  $AR(1)$  univariante nos estamos limitando a correlaciones del tipo exponenciales). Sin embargo los modelos de series temporales multivariantes son más descriptivos del proceso y, también a nuestro parecer, la predicción es más sencilla.

Como conclusión final, decir que con esta memoria hemos pretendido

mostrar que los modelos estadísticos espacio temporales pueden resultar muy útiles en el análisis de campos visuales y en particular de campos visuales para la detección del glaucoma. En cuanto a las dos metodologías utilizadas, decir que no nos podemos decantar por ninguna de ellas, ya que cada una tiene sus ventajas e inconvenientes. Nuestra experiencia en este trabajo nos lleva a aconsejar la utilización, ante cada problema concreto, de la metodología que mejor se adapte a sus características y al objetivo que nos hayamos propuesto.

# Bibliografía

- [1] P. Airaksinen and A. Tuulonen, *personal communication, Oulu, Finland*, 261–271, 1988.
- [2] H. Akaike, *Fitting autoregressive models for prediction*, *Annals of the Institute of Statistical Mathematics*, 21:243–247, 1969.
- [3] H. Akaike, *Autoregressive model fitting for control*, *Annals of the Institute of Statistical Mathematics*, 23:163–180, 1971.
- [4] H. Akaike, *A new look at the statistical model identification*, *IEEE Transactions on Automatic Control*, AC-19:716–723, 1974.
- [5] R. Aykroyd and P. Green, *Global and local priors, and the location of lesions using gamma-camera imagery*, *Philosophical Transactions of the Royal Society of London, Ser. A*, 337:323–342, 1991.
- [6] J. Bai and P. Perron, *Estimating and testing linear models with multiple structural changes*, *Manuscript, Dept. of Economics*, MIT, 1995.
- [7] J. Bai and P. Perron, *Testing for and estimation of multiple structural changes*, *Econometrica*, 66:47–79, 1998.
- [8] A. Banerjee, R. Lumsdaine, and J. Stock, *Recursive and sequential tests of the unit root and trend break hypothesis. theory and international evidence*, *Journal of Business and Economic Statistics*, 10:271–287, 1992.

- [9] S. Barker and P. Rayner, *Unsupervised image segmentation using markov random field models*, *Pattern Recognition*, 33:587–602, 2000.
- [10] J. Barry, M. Crowder, and P. Diggle, *Parametric estimation of the variogram*, Tech. Rep. ST-97-06, Lancaster University, UK, 1997.
- [11] S. Bayomog, *Estimation of a markov field dynamic*, *Prépublication du SAMOS, Paris, France*, (52), 1996.
- [12] B. Bengtsson and A. Heijl, *Evaluation of a new perimetric threshold strategy, SITA, in patients with manifest and suspect glaucoma*, *Acta Ophthalmologica Scandinavica*, 76:268–272, 1998.
- [13] B. Bengtsson and A. Heijl, *SITA Fast, a new rapid perimetric threshold test. description of methods and evaluation in patients with manifest and suspect glaucoma*, *Acta Ophthalmologica Scandinavica*, 76:431–437, 1998.
- [14] B. Bengtsson, A. Heijl, and J. Olsson, *Evaluation of a new threshold visual field strategy, SITA, in normal subjects*, *Acta Ophthalmologica Scandinavica*, 76:165–169, 1998.
- [15] B. Bengtsson, J. Olsson, A. Heijl, and H. Rootzen, *A new generation of algorithms for computerized threshold perimetry, SITA*, *Acta Ophthalmologica Scandinavica*, 75:368–375, 1997.
- [16] R. Bennett, *Spatial time series: Analysis-forecasting-control*, Pion, 1979.
- [17] J. Besag, *Statistical analysis of non-lattice lattice data*, *The Statistician*, 24:179–195, 1975.
- [18] J. Besag, *On the statistical analysis of dirty pictures*, *J. R. Statist. Soc. B*, 48:259–302, 1986.

- [19] S. Bochner, *Harmonic Analysis and the Theory of Probability*, University of California Press, Berkeley and Los Angeles, CA, 1955.
- [20] P. Bogaert and G. Christakos, *Spatiotemporal analysis and processing of thermometric data over belgium*, *Jou. Geoph. Res.*, 102(D22):25831–25846.
- [21] G. Box and G. Jenkins, *Time Series Analysis Forecasting and Control*, Holden-Day, San Francisco, 1970.
- [22] P. Brown, P. Diggle, M. Lord, and P. Young, *Space-time calibration of radar rainfall data*, *Appl Statist*, 50:221–241, 2001.
- [23] R. Brown, J. Durbin, and J. Evans, *Techniques for testing the constancy of regression relationships over time*, *Journal of the Royal Statistical Society, Series B*, 37:149–192, 1975.
- [24] A. Buja, T. Hastie, and R. Tibshirani, *Linear smoothers and additive models*, *Annals of Statistics*, 17:453–555, 1989.
- [25] J. Cai and Z.-Q. Liu, *Pattern recognition using markov random field models*, *Pattern Recognition*, 35:725–733, 2002.
- [26] J. Campos, N. Ericsson, and D. Hendry, *Cointegration tests in the presence of structural breaks*, *Journal of Econometrics*, 70:187–220, 1996.
- [27] L. D. Cesare, D. Myers, and D. Posa, *Estimating and modeling space-time correlation structures*, *Statistics & Probability Letters*, 51:9–14, 2001.
- [28] L. D. Cesare, D. Myers, and D. Posa, *Product-sum covariance for space-time modeling: An environmental application*, *Environmetrics*, 12:11–23, 2001.

- [29] B. Chalmond, *An iterative gibbsian technique for reconstruction of m-ary image*, *Pattern Recognition*, 15:115–129, 1989.
- [30] Y. Champagnat, F. Goussard and J. Idier, *Unsupervised deconvolution of sparse spike train using stochastic approximation*, *IEEE Trans. Signal Processing*, 44, 1996.
- [31] R. Chellappa and A. Jain, *Markov Random Fields. Theory and Application*, Academic Press, Inc., 1991.
- [32] R. Chen and S. Tsay, *Nonlinear additive ARX models*, *American Statistical Association*, 88(423):955–967, 1993.
- [33] G. Chow, *Tests of equality between sets of coefficients in two linear regressions*, *Econometrica*, 28:591–605, 1960.
- [34] G. Christakos, *Modern Spatiotemporal Geostatistics*, Oxford University Press, New York, 2000.
- [35] G. Christakos and D. Hristopulos, *Spatiotemporal environment health modelling: A tractatus stochasticus*, Kluwer, Boston, 1998.
- [36] G. Christakos, D. Hristopulos, and P. Bogaert, *On the physical geometry concept at the basis of space/time geostatistical hydrology* *Advances in Water Resources*, 23:799–810, 2000.
- [37] C. Chu and H. White, *A direct test for a changing trend*, *Journal of Business and Economic Statistics*, 10:289–299, 1992.
- [38] N. Cressie, *Statistics for spatial data*, Wiley Series in Probability and Mathematical Statistics, 1993.
- [39] N. Cressie and H. Huang, *Classes of nonseparable, spatio-temporal stationary covariance functions*, *JASA*, 94:1330–1340, 1999.

- [40] N. Cressie and J. Majure, *Spatio-temporal statistical modeling of livestock waste in streams*, *Journal of Agricultural, Biological, and Environmental Statistics*, 2:24–47, 1997.
- [41] N. Davies, C. Triggs, and P. Newbold, *Significance levels of the box-pierce portmanteau statistics in finite samples*, *Biometrika*, 64:517–522, 1977.
- [42] A. Davison and D. Hinkley, *Bootstrap Methods and their Applications*, Cambridge University Press, 1997.
- [43] G. de Fouquet, *Simulation conditionnelle de fonctions aléatoires: cas gaussien stationnaire et schéma linéaire*, tech. rep., Cours C-151, Ecole des Mines de Paris, November 1993.
- [44] A. Dempster, N. Laird, and D. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, *J.R.S.S., Series B*, 39:1–38, 1977.
- [45] D. Dickey and W. Fuller, *Distribution of the estimators for autoregressive time series with unit root*, *Journal of the American Statistical Association*, 74:427–431, 1979.
- [46] J. Diebolt and E. Ip, *Stochastic EM: method and application*, in *Markov Chain Monte Carlo in practice*, W. Gilks, S. Richardson, and D. J. Spiegelhalter, eds., Chapman and Hall, London, 1996, 259–273.
- [47] P. Diggle, *Statistical analysis of spatial point patterns*, Academic Press, 1983.
- [48] P. Diggle, K. Liang, and S. L. Zeger, *Analysis of longitudinal data*, no. 13, Oxford statistical science series, 1994.

- [49] P. Diggle and P. Ribeiro, *Model-based geostatistics*, in 14th. SINAPE-Simpósio Nacional de Probabilidade e Estatística., Associação Brasileira de Estatística., 2000.
- [50] P. Diggle, J. Tawn, and R. Moyeed, *Model based geostatistics. (with discussion)*, *Applied Statistics*, 47:299–350, 1998.
- [51] J. Domingo, G. Ayala, A. Simó, E. D. Ves, L. Martínez-Costa, and P. Marco, *Irregular motion recovery in fluorescein angiograms* *Pattern Recognition Letters*, 18(8):805–821, 1997.
- [52] J. Dufour, *Recursive stability analysis of linear regression relationships* *Journal of Econometrics*, 19:31–76, 1982.
- [53] R. Engle and C. Granger, *Co-integration and error correction: Representation, estimation and testing*, *Econometrica*, 55:251–276, 1987.
- [54] R. Engle and B. Yoo, *Forecasting and testing in cointegrated systems* *Journal of Econometrics*, 35:143–159, 1987.
- [55] N. Fountis and D. Dickey, *Testing for a unit root nonstationarity in multivariate autoregressive time series* *The Annals of Statistics*, 17:419–428, 1989.
- [56] M. Fuentes, *Testing for separability of spatial-temporal covariance functions*, tech. rep., North Carolina State University, Institute of Statistics Mimeo Series 2545, 2003.
- [57] W. Fuller, *Introduction to Statistical Time Series* New York: John Wiley, 1976.
- [58] B. S. G. Huerta and J. Stroud, *Space-time analysis of mexico city ozone levels*, *To appear*.

- [59] A. Gelfand and B. Carlin, *Maximum-likelihood estimation for constrained- or missing-data models*, *Canadian Journal of Statistics*, 21:303–311, 1993.
- [60] D. Geman, *Bayesian image analysis by adaptive annealing*, in IGARSS'85, Amherst, USA, Octubre 1985, Springer-Verlag, 269–277.
- [61] S. Geman and D. Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*, *IEEE Trans. PAMI*, 6:721–741, 1984.
- [62] S. Geman and D. McClure, *Statistical methods for tomographic image reconstruction*, in Proc. 46th Sess. Inst. Stat. Inst. Bulletin ISI, vol. 52, 1987.
- [63] C. Geyer, *Likelihood inference for spatial point processes*, in Stochastic geometry, O. Barndorff-Nielsen, W. Kendall, and M.Ñ. M. Lieshout, eds., Chapman and Hall/CRC, Boca Raton, FL, 1999, 79–140.
- [64] C. Geyer and E. Thompson, *Constrained monte carlo maximum likelihood for dependent data, (with discussion)*, *Journal of the Royal Statistical Society, Series B*, 54:657–699, 1992.
- [65] W. Gilks, S. Richardson, and D. Spiegelhalter, *Introducing markov chain monte carlo*, in Markov Chain Monte Carlo in Practice, R. S. Gilks W.R. and S. D. J., eds., Chapman and Hall, London, 1996, 1–20.
- [66] T. Gneiting, *Nonseparable, stationary covariance functions for space-time data*, *Journal of the American Statistical Association.*, 97:590–600, 2002.

- [67] T. Gneiting, Z. Sasvári, and M. Schalther, *Analogies and correspondences between variograms and covariance functions*, *Adv. Appl. Prob.*, 33:617–630, 2001.
- [68] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*, Oxford University Press, 1997.
- [69] P. Goovaerts and C. Chiang, *Temporal persistence of spatial patterns for mineralizable nitrogen and selected soil properties*, *Soil Sci. Soc. America Jour.*, 2:372–381, 1993.
- [70] C. Granger, *Prediction with a generalized cost of error function*, *Operations Research Quarterly*, 20:199–207, 1969.
- [71] C. Granger, *Some properties of time series data and their use in econometric model specification*, *Journal of Econometrics*, 16:121–130, 1981.
- [72] C. Granger and P. Newbold, *Forecasting Economic Time Series*, New York: Academic Press, 2nd ed., 1986.
- [73] G. Gravier, M. Sigelle, and G. Chollet, *A markov random field model for automatic speech recognition*, in Proc. 15th Int. Conference on Pattern Recognition, A. Sanfeliu, J. Villanueva, M. Vanrell, R. Alquezar, T. Huang, and J. S. (eds.), eds., vol. 3, 2000, 258–261.
- [74] P. Green, *Bayesian reconstruction from emission tomography data using a modified EM algorithm*, *IEEE Transactions in Medical Imaging*, 9:84–93, 1990.
- [75] A. Gregory, J. Mason, and D. Watt, *Testing for structural breaks in cointegrated relationships*, *Journal of Econometrics*, 71:321–341, 1996.
- [76] P. Guttorp, W. Meiring, and P. Sampson, *A space-time analysis of ground-level ozone data*, *Environmetrics*, 5:241–254, 1994.

- [77] P. Guttorp, P. Sampson, and K. Newman, *Non-parametric estimation of spatial covariance with application to monitoring network evaluation*, in *Statistics in environmental and earth sciences*, A. Walden and P. Guttorp, eds., London, 1992, Edward Arnold, 39–51.
- [78] X. Guyon, *Random fields on a network. Modeling statistics and applications*, Springer-Verlag, 1995.
- [79] X. Guyon and C. Hardouin, *Markov chain field dynamics: models and statistics*, *Statistics*, 13(4):339–363, 2002.
- [80] T. Hainsworth and K. Mardia, *A markov random field restoration of image sequences*, in *Markov random fields. Theory and application*, R. Chellappa and A. Jain, eds., Academic Press Inc, Oxford, 1991, 409–446.
- [81] M. Halley, *The field analyzer primer*, Allergan Humphrey, San Leandro, CA, 1986.
- [82] J. Hamilton, *Rational-expectations econometric analysis of changes in regime. an investigation of the term structure of interest rates*, *Journal of Economic Dynamics and Control*, 12:385–423, 1988.
- [83] J. Hamilton, *A new approach to the economic analysis of nonstationary time series and the business cycle*, *Econometrica*, 57:357–384, 1989.
- [84] J. Hamilton, *Analysis of time series subject to changes in regime*, *Journal of Econometrics*, 45:39–70, 1990.
- [85] J. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey, 1994.
- [86] M. Handcock and J. Wallis, *An approach to statistical spatial-temporal modeling of meteorological fields*, *JASA*, 89(426):368–390, 1994.

- [87] E. Hannan, *Multiple Time Series*, New York: John Wiley, 1970.
- [88] A. Harvey, *Time Series Models*, Prentice Hall, 2nd ed., 1993.
- [89] A. Harvey, *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, 1996.
- [90] J. Haslett and A. Raftery, *Space-time modelling with long memory dependence: assessing ireland's wind power resource*, *Appl. Statist.*, 38(1):1–50, 1989.
- [91] T. Hastie and R. Tibshirani, *Local likelihood estimation*, *Journal of the American Statistical Association*, 82:559–567, 1987.
- [92] T. Hastie and R. Tibshirani, *Generalized additive models*, vol. 43, Chapman & Hall, (Monographs on statistics and applied probability), 1990.
- [93] A. Heijl, *The humphrey field analyzer, construction and concepts* in Proceedings of the 6th International Perimetric Society Meeting, 1984, Heijl and Greve, eds., no. 42 in Documenta Ophthalmol. Proc. series, Dordrecht, 1985, Junk Publishers, 77–84.
- [94] A. Heijl, G. Lindgren, and J. Olsson, *Normal variability of static perimetric threshold values across the central visual field* *Arch Ophthalmol*, 105:1544–1549, 1987.
- [95] A. Heijl, G. Lindgren, and J. Olsson, *The effect of perimetric experience in normal subjects*, *Arch Ophthalmol*, 107:81–86, 1989.
- [96] D. Hendry and A. Neale, *A monte carlo study of the effects of structural breaks on test for unit roots*, in *Economic Structural Change*, P. Hackl and A. Westlung, eds., Springer Verlag, New York, 1991, 95–119.

- [97] D. Higdon, *A process-convolution approach to modeling temperatures in the north atlantic ocean*, *Environ. Ecol. Statist.*, 5:173–190, 1999.
- [98] D. Higdon, J. Swall, and J. Kern, *Non-stationary spatial modeling*, in Bayesian Statistics, J. Bernardo, J. Berger, A. Dawid, and A. Smith, eds., Oxford University Press, Oxford, 1999, 761–768.
- [99] J. Hosking, *The multivariate pormanteau statistic*, *Journal of the American Statistical Association*, 75:602–608, 1980.
- [100] M. Ibáñez and A. Simó, *Bayesian detection of the fovea in eye fundus angiographies*, *Pattern Recognition Letters*, 20:229–240, 1999.
- [101] M. Ibáñez and A. Simó, *Spatio-temporal modelling of perimetric test data*, Tech. Rep. 55, Department of Mathematics. University Jaume I. Castellón. Spain, 2003.
- [102] A. Jalobeanu, L. Blanc-éraud, and J. Zerubia, *Hyperparameter estimation for satellite image restoration using a MCMC maximum-likelihood method*, *Pattern Recognition*, 35:341–352, 2002.
- [103] E. Järpe, *A new ising-type model for spatio temporal interactions* *Manuscript*, Submitted for publication.
- [104] S. Johansen, *Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models*, *Econometrica*, 59:1551–1580, 1991.
- [105] C. Johnson and L. Shapiro, *A comparison of MOBS and staircase test procedures in automated perimetry*, *J Opt Soc Am Tech Dig Ser*, 7:98–102, 1989.
- [106] R. Jones and Y. Zhang, *Models for continuous stationary space-time models*, in Lecture Notes in Statistics. In Modelling Longitudinal

- and Spatially Correlated Data., T. Gregorie, D. Brillinger, P. Diggle, E. Russek-Cohen, W. Warren, and R. Wolfinger, eds., Springer, New York, 1997, 289–298.
- [107] A. Journel, *Geostatistics for conditional simulation of ore bodies*, *Economic Geology*, 69:673–687, 1974.
- [108] L. Joyeux, S. Boukir, B. Besserer, and O. Buisson, *Reconstruction of degraded image sequences. application to film restoration*, *Image and Vision Computing*, 2001.
- [109] R. Kalman, *A new approach to linear filtering and prediction problems*, *Journal of Basic Engineering*, 82:34–45, 1960.
- [110] R. Kalman and R. Bucy, *New results in linear filtering and prediction theory*, *Journal of Basic Engineering*, 83:95–108, 1961.
- [111] C. Kao and S. Ross, *A CUSUM test in the linear regression model with serially correlated disturbances*, tech. rep., Syracuse University, 1992.
- [112] N. Karssemeijer, *Application of bayesian methods to segmentation in medical images*, in *Lecture Notes in Statistics. Stochastic Models, Statistical Methods, and Algorithms in Image Analysis*, F. Barone, P. and M. Piccioni, eds., vol. 74, Springer-Verlag, 1992.
- [113] Z. Kato, J. Zerubia, and M. Berthod, *Unsupervised parallel image classification using a hierarchical markovian model*, Tech. Rep. Research Report No 2528, 1995.
- [114] Z. Kato, J. Zerubia, and M. Berthod, *Unsupervised parallel image classification using markovian models*, *Pattern Recognition*, 32:591–604, 1999.

- [115] G. Kauerman and J. Opsomer, *Local likelihood estimation in generalized additive models*, Preprint 00-8. Department of Statistics, Iowa State University, 1998.
- [116] J. Kelsall and P. Diggle, *Spatial variation in risk of disease: a nonparametric binary regression approach*, *Appl. Statist.*, 47(4):559–573, 1998.
- [117] I. Kim, *Detecting the number of structural breaks*, *Economics Letters*, 57:145–148, 1997.
- [118] I. Kim and G. Maddala, *Multiple structural breaks and unit roots in exchange rates*, Paper presented at the Econometric Society Meeting at New Orleans, 1991.
- [119] D. Krige, *A statistical approach to some mine valuations and allied problems at the witwatersrand*, Master's thesis, University of Witwatersrand, 1951.
- [120] H. Krolzig, *Markov Switching vector autoregressions. Modelling, statistical inference and application to business cycle analysis*, Lecture Notes in Economics and Mathematical Systems, Springer, 1997.
- [121] H. Krolzig, *Predicting markov-switching vector autoregressive processes*, *Journal of Forecasting*, to appear.
- [122] P. Kyriakidis and A. Journel, *Geostatistical space-time models: a review*, *Mathematical Geology*, 31:651–684, 1999.
- [123] S. Lakshamanan and Derin, *Simultaneous parameter estimation and segmentation of gibbs random fields using simulated annealing* *IEEE Trans. PAMI*, 11, 1989.
- [124] C. Lantuéjoul, *Geostatistical Simulation. Models and Algorithms*, Springer-Verlag, 2002.

- [125] M. Lavine and S. Lozier, *A markov random field spatio-temporal analysis of ocean temperature*, *Environ. Ecol. Statist.*, 6:249–273, 1999.
- [126] S. Li, *Markov Random Field Modeling in Computer Vision*, Springer, 1995.
- [127] W. Li and A. McLeod, *Distribution of the residual autocorrelations in multivariate ARMA time series models*, *Journal of the Royal Statistical Society, B*, 43:231–239, 1981.
- [128] O. Linton, *Efficient estimation of generalized additive nonparametric regression models*, *Econometric Theory*, 16:502–523, 2000.
- [129] O. Linton and J. Nielsen, *A kernel method of estimating structured nonparametric regression based on marginal integration*, *Biometrika*, 82:93–100, 1997.
- [130] G. Ljung and G. Box, *On a measure of lack of fit in time series models* *Biometrika*, 65:297–303, 1978.
- [131] H. Lucke, *Improved acoustic modeling for speech recognition using 2D markov random fields*, in ICASSP'95. Detroit, MI, May 1995, 540–543.
- [132] H. Lutkepohl, *Introduction to Multiple Time Series Analysis*, Springer-Verlag, New York, 2nd ed., 1993.
- [133] C. Ma, *Spatio temporal covariance functions generated by mixtures* *Mathematical Geology*, 34(8), 2002.
- [134] C. Ma, *Families of spatio-temporal stationary covariance models* *J. Stat. Plann. Infer.*, to appear.
- [135] C. Ma, *Spatio-temporal stationary covariance models* *J. Multivariate Analysis*, to appear.

- [136] G. Maddala and I.-M. Kim, *Unit roots, cointegration and structural change*, Themes in modern econometrics, Cambridge University Press, 1998.
- [137] K. Mardia, *Measures of multivariate skewness and kurtosis with applications*, *Biometrika*, 57:519–520, 1970.
- [138] K. Mardia and C. R. Goodall, *Spatial-temporal analysis of multivariate environmental data*, in *Multivariate environmental statistics*, G. Patil and C. Rao, eds., Elsevier, 1993, 347–386.
- [139] J. Mateos, K. Katsaggelos, and M. R., *A bayesian aproach for the estimation and transmission of regularization parameters for reducing blocking artifacts*, *IEEE Trans. Image Processing*, 9:1200–1215, 2000.
- [140] B. Matérn, *Spatial variation*, tech. rep., Meddelanden fran Statens Skogsforsningsinstitut, Stockholm, 1960.
- [141] B. Matérn, *Spatial Variation*, Lecture Notes in Statistics, Springer-Verlag, New York, 2nd ed., 1980. First edition published in Meddelanden fran Statens Skogsforskningsinstitut, Band 49, No 5, 1960.
- [142] G. Matheron, *Traité de géostatistique appliquée. Tome I. Mémoires du Bureau de reserches Géologiques et Minières*, Editions Technip, Paris, 1962.
- [143] G. Matheron, *The intrinsec random functions and their applications* *Advances in Applied Probability*, 5:439–468, 1998.
- [144] R. Meinhold and N. Singpurwalla, *Understanding the kalman filter*, *The American Statistician*, 37(2):123–127, 1983.
- [145] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, *Markov random field and fuzzy logic modeling in sonar imagery: Application to the*

- classification of underwater floor, Computer Vision and Image Understanding: CVIU*, 79(1):4–24, 2000.
- [146] R. Molina, K. Katsaggelos, and J. Mateos, *Bayesian and regularization methods for hyperparameter estimation in image restoration*, *IEEE Trans. Image Processing*, 8:231–246, 1999.
- [147] F. Mouria-Beji, *A hierarchical bayesian model for continuous speech recognition*, *Pattern Recognition Letters*, 23:773–781, 2002.
- [148] M. Muller, *Estimation and testing in generalized partial linear models - a comparative study*, *Statistics and Computing*, to appear, 2002.
- [149] D. Peña, *Estadística. Modelos y métodos*, vol. 2. Modelos lineales y series temporales, Alianza Editorial, 2nd ed., 1989.
- [150] S. Nielsen, *On simulated EM algorithms*, *J. of Econometrics*, 96:267–292, 2000.
- [151] X. Niu, *Nonlinear additive models for environmental time series, with applications to ground-level ozone data analysis* *Journal of the American Statistical Association*.
- [152] H. Noda, M. Shirazi, and E. Kawaguchi, *MRF-based texture segmentation using wavelet decompose images*, *Pattern Recognition*, 35, 2002.
- [153] J. Olsson, *Thesis: Statistics in perimetry. department of mathematical statistics, Lund University*, 1991.
- [154] J. Olsson and H. Rootzen, *An image model for quantal response analysis in perimetry*, *Scandinavian Journal of Statistics*, 21:375–387, 1994.
- [155] J. Opsomer, *Asymptotic properties of backfitting estimators*, *Journal of Multivariate Analysis*, 73:166–179, 2000.

- [156] J. Opsomer and D. Ruppert, *Fitting a bivariate additive model by local polynomial regression*, *Annals of Statistics*, 25:186–211, 1997.
- [157] P. Perron, *The great crash, the oil price shock and the unit root hypothesis*, *Econometrica*, 57:1361–1401, 1989.
- [158] P. Phillips, *Bayesian model selection and prediction with empirical applications*, *Economics Letters*, 69:289–365, 1995. with comments.
- [159] P. Phillips and S. Ouliaris, *Asymptotic properties of residual based tests for cointegration*, *Econometrica*, 58:165–193, 1990.
- [160] W. Pieczynski, *Statistical image segmentation*, in GKPO'92, May 1992.
- [161] W. Ploberger, W. Kramer, and R. Alt, *A modification of the CUSUM test in linear regression model with lagged dependent variables*, *Empirical Economics*, 14:65–75, 1989.
- [162] D. Poskitt and A. Tremayne, *Diagnostic tests for multiple time series models*, *Annals of Statistics*, 10:114–120, 1982.
- [163] M. Priestley, *Non-linear and non-stationary time series analysis*, Academic Press, second ed., 1997.
- [164] R. Quandt, *The estimation of parameters of linear regression system obeying two separate regimes*, *Journal of the American Statistical Association*, 55:873–880, 1958.
- [165] F. Quintana, J. Liu, and G. del Pino, *Monte carlo EM with importance reweighting and its applications in random effects models*, *Computation Statistics and Data Analysis*, 29:429–444, 1999.
- [166] P. Rappoport and L. Reichlin, *Segmented trends and nonstationary time series*, *Economic Journal*, 99, (supplement):168–177, 1989.

- [167] G. Reinsel, *Elements of Multivariate Time Series Analysis* Springer Series in Statistics, Springer, second ed., 1997.
- [168] B. Ripley, *Spatial Statistics*, Chapman and Hall, New York, 1981.
- [169] I. Rodríguez-Iturbe and J. Mejía, *The design of rainfall networks in time and space*, *Water Resources Res.*, 10(4):713–728, 1974.
- [170] X. Roggen, K. Herman, L. V. Malderen, M. Devos, and W. Spileers, *Different strategies for humphrey automated perimetry; FASTPAC, SITA standar ans SITA Fast in normal subjects and glaucoma patients*, *Bull Spc. belge Ophtalmol.*, 279:23–33, 2001.
- [171] S. Rouhani and D. Myers, *Problems in space-time kriging of geohydrological data*, *Mathematical Geology*, 22:611–623, 1990.
- [172] P. Ruud, *Extensions of estimation methods using the EM algorithm*, *Journal of Econometrics*, 49:305–341, 1991.
- [173] F. Salzenstein and W. Pieczynski, *Parameter estimation in hidden fuzzy markov random fields and image segmentation*, *CVGIP: Graphical Models and Image Processing*, 59:205–220, 1997.
- [174] P. Sampson, D. Damian, and P. Guttorp, *Advances in modeling and inference for environmental processes with nonstationary spatial covariance*, *NRCSE Technical Report Series (NRCSE-TRS No 061)*, 2001.
- [175] P. Sampson and P. Guttorp, *Nonparametric estimation of nonstationary spatial covariance structure*, *JASA*, 87(417):108–119, 1992.
- [176] B. Sansó and L. Guenni, *Venezuelan rainfall data analysed by using a bayesian space-time model*, *Appl. Statist*, 48(3):345–362, 1999.
- [177] B. Sansó and L. Guenni, *A non-stationary multi-site model for rainfall*, *J. Am. Statist. Ass.*, 95:1089–1100, 2000.

- [178] S. Saquib, C. Bouman, and K. Sauer, *M.l. parameter estimation for markov random fields, with applications to bayesian tomography IEEE Trans. Image Processing*, 7, 1998.
- [179] M. Schalter, *Introduction to positive definite functions and to unconditional simulation of random fields*, Tech. Rep. ST-99-10, Department of Mathematics and Statistics. Faculty of Applied Sciences, Lancaster University, UK, 1999.
- [180] G. Seber, *The linear hypothesis: a general theory*, Griffin's Statistical Monographs and Courses, No 19, second ed., 1980.
- [181] G. Seber, *Multivariate observations*, Wiley Series in probability and mathematical Statistics, 1984.
- [182] L. Shapiro and C. Johnson, *Quantitative evaluation of manual kinetic perimetry using computer simulation*, *Applied Optics*, 29(10):1445–1450, 1990.
- [183] L. Shapiro, C. Johnson, and L. Kennedy, *KRAKEN: A computer simulation procedure for static, kinetic, suprathreshold static and heuristic perimetry*, in VIII International Visual Field Symposium, Documenta Ophthalmologica Proceeding Series, Vancouver B.C., 1998, in press.
- [184] M. Shinozuka and C. Jan, *Digital simulation of random processes and its applications*, *J. of Sound and Vib.*, 25(1):111–128, 1972.
- [185] Y. Simard and D. Marcotte, *Assessing similarities and differences among maps: A study of temporal changes in distribution of northern shrimp (*pandalus borealis*) in the gulf of st. lawrence*, in A. Soares, ed., *Geostatistics Tróia '92*, 2. Kluwer Academic Publ., Dordrecht:865–874, 1993.

- [186] A. Simó and E. D. Ves, *Segmentation of macular fluorescein angiographies*, *Pattern Recognition*, 34(4):795–809, 2001.
- [187] C. Sims, *Macroeconomics and reality*, *Econometrica*, 48:1–48, 1980.
- [188] S. Sperlich, O. Linton, and W. Hardle, *Integration and backfitting methods in additive models - finite sample properties and comparison*, *Test*, 8:419–459, 1999.
- [189] P. Spry, A. Bates, C. Johnson, and B. Chauhan, *Simulation of longitudinal threshold visual field data*, *Investigative Ophthalmology & Visual Science*, 41(8):2192–2200, 2000.
- [190] M. Stein, *Interpolation of Spatial Data*, Springer-Verlag. New York, 1999.
- [191] M. Stein, *Space-time covariance functions*, Tech. Rep. 4, The University of Chicago. Center for Integrating Statistical and Environmental Science (CISES), 2003.
- [192] J. Stock and M. Watson, *Testing for common trends*, *Journal of the American Statistical Association*, 83:1097–1107, 1988.
- [193] M. Teles, *Bayesian estimation of ventricular contours in angiographic images*, *IEEE Trans. on Medical Imaging*, 11(3):416–429, 1992.
- [194] STATPAC, *Users guide*, Allergan Humphrey, San Leandro CA, 1986.
- [195] J. Tuckey, *Exploratory Data Analysis*, Addison-Wesley. Reading, MA, 1977.
- [196] A. Turpin, A. McKendrick, C. Johnson, and A. Vingrys, *Development of efficient threshold strategies for frequency doubling technology perimetry using computer simulation*, *Invest Ophthalmol Vis Sci*, 43(2):322–331, 2002.

- [197] E. Vesti, P. Spry, B. Chauhan, and C. Johnson, *Sensitivity differences between real-patient and computer-simulated visual fields*, *Journal of Glaucoma*, 11:35–45, 2002.
- [198] V. Vyas and G. Christakos, *Spatiotemporal analysis and mapping of sulfate deposition data over eastern USA*, *Atmosph Environment*, 31(21):3623–3633, 1997.
- [199] H. Wackernagel, *Multivariate Geostatistics*, Springer-Verlag. New York, second ed., 1998.
- [200] L. Waller, B. Carlin, H. Xia, and A. Gelfand, *Hierarchical spatio-temporal mapping of disease rates*, *JASA*, 92(428):607–617, 1997.
- [201] I. Weir, *Fully bayesian reconstruction from single photon emission computed tomography data*, *Journal of the American Statistical Association*, 92(437):49–60, 1997.
- [202] I. Weir and P. Green, *Modeling data from single-photon emission computed tomography*, in *Statistics and Images*, K. Mardia, ed., 1994, 313–338.
- [203] P. Whittle, *On stationay processes in the plane*, *Biometrika*, 41:434–449, 1954.
- [204] J. Wild, I. Pacey, S. Hancock, and I. Cunliffe, *Between algorithm, between individual differences in normal perimetric sensitivity: Full threshold, FASTPAC, and SITA*, *Invest. Ophthalmol. Vis. Sci.*, 40:1152–1161, 1999.
- [205] G. Winkler, *Image Analysis, Random fields Models and Dynamic Monte Carlo Methods*, Springer-Verlag, 1996.

- [206] K. Yao, M. Mignotte, C. Collet, P. Galerne, and G. Burel, *Unsupervised segmentation using a self-organizing map and a noise model estimation in sonar imagery*, *Pattern Recognition*, 33:1575–1584, 2000.
- [207] L. Younes, *Estimation and annealing for gibbsian fields*, *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques*, 24:269–294, 1988.
- [208] L. Younes, *Parameter estimation for imperfectly observed gibbs fields and some comments on chalmond's EM gibbsian algorithm*, no. 47 in LNS, Springer, 1991, 240–258.
- [209] E. Zivot and D. Andrews, *Further evidence on the great crash, the oil-price shock and the unit root hypothesis*, *Journal of Business and Economic Statistics*, 10:251–287, 1992.