

**Series Monográficas
del Instituto Nacional de Meteorología**

(2004)

Redes Probabilísticas y Neuronales en las Ciencias Atmosféricas

José M. Gutiérrez

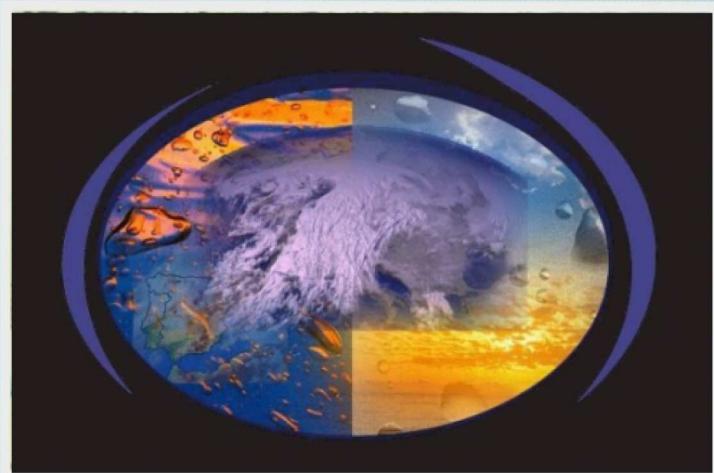
Rafael Cano

Antonio S. Cofiño

Carmen M. Sordo

Publicado por el Ministerio de Medio Ambiente, Madrid

ISBN: 84-8320-281-6



**Universidad de Cantabria
Instituto Nacional de Meteorología**



**MINISTERIO
DE MEDIO AMBIENTE**

DIRECCIÓN GENERAL
DEL INSTITUTO NACIONAL
DE METEOROLOGÍA

Índice general

Prefacio	ix
Acrónimos y Terminología Utilizados	xii
I Introducción	1
1. Modelos y Datos Atmosféricos	3
1.1. Introducción	3
1.2. Las Ecuaciones de la Atmósfera	4
1.2.1. Filtrado de Soluciones Triviales	7
1.3. Resolución Numérica de las Ecuaciones	8
1.3.1. Asimilación de Datos	8
1.3.2. Parametrización de Procesos Físicos	10
1.4. Variabilidad Atmosférica	11
1.4.1. Escalas Atmosféricas	11
1.4.2. Patrones de Oscilación Atmosférica y Teleconexión	12
1.5. Tipos de Modelos Numéricos del Tiempo	14
1.5.1. Modelos Globales de Circulación General	15
1.5.2. Modelos Regionales	16
1.5.3. Modelos Mesoscales	17
1.6. No Linealidad y Caos en la Atmósfera	18
1.6.1. Predecibilidad y Predicción por Conjuntos	21
1.7. Estado Actual de la Predicción Operativa	24
1.7.1. Nowcasting. Predicción Inmediata	24
1.7.2. Predicción a Corto Plazo	24
1.7.3. Predicción a Medio Plazo por Conjuntos	26
1.7.4. Predicción Estacional por Conjuntos	28
1.7.5. Predicción Climática	30

1.7.6. Tipos de Predicción: Determinista y Probabilística	30
1.8. Futuro de la Predicción Numérica	31
1.9. Datos Climatológicos y Meteorológicos	31
1.9.1. Redes de Observación	32
1.9.2. <i>MeteoLab</i> : Observaciones	34
1.9.3. Datos Paleoclimáticos	38
1.9.4. Simulaciones de Modelos Numéricos	39
1.9.5. <i>MeteoLab</i> : Patrones Atmosféricos	41
II Técnicas Estadísticas Estándar	45
2. Técnicas Estadísticas. Análisis y Exploración de Datos	47
2.1. Introducción	47
2.2. Probabilidad y Variables Aleatorias	48
2.2.1. Funciones de Probabilidad	49
2.2.2. Probabilidades Conjuntas	52
2.2.3. Probabilidades Marginales y Condicionadas	54
2.3. Generadores de Tiempo (Weather Generators)	56
2.3.1. Generadores Estocásticos de Precipitación	56
2.3.2. Modelos de Intensidad del Evento	59
2.4. Dependencia e Independencia de Variables	61
2.4.1. Correlación e Información Mutua	64
2.5. Componentes Principales y EOF	68
2.5.1. <i>MeteoLab</i> : Componentes Principales y EOF	74
2.5.2. Elección del Número de Componentes	79
2.5.3. Efectos de la Escala Temporal	81
2.6. Técnicas de Agrupamiento	82
2.6.1. Técnicas Jerárquicas	83
2.6.2. Técnicas Particionales	85
2.6.3. <i>MeteoLab</i> : Técnicas de Agrupamiento	89
3. Técnicas Estadísticas. Modelización y Predicción	91
3.1. Introducción	91
3.2. Downscaling Estadístico	92
3.3. Técnicas Estadísticas para Series Temporales	94
3.3.1. Modelos Autoregresivos	95
3.3.2. <i>MeteoLab</i> : Modelos Autoregresivos	96
3.3.3. Series Caóticas. Técnicas de Inmersión	97
3.3.4. Modelos de Markov	100
3.4. Regresión	101
3.4.1. Regresión Lineal	101
3.4.2. Regresión Lineal Múltiple	104
3.4.3. Regresión No Lineal	106
3.4.4. <i>MeteoLab</i> : Regresión	107
3.5. Correlación Canónica	109

3.5.1. Predicción con Correlación Canónica	111
3.5.2. <i>MeteoLab</i> : Correlación Canónica	111
3.5.3. Efectos de la Escala Temporal	115
3.6. Aplicación al Downscaling. MOS y Perfect Prog	117
3.6.1. Modelos Globales y Locales. Dependencia Espacial	118
3.7. Técnicas Locales de Análogos	120
3.7.1. Influencia del Período de Reanálisis Utilizado	121
3.7.2. <i>MeteoLab</i> : Técnicas Locales de Análogos	122
3.8. Comparación de Técnicas en el Corto Plazo	124
 III Técnicas de Aprendizaje Automático	 127
4. Redes Probabilísticas (Bayesianas)	129
4.1. Introducción	129
4.2. Sistemas Inteligentes	130
4.3. Sistemas Inteligentes Probabilísticos	132
4.4. Redes Probabilísticas	135
4.4.1. Codificación de Dependencias. Criterios de Separación	135
4.4.2. Factorización de la Probabilidad	137
4.5. Redes Bayesianas. Definición y Tipos	138
4.6. Razonamiento Probabilístico. Inferencia	139
4.6.1. <i>MeteoLab</i> : Modelo de Meteoros (I)	140
4.7. Algoritmos de Aprendizaje Automático	146
4.7.1. Algoritmo K2	147
4.7.2. <i>MeteoLab</i> : Modelo de Meteoros (II)	148
4.7.3. El Algoritmo de Aprendizaje B	150
4.7.4. Algoritmos de Aprendizaje Local. LK2 y LB	151
4.7.5. <i>MeteoLab</i> : Modelado de Meteoros (III)	153
4.8. Aplicaciones	155
4.8.1. Modelos Conceptuales	156
4.8.2. Dependencia Espacial en la Red Pluviométrica	160
4.8.3. Generadores de Tiempo (Weather Generators)	162
4.8.4. Predicción Probabilística Local (Downscaling)	165
4.8.5. Aplicación al Relleno de Lagunas	169
5. Redes Neuronales	175
5.1. Introducción	175
5.2. Estructura de una Red Neuronal	176
5.3. Aprendizaje y Validación	178
5.4. Perceptrones (Redes de una Capa)	180
5.4.1. Algoritmo de Aprendizaje. La Regla Delta	180
5.4.2. Mejoras y Modificaciones	182
5.5. Redes Multi-Capa	186
5.5.1. El Algoritmo de Retro-propagación	187
5.5.2. <i>MeteoLab</i> : El Sistema de Lorenz	193

5.6.	Redes de Función de Base Radial	194
5.6.1.	Aprendizaje de los Parámetros	195
5.6.2.	<i>MeteoLab</i> : Funciones de Base Radial	196
5.7.	Redes Competitivas	198
5.8.	Redes Auto-Organizativas (SOM)	201
5.8.1.	<i>MeteoLab</i> : Predicción con Redes Neuronales	206
5.9.	Predicción a Plazo Medio con SOM	209
5.10.	Predicción Estacional	215
IV	Apéndices	223
6.	Implementación Operativa. El Sistema PROMETEO	225
6.1.	Introducción	225
6.2.	Sistema Operativo en el INM	226
6.2.1.	Configuración e Inicialización del Sistema	227
6.2.2.	Explotación Operativa	229
6.2.3.	Acceso Web a las Predicciones	230
6.3.	Casos de Estudio	231
6.4.	Validación. Retroalimentación del Sistema	235
6.4.1.	Validación Operativa de Prometeo	236
7.	Verificación de Sistemas de Predicción Probabilística	241
7.1.	Introducción	241
7.2.	Tipos de Predicciones	242
7.2.1.	Predicciones Deterministas	243
7.2.2.	Predicciones Probabilísticas	243
7.3.	Aspectos de la Calidad de una Predicción	244
7.4.	Verificación de Predicciones Deterministas	248
7.4.1.	Predicciones Binarias	250
7.5.	Verificación de Predicciones Probabilísticas	253
7.5.1.	Brier Score	253
7.5.2.	Brier Skill Score	255
7.6.	Predicciones Binarias Probabilísticas	256
7.6.1.	Curvas ROC (Relative Operating Characteristics)	256
7.6.2.	Valor Económico	259
7.7.	<i>MeteoLab</i> : Validación	262
Bibliografía	265	
Indice de Palabras	276	

Prefacio

Durante las últimas décadas se ha producido un crecimiento vertiginoso de la capacidad de cálculo y almacenamiento de los ordenadores, así como un abaratamiento de esta tecnología. Estos hechos han desencadenado un crecimiento vertiginoso de cantidad de información que se almacena diariamente en distintas bases de datos (simulaciones científicas, observaciones, etc.), y ha hecho más fácil y rápido el acceso masivo a las mismas (por ejemplo, en tiempo real a través de Internet o de GRID; ver Foster and Kesselman (2003)). Estas bases de datos contienen conocimiento útil para numerosos problemas, y se requieren técnicas especiales para explorarlas y analizarlas de forma eficiente. En el ámbito de la Meterología se han utilizado las técnicas estadísticas de forma sistemática para abordar distintos problemas de modelización y predicción a partir de observaciones y/o de salidas de modelos numéricos (ver, por ejemplo, Ayuso, 1994). Sin embargo, el gran volumen de datos del que se dispone hoy día hace que estas técnicas resulten inapropiadas en numerosos problemas de interés.

La necesidad de métodos eficientes y automáticos para explorar bases de datos ha motivado un rápido avance en áreas del ámbito de la *Estadística* y, más concretamente, de la *Inteligencia Artificial*. El objetivo es desarrollar *métodos de aprendizaje* que operen de forma automática a partir de un conjunto de datos, preprocesando de forma rápida y fiable la información para capturar distintos patrones de conocimiento (reglas, grafos de dependencias, modelos neuronales, etc.) que sean apropiados para resolver un problema; de esta forma se compacta la información disponible, haciéndola manejable. Se trata también de que estas técnicas operen de forma automática, precisando de la mínima intervención humana. Durante los últimos años se ha producido un notable desarrollo de este tipo de herramientas en distintas áreas de conocimiento: aprendizaje automático (machine learning), computación neuronal, estadística, computación paralela, bases de datos, etc. En la década de los noventa se popularizó el término *Minería de Datos* (Data Mining) para referirse a este área interdisciplinar, y *Aprendizaje Estadístico* (Statistical Learning) para referirse a los métodos estadísticos desarrollados con el objetivo de facilitar el aprendizaje automático.

Dos de las técnicas más populares en este campo son las redes probabilísticas y las redes neuronales. Las redes probabilísticas son modelos apropiados

para el tratamiento de problemas con incertidumbre, y utilizan técnicas estadísticas modernas de inferencia y estimación para ajustar los parámetros a los datos y obtener conclusiones en base a los modelos resultantes. Por otra parte, las redes neuronales son modelos de inspiración biológica que “aprenden” tratando de reproducir la realidad codificada en un conjunto de datos. Estas técnicas responden a dos líneas de investigación distintas dentro del ámbito del aprendizaje automático y de la minería de datos, proporcionando una visión global de este campo.

Este libro se propone presentar una visión actual de estas herramientas y mostrar su relación con las técnicas estadísticas estándar utilizadas en las ciencias atmosféricas. Para ello, se analizarán desde ambas perspectivas distintos problemas de interés, como la predicción local de fenómenos meteorológicos, la mejora de resolución de las salidas de los modelos numéricos (downscaling), la generación de series climatológicas mediante simuladores estocásticos de tiempo (weather generators), el relleno de lagunas en los datos, la predicción por conjuntos, etc.

Este libro está organizado en tres partes. La primera parte es introductoria y tiene un único capítulo (Cap. 1), que presenta una introducción general a la predicción numérica del tiempo, analizando los distintos tipos de modelos utilizados hoy día en la predicción operativa, incluyendo las técnicas de predicción por conjuntos. En este capítulo también se describen los datos disponibles relativos a observaciones y reanálisis de modelos numéricos.

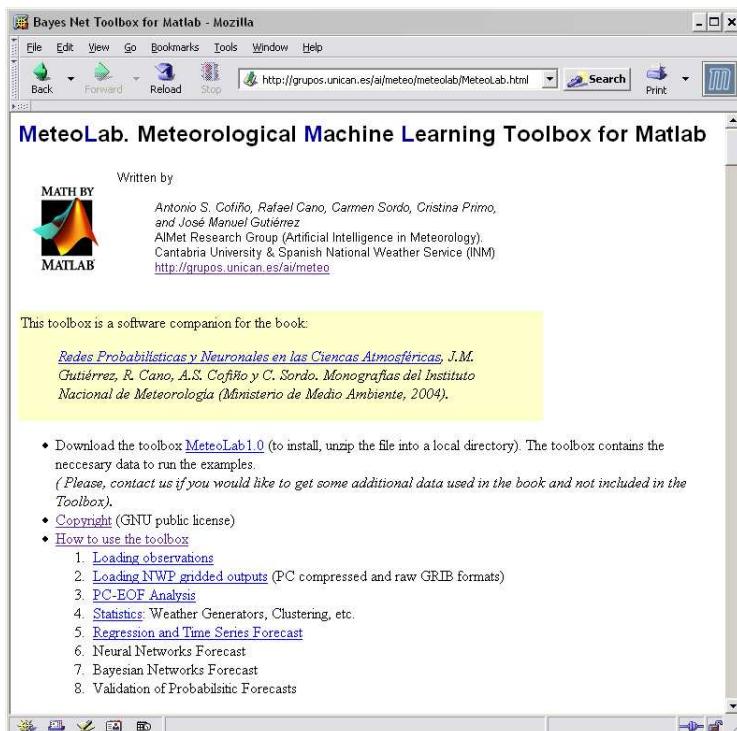
La segunda parte del libro está dedicada a los métodos estadísticos estándar que han sido profusamente utilizados en las ciencias atmosféricas; esta parte no trata de ser una guía completa del tema, pues existen excelentes textos para ello (ver, por ejemplo von Storch and Zwiers, 1999), sino describir de forma breve y práctica los métodos más habituales y los problemas donde se aplican. Los Capítulos 2 y 3 describen las técnicas estadísticas estándar de exploración y análisis de datos, y de modelización y predicción, respectivamente.

La tercera parte del libro se centra en las técnicas de aprendizaje automático, que generalizan algunas de las técnicas estándar anteriores. Los Capítulos 4 y 5 describen el fundamento teórico y las aplicaciones prácticas de dos técnicas recientes: redes probabilísticas y redes neuronales. Estas técnicas ilustran los nuevos conceptos y métodos de aprendizaje supervisado y no supervisado que se vienen aplicando desde la década de los ochenta para trabajar con problemas complejos de forma eficiente y sencilla.

La última parte incluye dos apéndices. En el primero de ellos (Cap. 6) se describe una aplicación operativa de predicción basada en los métodos descritos en el libro, y que está operativa en el INM. Por último, en el Cap. 7, se analiza el problema de la validación de predicciones probabilísticas, donde se introducen algunas medidas que son utilizadas a lo largo del libro.

Al final del libro se incluye una extensa bibliografía para consultas adicionales, así como un índice para la búsqueda de palabras clave en los temas tratados. La página Web del grupo (grupos.unican.es/ai/meteo) ofrecerá información actualizada de los distintos temas tratados en el libro.

Como complemento a este libro se ha desarrollado una Toolbox de Matlab (*MeteoLab. Meteorological Machine Learning Toolbox for Matlab*) para la aplicación de técnicas estadísticas y de aprendizaje automático en problemas de meteorología y climatología. Los programas incluidos en la Toolbox permiten cargar y visualizar de forma sencilla datos de redes de observaciones y salidas de modelos numéricos (reanálisis y predicciones). Posteriormente, se pueden analizar los datos aplicando técnicas estadísticas estándar como componentes principales (PC/EOF) para la eliminación de redundancia, técnicas de agrupamiento jerárquico e iterativo, técnicas de predicción con modelos de regresión o correlación canónica (CCA), etc. También se incluyen programas para aplicar de forma sencilla técnicas de redes probabilísticas y neuronales. Estos programas, junto con información adicional y conjuntos de datos, pueden obtenerse de la dirección web <http://grupos.unican.es/ai/meteo/MeteoLab.html>. Para un completo entendimiento de los programas, se requiere un conocimiento básico del programa Matlab (ver, por ejemplo, Middleton, 2000).



Por último, queremos agradecer la ayuda y comentarios de todos nuestros compañeros, especialmente de Cristina Primo y Bartolomé Orfila, así como el apoyo institucional del Instituto Nacional de Meteorología (INM) y de la Universidad de Cantabria (UC) para el desarrollo de este trabajo.

El libro está especialmente dedicado a toda la gente que nos ha enseñado a aprender lo que sabemos y a valorar lo que tenemos, especialmente a los Profesores Enrique Castillo Ron y Miguel Angel Rodríguez.

José M. Gutiérrez
Rafael Cano
Antonio S. Cofiño
Carmen M. Sordo

Santander, junio de 2004

Acrónimos y Terminología Utilizados

A continuación se describen algunos acrónimos y terminología utilizados en los distintos capítulos de este libro.

- *ARPS*. Advanced Regional Prediction System, www.caps.ou.edu/ARPS.
- *BS*. Brier Score; error cuadrático medio de la probabilidad predicha menos la probabilidad observada (ver Cap. 7).
- *BSS*. Brier Skill Score; índice de pericia que se obtiene como $1 - BSP/BSR$, donde *BSP* es el *BS* de la predicción y *BSR* es el *BS* de un sistema de referencia, normalmente la climatología. Ver Cap. 7.
- *COAMPS*. US Navy's Coupled Ocean/Atmosphere Mesoscale Prediction System, www.nrlmry.navy.mil/projects/coamps.
- *CP*. Componente principal. Ver Sec. 2.5.
- DEMETER. Proyecto Europeo “Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction”.
- *Downscaling*. Mejora de resolución (o interpolación) de una predicción de un modelo numérico efectuada en una rejilla.
- *ECMWF*. Centro europeo para predicción a plazo medio (European Center for Medium-Range Weather Forecast), www.ecmwf.org.
- *ENSO*. Patrón de oscilación de El Niño-Pacífico Sur (El Niño-Southern Oscillation). Ver Sec. 2.5.
- *EOF*. Función empírica ortogonal (Empirical Orthogonal Function). Término sinónimo utilizado para el análisis de Componentes Principales en la comunidad de las ciencias atmosféricas. Ver Sec. 2.5.

- *ERA*. Reanálisis del ECMWF. Un reanálisis es un experimento de simulación numérica de la circulación atmosférica para un largo período de tiempo con un mismo modelo numérico, asimilando toda la información disponible. Ver Sec. 1.9.4.
- *EPS*. Sistema de predicción por conjuntos (Ensemble Prediction System). Ver Sec. 1.6.1.
- *FPC*. Función de probabilidad conjunta $p(x_1, \dots, x_n)$.
- *GCOS*. Global Climate Observing System. www.wmo.ch/web/gcos.
- *GSN*, GCOS Surface Network. Ver Sec. 1.9.
- *HIRLAM*. Modelo de área limitada de alta resolución (High Resolution Limited Area Model) hirlam.knmi.nl.
- *INM*. Instituto Nacional de Meteorología, www.inm.es.
- *LAN*. Red de área local (Local Area Network).
- *MM5*. Penn State/NCAR Mesoscale Model, box.mmm.ucar.edu/mm5.
- *NAO*. Oscilación del Atlántico Norte (North Atlantic Oscillation). Ver Sec. 2.5.
- *NCEP*. National Centers for Environmental Prediction: www.ncep.noaa.gov.
- *NHC*. National Hurricane Center, www.nhc.noaa.gov.
- *RMSE*. Error cuadrático medio (Root Mean Square Error).
- *ROC, curva*. Relative Operating Characteristic. Curva obtenida a partir de los aciertos y las falsas alarmas de un sistema de predicción binario. Ver Cap. 7.
- *RSM*. Modelo regional espectral del NCEP (Regional Spectral Model).
- *SLP*. Presión a nivel del mar (Sea Level Pressure).
- *SOM*. Redes auto-organizativas (Self-Organizing Maps) para agrupamiento y visualización de datos. Ver sec. 5.8.
- *SST*. Temperatura de la superficie del agua del mar (Sea Surface Temperature).
- *UTC*. Coordenadas universales de tiempo (Universal Time Coordinates).
- *WAN*. Red de área extendida (Wide Area Network), en contraposición a las LAN.

A lo largo de todo el libro se utiliza la notación X para referirse a una variable y \mathbf{X} para referirse a un vector, mientras que las minúsculas se utilizan para referirse a valores concretos de las mismas; por ejemplo, $X = x$. El símbolo $\langle \rangle$ se utiliza para denotar el valor medio de una variable para un período dado, y $\hat{o}(t)$ para denotar un valor estimado (predicho) para $o(t)$.

En este libro se ha optado por utilizar el término *predicibilidad*, en lugar de *predictibilidad*, debido a su mayor implantación en la comunidad científica en este campo.

El término *patrón* se utiliza para referirse a una estructura numérica con una cierta organización de sus elementos (por ejemplo, una estructura matricial o tensorial relativa a una rejilla sobre una región geográfica concreta). Para su tratamiento matemático los patrones serán considerados vectores, descomponiendo su estructura de forma apropiada, y componiéndola de nuevo cuando sea necesario.

Parte I

Introducción

CAPÍTULO 1

Modelos y Datos Atmosféricos

1.1. *Introducción*

El esfuerzo investigador llevado a cabo en las tres últimas décadas ha permitido un gran avance en el desarrollo de modelos de circulación atmosférica que incorporan las parametrizaciones de los fenómenos físicos relevantes, adecuados a las escalas espaciales a que dichos modelos se aplican. Por otra parte, la aplicación operativa de estos modelos ha sido posible gracias a la disponibilidad de mejores y más complejos sistemas de observación (subsistema de observación terrestre: SYNOP, SHIP, TEMP, AIREP, DRIBU, etc. y subsistema espacial: SATEM, SATOB, etc., disponibles varias veces al día). Estos datos observacionales pueden ser asimilados en los modelos gracias al desarrollo y uso sucesivo de métodos de interpolación óptima y métodos de asimilación variacionales 3D y 4D, puestos a punto con el fin de establecer con la menor incertidumbre posible las condiciones iniciales a partir de las cuales se integran los modelos de predicción.

Como resultado de este trabajo, se dispone en la actualidad de eficientes modelos atmosféricos que se utilizan para la elaboración de un amplio abanico de predicciones con distintas resoluciones espaciales y distintos alcances temporales. Por una parte, los modelos de mayor resolución y de área limitada se utilizan principalmente para la predicción a corto y medio plazo, mientras que, en el otro extremo, los modelos acoplados océano-atmósfera de circulación general se aplican para la predicción de anomalías en la predicción estacional y para la preparación de escenarios climáticos en función de diversos supuestos de forzamiento radiativo (duplicación del nivel de CO_2 , etc.).

En este capítulo se describen brevemente las características más importantes de los modelos y procesos involucrados en la predicción numérica del tiempo. Las salidas de estos modelos se utilizan en el resto del libro como datos de entrada, o predictores, en los distintos métodos propuestos. En la Sección 1.2 se describen brevemente los principales procesos físicos que intervienen en la dinámica atmosférica. En la Sec. 1.3 se analizan algunos aspectos y limitaciones de la resolución numérica de estas ecuaciones. Por ejemplo, la Sec. 1.3.1 analiza el problema de la asimilación de datos, y la Sec. 1.3.2 describe algunas de las parametrizaciones básicas que se suelen utilizar en los modelos (y en la discretización del esquema numérico) para incluir los procesos físicos de menor escala que escapan a la resolución de la discretización (turbulencia, etc.). La Sec. 1.4 analiza los fenómenos y escalas que influyen en la variabilidad atmosférica. En la Sec. 1.5 se describen las características (resolución, alcance, etc.) de los principales tipos de modelos utilizados en predicción operativa. En la Sec. 1.6 se analiza el problema de la no linealidad y el caos, y la Sec. 1.6.1 describe el esquema práctico de la predicción por conjuntos utilizado para tener en cuenta este problema. La Sec. 1.7 describe el estado actual de las predicciones operativas que se obtienen diariamente con todos los modelos y técnicas anteriormente descritos; por otra parte, en la Sec. 1.8 se comentan brevemente las líneas de mayor interés en este campo de cara al futuro. Finalmente, la Sección 1.9 describe los datos observacionales (series climatológicas) y atmosféricos (simulaciones de modelos) que se utilizan en este libro.

1.2. Las Ecuaciones de la Atmósfera

En esta sección se presenta una breve descripción de las ecuaciones que controlan la dinámica de la atmósfera, con el único objetivo de ofrecer al lector una visión general sobre la complejidad de los modelos involucrados en la predicción numérica del tiempo. Para un estudio detallado, existen numerosos libros de meteorología dinámica con excelentes explicaciones y ejemplos de las deducciones y descripciones físicas de las ecuaciones (ver, por ejemplo, Holton, 1992; Barry and Chorley, 1998; Kalnay, 2003).

Desde un punto de vista físico, la atmósfera puede ser considerada como una mezcla de gases y agua en sus distintos estados. Esta mezcla está en movimiento dentro de un campo gravitatorio sobre una esfera en rotación y calentada por el Sol. En este sistema se deben cumplir la ecuación de estado de los gases y las leyes de conservación de energía, masa y momento. La mayoría de estas leyes se expresan con ecuaciones que relacionan las derivadas totales de ciertas magnitudes físicas en el tiempo. Estas derivadas son normalmente descompuestas en sus términos local y advectivo. Por ejemplo, considerando el campo de

velocidades $\mathbf{v} = (u, v, w)$ asociado al sistema de referencia (x, y, z) se tendrá:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial z} \frac{dz}{dt} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f \quad (1.1)$$

donde $f(x, y, z, t)$. Esta ecuación indica que la derivada total en el tiempo (Lagrangiana o individual) de una función, viene dada por la derivada local en el tiempo (parcial, o Euleriana) mas los cambios debidos al término de advección.

- Según la *ecuación de conservación de masa*, la variación total en el tiempo de la masa de una parcela de aire es nula ($dM/dt = 0$). La masa de aire contenida en un volumen $\Delta x \Delta y \Delta z$ es:

$$M = \rho \Delta x \Delta y \Delta z \quad (1.2)$$

donde ρ es la densidad del aire. Considerando la ecuación

$$\frac{1}{M} \frac{dM}{dt} = 0, \quad (1.3)$$

sustituyendo (1.2) y aplicando (1.1) se tiene la ecuación de conservación de masa

$$\frac{1}{\rho} \frac{d\rho}{dt} + \nabla \cdot \mathbf{v} = 0 \Leftrightarrow \frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}), \quad (1.4)$$

ya que $\partial u / \partial x = \Delta x^{-1} d\Delta x / dt$.

- La *ecuación de conservación de vapor de agua* determina que la variación total de vapor de agua en una parcela de aire se debe exclusivamente a procesos internos de evaporación E (fuente) y/o condensación C (sumidero):

$$\frac{dq}{dt} = E - C \quad (1.5)$$

donde q es la proporción en masa de vapor de agua en la parcela de aire (g/kg). Si esta ecuación se multiplica por ρ , se expande según (1.1), y se suma a la ecuación de conservación de masa (1.4) multiplicada por q , queda

$$\frac{\partial \rho q}{\partial t} = -\nabla \cdot (\rho \mathbf{v} q) + \rho (E - C). \quad (1.6)$$

De la misma forma que se ha incluido una ecuación de conservación para el vapor de agua, se podría incluir cualquier otra ecuación de conservación para otros elementos como agua líquida, ozono, etc., mientras también se incluyan sus correspondientes *fuentes y sumideros*.

- La *ecuación de estado de los gases ideales* aplicada a la atmósfera impone la siguiente relación entre las variables de estado termodinámicas:

$$p \alpha = R T, \quad (1.7)$$

donde p es la presión (mb ó Hpa), T la temperatura ($^{\circ}C$ ó K), R es la constante de los gases ideales, y α es el volumen específico (m^3/kg), inverso de la densidad ρ (kg/m^3).

- *Conservación de energía:* El foco principal de calor para la atmósfera es la superficie terrestre (calentada por el Sol), entendiendo como tal tierra y océano. Este calor absorbido por la atmósfera se emplea en variar su temperatura, su densidad o ambas cosas a la vez; es decir, si se aplica una tasa de calor Q por unidad de masa a una parcela de aire ($cal/s\,g$), esta energía es empleada en aumentar la energía interna $C_v T$ y producir un trabajo de expansión

$$Q = C_v \frac{dT}{dt} + p \frac{d\alpha}{dt}, \quad (1.8)$$

donde los coeficientes de calor específico a volumen constante (C_v) y a presión constante (C_p) se relacionan mediante $C_p = C_v + R$. Haciendo uso de la ecuación de estado se puede obtener una forma alternativa de la ecuación de conservación de energía:

$$Q = C_p \frac{dT}{dt} - \alpha \frac{dp}{dt}. \quad (1.9)$$

- *Conservación del Momento:*

$$\frac{d\mathbf{v}}{dt} = -\alpha \nabla p - \nabla \phi + \mathbf{F} - 2\boldsymbol{\Omega} \times \mathbf{v}; \quad (1.10)$$

La aceleración sobre la unidad de masa es debida a cuatro fuerzas: gradiente de presión ($-\alpha \nabla p$), gravedad aparente ($-\nabla \phi$), rozamiento (\mathbf{F}) y Coriolis ($-2\boldsymbol{\Omega} \times \mathbf{v}$).

Por tanto, resumiendo lo anterior, se tienen siete ecuaciones y siete incógnitas: $\mathbf{v} = (u, v, w)$, T , p , $\rho = 1/\alpha$ y q :

$$\frac{d\mathbf{v}}{dt} = -\alpha \nabla p - \nabla \phi + \mathbf{F} - 2\boldsymbol{\Omega} \times \mathbf{v} \quad (1.11)$$

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}) \quad (1.12)$$

$$p \alpha = RT \quad (1.13)$$

$$Q = C_p \frac{dT}{dt} - \alpha \frac{dp}{dt} \quad (1.14)$$

$$\frac{\partial \rho q}{\partial t} = -\nabla \cdot (\rho \mathbf{v} q) + \rho (E - C) \quad (1.15)$$

A las ecuaciones anteriores se las suele denominar ecuaciones primitivas, y representan el sistema que gobierna la dinámica de la atmósfera. El objetivo

de la predicción numérica del tiempo es obtener el estado de la circulación atmosférica en un tiempo futuro a partir de la condición inicial actual. Para ello es necesario disponer de un modelo numérico capaz de integrar las ecuaciones y que incluya los intercambios energéticos (en la capa límite) más importantes (radiación, turbulencia, calor latente, etc.). Dada la no linealidad y complejidad de las ecuaciones primitivas, en la práctica se suele recurrir a distintas aproximaciones que simplifican la resolución numérica y eliminan inestabilidades numéricas.

1.2.1. *Filtrado de Soluciones Triviales*

Las simplificaciones más básicas de las ecuaciones anteriores ayudan a entender las características de los tipos de ondas básicos presentes en la atmósfera, que también aparecen en las soluciones más generales. Por ejemplo, suponiendo que el movimiento es *adiabático*, *rectilíneo*, y *sin gravedad*, se obtienen como soluciones las *ondas sonoras* puras que se propagan a través de la compresión del aire. Si se considera la aproximación *hidrostática* (sin velocidad vertical), las ondas sonoras sólo se propagan horizontalmente (ondas de Lamb). Por otra parte, si se supone que no hay movimiento horizontal, pero sí desplazamientos verticales (el movimiento se considera *adiabático* y *rectilíneo*), entonces el resultado son las *ondas gravitatorias externas*: ‘oleaje’ en la superficie libre; y las *ondas gravitatorias internas*: cualquier partícula de fluido desplazada de su nivel de equilibrio oscila verticalmente con una frecuencia característica denominada frecuencia de Brunt-Vaisala (en la atmósfera, en condiciones normales, las oscilaciones típicas son de 100 s). Por otra parte, considerando un *flujo inicial zonal* ($v = w = 0$) y un campo uniforme de densidad, cualquier perturbación al flujo zonal supone un cambio de latitud y por tanto un cambio en la fuerza de Coriolis, que siempre actúa tratando de restaurar el flujo zonal produciendo las llamadas *ondas de Rossby* que se propagan alrededor del globo con un periodo de desarrollo de varios días a una semana y con una longitud de onda del orden de 3000 km para latitudes medias ($\varphi = 45^\circ$).

En general interesa que las soluciones obtenidas para la circulación de la atmósfera en un problema operativo muestren el comportamiento de interés meteorológico, por lo que es deseable eliminar algunas de las ondas básicas que no intervienen en esta dinámica, pero que pueden producir inestabilidades en el modelo numérico al ser propagadas (esto ocurre, por ejemplo con las ondas sonoras y gravitatorias). Este procedimiento de selección de tipos de ondas se denomina *filtrado* y consiste en eliminar parcialmente las fuerzas restauradoras que originan la onda que se desea filtrar. Por ejemplo, la *aproximación hidrostática* impide la propagación vertical de las ondas sonoras.

Por tanto, los modelos operativos utilizados en la predicción numérica del tiempo son aproximaciones más o menos completas de las ecuaciones primitivas de la atmósfera.

1.3. *Resolución Numérica de las Ecuaciones*

Los modelos atmosféricos de circulación se resuelven utilizando técnicas numéricas que discretizan el espacio y el tiempo. En coordenadas cartesianas se considera una rejilla 4D sobre la que se aplican técnicas de elementos finitos, mientras que en coordenadas esféricas se aplican técnicas espectrales que consideran un número finito de armónicos esféricos en la descripción del sistema (para introducción a estos métodos ver Kalnay, 2003). En ambos casos, la precisión del modelo está fuertemente influenciada por la resolución espacial (dada directamente por el tamaño de rejilla o por el período del modo de mayor frecuencia). Sin embargo, aumentar la resolución del modelo es extremadamente costoso ya que, por ejemplo, duplicar la resolución en el espacio tridimensional también requiere reducir a la mitad el paso de tiempo para satisfacer las condiciones de estabilidad computacional. Por tanto, el coste computacional total de duplicar la resolución crece con un factor de $2^4 = 16$. Las técnicas modernas de discretización intentan obtener un incremento en la precisión sin tanto coste computacional; estas técnicas son los esquemas semi-implícitos y semi-lagrangiano en el tiempo. Además, estos esquemas poseen condiciones de estabilidad menos estrictas. Aún así, existe una constante necesidad de aumentar la resolución para obtener una predicción mejor y más detallada. Esta tarea es la mayor aplicación de los super-ordenadores disponibles (www.top500.org).

A pesar de los avances logrados en cuanto a la modelización física de la atmósfera y a la resolución numérica de los modelos, la notable mejora en la capacidad de predicción se debe principalmente a otros factores. Por ejemplo, Kalnay (2003) describe los siguientes factores:

- Por un lado, el aumento de la potencia de los super-ordenadores permitiendo resoluciones mucho más finas y menos aproximaciones en los modelos atmosféricos operacionales.
- El aumento de la disponibilidad de datos, especialmente procedentes de satélites y aviones sobre los océanos y el hemisferio Sur.
- El uso de métodos más exactos de asimilación de datos, lo que resulta en una mejor condición inicial para los modelos.
- La mejora en la representación de procesos físicos de escala pequeña en los modelos (nubes, precipitación, transferencia de calor en régimen turbulento, humedad, radiación, etc.).

En las siguientes secciones se analizan estos temas en más detalle.

1.3.1. *Asimilación de Datos*

La predicción numérica es, en buena parte, un problema de condiciones iniciales: dada una estimación actual del estado de la atmósfera, un modelo

numérico simula su evolución, para obtener una predicción en un estado futuro. Esta condición inicial se establece a partir de la interpolación sobre los puntos de rejilla del modelo de las observaciones disponibles; este proceso de obtención de la condición inicial a partir de las observaciones se denomina *asimilación*. El principal problema de este proceso es que la cantidad de datos disponibles no es suficiente para inicializar el modelo en todos sus grados de libertad (por ejemplo, un modelo con una resolución típica de 1° de resolución horizontal y 20 niveles verticales podría tener $180 \times 360 \times 20 = 1.3 \times 10^6$ puntos de rejilla, en cada uno de los cuales están definidas 7 variables, con lo que tendríamos aproximadamente 10^7 grados de libertad). Para una ventana temporal de ± 3 horas, existen normalmente entre 10^4 y 10^5 observaciones de la atmósfera, dos órdenes de magnitud menor que el número de grados de libertad del modelo. Mas aún, la distribución espacial y temporal de las observaciones no es uniforme, existiendo regiones en Eurasia y Norteamérica con muchos datos, y regiones en el Hemisferio Sur con pocos datos. Por tanto, se hace necesario usar información adicional (llamado fondo, first-guess o información a priori) para preparar las condiciones iniciales de la predicción. Inicialmente se usaba la climatología como first-guess, pero al mejorar la pericia de las predicciones, se utiliza una predicción a corto plazo como first-guess en los sistemas de asimilación de datos operacionales (ciclos de análisis).

Para los modelos globales, el first-guess es la predicción del modelo a las 6 horas, \mathbf{x}_p (un array 4-dimensional) que es interpolada a los puntos de observación mediante un operador $H(\mathbf{x}_p)$ y convertida al mismo tipo que las variables observadas \mathbf{y}_0 . Las diferencias entre las observaciones y el first-guess $\mathbf{y}_0 - H(\mathbf{x}_p)$ son los *incrementos observacionales* o *mejoras*, y el análisis \mathbf{x}_0 se obtiene añadiendo las mejoras al first-guess del modelo con unos pesos W que son determinados en base a las covarianzas de los errores estadísticos de la predicción y observación:

$$\mathbf{x}_0 = \mathbf{x}_p + W[\mathbf{y}_0 - H(\mathbf{x}_p)] \quad (1.16)$$

Los diferentes esquemas de análisis están basados en este método:

- Interpolación óptima, donde la matriz de pesos se determina minimizando los errores en cada punto de rejilla.
- Métodos variales 3D y 4D, con funciones de coste proporcionales al cuadrado de la distancia entre el análisis, el first-guess y las observaciones (sobre un intervalo temporal o ventana de asimilación para el caso 4D).
- Más recientemente, los filtros de Kalman extendidos (Judd, 2003).

A la vista de lo anterior se puede interpretar que el ciclo de asimilación de datos es una continua integración del modelo que se va perturbando con las observaciones de tal forma que permanezca lo más cerca posible del estado real

de la atmósfera. La función del modelo es transportar información de zonas con muchos datos, a zonas con pocos datos y ofrecer una estimación del estado de la atmósfera.

Los errores que se cometan en el proceso de asimilación imponen una incertidumbre en el plazo de predicción, dada por la no linealidad de la atmósfera (y, por tanto, de los modelos que aproximan su dinámica). Este problema se vuelve importante en el plazo medio y en la predicción estacional y obliga a formular la predicción en términos probabilísticos. La predicción por conjuntos (*ensemble forecast*), que se analiza en una sección posterior, es una técnica práctica para mejorar la predicción teniendo en cuenta esta incertidumbre.

1.3.2. Parametrización de Procesos Físicos

A pesar de que el constante incremento de la capacidad de cálculo permite integrar modelos numéricos cada vez con más resolución, existen muchos procesos atmosféricos que no pueden ser resueltos de manera explícita a la escala que se utiliza en los modelos (algunos ocurren a escala molecular, como la radiación, evaporación, rozamiento y turbulencia); por tanto, estos procesos no están contemplados en las ecuaciones de los modelos que se usan y se vuelven importantes en escalas espaciales pequeñas (por debajo de las decenas de km). En estos casos, estos procesos juegan un papel crucial en el balance energético global afectando, por supuesto, a los procesos a gran escala. Por esa razón, la interacción entre escalas no puede ser ignorada por los modelos, y se recurre a la parametrización. Así, cuando las ecuaciones son discretizadas sobre un tamaño de rejilla dado (de unos pocos a varios cientos de kilómetros), se hace necesario añadir términos “fuentes” y “sumideros” para tener en cuenta el balance energético asociado a los fenómenos que escapan a la dinámica formulada en el modelo numérico.

La parametrización consiste en reformular a gran escala los efectos de pequeña escala (ver, por ejemplo, Beniston, 1998). Una dificultad añadida es que no siempre está claro a qué escala se asigna un fenómeno determinado; por ejemplo, un modelo de 50 km no resuelve la circulación de brisa, pero tampoco la ignora completamente, de manera que el proceso es doblemente difícil en estos casos. Como ejemplo, se podría considerar la ecuación de conservación de vapor de agua en coordenadas de presión dada por

$$\frac{\partial \bar{q}}{\partial t} + \bar{u} \frac{\partial \bar{q}}{\partial x} + \bar{v} \frac{\partial \bar{q}}{\partial y} + \bar{w} \frac{\partial \bar{q}}{\partial p} = \bar{E} - \bar{C} + \frac{\partial \bar{w}' q'}{\partial p} \quad (1.17)$$

donde q es la proporción de vapor de agua y masa de aire seco, x e y coordenadas horizontales, p es la presión, t el tiempo, u y v son las componentes de la velocidad horizontal del aire (viento), $w = \frac{dp}{dt}$ es la velocidad vertical en coordenadas de presión, y el producto de las variables prima representa el transporte turbulento de humedad desde las escalas no resueltas por la rejilla

usada en la discretización; con la barra horizontal (\bar{q}) se representan promedios espaciales sobre la rejilla del modelo. A la parte izquierda de (1.17) se la denomina “dinámica” del modelo, y se calcula explícitamente. La parte derecha de (1.17) se denomina “física” del modelo. La ecuación de la humedad incluye los efectos de los procesos físicos tales como evaporación (\bar{E}) y condensación (\bar{C}), y transferencias turbulentas de humedad que tienen lugar a escalas pequeñas que no pueden ser resueltas explícitamente por la “dinámica”.

1.4. Variabilidad Atmosférica

Una de las pruebas más simples para verificar el correcto funcionamiento de los modelos de predicción numérica del tiempo consiste en contrastar el grado de exactitud con el que aproximan la variabilidad climatológica de distintos fenómenos de interés, a distintas escalas espaciales y temporales.

1.4.1. Escalas Atmosféricas

Los fenómenos más relevantes de la atmósfera ocurren a distintas escalas espacio-temporales, desde escalas sinópticas con fluctuaciones en la circulación asociadas a grandes sistemas de presión, con longitudes de onda en torno a 5000 km y periodos entorno a 5 días, a escalas mesoescalares con fluctuaciones locales en la circulación derivadas de la interacción entre calentamientos diferenciales inducidos por el ciclo diurno e inhomogeneidades en superficie, con longitudes de onda en torno a 200 km (ver Orlanski, 1975; Jansá, 1990, para una descripción detallada).

Nombre	Escala	Duración	Fenómenos Atmosféricos
Macro-a	15.000 km	1 mes	Circulación general, ondas largas.
Macro-b	5.000 km	1 semana	Depresiones y anticiclones.
Macro-g	500 km	2 días	Frentes, huracanes.
Meso-b	100 km	6 horas	Brisas, ondas de montaña, chorros de bajo nivel, complejos convectivos, isla térmica.
Meso-g	5 km	1 hora	Tormentas, TAC.
Meso-d	500 m	30 mín.	Cúmulos, tornados, vientos catabáticos.
Micro-b	50 m	5 mín.	Penachos, estelas, trombas, tolvaneras.
Micro-g	2 m	1 seg.	Ondas sonoras, turbulencia.

Figura 1.1: Escalas meteorológicas (adaptado de Orlanski, 1975).

Los modelos numéricos de circulación atmosférica tratan de reproducir con la mayor precisión posible la variabilidad de estos fenómenos, a las distintas escalas a que los aproximan. Por ejemplo, si se está interesado en la precipitación diaria, habría que trabajar con un modelo numérico que reprodujese la variabilidad climatológica observada para este fenómeno. En algunos casos, la variabilidad de fenómenos diarios o mensuales puede estar afectada por patrones atmosféricos de gran escala cuya dinámica se desarrolla en escalas temporales estacionales e incluso interanuales. El estudio de estos patrones es, por tanto, de gran importancia para el conocimiento de la dinámica atmosférica y de su variabilidad.

1.4.2. Patrones de Oscilación Atmosférica y Teleconexión

En algunas regiones del globo la variabilidad de la atmósfera está dominada en gran medida a un cierto patrón atmosférico de escala sinóptica que oscila en el tiempo con períodos estacionales o interanuales. Algunos de estos patrones son bipolares y su estado (positivo o negativo) se puede caracterizar en base a un índice. Algunos ejemplos notorios de estos patrones son:

- *NAO* (North Atlantic Oscillation) (Wanner et al., 2001): Característica de latitudes medias y asociada en términos relativos a un patrón alta/baja presión en Islandia - baja/alta presión en las Azores. Este patrón suele definirse a través de un índice que caracteriza su fase y que es obtenido como la diferencia de presiones a nivel del mar en ambos puntos, o como la componente principal asociada a la primera EOF de la presión en 500 mb del Atlántico Norte (ver Capítulo 2, Fig. 2.12). Se han descubierto distintas interrelaciones entre este patrón y distintas fenomenologías sobre Europa como, por ejemplo, la precipitación (van Oldenborgh et al., 2000) o la temperatura en superficie (Pozo-Vázquez et al., 2000) sobre Europa o, más específicamente, sobre la Península Ibérica (Esteban-Parra et al., 2003).

La NAO posee una marcada variabilidad interanual lo cual permite una hipotética predicción estacional. A causa de su notable impacto sobre el tiempo y clima de Europa, existe un creciente interés en determinar su predecibilidad estacional e interanual. Por desgracia la NAO es un fenómeno de latitudes medias muy ruidoso, con un espectro de potencias casi plano en estas escalas temporales, muy parecido a un ruido blanco, de manera que incluso las predicciones posibles con modelos estadísticos, reproducen sólo un pequeño rango de la varianza total (Wunsch, 1999). Por tanto, un problema actual de gran importancia es la reproducción y predicción de este patrón con modelos de circulación atmosférica.

- *ENSO* (El Niño-Southern Oscillation) (Philander, 1990): Es un fenómeno tropical del Pacífico Sur (ver Capítulo 2, Fig. 2.12) caracterizado por un

calentamiento del agua de la superficie del Pacífico central y este y por la variación a gran escala asociada del sistema de presión atmosférica del trópico (denominado Oscilación del Sur, SO). Este patrón se caracteriza tradicionalmente basándose en las diferencias en las anomalías en la presión del aire entre Tahiti y Darwin. La variabilidad de este índice se muestra en la Fig. 1.2.

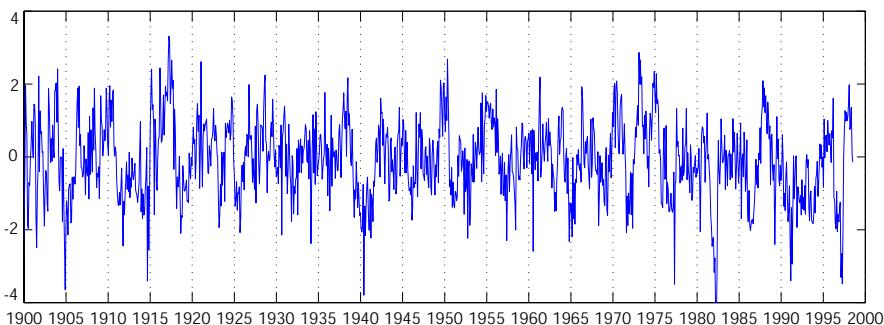


Figura 1.2: Índice de Oscilación del Sur (SOI) durante el siglo XX. El índice de un mes se define como $(T - D)/S$, donde T y D son las anomalías mensuales en Taití y Darwin, respectivamente y S es la varianza de $T - D$ para el mes dado.

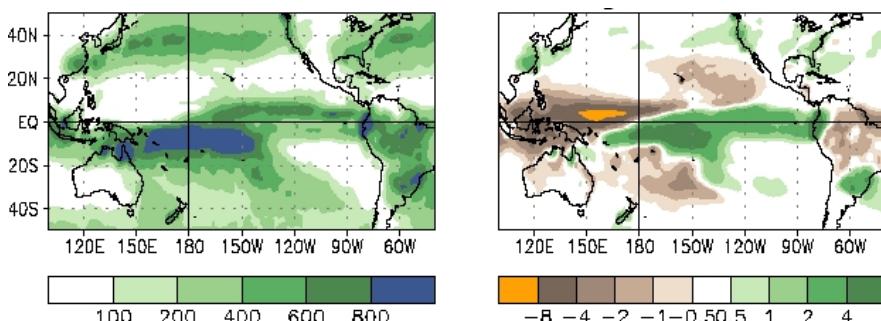


Figura 1.3: Precipitación en el Pacífico durante el episodio fuerte de El Niño (Enero-Marzo de 1998). La figura izquierda muestra la precipitación total y la figura derecha la desviación en tanto por uno respecto del valor medio (FUENTE: NCEP).

El ciclo ENSO tiene un periodo medio de cuatro años aproximadamente, aunque en el registro histórico el periodo ha variado entre dos y siete años. Entre 1980 y 2000 destacó un ciclo ENSO muy activo, con cinco episodios de El Niño,(1982/83, 1986/87, 1991-1993, 1994/95, y 1997/98) y tres episodios de La Niña (1984/85, 1988/89, 1995/96). En este periodo también tuvieron lugar dos de los episodios más fuertes del siglo (1982/83 y 1997/98), así como dos periodos consecutivos de las condiciones de El Niño durante 1991-1995 sin una intervención de episodio frío. El episodio más fuerte y reciente de La Niña fue en 1988/89, y entre las consecuen-

cias más graves del fenómeno de El Niño, están un fuerte incremento de las lluvias producidas sobre el sur de los EEUU y en Perú, lo que provoca destructivas inundaciones, así como sequías en el Pacífico Oeste, a veces asociadas a grandes incendios en Australia. Por ejemplo, la Fig. 1.3 muestra los efectos de El Niño 1998 sobre la precipitación en el Pacífico.

En la Tabla 1.1 se muestran algunos de los patrones de oscilación más importantes. Además, se han descubierto distintas *teleconexiones* entre estos patrones y la variabilidad fenomenológica observada en regiones del globo distantes y que parecen estar interrelacionadas de forma compleja (ver, por ejemplo, Mo and Livezey, 1986). Por ejemplo, existen evidencias de que el ciclo ENSO, además de estar implicado de forma directa con las precipitaciones en Sudamérica y sequías en Australia, está relacionado con episodios de lluvias monzónicas en Asia (Hendon, 2001); ésto supone la existencia de una teleconexión entre ambos fenómenos.

Índice	Descripción
<i>EA</i>	Patrón del Atlántico Este
<i>EAJET</i>	Patrón del chorro del Atlántico Este
<i>WP</i>	Patrón del Pacífico Oeste
<i>EP</i>	Patrón del Pacífico Este
<i>NP</i>	Patrón del Pacífico Norte
<i>PNA</i>	Patrón Pacífico-NorteAmérica
<i>EA/WR</i>	Patrón del Atlántico Este-Rusia Oeste
<i>SCA</i>	Patrón Escandinavia
<i>TNH</i>	Patrón Tropical del Hemisferio Norte
<i>POL</i>	Patrón Polar-Eurasia
<i>PT</i>	Patrón de Transición del Pacífico
<i>SZ</i>	Patrón Subtropical zonal
<i>ASU</i>	Patrón de Verano de Asia

Tabla 1.1: Algunos de los patrones atmosféricos más importantes (para más detalle ver www.cpc.ncep.noaa.gov/products).

1.5. Tipos de Modelos Numéricos del Tiempo

Como ya se ha comentado, la estabilidad numérica de los métodos de integración obliga a mantener un compromiso entre la resolución espacial y el alcance de los modelos numéricos operativos, para que éstos puedan ser integrados en un tiempo limitado. Las integraciones con gran resolución espacial requieren un paso de integración pequeño limitando el alcance operativo de las mismas a unos pocos días, mientras que las integraciones de modelos de baja resolución pueden prolongarse en el tiempo hasta escalas mensuales o

interanuales. Así, cada tipo de modelo caracteriza un conjunto de fenómenos asociados con la escala espacial de trabajo, desde fenómenos sinópticos de miles de kilómetros y períodos de días, hasta fenómenos mesoescalares de escalas de kilómetros y períodos de horas (ver Fig. 1.1).

Este hecho ha motivado que operativamente se consideren distintas configuraciones o tipos de modelos numéricos según el alcance y resolución de la predicción deseada. A continuación se describen las características de algunas de éstas configuraciones.

1.5.1. Modelos Globales de Circulación General

Los modelos globales se integran sobre todo el globo por lo que su tratamiento numérico se realiza en coordenadas esféricas. Por tanto, la resolución horizontal de estos modelos viene caracterizada por el número de modos esféricos que se consideren en el desarrollo de las soluciones; así, un modelo truncado a 144 modos se denomina $T144$ y tiene una resolución horizontal de 2.5 grados, que corresponden aproximadamente a 250 km en nuestra latitud (ver Fig. 1.4)). Este truncamiento también influye en la resolución vertical (número de niveles de presión) que habrá de ser consistente con la resolución espacial y temporal. Así, el modelo global operativo en la predicción a corto plazo del ECMWF (European Center for Medium-Range Weather Forecast, www.ecmwf.int) es un $T511L60$ con 60 niveles de presión, mientras que el modelo utilizado para la predicción estacional es un $T95L40$.

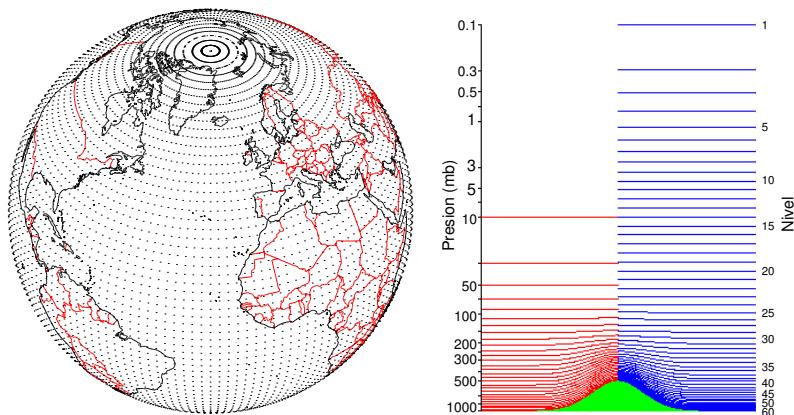


Figura 1.4: Rejilla global de 2.5° de resolución horizontal utilizada por modelos de circulación general sobre todo el globo; el tamaño de la rejilla es $144 \times 73 = 10512$ puntos. (derecha) Dos perfiles verticales con 31 y 60 niveles de altura geopotencial (expresados en milibares mb , y en números de nivel del modelo, respectivamente). La altura máxima mostrada ($0.1 mb$) es de aproximadamente 64 km.

Dada su escasa resolución espacial, estos modelos utilizan la aproximación hidrostática de la atmósfera (los movimientos verticales están parametrizados) y capturan su dinámica sinóptica. Normalmente son utilizados en la predicción mensual y estacional, y también en las simulaciones de escenarios de cambio climático. En algunos casos, los modelos de circulación general son acoplados con modelos oceánicos para tener caracterizado este término importante del forzamiento de la atmósfera; en otros casos, dado que la evolución de la atmósfera es la componente rápida del sistema, la temperatura del agua se supone constante (por ejemplo en la predicción mensual).

Otros términos como la orografía, el uso del suelo, la cubierta de hielo, etc. se parametrizan a escala de la rejilla del modelo y, por tanto, la orografía está muy suavizada y sólo refleja de forma grosera los principales sistemas montañosos del planeta.

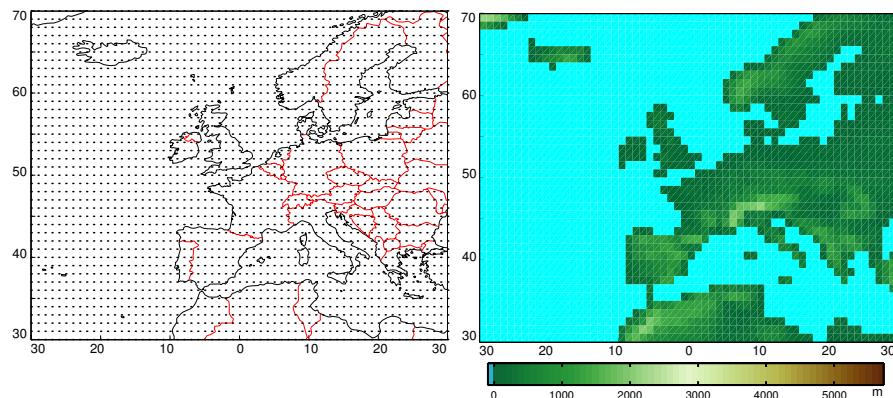


Figura 1.5: Rejilla horizontal de resolución 1° de longitud y latitud sobre Europa (izquierda). Orografía del modelo para la resolución dada (derecha).

1.5.2. Modelos Regionales

Una solución para aumentar la resolución del modelo sin incrementar el coste computacional es considerar rejillas limitadas a zonas geográficas de especial interés. Por ejemplo, la Fig. 1.5(a) muestra una rejilla de 1° de resolución en longitud y latitud centrada en Europa. Debido a su mayor resolución, los modelos regionales tienen una mayor exactitud para reproducir fenómenos de pequeña escala como las tormentas, y también tienen mejores forzamientos orográficos que los modelos globales (ver Fig. 1.5(b)). A pesar de que estos modelos han utilizado históricamente la aproximación hidrostática de la atmósfera, la tendencia actual es utilizar modelos no hidrostáticos en este tipo de predicciones. Por otro lado, al no ser globales, estos modelos tienen la desventaja de no ser “autocontenidos” y, aparte de las condiciones iniciales, requieren condiciones de contorno en las fronteras del dominio. Estas condiciones de contorno

necesitan ser lo más precisas posibles y por ello se toman interpolando la salida de un modelo global. Normalmente las condiciones de contorno se actualizan durante el transcurso de la predicción para imponer sobre el modelo regional la dinámica sinóptica simulada por el modelo global.

En algunos casos, se define un anidamiento de rejillas de tamaño decreciente y resolución creciente y las integraciones se realizan de forma *anidada* aprovechando las salidas de una rejilla como condiciones de contorno de la siguiente. Por ejemplo, el NCEP anida un modelo de baja resolución (*ETA-12*) con distintas rejillas de alta resolución (8 km) sobre zonas de interés (ver Fig. 1.6).

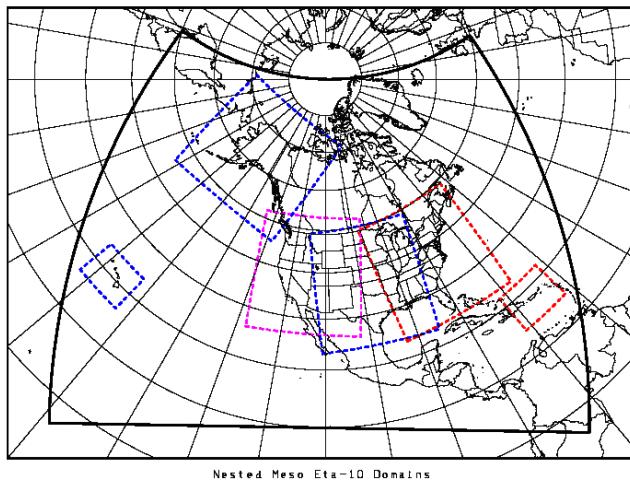


Figura 1.6: Rejillas anidadas con el modelo ETA-12 (línea sólida): Western U.S., Central U.S., Alaska, Hawaii, Puerto Rico. (FUENTE: NCEP/NOAA).

Este tipo de modelos son los utilizados en la predicción operativa a corto y medio plazo, donde cada servicio meteorológico se centra en su región de influencia, integrando un modelo regional con una resolución limitada principalmente por la capacidad de cómputo.

1.5.3. Modelos Mesoscales

Más recientemente, la resolución de algunos modelos regionales ha aumentado hasta llegar a unos pocos kilómetros, con objeto de mejorar la resolución de fenómenos convectivos locales (tormentas) y otros procesos físicos de pequeña escala. Dada su escasa gran resolución horizontal, la formulación de estos modelos suele darse en coordenadas cartesianas y no usan la aproximación hidrostática, la cual deja de tener validez para escalas horizontales menores de 10 km. En este caso, las parametrizaciones juegan un papel fundamental

y su calibración para la zona geográfica de interés es determinante para el buen funcionamiento del modelo. Así se han desarrollado varios modelos no hidrostáticos que se utilizan rutinariamente para la predicción de fenómenos de mesoscala. Los más usados son ARPS (Advanced Regional Prediction System), MM5 (Penn State/NCAR Mesoscale Model, Version 5), RSM (NCEP Regional Spectral Model) y COAMPS (US Navy's Coupled Ocean/Atmosphere Mesoscale Prediction System).

En la Fig. 1.7 se muestra la orografía de la Península Ibérica con una resolución de 0.2° (aprox. 20 km) y la orografía de la Cornisa Cantábrica con una resolución de 0.0083° (aprox. 1 km). Esta última resolución es la que permite caracterizar de forma apropiada la orografía de esta zona.

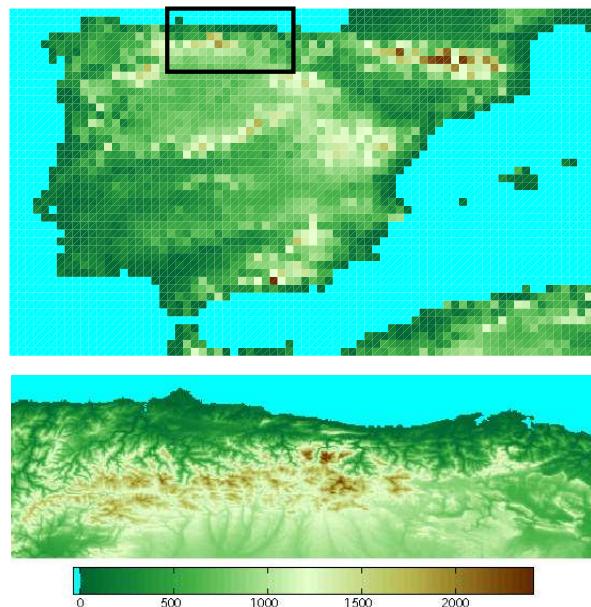


Figura 1.7: Orografía de la Península Ibérica correspondiente a un modelo de 0.2° (superior). Ampliación a 0.0083° de la zona norte (42.5-43.75 N, 3-7 O).

1.6. No Linealidad y Caos en la Atmósfera

Tradicionalmente, la dinámica de los sistemas no lineales ha sido estudiada a través de sistemas linealizados equivalentes, aproximados bajo distintos supuestos (pequeñas oscilaciones, etc.). Sin embargo, el descubrimiento de dinámicas “extrañas” asociadas a estos sistemas (Lorenz, 1963) provocó un enorme interés en el estudio de la dinámica no lineal (ver Strogatz, 2001, para una introducción). Uno de los fenómenos más sorprendentes es el conocido

como *caos determinista*, relacionado con la sensibilidad de estos sistemas a pequeñas perturbaciones (por ejemplo, en las condiciones iniciales de una trayectoria). Este fenómeno produce un comportamiento aparentemente impredecible y errático del sistema. Otro hecho destacable es que la dinámica caótica se desarrolla sobre un soporte fractal en el espacio de fases, llamado *atractor* (ver Grassberger and Procaccia, 1983, para más detalles). Así, geometría fractal y dinámica caótica están íntimamente ligados (Mandelbrot, 2004).

Los modelos de circulación atmosférica contienen términos no lineales, por lo que la teoría del caos ha tenido notable repercusión en este campo en las tres últimas décadas. La sensibilidad de los sistemas caóticos a pequeñas perturbaciones hacen que sólo puedan ser predichos hasta un cierto horizonte, ya que cualquier modelo aproximado conllevará un error que se propagará exponencialmente hasta que la dinámica del sistema aproximado difiera de la dinámica real. Este hecho supone una limitación teórica al alcance de predicción de los modelos numéricos de predicción del tiempo. Dependiendo del tipo de aproximaciones físicas y escalas que se consideren en el modelo numérico la no linealidad del modelo resultante impondrá un umbral menor o mayor en el alcance de la predicción (ver Lorenz, 1963, para más detalles).

El caos fue analizado por primera vez en las ecuaciones de un modelo atmosférico simplificado:

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = r x - y - x z \\ \dot{z} = x y - b z \end{cases} \quad (1.18)$$

Lorenz (1963) dedujo estas ecuaciones cuando estudiaba la posibilidad de predecir estructuras convectivas, tan frecuentes en la naturaleza y en tan diversas escalas. Para ello tomó como punto de partida las formaciones convectivas hidrodinámicas de Rayleigh-Bénard, por ser fáciles de generar y reproducir en el laboratorio. Basta con calentar por debajo una capa muy fina de un fluido viscoso, como la glicerina. El fluido caliente formado en la parte baja tiende a elevarse creando corrientes convectivas. Este proceso es muy parecido a lo que sucede en la atmósfera terrestre. Para un calentamiento suficientemente intenso la evolución temporal de estas corrientes convectivas tiene un comportamiento caótico; es decir, pequeñas perturbaciones en el sistema producen cambios impredecibles en su dinámica tras un tiempo suficiente. Las ecuaciones que describen este experimento tienen en cuenta los tres efectos más importantes que aparecen: La fuerza debida al gradiente térmico, la viscosidad y la difusión térmica. Lorenz observó que estas ecuaciones eran intratables numéricamente, pues no podía saber si el comportamiento errático era debido a una inestabilidad numérica o existía verdaderamente en el modelo. Propuso entonces una ecuación más simple, suponiendo que el campo de velocidades era el gradiente de una función Ψ y que existían simetrías espaciales. El modelo resultante (1.18) presenta dinámicas caóticas para ciertos valores de los parámetros (para otros valores su comportamiento es periódico). Por ejemplo, para los valores

de los parámetros $\sigma = 10$, $b = 8/3$, y $r = 28$, el sistema presenta una dinámica caótica. Partiendo de la condición inicial $(x(0), y(0), z(0)) = (-10, -5, 35)$ y utilizando un método de integración de Runge-Kutta de cuarto orden y paso de integración $\Delta t = 10^{-2}$ se obtiene la evolución del sistema en el espacio de fases que se encuentra dibujado en la Fig. 1.8(a). Por otra parte, las Fig. 1.8(b)-(d) muestran la evolución en el tiempo de las tres variables.

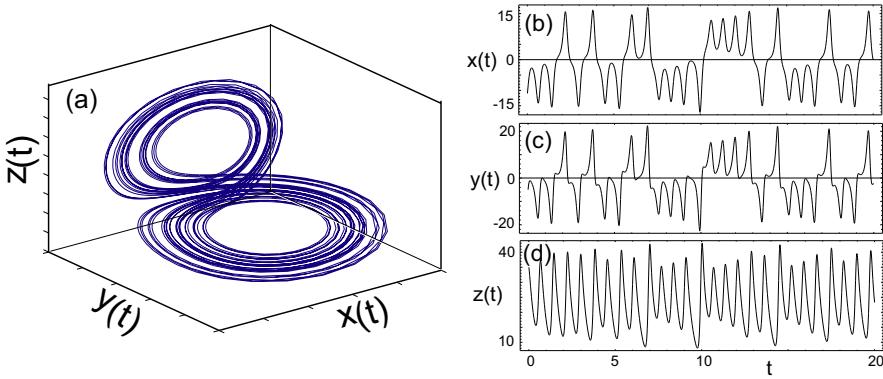


Figura 1.8: (a) Atractor del sistema de Lorenz. (b)-(d) Orbitas descritas por cada una de las tres variables del sistema.

El principal problema de este simple y pedagógico sistema de Lorenz es que no tiene en cuenta la componente espacial de la dinámica atmosférica. Sin embargo, también han sido desarrollados diversos modelos simplificados para analizar la dinámica no lineal espacio-temporal de la atmósfera. Por ejemplo, Lorenz (1986) propuso un sistema de 5 variables basado en un truncamiento en tres modos de las ecuaciones de aguas someras, y posteriormente también propuso sistemas de mayor dimensión, acoplando cadenas de sistemas de Lorenz. Las cadenas de sistemas acoplados (especialmente, de forma difusiva) han sido profusamente utilizados en distintos campos para modelizar la dinámica espacio-temporal. Por ejemplo, se podría considerar un sistema compuesto por una cadena de L sistemas de Lorenz acoplados de forma difusiva de la siguiente forma:

$$\begin{cases} \dot{x}_l = \sigma(y_l - x_l) \\ \dot{y}_l = x_l - y_l - x_l z_l + D(y_{l+1} + y_{l-1} - 2y_l), \quad l = 1, \dots, L. \\ \dot{z}_l = x_l y_l - bz_l \end{cases}$$

donde D es el parámetro de acoplamiento difusivo. En este caso, el espacio está representado de forma discreta por los L osciladores. La Fig. 1.9 muestra la dinámica de estos sistemas espacio-temporales para distintos valores de los parámetros.

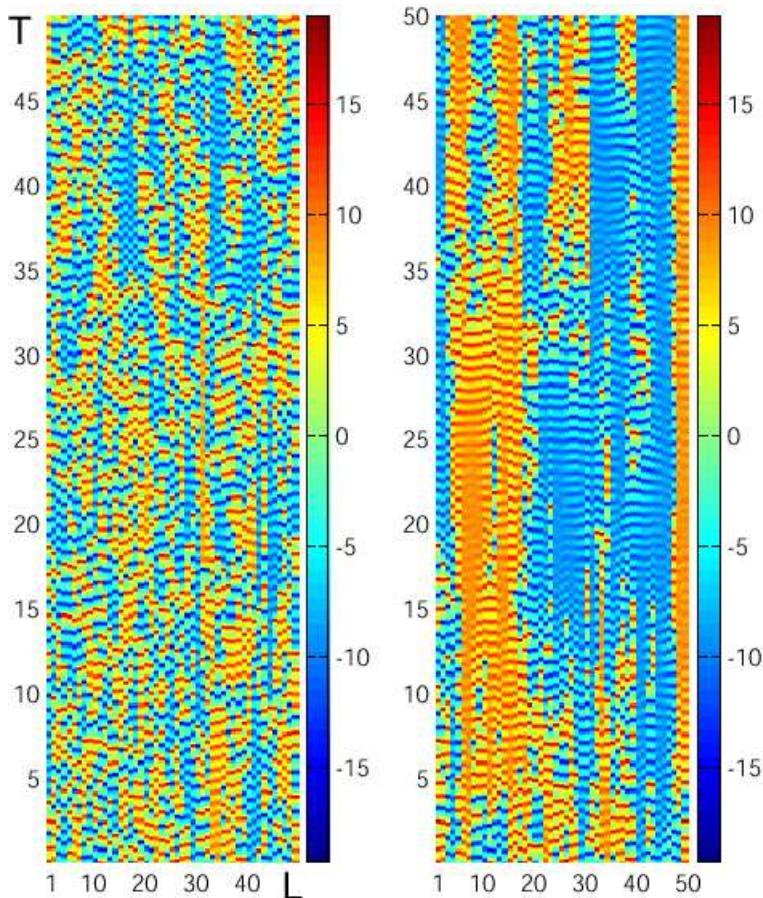


Figura 1.9: Evolución espacio-temporal de 50 osciladores de Lorenz acoplados difusivamente con parámetros (izquierda) $D = 0.25$ y (derecha) $D = 0.5$.

1.6.1. Predecibilidad y Predicción por Conjuntos

La no linealidad de los modelos de circulación y su consiguiente sensibilidad a las distintas fuentes de incertidumbre y error (el estado inicial de la atmósfera, la no exactitud del modelo numérico, etc.), imponen una limitación al alcance de las predicciones deterministas ya que estos errores producen fluctuaciones en el modelo que crecen exponencialmente en el tiempo. Este hecho ha obligado a formular el problema de la predicción en términos probabilísticos (Lorenz, 1996). Analizado de forma general, el origen de fluctuaciones en estos modelos puede provenir de errores en la condición inicial (caos), de la inexactitud del modelo (ruido) o de las parametrizaciones, incluida la orografía (desorden). Para tener en cuenta estos factores, en el último decenio se han desarrollado distintos métodos de *predicción por conjuntos* (ensemble forecast).

Estos métodos estiman la probabilidad a posteriori de las distintas variables atmosféricas, a partir de un conjunto de predicciones obtenidas integrando el modelo atmosférico con distintas realizaciones de las fuentes de incertidumbre.

Las fluctuaciones producidas por *errores en la condición inicial* han sido el caso más estudiado en la predicción meteorológica. Se suele analizar aplicando un método de predicción por conjuntos a modelos numéricos con muchos modos espaciales (Anderson, 1996); para ello se obtiene un conjunto de condiciones iniciales perturbando ligeramente el estado inicial asimilado a partir de las observaciones ; a continuación se integra el modelo numérico a partir de cada condición inicial, obteniéndose los miembros de la predicción por conjuntos. El coste computacional de este proceso impone la necesidad de equipos más rápidos y potentes, y limita el número de miembros que pueden ser tratados de forma operativa (unas decenas). Los conjuntos se pueden formar de muy diversas formas, siendo el método de vectores singulares (Buizza and Palmer, 1995) o el método de breeding (Pu et al., 1997) los más utilizados para la generación de miembros, a partir de perturbaciones específicas de las condiciones iniciales. El método de vectores singulares ha sido adoptado por el ECMWF y el método de breeding por el NCEP. Ambas técnicas presentan características distintas pero sus resultados finales, medidos en valores económicos, no son muy distintos. El método de breeding no requiere grandes esfuerzos computacionales, y es técnicamente muy simple. La perturbación se obtiene dinámicamente a partir de una trayectoria que es continuamente re-escalada para que contenga los modos espaciales de fluctuación de los errores. El método de vectores singulares es computacionalmente mucho más costoso y técnicamente más complejo, pues calcula las fluctuaciones lineales más significativas ocurridas en un alcance de predicción dado propagando el operador lineal asociado al modelo adelante y atrás en el tiempo.

En un sistema caótico de baja dimensionalidad el método de breeding produciría conjuntos muy poco significativos, pues a tiempos no muy largos se recuperarían vectores muy próximos al vector de Liapunov. Es por esto que los seguidores del método de vectores singulares enfatizan sobre la escasa representatividad estadística de los conjuntos producidos por métodos de breeding. Sin embargo, recientemente se ha mostrado con simulaciones de modelos meteorológicos simples que tal representatividad estadística es mucho más amplia de lo que sería un modelo de baja dimensionalidad (Corazza et al., 2003). Muestran además cómo las fluctuaciones aparecen espontáneamente y de forma muy local. Este hecho sólo puede ser debido a que en realidad un modelo meteorológico no puede ser representado con pocos grados de libertad. Por tanto, ha de pensarse en términos de caos espacio temporal.

En la práctica, este tipo de predicción por conjuntos se realiza, tanto para la predicción a medio plazo (entre 3 y 10 días), como para la predicción estacional. En este período la predicción se entiende en sentido probabilístico

y cobra importancia el concepto de predecibilidad, y su caracterización para cada situación (unas situaciones son más predecibles que otras). En estos casos, la predicción por conjuntos proporciona una solución práctica, ya que los miembros del conjunto de predicciones, proporcionan descripciones equiprobables del patrón atmosférico previsto, a partir de las cuales se puede obtener una predicción probabilística, y una estimación de su predecibilidad.

Por otra parte, la incertidumbre causada por la *no exactitud del modelo* es considerada de forma práctica aplicando distintos métodos; por una parte, algunas técnicas consideran perturbaciones en las parametrizaciones (incluida la orografía). Por otra parte, otras técnicas realizan una predicción por conjuntos uniendo las predicciones de distintos modelos (*predicción multi-modelo*). Quizás la iniciativa más destacada en este ámbito sea la llevada a cabo en el proyecto DEMETER (Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction) que está integrado por seis modelos globales de predicción por conjuntos con distintos esquemas de perturbación (www.ecmwf.int/research/demeter/) con los que se aborda el problema de la predicción mensual y estacional (Palmer et al., 2004).

Pese a las dificultades teóricas y computacionales asociadas con este tipo de técnicas, los avances producidos han dado un enorme impulso a la predicción meteorológica en el plazo medio (entre 4 y 15 días) y a la predicción mensual y estacional. Por ello, estos métodos se han convertido en un componente central en los grandes centros meteorológicos mundiales, incluyendo el ECMWF en Europa y el NCEP en EEUU.

Aunque la predicción probabilística no resulta familiar para los usuarios finales de las predicciones meteorológicas (sectores energético, hidrológico, agricultura, etc.), ésta posee el valor añadido de permitir definir una estrategia de toma de decisiones en base a los costes de protegerse de los eventuales efectos de una predicción adversa, y a las pérdidas que se producirían si no se toman las medidas de protección. Este esquema de valoración se obtiene a partir del llamado “valor económico” del modelo, que muestra la ganancia del modelo frente a la predicción climatológica para distintos valores de la razón coste/pérdidas (ver Cap. 7). De esta forma, es más sencillo obtener el valor de las predicciones meteorológicas, para los distintos sectores socio-económicos de aplicación, permitiendo una mayor difusión y un uso más racional de los productos de predicción meteorológica por parte de los usuarios.

Distintas validaciones de la predicción probabilística por conjuntos, han mostrado su superioridad frente a la predicción determinista tradicional (?). Sin embargo, a pesar de que cualitativamente la dispersión del conjunto parece apropiada, como medida de predecibilidad, todavía no se ha obtenido ningún resultado concluyente relacionando estos conceptos.

1.7. *Estado Actual de la Predicción Operativa*

Las predicciones que se preparan diariamente para el público en el Instituto Nacional de Meteorología (INM, www.inm.es) y en otros centros meteorológicos europeos (MeteoFrance, UK MetOffice, etc.) utilizan las salidas de los modelos numéricos anteriores. Estas predicciones, denominadas operativas, pueden ser de distintos tipos dependiendo de la escala espacial y temporal de su aplicación. Tradicionalmente, las predicciones se han obtenido integrando un único modelo numérico a partir de una condición inicial dada. Sin embargo, en la década de los noventa las técnicas de predicción por conjuntos han permitido considerar los efectos causados por las distintas fuentes de incertidumbre en la evolución del modelo, produciendo predicciones probabilísticas. Hoy día se utilizan en la predicción operativa distintos esquemas de *predicción por conjuntos* que perturban las condiciones iniciales, o combinan las salidas de distintos modelos (*multi-modelo*), para obtener un conjunto de predicciones con el que anticipar la evolución de la atmósfera. A continuación se muestra una descripción del abanico de predicciones que se realizan en distintos centros meteorológicos de forma operativa.

1.7.1. *Nowcasting. Predicción Inmediata*

Marca el primer umbral de predicción y se refiere a la predicción a muy corto plazo (minutos/horas). En este alcance, el radar, las imágenes de satélite y las labores de vigilancia llevadas a cabo por observadores y predictores humanos desempeñan el papel principal, relegando a los modelos numéricos a un segundo plano. El objetivo principal de este tipo de predicción es anticipar la intensidad de eventos extremos (fuertes tormentas, etc.), así como la localización geográfica de los mismos. En este libro no se abarca este alcance de predicción, pues supone el estudio de técnicas especiales para este tipo de fenómenos.

1.7.2. *Predicción a Corto Plazo*

Se entiende en un rango de entre 1 y 3 días de alcance. En un principio se realiza a nivel global con una resolución horizontal que oscila entre 0.5 y 1° . Por ejemplo, el ECMWF proporciona salidas cada 6 horas sobre todo el globo con 0.6° (aprox. 60 km) a partir de un modelo de circulación general T511. Este centro es una organización multinacional que aglutina esfuerzos de los distintos países europeos para realizar una predicción base que sirva de soporte para el resto de servicios meteorológicos. Las salidas de los modelos del ECMWF se utilizan en los distintos servicios meteorológicos europeos como condiciones de contorno para modelos regionales de mayor resolución. Por ejemplo, el producto final de predicción operativa del INM para España se obtiene aplicando el modelo regional HIRLAM, con una resolución espacial de 0.2° , a las salidas del ECMWF (ver Fig. 1.10).

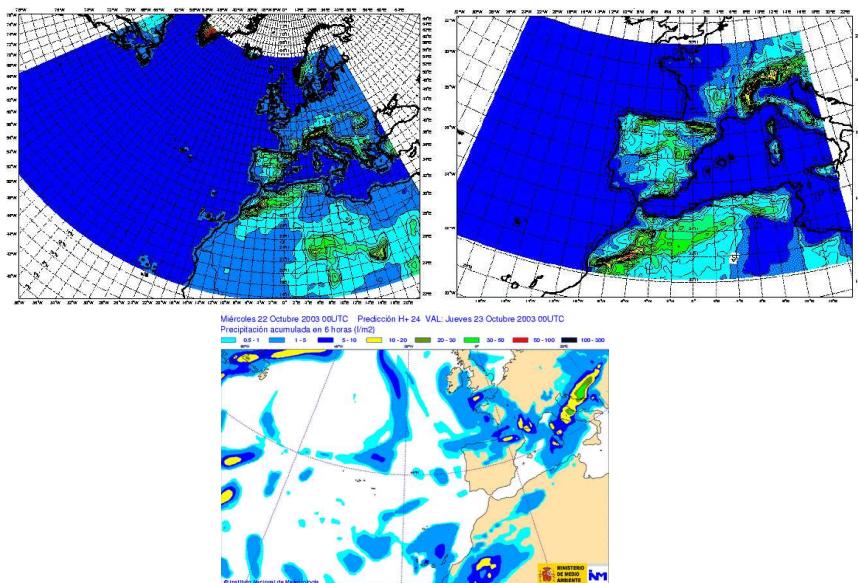


Figura 1.10: (superior) Orografía de los modelos HIRLAM de 0.5° y 0.2° utilizados por el INM en la predicción operativa. (inferior) Salida de precipitación para un alcance de 24 horas (D+1) del modelo con una resolución de 0.5° (FUENTE: Instituto Nacional de Meteorología).

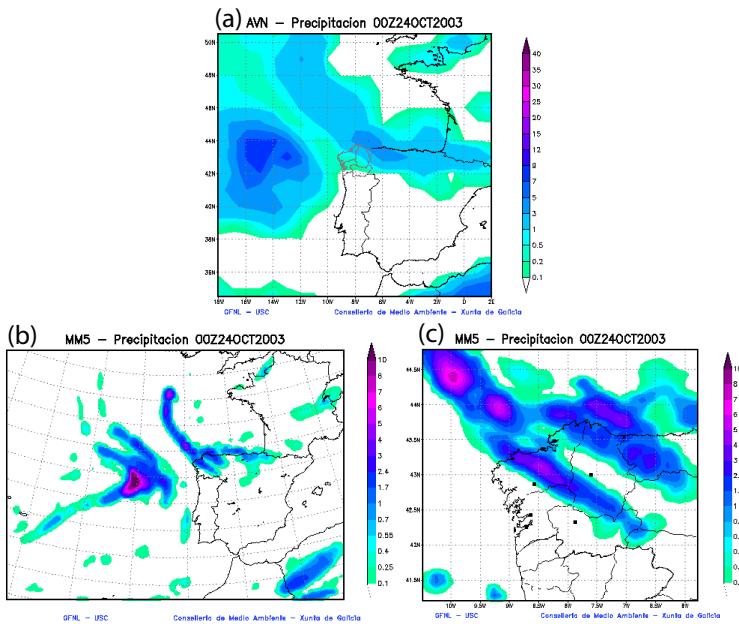


Figura 1.11: Campos de precipitación previstos para D+1 por (a) el modelo AVN del NCEP; y el modelo MM5 utilizado por Meteo Galicia con (b) 30 y (c) 10 km de resolución. (FUENTE: Página web de Meteo Galicia).

De forma análoga, en EEUU el NCEP (National Centers for Environmental Prediction, www.ncep.noaa.gov) realiza una predicción global que distribuye libremente a través de Internet, y que es luego utilizada por distintos centros regionales en diversas aplicaciones. Por ejemplo la Fig. 1.11 muestra el campo de precipitación estimado por el modelo del NCEP y los campos estimados por el modelo regional MM5 utilizado en *Meteo Galicia meteo.usc.es* con 30 y 10 km de resolución.

En España, son numerosos los grupos de investigación que llevan a cabo integraciones a corto plazo sobre regiones concretas de nuestra geografía. Por ejemplo la red Ibérica MM5 redibericamm5.uib.es aglutina a los grupos de investigación que utilizan en modelo MM5 para distintos estudios de modelización y predicción.

1.7.3. Predicción a Medio Plazo por Conjuntos

Abarca el período comprendido entre los 4 y 15 días de alcance. En este período el sistema tradicional de predicción determinista comienza a perder predecibilidad fruto de la no linealidad del modelo y los efectos de la incertidumbre. Por tanto, en este período, la predicción se entiende en sentido probabilístico y se utilizan técnicas de predicción por conjuntos para tener en cuenta estos efectos. La predicción por conjuntos se lleva a la práctica integrando el sistema varias veces utilizando un número arbitrario de condiciones iniciales distintas que se obtienen perturbando de forma apropiada la condición inicial obtenida a partir de las observaciones (ver Sec. 1.6.1 para más detalles). Como resultado, se obtiene un conjunto de campos previstos para un mismo alcance, que han de procesarse de forma apropiada para obtener una predicción (numérica o probabilística).

En cuanto a los sistemas operativos globales de predicción por conjuntos, sólo los grandes centros de predicción numérica tienen la capacidad de cálculo necesaria para llevar a cabo las costosas integraciones. Por una parte, desde Noviembre de 2000, el ECMWF ha puesto en marcha un modelo de predicción por conjuntos a plazo medio (con alcance de 10 días) basado en 50 integraciones con condiciones iniciales perturbadas y una integración con condiciones sin perturbar de un modelo *T255L40* (aprox. 80 km de resolución en latitudes medias). En este caso, las perturbaciones se obtienen con el método de “singular vectors” (ver Sec. 1.6.1). Por ejemplo, la Fig. 1.12 muestra la presión media a nivel del mar para los nueve primeros miembros del “ensemble” previstos el 29/12/2003 con un alcance de 72 horas. En este ejemplo ya se pueden apreciar sutiles diferencias entre unos miembros y otros incluso para este alcance de tres días que estaría en la frontera entre predicción a corto y medio plazo. Por otra parte, la Fig. 1.13 muestra la predicción probabilística obtenida a partir del EPS, que resume el efecto de la incertidumbre sobre los distintos patrones de circulación atmosférica previstos sobre un evento dado ($precip > 10\text{ mm}$).

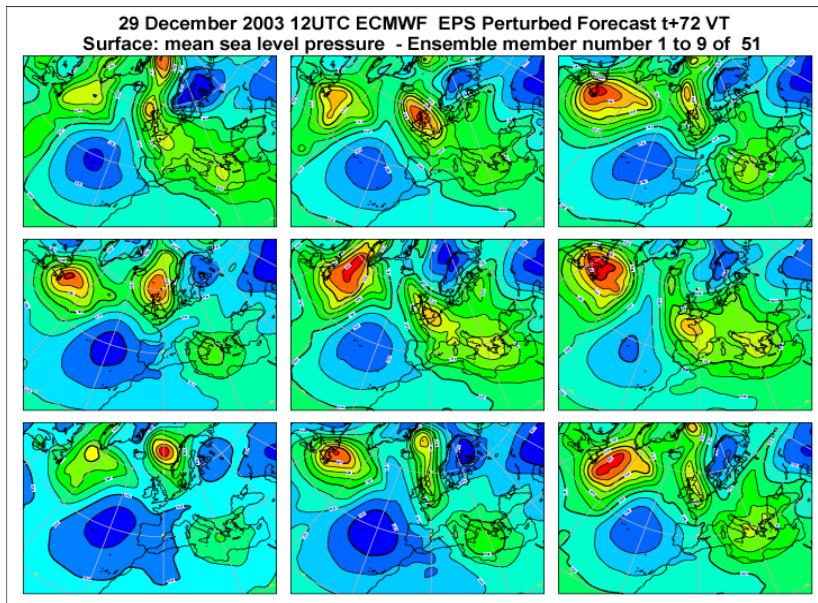


Figura 1.12: Nueve primeros miembros de la predicción por conjuntos del ECMWF previstos el día 29/12/2003 con un alcance de 72 horas. Se muestra la presión a nivel del mar. (FUENTE: ECMWF).

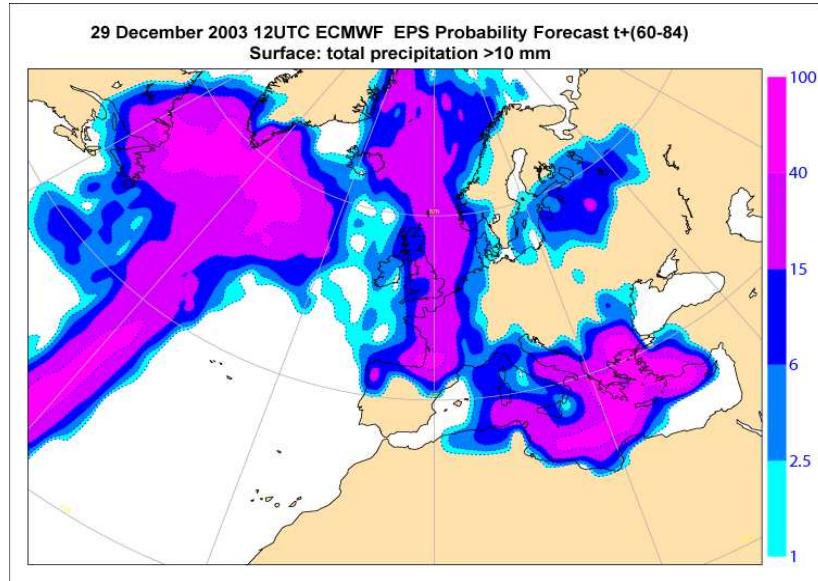


Figura 1.13: Predicción probabilística del evento *precip.* > 10 mm obtenida con los 50 miembros del EPS (FUENTE: ECMWF).

Por otra parte, el NCEP también produce de forma operativa predicciones por conjuntos a medio plazo (hasta 15 días) utilizando el Global Forecast System model (GFS) (ver www.nco.ncep.noaa.gov/pmb/products/).

Paralelamente al crecimiento de los recursos computacionales, especialmente la capacidad de cómputo, la resolución de los modelos se va incrementando, permitiendo mejorar la resolución de las costosas simulaciones por conjuntos de alcance medio. Por ejemplo, la resolución del modelo actual del ECMWF es cuatro veces superior a la del modelo anterior puesto en marcha en Diciembre de 1992, y se ha constatado que este aumento de resolución ha mejorado sustancialmente la predicción probabilística de la precipitación (ver Buizza et al., 2001, para más detalles). Por tanto, es previsible que en el futuro cercano la predicción a corto plazo y la predicción por conjuntos a plazo medio se fundan en una sola predicción, extendiendo la aplicabilidad de la predicción por conjuntos (de hecho, se están dando ya las primeras experiencias de extender la predicción por conjuntos al corto plazo, buscando técnicas perturbativas apropiadas).

1.7.4. *Predicción Estacional por Conjuntos*

En la predicción estacional ya no se trata de predecir el estado real de la atmósfera en un instante de tiempo dado, sino la anomalía en la circulación (desviación respecto del comportamiento promedio) para un trimestre. En este supuesto, la condición inicial (el estado inicial de la atmósfera) no es el factor más importante para la identificación de anomalías en la evolución de la circulación atmosférica, sino que existen otros factores más determinantes, como la temperatura del agua del océano. Por tanto, los modelos numéricos apropiados para este tipo de predicción son modelos numéricos acoplados atmósfera-océano; en algunos casos, también se consideran modelos atmosféricos que tienen forzadas las temperaturas del agua del mar. Por ejemplo, el ECMWF integra mensualmente y hasta 180 días un modelo de predicción por conjuntos denominado System-II para la elaboración de predicciones de anomalías mensuales y estacionales (en este caso, las distintas condiciones iniciales se obtienen perturbando positiva y negativamente las temperaturas del agua del mar). El Ejemplo 1.1 muestra un caso práctico de predicción con este sistema. El modelo surgió de la experiencia llevada a cabo en el proyecto DEMETER (*Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction*) en el que se construyó un sistema *multi-modelo* de predicción estacional compuesto por seis modelos globales de predicción por conjuntos con distintos esquemas de perturbación (www.ecmwf.int/research/demeter/). De hecho, se dispone de un re-análisis de este sistema multi-modelo que abarca el mismo período de ERA-40.

Ejemplo 1.1 (Predicción de El Niño). Uno de los ejemplos más conocidos de predicción estacional es el fenómeno de El Niño; en concreto, la zona Niño-3 comprende un área del Océano Pacífico entre las latitudes $5N - 5S$ y longitudes $90W - 150W$. El promedio de la temperatura del agua del mar en esta región se suele utilizar como indicador de la actividad de El Niño. La Figura 1.14 muestra las predicciones realizadas por el modelo estacional, System-II del ECMWF en (a) abril y (b) agosto, para cada uno de los miembros del conjunto de predicciones. En ambas figuras se muestra el valor real ocurrido en trazo discontinuo. En la figura (a) casi todos los miembros de la predicción indican una anomalía positiva en el futuro. En cambio, en la figura (b) unos miembros del conjunto indican anomalía positiva, mientras otros indican lo contrario. En este caso, la situación se supone menos predecible y la confianza asociada a la predicción es baja. En la figura inferior se muestra la probabilidad obtenida a partir del conjunto de predicciones, para una anomalía positiva de la temperatura en la superficie del mar durante el trimestre septiembre-noviembre de 2003. ■

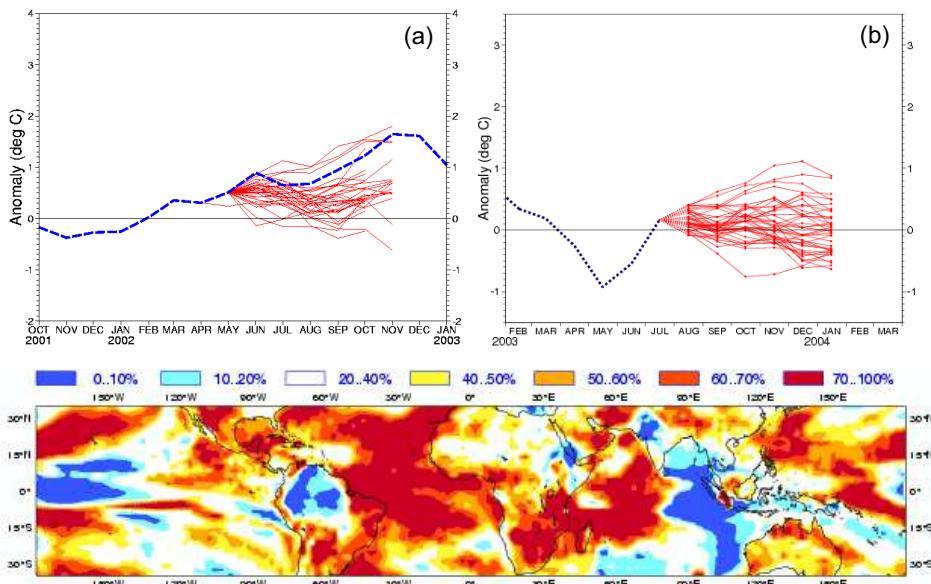


Figura 1.14: Predicción por conjuntos del fenómeno de El Niño con un conjunto de 51 miembros. (a) La predicción realizada en abril muestra una anomalía positiva hasta octubre; (b) los miembros de la predicción realizada en julio no coinciden en una misma predicción. La figura inferior muestra la probabilidad (en tanto por ciento) predicha el 1 de agosto de que la temperatura a 2m se encuentre en el tercil superior (anomalía positiva) en el período SON 2003. (FUENTE: ECWMF)

1.7.5. Predicción Climática

Finalmente, dentro de este rango, varios centros climáticos europeos (Centro Hadley en el Reino Unido, Instituto Max Plank en Alemania, Centro Nacional de Investigación de Méteo-France) llevan a cabo integraciones durante 100 o más años bajo diversas hipótesis de forzamiento radiativo para la realización de estudios de escenarios climáticos (por ejemplo, los efectos de la duplicación de las emisiones de CO_2 en la atmósfera). En estos casos, los modelos son globales y están acoplados con modelos oceánicos, por lo que su resolución es pequeña (entre 2.5° y 5° , es decir 250-500 km aproximadamente).

1.7.6. Tipos de Predicción: Determinista y Probabilística

Cada uno de los tipos de predicción anteriores está influido en mayor o menor medida por la no linealidad de la atmósfera, que impone un horizonte a la predicción determinista y hace necesario plantear la predicción en términos probabilísticos a partir del medio plazo. La no linealidad también limita la resolución espacial con que puede obtenerse una predicción para un alcance dado (a medida que aumenta el alcance, decrece la resolución espacial de los fenómenos que se pueden predecir). La Figura 1.15 ilustra estas relaciones y detalla el tipo de predicción (determinista o probabilística) adecuada para cada uno de los alcances. La predicción en términos probabilísticos también puede ser una elección, por ejemplo en el corto plazo, para cuantificar la incertidumbre de las predicciones (ver Gutiérrez et al., 2004).



Figura 1.15: Relación entre el alcance de las predicciones y su resolución espacial.

1.8. *Futuro de la Predicción Numérica*

Desde la perspectiva actual, el futuro de la predicción numérica se ve centrado en los desarrollos teóricos y prácticos para comprender los efectos de la incertidumbre y su propagación en los modelos numéricos; ésto incluye al proceso de asimilación y a los métodos de perturbación para predicción por conjuntos. Los siguientes problemas marcarán el futuro próximo de la predicción numérica (ver también Kalnay (2003)):

- Nuevas técnicas de asimilación de datos capaces de extraer la máxima información de los sistemas de observación, especialmente de satélites y radares.
- Modelos numéricos de fenómenos convectivos locales (tormentas) a corto alcance, con capacidad predictiva de fenómenos adversos.
- Nuevos métodos para modelizar la evolución de la incertidumbre (errores) en los modelos numéricos: perfeccionamiento de las técnicas de predicción por conjuntos y nuevas técnicas de evolución estocástica.
- Sistemas de alta resolución que acoplen la atmósfera con los procesos de suelo para aplicaciones en hidrología, agricultura, etc., con capacidad para pronosticar localmente distintas variables de interés social y productivo.
- Un mayor uso de modelos atmósfera-océano-tierra en los cuales anomalías de larga duración tales como temperatura de la superficie del mar y humedad de suelo, lleven a mejores predicciones estacionales y climáticas.
- Aplicaciones de los modelos numéricos en sectores productivos relacionados con la actividad humana: agricultura, energía, prevención de riesgos naturales, etc., y en problemas que afectan a la salud tales como contaminación atmosférica, transporte de contaminantes y radiación ultravioleta.

En este libro se abordan algunos de estos problemas utilizando nuevas técnicas estadísticas de aprendizaje automático; en concreto se analiza la predicción local mediante técnicas híbridas y se aplican los métodos desarrollados para los sistemas de predicción por conjuntos.

1.9. *Datos Climatológicos y Meteorológicos*

Los modelos numéricos y las distintas técnicas estadísticas de post-proceso se valen de las observaciones de variables meteorológicas de la red mundial de observatorios, e incluso de datos paleoclimáticos obtenidos indirectamente a partir de mediciones de hielo, anillos de árboles, etc. En esta sección se describen las fuentes de esta cantidad ingente de información, que será utilizada en el resto del libro. Asimismo, se muestra la forma de utilizar la Toolbox *MeteoLab* para trabajar con estos conjuntos de datos.

1.9.1. Redes de Observación

Las observaciones rutinarias de la atmósfera (en especial en superficie) han ido extendiéndose de forma continua desde la creación de la OMM en 1950, propiciando el marco intergubernamental para el establecimiento de redes internacionales de observación. El GCOS (Global Climate Observing System) es el sistema coordinado que recopila y comprueba las observaciones de variables atmosféricas y oceánicas (en superficie) a nivel global. El núcleo de las observaciones superficiales consta de aproximadamente 10000 observatorios que realizan observaciones al menos cada seis horas, y a menudo horariamente. Las variables observadas son: presión atmosférica, viento, temperatura del aire y humedad relativa. Además existen unas 1000 estaciones de radiosonda y más de 3000 aviones que realizan observaciones en varios niveles de la atmósfera. Unos 7300 barcos, 600 boyas a la deriva, 300 boyas fijas y 600 plataformas proporcionan observaciones en los océanos. Estos datos son utilizados diariamente por los modelos numéricos de circulación atmosférica, asimilando el estado de la atmósfera inicial en base a un conjunto de estas observaciones. Por ejemplo, la Figura 1.16 muestra las observaciones utilizadas un día particular en el ECMWF para el proceso de asimilación (inicialización) de su modelo operativo.

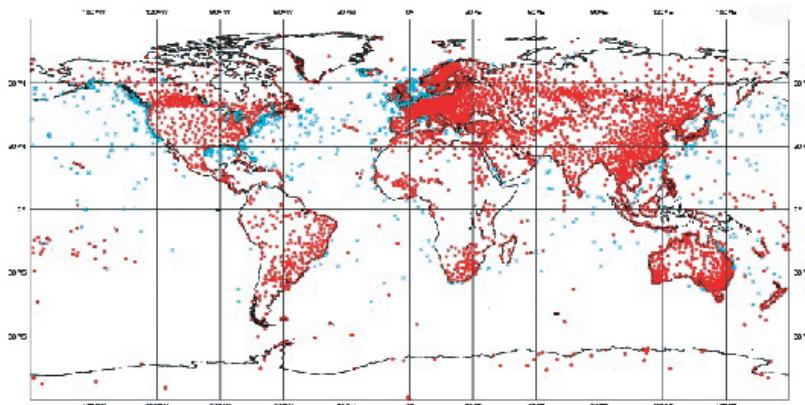


Figura 1.16: Observaciones utilizadas en la asimilación de datos del modelo numérico del ECMWF. (FUENTE. Página Web del ECMWF).

La longitud de las series de datos almacenadas es muy variable y oscila entre un par de cientos de años (para estaciones históricas), hasta unos pocos años, o meses. La información de un subconjunto de estas estaciones de superficie (GSN, GCOS Surface Network) puede consultarse directamente en internet (www.ncdc.noaa.gov/servlets/gsn).

El GCOS también dispone de un subsistema espacial de observación, integrado por cuatro satélites en órbita polar y cinco geoestacionarios, con el

objetivo principal de completar la cobertura de las observaciones de la red en superficie. El papel de las observaciones procedentes de satélites es cada vez más importante y la asimilación de estos datos en los modelos numéricos es un área de intensa investigación.

Por otra parte, los distintos servicios meteorológicos nacionales disponen de su propia red de observación (algunas de cuyas estaciones se integran en el GCOS). Estas redes, mucho más densas, abarcan gran variedad de variables climatológicas. Por ejemplo, el Instituto Nacional de Meteorología (INM) dispone de una red de observatorios que abarca más de 6000 puntos geográficos en las distintas cuencas hidrográficas Españolas (ver Fig. 1.17(b)) con mediciones diarias de precipitación y meteoros (tormenta, nieve, granizo, niebla, lluvia, calima, rocío y escarcha) en el período 1900-2000 (las series tienen longitudes variables y contienen lagunas, pero en promedio se dispone de más de 30 años de información). Las temperaturas máxima y mínima se observan en 2281 puntos. Además de la red secundaria, existen algunas estaciones en las que se realizan observaciones por parte de personal cualificado. Estas estaciones corresponden a la red principal y en la actualidad consta de 225 puntos de observación en los que se mide insolación, evaporación, recorrido, dirección y racha máxima de viento, temperaturas medias, y otras variables. Las observaciones se realizan diariamente, aunque en la red principal se realizan observaciones con mayor frecuencia (cada 10 min, cuatro veces al día, etc.). En total se tiene un gran volumen de información que es necesario tratar de forma eficiente.

A continuación se detallan las observaciones disponibles; salvo indicación expresa, las observaciones son diarias y hacen referencia a un periodo de 24 horas comprendidas de 07 UTC a 07 UTC del día siguiente:

- Observaciones en la red secundaria (Fig. 1.17(d)):
 1. *Temperaturas extremas* máxima y mínima a 2 m del suelo (en °C).
 2. *Precipitación* diaria acumulada (en $mm \equiv l m^{-2}$).
 3. Ocurrencia de *meteoros*: niebla, tormenta, granizo, nieve, escarcha y rocío.
- Observaciones en la red principal (Fig. 1.17(c)). Además de incluir las observaciones anteriores, posee una serie adicional de mediciones:
 1. *Racha máxima de viento*, medida de 0 a 24 horas a 10m del suelo (en $km h^{-1}$). La dirección de la racha máxima se mide en grados (tomando como origen el Norte).
 2. *Insolación* medida de 0 a 24 horas (en horas diarias de sol). Si es relativa se mide en % respecto a la máxima teórica del día.
 3. *Evaporación potencial*, considerada como la máxima evaporación posible, medida de 0 a 24 horas (en mm) (existen algunas medidas en la red secundaria).

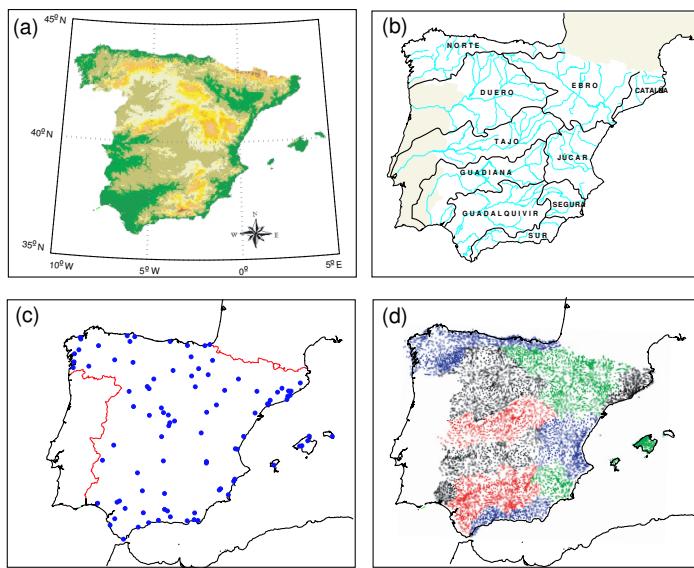


Figura 1.17: (a) Orografía de España peninsular y Baleares y sus (b) cuencas hidrográficas principales. (c) Red principal de estaciones del INM, (d) red secundaria (pluviométrica y termopluviométrica).

4. En la red principal existen otras muchas variables que se podrían incluir entre las observaciones. Algunas de ellas son: Intensidad máxima y duración de precipitación, recorrido del viento, cantidad y tipo de nubosidad, humedad media, y visibilidad.

1.9.2. MeteoLab: *Observaciones*

En esta sección se introduce por primera vez la toolbox *MeteoLab* desarrollada utilizando el programa de cálculo científico *Matlab* (www.mathworks.com). *MeteoLab* es el acrónimo de Meteorological Laboratory for Machine Learning. Esta toolbox ha sido desarrollada por los autores del libro y se distribuye libremente a través de la Web grupos.unican.es/ai/meteo (ver el prefacio del libro para más detalles). Esta toolbox incluye comando sencillos para leer datos de observaciones y de salidas de modelos numéricos (en formato GRIB), así como otros comandos que implementan los distintos métodos descritos en este libro (tanto lineales, basados en las técnicas estadísticas estándar, como no lineales, basados en las técnicas modernas de aprendizaje automático).

Lo primero que se ha de hacer al iniciar cada sesión es ejecutar el script de inicio `init.m` del directorio `MLToolbox/MeteoLab`. A continuación, se puede cargar los datos correspondientes a las observaciones de una variable en un

conjunto de estaciones para un período de tiempo determinado. La Toolbox incluye un conjunto de datos de precipitación, temperatura y presión en un conjunto de estaciones europeas de la red GSN (GCOS Stations Network), descrita en la Sec. 1.9.1. Por ejemplo, para cargar los datos de precipitación disponibles en las estaciones de Europa se procedería de la siguiente forma:

```
>> Example1.Network={'GSN'};  
>> Example1.Stations={'Europe.stn'};  
>> Example1.Variable={'Precip'};  
>> [data,Example1]=loadStations(Example1,'ascfile',1);
```

MeteoLab trabaja con las redes de observaciones a través de una estructura que contiene tres campos: **Network**, **Stations** y **Variable**; en el ejemplo anterior esta estructura se llama **Example1**. La función **loadStations** devuelve los datos **data** de las estaciones, así como la estructura completada con las características de los datos leídos (latitud y longitud de las estaciones, etc.):

- **Network** indica el nombre de la red que se va a utilizar. Existe un conjunto de redes predefinidas en el fichero **Networks.txt** del directorio **MeteoLab/Observations**, pero cada usuario puede incluir de forma sencilla sus propias redes de datos. Cada red tiene asociado un directorio que contiene un fichero maestro **Master.txt** con la información sobre las estaciones existentes (identificador, nombre de la estación, longitud, latitud, y meta-datos); este fichero se halla en el directorio **ObservationsData**. En concreto, existen datos de la red GSN (que se utilizan en los distintos ejemplos del libro), así como una red de propósito general **MyStations** para que cada usuario pueda añadir y cargar fácilmente sus propios datos de observaciones. Para ello, basta con seguir las instrucciones que se adjuntan en el directorio **ObservationsData/MyStations**.
- **Stations** indica un fichero que contiene los indicativos del subconjunto de estaciones que deseen cargarse (es decir, el subconjunto de las estaciones del **Master** que quieran cargarse en una aplicación concreta); por ejemplo, en el directorio de la red GSN están definidos por defecto los ficheros **Europe.stn** y **Spain.stn**, que contienen los indicativos de las estaciones europeas y españolas, respectivamente. Este fichero puede estar en el directorio local donde se esté ejecutando *Meteolab*, o en el directorio de la red correspondiente (el indicado en **Networks.txt**).
- **Variable**, es el nombre del subdirectorio de **data** (dentro del directorio de la red correspondiente en **ObservationsData**) donde se encuentran los datos para la variable de interés. Los ficheros pueden ser de formato texto, o binarios. Se han incluido dos ficheros **readObservations** y **saveObservations** en **ObservationsData/MyStations/** explicando la forma de grabar y leer datos en el formato correcto.

Una vez definida la estructura del ejemplo, la función `loadStations` permite cargar los datos indicados. Como salida se obtiene la matriz de datos `data`, que tiene tantas columnas como número de estaciones, y tantas filas como días disponibles en la base de datos (rellenando con `NaN` los días para los que no hay observación); y la estructura completada `Example1` en la que se incluyen varios campos nuevos:

```
>> Example1
Example1 =
    Network: {'GSN'}
    Stations: {'Europe.stn'}
    Variable: {'Precip'}
    Info: [1x1 struct]
    StartDate: '01-Jan-1879'
    EndDate: '31-Dec-2001'
    StepDate: '24:00'
```

El campo `Info` es una estructura que contiene las características relativas a cada una de las estaciones (identificativo, nombre, longitud y latitud, fecha de inicio de los datos, fecha de fin, número de datos, lagunas). Por ejemplo, podemos consultar el nombre de la primera estación `Example1.Info.Name(1,:)`, etc. También es posible dibujar la localización de las estaciones cargadas utilizando el comando `drawStations`:

```
>> drawAreaStations(Example1)
```

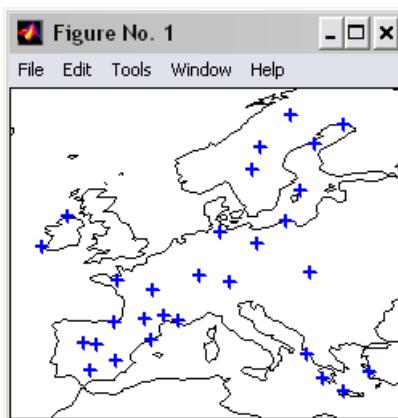


Figura 1.18: Localización de las estaciones europeas de la red GSN incluidas en *MeteoLab*.

También se pueden dibujar las estaciones de forma interactiva con la función `drawObservations` (ver Fig. 1.19, derecha):

```
>> drawObservations(data,Example1)
```

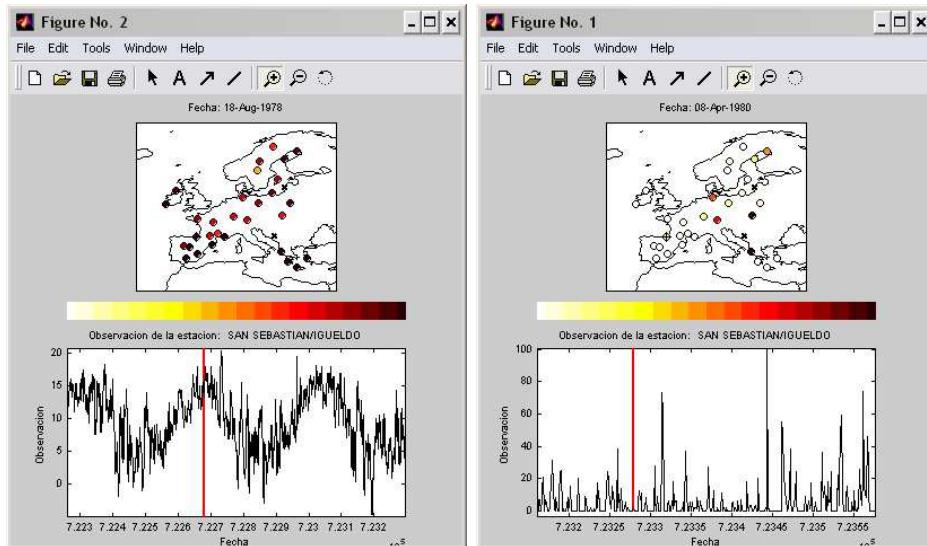


Figura 1.19: (izquierda) Datos de temperatura máxima de las estaciones europeas de la red GSN; (derecha) datos de precipitación.

El comando `loadStations` tiene un parámetro opcional, `dates`, que permite indicar el rango de fechas en que se desean cargar los datos; en caso de ser omitido se cargan las observaciones de todas las fechas extremas registradas (como en el ejemplo anterior). El siguiente ejemplo muestra la forma de especificar un rango de fechas y cargar los datos sólo en ese período. De esa forma, los parámetros `StartDate` y `EndDate` de la estructura definida en el ejemplo (`Example1` en este caso) abarcarán el período considerado para todas las estaciones:

```
>> date={'1-Jan-2000','31-Dec-2000'};
>> [data,Example1]=loadStations(Example1,'dates',date,'ascfile',1);
>> [Example1.StartDate; Example1.EndDate]
ans =
    01-Jan-2000
    31-Dec-2000
```

De forma similar se pueden cargar datos para otras redes de observación y representar gráficamente los valores. Por ejemplo, la red denominada **MyStations** contiene datos diarios de temperatura máxima, racha máxima de viento, y precipitación de tres ciudades españolas: Oviedo, Barcelona y Salamanca, así como datos horarios de temperaturas en Oviedo. Los datos de esta red de estaciones pueden ser cargados de la forma antes explicada.

Por otra parte, si se dispone de datos de otra red de estaciones (por ejemplo datos de la red secundaria de estaciones del INM), se puede adaptar fácilmente a la estructura de *MeteoLab* para poder ser cargados desde la Toolbox. El fichero **readme.txt** del directorio **ObservationsData/MyStations** describe la forma de configurar nuevas redes de observación en el programa. Por ejemplo, la Fig. 1.20(izquierda) muestra los datos de racha máxima de la principal de estaciones del INM:

```
>> Example2.Network={'INM'};
>> Example2.Stations={'completas.stn'};
>> Example2.Variable={'Vx'};
>> [data,Example2]=loadStations(Example2,'zipfile',1);
>> drawAreaStations(Example2);
>> drawObservations(data,Example2);
```

De forma similar, se puede cargar la precipitación en las estaciones de la red secundaria en el cantábrico (ver Fig. 1.20, derecha). El fichero **Examples.m** del directorio **MeteoLab/Observations** contiene ejemplos adicionales del uso de los comandos para cargar observaciones.

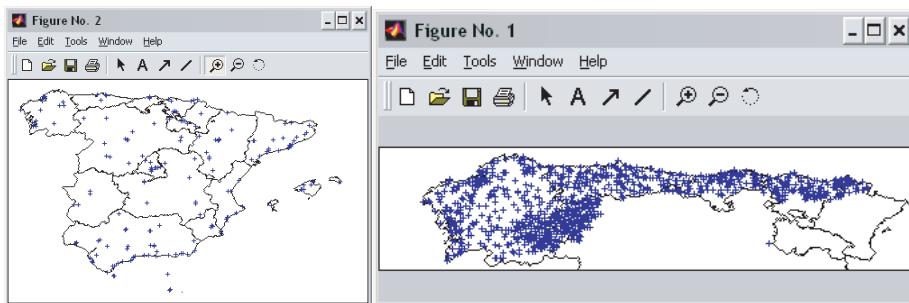


Figura 1.20: Datos de precipitación en las estaciones de (izquierda) la red principal del INM y (derecha) la red secundaria del INM de la Cuenca Norte.

1.9.3. *Datos Paleoclimáticos*

Los datos instrumentales son medidas directas representativas de las variables meteorológicas; sin embargo, las series más largas no pasan de unos pocos cientos de años. Para llevar a cabo estudios climáticos a tiempos muy largos

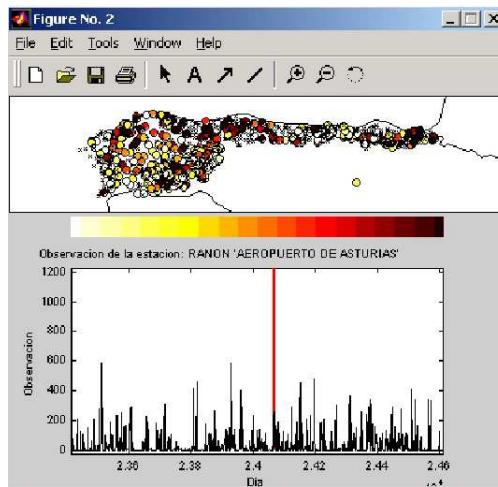


Figura 1.21: Datos de precipitación en las estaciones de la red secundaria del INM de la Cuenca Norte.

(del orden de miles de años) es necesario disponer de series de datos más largas. Por ello, se han desarrollado distintas técnicas indirectas de medición a partir de indicadores de la naturaleza (datos “proxy”). Los más utilizados se obtienen de los árboles (fisiología de los anillos, análisis de isótopos, etc.), las catas de hielo (composición de isótopos, acumulación y estratificación, etc.), los corales (crecimiento, etc.), sedimentos oceánicos, fósiles, etc. Por ejemplo las extracciones del hielo profundo de glaciares contienen muestras de burbujas de aire, polvo, polen, o isótopos de oxígeno, que sirven para reconstruir el clima pasado, del área donde fue recogida la muestra.

1.9.4. Simulaciones de Modelos Numéricos

Los campos resultantes de las integraciones de los modelos numéricos del tiempo caracterizan el estado de la atmósfera, tanto los análisis (estado inicial asimilado de la atmósfera), como las predicciones, y en muchos casos están disponibles para la comunidad investigadora. Por otra parte, también están disponibles los campos producidos por diversos proyectos de reanálisis en los que se ha integrado off-line un mismo modelo para un largo período de tiempo representativo de la climatología. Por ejemplo, el primer proyecto de reanálisis global llevado a cabo en el ECMWF se denominó ERA-15 y proporciona los campos de análisis y las predicciones a corto plazo cada seis horas obtenidas con un modelo $T106L31$ (1.125 grados de resolución) para el período comprendido entre Diciembre-1978 y Febrero-1994. Recientemente, este proyecto ha sido extendido y ya se dispone de información para el período Septiembre 1957 - Agosto 2002 (ERA-40) obtenida con un modelo de mayor resolución ($T159L60$, aprox. 0.675 grados). El NCEP también dispone de un reanálisis llevado a ca-

bo con un modelo *T62L28* (1.875 grados de resolución) para el período 1958 hasta la actualidad (Kalnay et al., 1996); para más información sobre actualizaciones del reanálisis consultar wesley.wwb.noaa.gov/reanalysis2. La lista completa de variables disponibles para el proyecto ERA puede consultarse en www.ecmwf.org/research/era, mientras que en el caso del NCEP puede consultarse wesley.wwb.noaa.gov/reanalysis.html.

En este libro se utiliza información de cada uno de los 4 análisis diarios de ERA-40, a las 00, 06, 12, y 18 UTC para definir el estado de la atmósfera en un momento dado. Se dispone de variables en superficie y en niveles de presión, y de variables derivadas que no son integradas directamente en el modelo. Básicamente las variables utilizadas en altura son el geopotencial (*Z*), temperatura (*T*), velocidad del viento (*U* y *V*) y humedad relativa (*H*) en los niveles de presión de 1000, 925, 850, 700, 500, 300 y 200 milibares (mb). Además estas variables también se dan en superficie: presión media a nivel del mar (mslp), temperatura a 2 metros (T2m), velocidad del viento a 10 metros (*U10* y *V10*) y temperatura del punto de rocío (*Td*). En total son $5569 \times 4 \times 181 \times 360 \times 7 \times 5 \simeq 5 \times 10^{10}$ datos. Estos datos se encuentran codificados en formato GRIB (WMO, 1994) ocupando un total de 100 GBytes de información.

En los distintos ejemplos presentados en este libro se consideran distintas regiones y combinaciones de campos para definir el “estado de la atmósfera”. Por ejemplo, para caracterizar el patrón de circulación que afecta a la península Ibérica se pueden utilizar distintas rejillas y escalas temporales:

- *Modelo 1:* Rejilla de $2.5^\circ \times 2.5^\circ$ de longitud y latitud mostrada en la Figura 1.22(a) y que corresponde a la zona del Atlántico Norte, que influencia en gran medida el clima en la Península. En este caso, los patrones se obtienen combinado los campos *T*, *H*, *Z*, *U* y *V* a las 12h en los niveles 1000mb, 925mb, 850mb, 700mb, 500mb y 300mb:

$$\mathbf{x}_{12} = (T_{12}^{1000}, \dots, T_{12}^{300}, H_{12}^{1000}, \dots, H_{12}^{300}, \dots, V_{12}^{1000}, \dots, V_{12}^{300}), \quad (1.19)$$

donde X_i^j denota el campo de la variable *X* a la hora *i* en el nivel *j*.

- *Modelo 2:* Rejilla de $1.0^\circ \times 1.0^\circ$ mostrada en la Figura 1.22(b), cubriendo la zona de estudio. En este caso, se consideran las mismas variables y niveles de presión (altura), pero incluyendo una componente temporal (se toman los campos a las 06h y 30h). Esta componente temporal compensa la reducción de escala de la rejilla teniendo en cuenta efectos de borde y de contorno que podrían alcanzar la zona de la rejilla durante el período de interés.

$$\mathbf{x} = (\mathbf{x}_{00}, \mathbf{x}_{30}). \quad (1.20)$$

- *Modelo 3:* Rejilla de $1.0^\circ \times 1.0^\circ$ mostrada en la Figura 1.22(c). En este caso se considera un patrón atmosférico concreto para cada una de las doce cuencas hidrográficas españolas. Para ello se combinan los campos anteriores en un dominio temporal de mayor resolución: 06, 12, 18, 24,

y 30 UTC. En este caso se cubre el período de predicción con toda la información disponible:

$$\mathbf{x} = (\mathbf{x}_{06}, \mathbf{x}_{12}, \mathbf{x}_{18}, \mathbf{x}_{24}, \mathbf{x}_{30}). \quad (1.21)$$

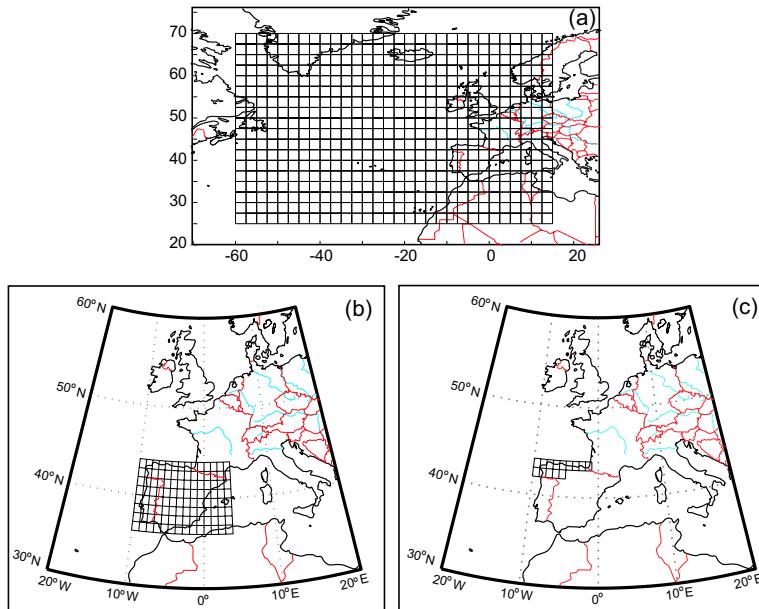


Figura 1.22: Distintas áreas que cubren la península Ibérica: (a) Rejilla de larga escala (macro- β) de $2.5^\circ \times 2.5^\circ$ de latitud y longitud correspondiente al Atlántico Norte; (b) rejilla peninsular (meso- α) $1^\circ \times 1^\circ$ de latitud y longitud; (c) rejilla meso- β $1^\circ \times 1^\circ$ para la cuenca Norte de la Península Ibérica (cada una de las doce cuencas tiene su propia rejilla).

En los casos anteriores se han definido patrones diarios que comprenden el rango horario entre las 6 y 30 horas, correspondiente al horario de medición de la mayoría de las observaciones del INM.

1.9.5. MeteoLab: Patrones Atmosféricos

La forma de trabajar con salidas de modelos numéricos en *MeteoLab* consiste en definir un “patrón atmosférico” como los descritos en la sección anterior; para ello se ha de indicar la región de interés y las variables, niveles y horas que se quieren considerar. El directorio *PatternsData* contiene las zonas definidas por defecto en la Toolbox (NAO e Iberia, correspondientes a los Modelos 1 y 2 de la Fig. 1.22); para cada zona es necesario incluir un fichero *domain.cfg* con la información relativa al dominio y variables atmosféricas (el patrón). Por ejemplo, el fichero *domain.cfg* del directorio *PatternsData/NAO* contiene la información, que corresponde al Modelo 1 de la Fig. 1.22(a):

```

lon=-60:2.5:15
lat=25:2.5:70
nod=1:589

lvl=0
tim=12
par=151

src=$METEOLAB_HOME/..../NWPData/ERA40/SurfaceGlobe25_SST_SLP/
startDate=01-Sep-1957
endDate=31-Aug-2002

```

`lon` y `lat` son las coordenadas del dominio de trabajo (en este caso el Atlántico Norte), a partir del cual se entresaca un subdominio determinado en el parámetro `nod` (numerados de izquierda a derecha y de arriba abajo). La segunda parte del fichero define los datos a considerar: `lvl` son los niveles de presión (definidos en mb), `tim` las horas de análisis y `par` las variables (los códigos del ECMWF para las distintas variables se pueden consultar en www.ecmwf.int), 151 corresponde a la presión a nivel del mar. A continuación se indican las fechas de inicio y fin de los datos que se utilizan (la de este ejemplo se corresponden con las fechas de ERA40). Por último, se indica la ruta donde se hallan los datos, en formato GRIB (se puede utilizar la variable de entorno `$METEOLAB_HOME`, que se asigna al directorio de MeteoLab al iniciar la sesión con `init`).

La función `readDomain` lee la estructura del dominio definida en el fichero `domain.cfg` de la zona definida (las zonas disponibles están definidas en el fichero `MeteoLab/Patterns/Zones.txt`). Por ejemplo, una vez cargada se puede dibujar el dominio definido:

```

>> dmn=readDomain('Nao');
>> drawAreaPattern(dmn);

```

Dado el gran volumen de información necesario para definir los patrones atmosféricos, *MeteoLab* trabaja con componentes principales para comprimir los patrones; las componentes del patrón anterior han sido previamente generadas y se proporcionan con *MeteoLab* (ver Sec. 2.5). Así, se puede cargar un campo definido para un nivel de presión, una hora y un día determinados a partir de un número especificado de componentes principales (ver Fig. 1.23, derecha):

```

>> dates={'01-Aug-1992','01-Aug-1992'};
>> [field,info]=getFieldfromEOF(dmn,'ncp',50,...
>>     'var',151,'time',12,'level',0,'dates',dates);
>> drawGrid(field(1:end,:),dmn)

```

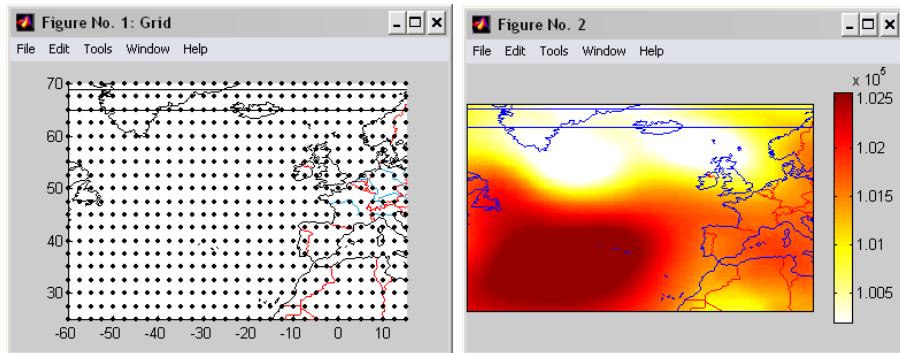


Figura 1.23: (izquierda) Nodos de grid correspondientes a la zona del Atlántico Norte; (derecha) datos reconstruidos de geopotencial en $1000mb$ a las $0Z$ para el día 1 de Agosto de 1992.

`field` es la matriz de los datos reconstruidos, de tamaño $m \times n$ donde m es el número de días y n el número de nodos del dominio. La función `getFieldfromEOF` admite el parámetro opcional '`NCP`' para indicar el número de componentes principales que desean utilizarse para extraer la información (si no se especifica se consideran todas las almacenadas); obsérvese que el campo resultante será tanto más aproximado al real cuanto más componentes se tomen (ver Sec. 2.5).

Trabajar con las salidas directas del modelo es una tarea compleja debido al formato binario especial en el que se manejan este tipo de datos (GRIB, netCDF, etc.). La función `getFieldfromGRIB` permite extraer los datos correspondientes a un patrón `domain.cfg` a partir de los ficheros grib almacenados en el directorio indicado en el `path` del dominio. Este comando devuelve una variable `patterns` que es una matriz de tamaño $m \times n$ con m el tamaño de la muestra (número de días) y n es el producto de número de variables extraídas, horas, niveles de presión y nodos. Si queremos obtener y dibujar un campo concreto de los que componen el patrón (una variable en un nivel de altura y a una hora), la función `findVarPosition` nos devuelve la posición que ocupan en el vector completo la información solicitada. En el siguiente ejemplo se muestra la aplicación de *MeteoLab* para leer y transformar varios campos definidos en el dominio `IberiaPrecip`, que define una rejilla sobre España en la que se consideran las variables precipitación de larga escala y precipitación convectiva. El siguiente código lee los patrones correspondientes al año 1999 y separa los dos campos de precipitación para sumarlos en un único campo que representa la precipitación total.

```
>> dmn=readDomain('IberiaPrecip');
>> ctl.fil='IberiaPrecip.ctl';
>> ctl.cam=dmn.src;
```

```
>> date=datevec(datenum('01-Jan-1999'):datenum('31-Dec-1999'));
>> [patterns,fcDate]=getFRCfromGRIB(ctl,dmn,date,00,00);
>> %Adding large scale and convective precip and drawing the fields
>> precip=1000*(patterns(:,findVarPosition(142,30,0,dmn))+...
>>           patterns(:,findVarPosition(143,30,0,dmn)));
>> precip=sum(precip,1); %Accumulated precipitation
>> drawGrid(precip,dmn);
```

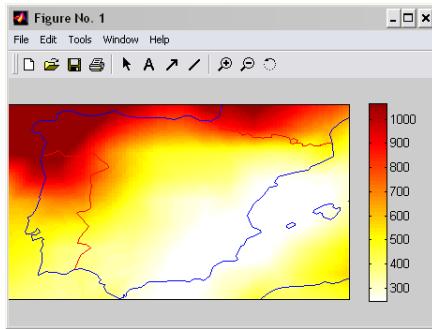


Figura 1.24: Precipitación acumulada (larga escala + convectiva) en *mm* de ERA40 durante el año 1999.

Por último, *MeteoLab* permite trabajar con dominios irregulares. Por ejemplo, la figura siguiente muestra el dominio y un campo asociados a la cuenca norte de la península Ibérica (ver PatternsData/Iberia/NortherBasin).

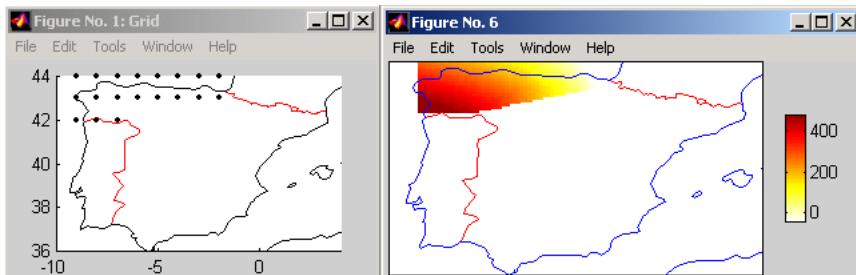


Figura 1.25: (izquierda) Nodos de grid correspondientes a la zona cuencaNorte; (derecha) datos reconstruidos de geopotencial en 1000mb a las 0Z para la cuenca Norte el día 1 de Enero de 1990.

Junto con los re-análisis de modelos numéricos que cubren períodos continuos de tiempo, también se han elaborado bases de datos de observaciones sin lagunas y homogeneizadas sobre rejillas. Por ejemplo Chen et al. (2002) describe la elaboración de una base de datos con medias mensuales de precipitación en superficie para una rejilla de 2.5° de resolución sobre el globo.

Parte II

Técnicas Estadísticas Estándar

CAPÍTULO 2

Técnicas Estadísticas. Análisis y Exploración de Datos

2.1. *Introducción*

En este capítulo se describen brevemente algunas herramientas estadísticas básicas para la exploración y el análisis de datos meteorológicos y climatológicos. Estas técnicas han sido utilizadas profusamente para explorar, comprender y simplificar los datos disponibles en un problema dado, analizando las relaciones de dependencia entre variables y eliminando la redundancia. En el capítulo siguiente se analizan técnicas estadísticas de modelización y predicción que operan sobre estos datos simplificados, permitiendo obtener modelos que expliquen las relaciones observadas en los datos. Otras técnicas estadísticas no lineales más avanzadas serán descritas a lo largo del libro. Para una información más detallada de la aplicación de técnicas estadísticas en este ámbito se remite al lector a Ayuso (1994), Wilks (1995) y von Storch and Zwiers (1999).

Este capítulo comienza con una breve introducción a la probabilidad y las variables aleatorias, ilustrada con múltiples ejemplos (Sec. 2.2). A continuación, la Sec. 2.3 muestra la aplicación práctica de estos conceptos para la simulación de series climatológicas (weather generators), manteniendo las principales características estadísticas de la serie original. La Sec. 2.4 analiza el concepto de dependencia entre variables desde un punto de vista lineal y no lineal. La Sec. 2.5 describe el método de componentes principales (y las correspondientes funciones ortogonales empíricas), utilizadas para eliminar la dependencia y redundancia en los datos, representando éstos en un espacio de dimensión inferior donde las variables (o componentes) son independientes. La Sec. 2.6 describe distintas técnicas de agrupamiento.

2.2. Probabilidad y Variables Aleatorias

En la mayoría de las ocasiones, además de los factores controlados por el experimentador en un problema dado, existen otros factores desconocidos que dan lugar a una incertidumbre y a un carácter aleatorio en el problema. La probabilidad es una herramienta apropiada en estas situaciones, donde se conocen los posibles resultados, pero no se sabe cual de ellos ocurrirá en la realización del mismo (ver, por ejemplo Lindley, 1987).

Por ejemplo, si se desea relacionar la *lluvia*, el *viento* y la *estación del año* en un punto geográfico concreto y sólo se dispone de registros históricos de observaciones, tendrá que considerarse el problema como un experimento aleatorio y definir la probabilidad en base a las frecuencias observadas. Por ejemplo, en la Tabla 2.1 se muestran las frecuencias observadas en el aeropuerto de Parayas (Santander) durante un período de diez años.

	Anual		Invierno		Primavera		Verano		Otoño	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	190	99	287	166	360	162	177	89
SE	64	57	24	18	6	4	1	9	33	26
SW	225	661	98	223	18	119	15	71	94	248
NW	288	825	49	150	95	277	108	251	36	147
Total	1591	2059	361	490	406	566	484	493	340	510

Tabla 2.1: Datos de precipitación: lluvia (Ll) o seco (S), estación del año, y dirección del viento (NE, SE, SW, y NW) registrados en Santander (aeropuerto de Parayas) entre 1979 y 1989 ($N = 3650$).

Para tener en cuenta las características de este problema y poder responder a preguntas del tipo “¿lloverá hoy?” es necesario disponer de alguna medida de incertidumbre, como la probabilidad, con la que podamos dar respuestas de la forma “la probabilidad de que llueva es p ”. Para ello, se parte del espacio muestral que incluye todos los posibles resultados del experimento (*sucesos elementales*). Una medida de probabilidad P es una función que asocia un número real en el intervalo $[0, 1]$ a los distintos sucesos, midiendo su grado de incertidumbre. Como es bien sabido, estas funciones han de cumplir los axiomas de normalización $P(\mathcal{M}) = 1$ y aditividad $P(\bigcup_i A_i) = \sum_i P(A_i)$, para cualquier sucesión infinita, A_1, A_2, \dots , de subconjuntos disjuntos. El primer axioma establece que, independientemente de nuestro grado de certeza, ocurrirá un elemento del espacio muestral. El segundo es una fórmula de agregación que se usa para calcular la probabilidad de la unión de subconjuntos disjuntos; establece que la incertidumbre de un cierto subconjunto es la suma de las incertidumbres de sus partes disjuntas (ver DeGroot, 1989, para una introducción rigurosa a la probabilidad).

De estos axiomas pueden deducirse las reglas de cálculo necesarias para operar con esta medida, obteniendo funciones de probabilidad marginales y condicionadas a partir de la probabilidad total.

Ejemplo 2.1 (Precipitación y viento en Santander). A partir de los datos de la Tabla 2.1, se puede definir una probabilidad que relaciona los posibles sucesos relativos a las combinaciones de estación del año, precipitación y viento en la ciudad de Santander. Lo primero es definir el espacio muestral, formado en este caso por todas las posibles combinaciones de elementos entre los conjuntos $\{\text{Invierno}, \text{Primavera}, \text{Verano}, \text{Otoño}\}$, $\{\text{Seco}, \text{Lluvia}\}$ y $\{\text{NE}, \text{SE}, \text{SW}, \text{NW}\}$, es decir,

$$\mathcal{M} = \{\{\text{Invierno}\}, \{\text{Invierno}, \text{Seco}, \text{NE}\}, \{\text{Otoño}, \text{Seco}, \text{SW}\}, \dots\}.$$

Así, en base a las frecuencias observadas dadas en la tabla se puede calcular la probabilidad de cualquier suceso:

$$\rightarrow P(I, NW, Ll) = freq(I, NW, Ll)/N = 150/3650 = 0.041$$

$$\rightarrow P(I, SE, Ll) = freq(I, SE, Ll)/N = 18/3650 = 0.004$$

$$\rightarrow P(I) = \sum_{v \in \{NE, SE, NW, SW\}} \sum_{ll \in \{S, Ll\}} freq(I, v, ll)/N = 0.236 \blacksquare$$

En este ejemplo se usan variables categóricas que toman como valor una determinada categoría (por ejemplo, primavera, verano, otoño o invierno) pero, en la práctica, los sucesos de mayor interés suelen estar asociados a alguna variable numérica relacionada con el experimento (por ejemplo la cantidad de precipitación, *Precip*). El concepto de *variable aleatoria* permite tratar estas situaciones asignando un número real a cada suceso del espacio muestral (por ejemplo, la cantidad de precipitación, en mm). Cuando el rango de la variable aleatoria es un número finito o contable de elementos, ésta se denomina discreta y, en caso contrario se denomina continua.

Dependiendo del tipo de problema, algunas variables pueden considerarse tanto discretas, como continuas. Por ejemplo, la precipitación puede considerarse como variable discreta, con estados $\{0, 1\}$ (correspondiente a la ausencia o presencia del evento de precipitación), pero también puede considerarse como variable continua; en este caso el rango sería el intervalo $[0, \infty)$ (correspondiente a la cantidad de precipitación).

Las siguientes funciones permiten caracterizar la probabilidad de que una variable tome distintos valores (o intervalos de valores) en su rango.

2.2.1. Funciones de Probabilidad

Para una variable discreta X , la *función de probabilidad* $p(x)$ determina la forma en que la probabilidad se distribuye entre los distintos estados de la variable:

$$P(x = a) = p(a) \Leftrightarrow P(a \leq x \leq b) = \sum_{a \leq x \leq b} p(x). \quad (2.1)$$

En cambio, si la variable es continua, la distribución de la probabilidad en los distintos intervalos del rango de la variable está caracterizada por su *función de densidad*, que sustituye las sumas por integrales:

$$P(a \leq x \leq b) = \int_a^b f(x) dx. \quad (2.2)$$

Para que la probabilidad total sea la unidad, estas funciones han de cumplir: $\sum_x p(x) = 1$, $\int_x f(x) = 1$.

Ejemplo 2.2 (Carácter dual de la precipitación). Si se consideran los datos de ausencia o presencia de precipitación se tendrá una variable binaria cuya función de probabilidad podrá obtenerse a partir de las frecuencias de los distintos estados de la variable, según lo visto en (2.1). Las Figuras 2.1(a) y (b) muestran la función de probabilidad de esta variable para las localidades de Almería y Santander, respectivamente, obtenidas a partir de series de valores como las mostradas en la Tabla 2.1. Por otra parte, también es frecuente discretizar la cantidad de precipitación considerando distintos umbrales para la misma; una discretización bastante habitual se establece en base a los umbrales: 0, 2, 10, y 30 mm. Las Figuras 2.1(c) y (d) muestran las correspondientes funciones de probabilidad para las estaciones anteriores.

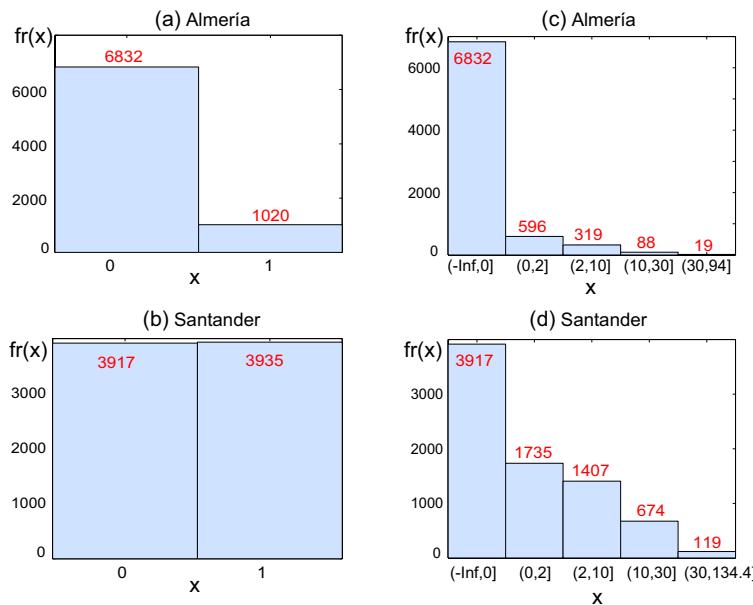


Figura 2.1: Variable binaria “ocurrencia de precipitación” en (a) Almería y (b) Santander. Variables discretas resultantes de una discretización por umbrales de la cantidad de precipitación en (c) Almería y (d) Santander. $fr(x)$ denota la frecuencia absoluta.

Para caracterizar la precipitación como variable continua (cantidad de precipitación), considerando los días en que ocurrió el evento, es necesario definir una función de densidad según (2.2). En este caso, como la función es continua es necesario suponer que la función pertenece a alguna familia paramétrica y estimar sus parámetros en base a los datos disponibles. En la práctica, para la cantidad de precipitación se suelen considerar funciones de densidad exponen-

cial y gamma para aproximar la función de densidad real

$$f_{exp}(x; \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}, \quad (2.3)$$

$$f_{gam}(x; a, b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-x/b}. \quad (2.4)$$

El parámetro μ es la media de los datos de precipitación (6.38 en el caso de Santander y 4.12 en Almería) y a y b son los parámetros de forma y escala de la distribución gamma, respectivamente ($\mu = ba$, $\sigma^2 = b^2a$); obsérvese que para $a = 1$ la función Gamma se reduce a una exponencial. La Fig. 2.2 muestra los histogramas empíricos de la cantidad de precipitación en Almería y Santander, así como las funciones de densidad exponenciales ajustadas a estos datos. En esta figura puede verse que estas funciones subestiman la cola de la función. Por otra parte, la Fig. 2.3 muestra las densidades exponencial y gamma ajustadas con los datos de precipitación en Santander. ■

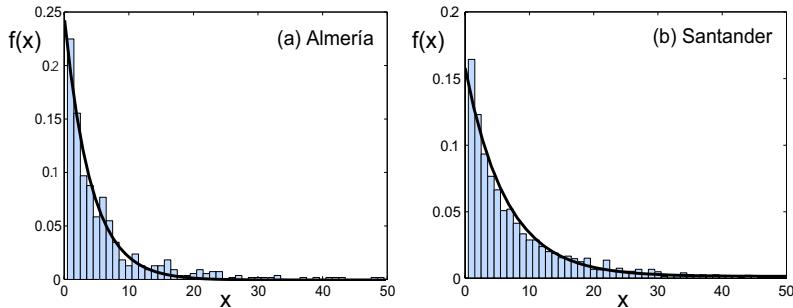


Figura 2.2: Funciones de densidad exponencial y probabilidades empíricas obtenidas a partir de los datos de (a) Almería y (b) Santander.

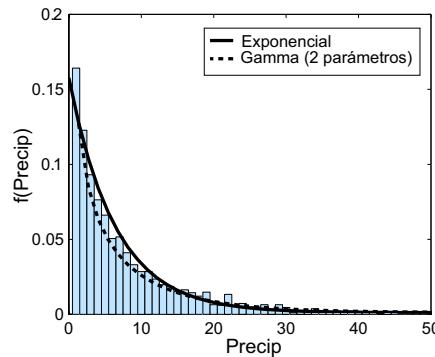


Figura 2.3: Funciones de densidad exponencial y gamma ajustadas a la cantidad de precipitación en Santander.

Una forma equivalente de definir la probabilidad de una variable aleatoria, que es independiente del carácter discreto o continuo de la misma, es su *función de distribución*. La función de distribución de X , denotada por $F(x)$, asigna a cada valor a la probabilidad acumulada $F(a) = P(X \leq a)$, que se traduce en:

$$F(a) = \sum_{x \leq a} P(x); \quad F(a) = \int_{-\infty}^a f(x) dx,$$

para variables discretas y continuas, respectivamente. Por tanto, $F(x)$ es monótona creciente, $F(-\infty) = 0$, $F(\infty) = 1$ y $P(a < X \leq b) = F(b) - F(a)$.

Por ejemplo, integrando la función de densidad exponencial (2.3) se obtiene su función de distribución analíticamente:

$$F(x) = 1 - e^{-\frac{x}{\mu}}. \quad (2.5)$$

La Fig. 2.4 muestra distintas funciones de distribución asociadas con las funciones de probabilidad y densidad anteriores.

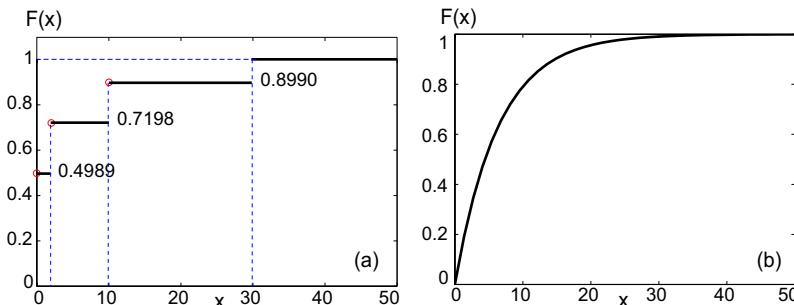


Figura 2.4: (a) Función de distribución asociada a la función de probabilidad de la Fig. 2.1(d). (b) Función de distribución asociada a la función de densidad de la Fig. 2.2(b).

Estas funciones son las que nos permiten simular de forma simple y eficiente muestras de valores de la variable, conservando las propiedades estadísticas básicas de las mismas. Este tema se analizará en detalle en la Sec. 2.3 para su aplicación a los generadores estocásticos de tiempo.

2.2.2. Probabilidades Conjuntas

Cuando la variable aleatoria es multidimensional, los conceptos anteriores siguen teniendo sentido, realizando las sumas e integrales en dominios multidimensionales. De esta forma, la función de probabilidad de una variable aleatoria discreta $\mathbf{X} = (X_1, \dots, X_n)$, denotada por $p(x_1, \dots, x_n)$, es una función que asigna un valor

$$\sum_{a_1 \leq x_1 \leq b_1} \dots \sum_{a_n \leq x_n \leq b_n} p(x_1, \dots, x_n) \quad (2.6)$$

a la probabilidad de que la variable tome valores en un rectángulo arbitrario $\{a_1 \leq x_1 \leq b_1, \dots, a_n \leq x_n \leq b_n\}$. En el caso de variables continuas, el tratamiento es análogo, pero la función ha de ser construida en base a integrales, como en el caso de variables unidimensionales.

Ejemplo 2.3 (Precipitación conjunta). Se considera la variable aleatoria discreta (X_1, X_2) , donde X_1 y X_2 son los datos de precipitación discretizados (utilizando los umbrales 0, 2, 10, 20 y 30 mm) para dos estaciones. A partir de (2.6) se puede calcular la función de probabilidad conjunta; en la figura 2.5 se muestran los resultados obtenidos tomando como variables: precipitación en Santander y Bilbao; Santander y Santiago y, finalmente, Santander y Almería. En estas figuras se puede observar la dependencia existente entre las variables Santander y Bilbao y, en menor medida, Santander y Santiago, mientras que no se observa ninguna relación entre Santander y Almería. El concepto de relaciones entre variables se analiza en detalle en la Sec. 2.4.1. ■

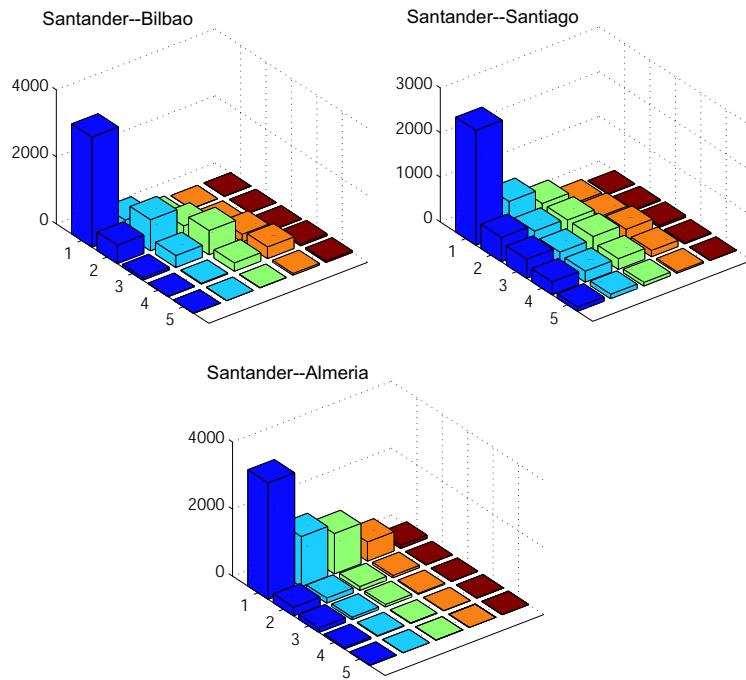


Figura 2.5: Funciones de probabilidad conjunta de la precipitación discretizada para distintos pares de ciudades.

2.2.3. Probabilidades Marginales y Condicionadas

Una vez que se tiene definida una función de probabilidad conjunta, es posible obtener información de cada variable por separado, o incluso la probabilidad de alguna variable cuando se dispone de cierta información sobre las otras. Estas ideas se formalizan en los conceptos de probabilidad marginal y probabilidad condicionada, que son fundamentales para aplicar los mecanismos de inferencia probabilística en la práctica.

A partir de una función de probabilidad $p(x_1, \dots, x_n)$ de una variable aleatoria discreta (X_1, \dots, X_n) , se define la función de *probabilidad marginal* de la variable X_i como

$$p(x_i) = P(X_i = x_i) = \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} p(x_1, \dots, X_i = x_i, \dots, x_n). \quad (2.7)$$

Por otra parte, si X e Y son dos conjuntos disjuntos del conjunto de variables tales que $p(y) > 0$, entonces la *probabilidad condicional*, o función de probabilidad condicionada de X dado $Y = y$, viene dada por

$$p(x|Y = y) = \frac{p(x, y)}{p(y)}. \quad (2.8)$$

La ecuación (2.8) implica que la función de probabilidad conjunta de X e Y puede escribirse como

$$p(x, y) = p(y)p(x|y). \quad (2.9)$$

La ecuación (2.9) puede generalizarse para conjuntos arbitrarios de variables, dando lugar a la llamada *regla de la cadena*:

$$p(x_1, x_2, \dots, x_n) = p(x_1)p(x_2|x_1) \dots p(x_n|x_1, \dots, x_{n-1}). \quad (2.10)$$

Esta fórmula es de gran interés para poder expresar una función de probabilidad como producto de funciones marginales y condicionadas que, en algunos casos, utilizando ciertas condiciones de independencia que posea un modelo dado, podrán simplificarse dando lugar a modelos probabilísticos más sencillos. Esta idea se desarrolla en detalle en el Cap. 4.

Ejemplo 2.4 (Probabilidades condicionadas). A partir de la Tabla 2.1, se puede fácilmente calcular la probabilidad de lluvia (marginal), o la probabilidad de lluvia una vez que se sabe que la racha de viento es del SW (condicionada):

$$\rightarrow P(\text{Lluvia}) = 2059/3650 = 0.564.$$

$$\rightarrow P(\text{Lluvia}|SW) = P(\text{Lluvia}, SW)/P(SW) = 661/886 = 0.746.$$

Las probabilidades anteriores indican que el conocimiento de la variable viento aporta información para el conocimiento de la variable lluvia, ya que la probabilidad de ésta aumenta (y disminuye) con direcciones concretas del viento. Este concepto de dependencia entre variables se desarrolla en la Sec. 2.4. ■

El siguiente ejemplo de Matlab ilustra estos conceptos aplicados a un conjunto de datos de precipitación del GSN descritos en la Sec. 1.9.2.

```
% Loading precipitation data for Spanish stations
Example1.Network={'GSN'};
Example1.Stations={'Spain.stn'};
Example1.Variable={'Precip'}; date={'1-Jan-1998','31-Dec-1998'};
[precip,Example1]=loadStations(Example1,'dates',date,'ascfile',1);
station=1; %Choosing the station to perform the study

% Precipitation as binary variable
binaryPrecip=ones(size(precip,1),1);
binaryPrecip(find(precip(:,station)>0),1)=2;

% Precipitation with 5 states
threshold=[-inf 0 2 10 20];
for i=1:size(precip,1)
    discretePrecip(i,1)=max(find(precip(i,station)>threshold));
end

% Drawing histograms
h1=hist(binaryPrecip,2);
h2=hist(discretePrecip,length(threshold));
figure;
subplot(2,1,1); bar(h1,'hist');
subplot(2,1,2); bar(h2,'hist');

% Exponential density function fitted to data
ind=find(precip(:,station)>0);
x=linspace(min(precip(:,station)),max(precip(:,station)));
mu=mean(precip(ind,station));
fExp=(1/mu)*exp(-x/mu);

% Gamma density function fitted to data
mu=mean(precip(ind,station));
sig2=var(precip(ind,station));
a=mu^2/sig2; b=sig2/mu;
fGam=(x.^(a-1).*exp(-x/b))/((b^a)*gamma(a));

% Using statistics Toolbox functions for a
% more efficient estimation of parameters
par=gamfit(precip(ind,station));
fGam2=gampdf(x,par(1,1),par(1,2));

figure;
plot(x,[fExp; fGam; fGam2],'k');
```

2.3. Generadores de Tiempo (*Weather Generators*)

Los generadores estocásticos de tiempo son técnicas estadísticas que simulan series o muestras de una variable meteorológica (por ejemplo, la precipitación diaria en Santander) a partir de su distribución y/o algún conjunto de estadísticos representativos: medias anuales o mensuales, varianzas, frecuencias de distintos eventos, etc. (ver Wilby and Wilks, 1999, para una introducción). Estos métodos se han utilizado en distintas aplicaciones hidrológicas y agrícolas para simular datos climatológicos a la escala temporal apropiada (por ejemplo, la escala diaria requerida por los modelos de crecimiento de cosechas). Recientemente, también se ha aplicado esta técnica para simular datos diarios a partir de las predicciones mensuales proporcionadas por los modelos numéricos estacionales o climáticos (Feddersen and Andersen, 2004).

La precipitación es una de las variables más importantes para aplicaciones prácticas en agricultura, hidrología, etc. (incluso otras variables como la evapotranspiración, etc., dependen de ella). Por tanto, esta variable desempeña un papel destacado en los generadores estocásticos de tiempo, que en su gran mayoría están adaptados para trabajar con ella. Por ejemplo, Richardson (1981) propuso un método que primero simula la precipitación y en función de ella la temperatura y radiación solar.

2.3.1. Generadores Estocásticos de Precipitación

La principal dificultad para trabajar con la precipitación es su carácter mixto discreto/continuo, pues sólo el evento $Precip = 0$ tiene probabilidad mayor que cero. Por tanto, es necesario distinguir entre una variable binaria “ocurrencia de precipitación” ($Precip = 0$) y una variable continua “cantidad de precipitación” (condicionada a $Precip > 0$). La mayoría de los generadores de tiempo tratan estas dos variables de forma separada, desarrollando métodos apropiados para simular tanto la ocurrencia/ausencia del evento, como su cantidad o intensidad.

Modelos de Ocurrencia del Evento

Las cadenas de Markov de primer orden fueron los primeros modelos aplicados para simular series binarias de precipitación (ausencia/ocurrencia), conservando la correlación temporal y los estadísticos básicos de la estación local estudiada. Gabriel and Neumann (1962) mostraron que estos modelos podían capturar la persistencia natural de la precipitación diaria (esto es, la autocorrelación de la serie), y también podían caracterizar otros aspectos importantes de la ocurrencia de precipitación: longitud de las rachas secas o lluviosas, etc. Estos modelos suponen que la probabilidad de ocurrencia para un determinado día sólo depende de la ocurrencia o no del día anterior. Por tanto, este modelo está caracterizado por las probabilidades:

$$p_{10} = P(Pret_t = 1 | Pret_{t-1} = 0), \quad p_{11} = P(Pret_t = 1 | Pret_{t-1} = 1). \quad (2.11)$$

donde Pre representa la variable binaria “ocurrencia de precipitación” ($Pre = 1$ indica la ocurrencia del evento y $Pre = 0$ la no ocurrencia). Estas probabilidades se pueden estimar a partir de las frecuencias correspondientes de la serie de datos de cada una de las estaciones que se deseen simular.

Esto nos lleva a un simple pero eficiente procedimiento para simular series de ocurrencia/ausencia de precipitación para una estación dada, conocidos p_{01} y p_{11} . Para ello, sólo es necesario utilizar una serie pseudo-aleatoria de números generados por el ordenador a_n , $n = 1, 2, \dots$. Estos números siguen una distribución uniforme en el intervalo $[0, 1]$; por tanto, para cada $n \geq 2$ se simulará un valor Pre_n de la siguiente forma:

- Si $Pre_{n-1}=0$, entonces $Pre = 1$ si $a_n \leq p_{10}$, y $Pre_n = 0$ en caso contrario.
 - Si $Pre_{n-1}=1$, entonces $Pre = 1$ si $a_n \leq p_{11}$, y $Pre_n = 0$ en caso contrario.

El primer día de la serie puede generarse teniendo en cuenta la proporción de días lluviosos y no lluviosos observados.

Este método permite simular series tan largas como se desee manteniendo los estadísticos fundamentales de la estación. Por otra parte, la estructura del método permite inferir algunas propiedades que tendrán las series simuladas. Por ejemplo, las rachas de días secos/lluviosos producidas por este modelo son independientes y siguen la distribución geométrica:

$$P(X = x) = p(1 - p)^{x-1}, \quad x = 1, 2, \dots \quad (2.12)$$

donde $p = p_{10}$ para el caso de rachas secas y $p = 1 - p_{11}$ para rachas lluviosas.

Ejemplo 2.5 (Weather Generators. Ocurrencia de Precipitación). Supongamos que se dispone de una serie de datos diarios de precipitación de una estación (por ejemplo la estación de Santander, en décimas de mm):

Esta serie tiene asociada la siguiente serie binaria de ocurrencia/ausencia de precipitación:

$0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1,$
 $0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1,$
 $1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,$
 $1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1.$

Las probabilidades (2.11) estimadas de este conjunto de datos resultan $p_{10} = 0.395$ y $p_{11} = 0.758$ (éstas probabilidades se han obtenido a partir de las frecuencias: $\text{frec}(Pre_t = 1, Pre_{t-1} = 1) = 47$, $\text{frec}(Pre_t = 1, Pre_{t-1} = 0) = 15$, $\text{frec}(Pre_t = 0) = 38$, $\text{frec}(Pre_t = 1) = 62$). Con estas probabilidades ya se puede comenzar a simular una serie de ocurrencia/ausencia de precipitación. Para ello, el valor del primer día se genera a partir del primer valor

pseudo-aleatorio a_1 : $Pre_1 = 0$ si $a_1 < p_0 = 0.38$ y $Pre_1 = 1$ en caso contrario (supongamos que el primer número aleatorio era 0.15 y que, por tanto $Pre_1 = 0$). A continuación se considera el siguiente número pseudo-aleatorio (supóngase que $a_2 = 0.486$); como el día anterior fue seco y como $a_2 \geq p_{11}$, se tiene un nuevo cero en la serie, y se continúa repitiendo el mismo proceso. El siguiente código de Matlab muestra la forma de realizar este proceso:

```

Example.Network={'MyStations'};
Example.Stations={'dailyPrecip.stn'};
Example.Variable={'common'};
[data,Example]=loadStations(Example,'ascfile',1);
data=data(:,1); % Selecting the first statiaon: Barcelona

Pre(find(data==0),1)=0;
Pre(find(data>0),1)=1;

c=zeros(2,2);
for k=2:size(Pre,1)
    i=Pre(k-1)+1;
    j=Pre(k)+1;
    c(i,j)=c(i,j)+1;
end

p10=c(1,2)/(c(1,1)+c(1,2)); %P(1|0)=P(1,0)/P(0);
p11=c(2,2)/(c(2,1)+c(2,2)); %P(1|1)=P(1,1)/P(1);
p0=sum(Pre(:,1)==0)/size(Pre,1);

N=size(Pre,1); %Length of the simulated series
x=zeros(N,1); u=rand;
%Simulating the first day
if u<=p0, x(1,1)=1;
else x(1,1)=0;
end

%Simulating the rest of the serie
for i=2:N,
    u=rand;
    if (x(i-1)==0) & (u<=p10)
        x(i,1)=1;
    end
    if (x(i-1)==1) & (u<=p11)
        x(i,1)=1;
    end
end

```

Algunos autores observaron que los modelos de Markov de primer orden no eran apropiados para ciertos climas (por ejemplo, pueden generar series con muy pocas rachas secas). Por ello, se introdujeron modelos de Markov de

mayor orden, condicionando la ocurrencia del evento a los k días previos. El problema de estos modelos es que el número de parámetros necesarios crece exponencialmente con el orden. Por tanto, a la hora de decidir entre modelos de distintos órdenes habrá que tener en cuenta la complejidad de cada uno de ellos (para ello se han utilizado criterios como el información de Akaike o el criterio de información Bayesiana).

Los modelos anteriores generan series binarias de precipitación en una única estación; sin embargo, cuando se está interesado en simular series espaciales (para un conjunto de estaciones), son necesarios modelos más generales (ver, por ejemplo Wilks, 1999b). La simulación de series con dependencia espacial a partir de los datos de varias estaciones es todavía un reto.

2.3.2. Modelos de Intensidad del Evento

Una vez que se ha simulado una serie binaria de ocurrencia/ausencia de la precipitación, interesa simular la cantidad de precipitación para los días lluviosos. En el Ejemplo 2.2 se describió la modelización de esta variable en base a una distribución exponencial o gamma; sin embargo, ninguno de los casos es satisfactorio del todo para representar tanto los valores cercanos a cero, como los valores extremos. Recientemente también se ha planteado el uso de una distribución exponencial mixta, que es una mezcla de dos distribuciones exponenciales de medias μ_1 y μ_2 , con función de densidad:

$$f(x) = \frac{\alpha}{\mu_1} \exp\left[\frac{-x}{\mu_1}\right] + \frac{1-\alpha}{\mu_2} \exp\left[\frac{-x}{\mu_2}\right] \quad (2.13)$$

media $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ y varianza $\sigma^2 = \alpha\mu_1^2 + (1-\alpha)\mu_2^2 + \alpha(1-\alpha)(\mu_1 - \mu_2)$. Este distribución proporciona mejores ajustes globales a los datos de precipitación diaria pues una de las exponenciales se centra en el cero y la otra en los valores extremos (ver Wilks, 1999a, para más detalles).

Una vez elegida la distribución más conveniente, es necesario simular un valor para cada uno de los días donde el modelo de Markov ha generado una ocurrencia de precipitación. Este problema se reduce a obtener una muestra aleatoria a partir de una distribución de probabilidad concreta.

Simulación de Variables Aleatorias

Los ordenadores proporcionan números pseudo-aleatorios distribuidos de forma uniforme. Sin embargo, en la práctica interesa obtener una muestra de números aleatorios distribuidos según una cierta función de densidad $f(x)$ (por ejemplo, exponencial, Gamma, doble exponencial, etc.). Para ello se utiliza la función de distribución correspondiente

$$F(x) = p(X \leq x) = \int_{-\infty}^x f(x)dx.$$

Seguidamente, a partir de una sucesión de números aleatorios $\{u_1, \dots, u_N\}$ que forman una muestra aleatoria de una distribución uniforme $U(0, 1)$, se puede obtener una muestra aleatoria $\{x_1, \dots, x_N\}$ distribuida según $f(x)$ resolviendo la ecuación $F(x_i) = u_i$, $i = 1, \dots, N$, que da $x_i = F^{-1}(u_i)$, donde $F^{-1}(u_i)$ es la inversa de la función de distribución evaluada en el punto u_i . Por ejemplo, la Figura 2.6 muestra una función de distribución $F(x)$ y los dos valores x_1 y x_2 correspondientes a los números u_1 y u_2 procedentes de la población uniforme $U(0, 1)$. Los fundamentos teóricos de la simulación de variables aleatorias se describen en detalle en Rubinstein (1981); Devroye (1986).

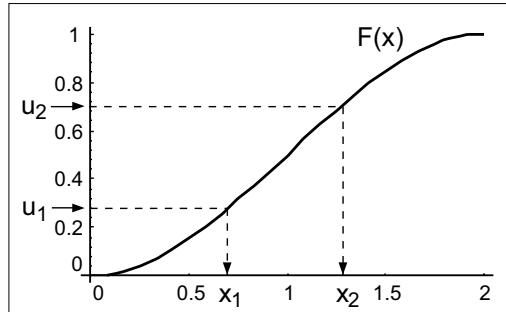


Figura 2.6: Generando muestras de una población con función de densidad $h(x)$ utilizando la función de distribución $H(x)$.

Ejemplo 2.6 (Weather Generators. Cantidad de Precipitación).

Supóngase que en el Ejemplo 2.5 se obtuvo la siguiente serie simulada de ocurrencia/ausencia de precipitación:

0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0,
 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
 1,
 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,

en la que se tienen 71 días lluviosos y 29 secos.

Para simular un valor de precipitación para los 71 días lluviosos se considera una función de densidad exponencial, con función de distribución:

$$F(x) = -\exp\left[\frac{-x}{\mu}\right] + 1. \quad (2.14)$$

Se estima el parámetro de esta distribución (la media μ) a partir de las cantidades de precipitación de los días en que ha llovido en la serie mostrada en el Ejemplo 2.5 ($\mu = 61.98$). Dado que

$$x = F^{-1}(y) = -\mu \ln(1 - y) \quad (2.15)$$

un muestra de 71 números aleatorios uniformes (y) determinará la muestra de cantidades de precipitación (x) requerida. Combinando este resultado con la

serie binaria se tiene la siguiente serie de precipitación resultante del proceso de simulación:

0, 0, 38.03, 126.3, 0, 0, 0, 0, 168.7, 19.0, 10.8, 127.8, 16.83, 0, 64.3, 0, 0, 0, 0, 0, 211.2, 67.7, 126.6, 0, 0, 0, 0.6, 9.1, 0, 0, 105.8, 0, 0, 34.8, 136.9, 82.2, 72.0, 26.3, 11.2, 10.4, 13.1, 34.0, 120.1, 41.7, 104.9, 38.2, 37.8, 37.1, 32.9, 143.7, 0.3, 21.8, 3.1, 73.2, 65.0, 252.5, 49.8, 31.6, 13.7375, 60.8, 81.9, 29.2, 0.6, 33.7, 86.8, 97.8, 156.5213, 115.4, 28.4, 60.1, 81.4, 13.3, 0, 145.7, 52.1, 0, 61.9, 0, 0, 16.5, 49.3, 166.2, 25.3, 66.0, 30.8, 0, 61.1, 74.4, 0, 0, 31.3, 33.0, 66.0, 112.6, 28.7, 34.3, 55.9 ■

El siguiente código ilustra el uso de la función `WeatherGen` de la Toolbox *Meteolab* para simular series binarias o continuas de precipitación para una estación dada. En el ejemplo se considera una única estación de la red de observaciones GSN descrita en la Sec. 1.9.1. Por una parte, la opción '`Markov`' permite obtener una serie discreta (no necesariamente binaria), y por otra la opción '`wg`' permite obtener una serie continua.

```

Example1.Network={'GSN'};
Example1.Stations={'Spain.stn'};
Example1.Variable={'Precip'};
date={'1-Jan-1998','31-Dec-1998'};
[data,Example1]=loadStations(Example1,'dates',date,'ascfile',1);
data=data(:,1); % Selecting the first station: San Sebastian.

umbral=[0,0.5,10,20,inf];

% Occurrence of precipitation (discrete variable) Markov process.
s1=weatherGen(data,umbral,2,10*size(data,1),'markov');
plot(s1(1:365,1))

% Precipitation amount (continuous variable)
[s2,r2]=weatherGen(data,umbral,2,10*size(data,1),'wg');
plot(s2(1:365,1)) % Symbolic (discrete)

subplot(2,1,1);
plot(data(1:end,1)) % Original
subplot(2,1,2)
plot(r2(1:size(data,1),1)) % Simulated

```

2.4. Dependencia e Independencia de Variables

El concepto de probabilidad condicionada proporciona la forma de definir la dependencia estadística de dos conjuntos de variables aleatorias. Si X e Y son dos subconjuntos disjuntos de variables aleatorias, se dice que X es independiente de Y (que denotaremos $I(X, Y)$) si y solamente si

$$p(x|y) = p(x), \quad (2.16)$$

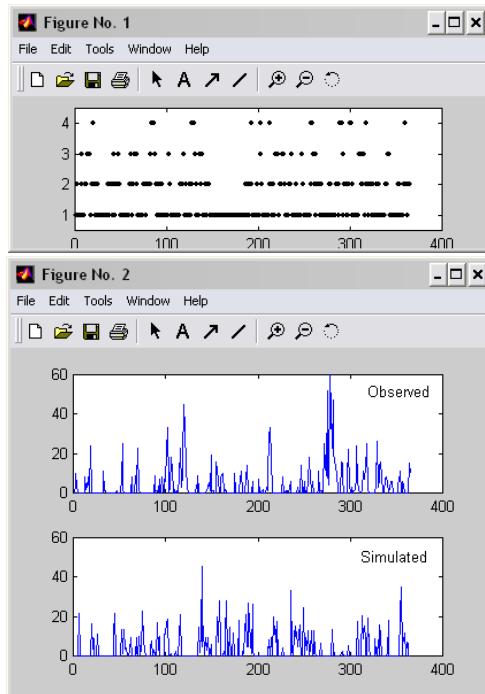


Figura 2.7: (arriba) Simulación de una variable de precipitación discreta (con cuatro estados); (debajo) simulación de la cantidad de precipitación. Se muestran las series original y simulada.

para todos los valores posibles x e y de X e Y , respectivamente; en otro caso, X se dice dependiente de Y (y se denota $D(X, Y)$).

Si X es independiente de Y , pueden combinarse (2.9) y (2.16) para obtener:

$$p(x, y) = p(x)p(y). \quad (2.17)$$

La ecuación (2.17) indica que si X es independiente de Y , entonces la función de probabilidad conjunta de X e Y es igual al producto de sus marginales.

Los conceptos de dependencia e independencia se refieren a dos subconjuntos de variables, pero se puede generalizar el concepto cuando hay implicados más de dos conjuntos. Si X , Y y Z son tres conjuntos disjuntos de variables, entonces X se dice condicionalmente independiente de Y dado Z , $I(X, Y|Z)$, si y sólo si

$$p(x|z, y) = p(x|z), \quad (2.18)$$

para todos los valores posibles x , y , z . En otro caso X e Y se dicen condicionalmente dependientes dado Z , $D(X, Y|Z)$. Una definición alternativa, pero equivalente, de la independencia condicional es

$$P(x, y|z) = P(x|z)P(y|z). \quad (2.19)$$

El concepto de independencia condicional $I(X, Y|Z)$ lleva en sí la idea de que una vez que es conocida Z , el conocimiento de Y no altera la probabilidad de X . En otras palabras, si Z ya es conocida, el conocimiento de Y no añade información alguna sobre X .

Ejemplo 2.7 (Dependencia e Independencia condicional). *En este ejemplo se consideran de nuevo los datos de la Tabla 2.1 y también los datos nuevos de la Tabla 2.2; éstos últimos corresponden a la precipitación, fase lunar y dirección del viento registrados en Santander, en el mismo periodo de tiempo que los considerados en la primera tabla.*

	Anual		Llena		C. Menguante		C. Creciente		Nueva	
	S	Ll	S	Ll	S	Ll	S	Ll	S	Ll
NE	1014	516	255	137	208	106	297	132	254	141
SE	64	57	12	12	16	16	22	12	14	17
SW	225	661	59	165	65	166	58	175	43	155
NW	288	825	51	192	77	231	82	225	78	177
Total	1591	2059	377	506	366	519	459	544	389	490

Tabla 2.2: Datos de precipitación (lluvia (Ll) o no lluvia (S)), fase lunar y dirección del viento registrados en Santander (aeropuerto de Parayas) entre 1979 hasta 1989 ($N = 3650$).

A partir de la tabla 2.2, se puede calcular:

- $P(Ll) = 2059/3650 = 0.564$.
- $P(Ll|CC) = 490/(490 + 389) = 0.557$.
- $P(Ll|LN) = 544/(544 + 459) = 0.542$.
- $P(Ll|CM) = 519/(519 + 366) = 0.586$.
- $P(Ll|LL) = 506/(506 + 377) = 0.573$.

A partir de estos resultados se puede comprobar que la lluvia en Santander es independiente de la fase lunar, $I(Lluvia, Fase lunar)$, como es de esperar.

Considerando ahora los datos registrados en la Tabla 2.1, se tiene:

- $P(Ll|Invierno) = 490/(490 + 361) = 0.576$.
- $P(Ll|Primavera) = 566/(566 + 406) = 0.582$.

la probabilidad de lluvia es independiente de la estación del año (invierno, primavera) dada. Sin embargo, si se consideran también la evidencia Viento = SW, se tiene:

- $P(LL|Invierno, SW) = 223/(223 + 98) = 0.695$.
- $P(LL|Primavera, SW) = 119/(119 + 18) = 0.869$.

la lluvia y la estación del año (invierno, primavera) se vuelven dependientes; es decir, la lluvia y la estación del año en Santander son independientes $I(Lluvia, Estacion)$, pero son condicionalmente dependientes dado el viento $D(Lluvia, Estacion|Viento)$. También se puede dar la situación contraria. ■

2.4.1. Correlación e Información Mutua

Los conceptos de covarianza y correlación caracterizan la relación lineal existente entre variables continuas de un cierto proceso. La covarianza tiene el defecto de depender de las escalas de los datos. Sin embargo, el coeficiente de correlación es adimensional y caracteriza el grado de la relación lineal existente; el coeficiente de correlación r está relacionado con la covarianza de las dos variables de la siguiente forma:

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right) = \frac{\sigma_{xy}}{\sigma_x \sigma_y},$$

y sus posibles valores están acotados en $-1 \leq r \leq 1$. Los valores extremos corresponden a relaciones lineales perfectas, ya sea en sentido creciente o decreciente, mientras que un coeficiente de correlación cero es indicativo de que no existe relación lineal entre ambas variables. Por tanto, el concepto de correlación es de gran importancia teórica y práctica para medir la influencia relativa de unas variables sobre otras. La forma exacta de esta relación vendrá dada, bajo ciertas hipótesis, por la recta de regresión que se describe brevemente en el siguiente capítulo.

El concepto de correlación está fuertemente limitado por su naturaleza lineal. Por ello, cuando las variables de un problema dado puedan estar relacionadas de una forma arbitraria, es necesario disponer de técnicas más generales para analizar su grado de dependencia. La entropía es una medida utilizada en la Teoría de la Información para medir el “desorden” (uniformidad o normalidad) de una variable. Dada una variable aleatoria X se define la entropía como:

$$H(X) = - \sum_{x \in X} p(x) \log_b p(x) = \sum_{x \in X} p(x) \log_b \frac{1}{p(x)} = E \left[\log_b \frac{1}{p(x)} \right] \quad (2.20)$$

donde la base del logaritmo determina la unidad en la que se mide la información de los datos. Si dicha base es 10 se mide en Dits, si es 2 en Bits, etc. A partir de la definición, considerando las propiedades de la probabilidad, se puede deducir que $H(x) \geq 0$, y será nula sólo cuando toda la probabilidad de la variable se concentre en un único valor. Otra medida muy importante es la entropía relativa o entropía diferencial, también llamada distancia de Kullback-Leibler, que mide la distancia entre dos posibles distribuciones de probabilidad, p y q , de una misma variable aleatoria X , y se define como:

$$Dkl(p, q) = \sum_{x \in X} p(x) \log_b \frac{p(x)}{q(x)} \quad (2.21)$$

como en el caso anterior se tiene que $Dkl(p, q) \geq 0$ y en este caso, $Dkl(p, q) = 0 \Leftrightarrow p(x) = q(x)$. Obsérvese que D no es un distancia métrica, pues se da el caso que $Dkl(p, q) \neq Dkl(q, p)$.

El concepto de información mutua deriva del concepto de entropía y proporciona un criterio de medida de dependencia. Mide la cantidad de información

que una variable contiene sobre otra, es decir, $I(X;Y)$ es la información que X contiene sobre Y. Se define la información mutua como:

$$IM(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_b \frac{p(x, y)}{p(x)p(y)} \quad (2.22)$$

es decir, se tiene que $IM(X, Y) = DKL(p(x, y), p(x)p(y))$, que será nula cuando $p(x, y) = p(x)p(y)$, es decir, cuando X e Y sean variables aleatorias independientes. También, a partir de las propiedades de la probabilidad se puede deducir que, $IM(X, Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X, Y)$.

Ejemplo 2.8 (El Fenómeno de El Niño). En este ejemplo se relacionan distintas variables meteorológicas medidas en diferentes puntos del Perú mediante las técnicas antes descritas. Las variables a considerar son la temperatura media mensual del agua del mar (SST), la media mensual de precipitación y el caudal mensual de un río. La SST esta medida en dos puntos situados frente a la costa peruana, más concretamente en Paita y Chicama; la precipitación corresponde a la de la ciudad de Piura, y el caudal del río es la del río Piura, que desemboca en esta misma ciudad. Estas localidades se encuentran en la zona Norte de Perú y están fuertemente afectadas por el fenómeno de El Niño. La Fig. 2.8 muestra visualmente las relaciones entre estas variables.

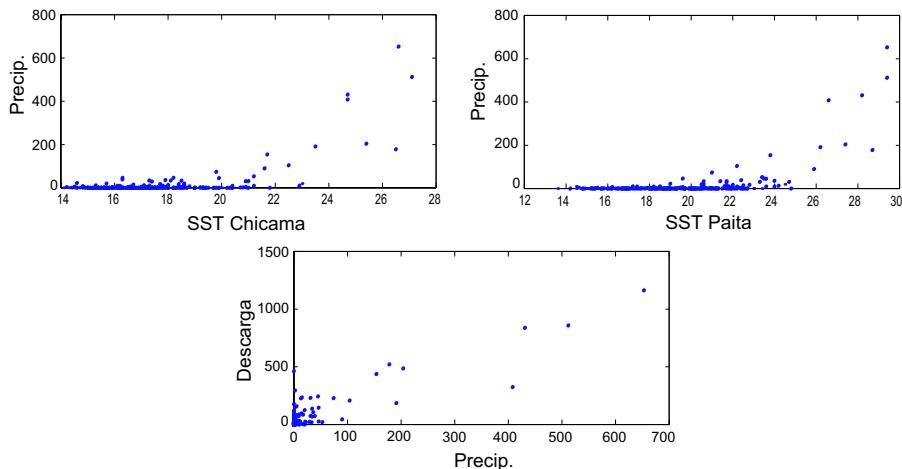


Figura 2.8: Precipitación en la ciudad de Piura frente a las SST en Chicama y Paita y el caudal del río Piura.

A continuación se procede a calcular cuantitativamente estas relaciones de dependencia. En primer lugar, se calcula la correlación dos a dos de todas las variables, obteniéndose una medida de la relación lineal existente entre ellas:

	Precip. Piura	SST Paita	SST Chicama	Desc. Piura
Precip. Piura	1	0.519	0.586	0.884
SST Paita	–	1	0.806	0.584
SST Chicama	–	–	1	0.636
Desc. Piura	–	–	–	1

Los resultados que se muestran en la tabla indican que la precipitación en Piura está muy correlacionada con el caudal del río Piura y las SST en Chicama y Paita también presentan una relación lineal bastante fuerte. Sin embargo, se observa que la precipitación en Piura no tiene una relación lineal importante con las SST. Para analizar la posibilidad de que exista alguna relación no lineal entre estas variables, se ha calculado la información mutua entre la precipitación en Piura y el resto de variables, para poder comparar con los resultados anteriores. Para ello, lo primero que se hace es discretizar los datos; tanto la SST de Paita, como la de Chicama y el caudal del río Piura se discretizan utilizando sus quintiles, mientras que para la precipitación se utilizan unos umbrales apropiados para esta variable, de modo que en el primer intervalo aparecen todos los eventos de no precipitación (datos iguales a 0), y en el último estarán los datos de precipitación correspondientes a eventos fuertes de El Niño. Se obtienen los siguientes valores:

	SST Paita	SST Chicama	Caudal Piura
Precip. Piura	2.054	1.868	2.059

Así, se puede concluir que la precipitación en Piura está muy relacionada con las otras tres variables, aunque un poco menos con la SST de Chicama. Teniendo en cuenta estos resultados, la relación existente con el caudal en Piura es lineal mientras que con las SST es no lineal. ■

En el siguiente ejemplo se muestra que las medidas anteriores muestran una visión estática “a priori” de las dependencias entre las variables, y que estas relaciones se pueden modificar al conocer información adicional sobre las variables involucradas. Este hecho hará necesario considerar modelos probabilísticos conjuntos para modelizar toda la complejidad de las relaciones entre un conjunto de variables (Cap. 4).

Ejemplo 2.9 (Precipitación en las Cuencas Españolas). En este ejemplo se han considerado los datos de precipitación en 105 estaciones de la red completa del INM sobre distintas cuencas hidrográficas. En la Fig. 2.9(a) se muestra la correlación entre la series de precipitación de cada par de estaciones. Se observa claramente, que las estaciones que corresponden a una misma cuenca hidrográfica presentan una correlación más fuerte (las correlaciones más fuertes se agrupan en torno a la diagonal), aunque algunas cuencas se entremezclan (por ejemplo las cuencas del Tajo, Guadiana, Guadalquivir y Sur).

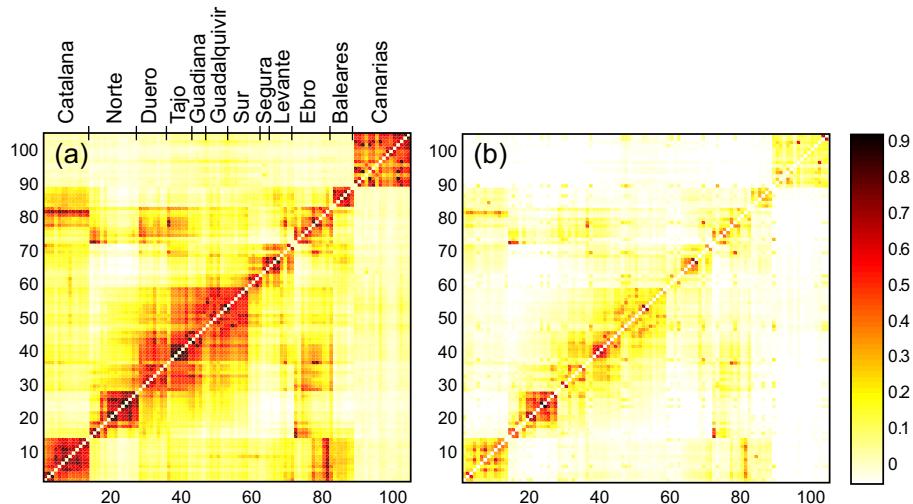


Figura 2.9: (a) Correlación y (b) información mutua entre los registros de precipitación de cada par de ciudades en una red de 105 estaciones en España.

En la Fig. 2.9(b) se presenta la información mutua entre cada par de estaciones; en este caso las relaciones son más precisas, pues existen menos estaciones que den una información clara sobre otras. Sin embargo, también se ve claramente la separación entre cuencas.

En la Fig. 2.10 se analiza en más detalle la correlación que presentan las series de datos de las estaciones correspondientes a la cuenca Norte y a la Duero. La Fig. 2.10(a) es una ampliación de la Fig. 2.9(a) correspondiente a la zona estudiada. Por otra parte, en la Fig. 2.10(b) se muestra la correlación para las series condicionadas a la no precipitación en la estación de Palencia; es decir, la correlación condicionada al evento “Precipitación en Palencia = 0 mm”. En este caso se observa que, el hecho de que no llueva en Palencia, prácticamente no altera la correlación entre las estaciones de la cuenca Norte, pero sí que altera la correspondiente a la cuenca Duero, como era de esperar, haciendo a las estaciones más independientes, unas de otras (presentan menor correlación entre ellas).

En la Fig. 2.10(c) se presentan las correlaciones condicionadas al evento “Precipitación en Palencia > 10 mm”. En este caso, el mayor cambio se produce con la aparición de correlaciones entre la cuenca Norte y Duero, principalmente con las tres primeras consideradas (estaciones de Cantabria y Vizcaya); por tanto, el hecho de que llueva en Palencia influye en el comportamiento de algunas estaciones de la cuenca Norte. También, es de destacar la desaparición de una fuerte correlación que aparecía en la Fig. 2.10(a) entre las estaciones 16 y 17 (Palencia y Valladolid). ■

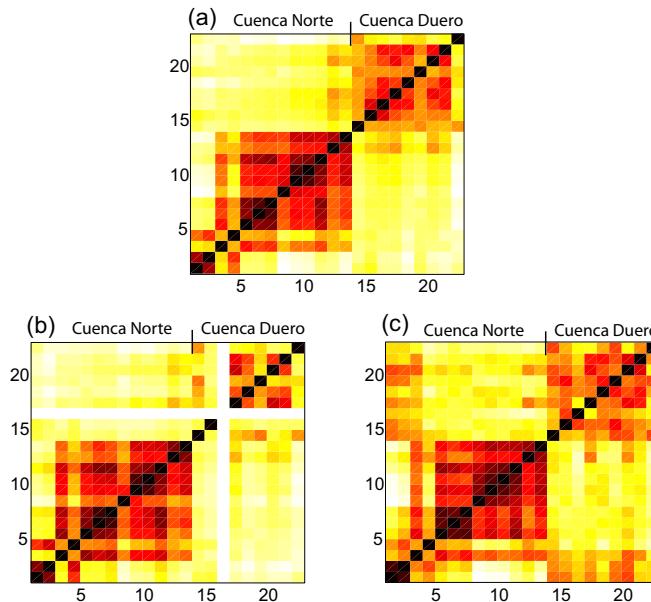


Figura 2.10: (a) Correlación entre estaciones en las cuencas Norte y Duero. (b) Correlación condicionada al caso *Precipitación en Palencia = 0 mm*. (c) Correlación condicionada a *Precipitación en Palencia > 10 mm*

2.5. Componentes Principales y EOF

En la sección anterior se ha analizado el problema de la dependencia y correlación entre las variables de diversos problemas prácticos. En algunas ocasiones interesa utilizar esta dependencia para poder representar los datos con un número menor de variables, independientes unas de otras. El análisis de Componentes Principales (CPs), también conocido en las ciencias atmosféricas como análisis de Funciones Ortogonales Empíricas (EOF), es una técnica estándar para eliminar la información redundante con la mínima pérdida de variabilidad. Esto se logra proyectando el conjunto de datos en un nuevo espacio de menor dimensión que el original, donde las nuevas variables (dimensiones) representen aquellas direcciones del espacio donde los datos tienen mayor varianza; en otras palabras, el análisis de CPs es una técnica eficiente para compimir datos (Preisendorfer and Molley, 1988). Este método es especialmente útil en espacios de alta dimensionalidad, donde los datos pueden estar correlacionados en sus distintas componentes y, por tanto, pueden contener mucha información redundante en su descripción. Un ejemplo típico en meteorología lo constituyen los campos de circulación atmosférica, dados por los valores de una o varias magnitudes (la presión a nivel del mar, etc.) en una rejilla sobre una cierta zona de interés. Dado que estas magnitudes están correlacionadas espacialmente, existirá una gran redundancia en esta forma de expresar los datos. La técnica de componentes principales reduce la dimensión del espacio

preservando el máximo de varianza de la muestra. Para ello, la base del nuevo espacio se forma con aquellos vectores donde la muestra proyectada presenta mayor varianza. Los vectores de esta base son de enorme utilidad en meteorología, pues los primeros de ellos pueden corresponder a patrones dominantes como la NAO (en el sentido de la variabilidad de la muestra que representan).

Se parte de una muestra de m datos

$$\mathbf{x}_k = (x_{k1}, \dots, x_{kn})^T, k = 1, \dots, m, \quad (2.23)$$

en un espacio n dimensional con base canónica $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. Se desea obtener un subespacio de dimensión $d < n$ dado por una nueva base $\{\mathbf{f}_1, \dots, \mathbf{f}_d\}$ (siendo cada \mathbf{f}_j una combinación lineal de los vectores \mathbf{e}_i de la base canónica). Para un vector \mathbf{x} se tendrá:

$$\mathbf{x} = \sum_{i=1}^n \mathbf{e}_i x_i \approx \bar{\mathbf{x}} = \sum_{i=1}^d \mathbf{f}_i \bar{x}_i. \quad (2.24)$$

El criterio para obtener este subespacio es que la muestra proyectada en el nuevo espacio $\{\bar{\mathbf{x}}_k, k = 1, \dots, m\}$ tenga la máxima varianza posible. Es decir, fijada una dimensión d , el problema consiste en encontrar los vectores \mathbf{f}_i que proyectan la muestra con varianza máxima. El cálculo matemático para obtener los vectores óptimos es sencillo y consiste en estimar la matriz de varianzas y covarianzas a partir de la muestra de datos. Los autovectores (o *Funciones Ortogonales Empíricas*, EOFs) de esta matriz son los nuevos vectores \mathbf{f}_i y los correspondientes autovalores indican la varianza explicada (la varianza de la muestra proyectada sobre el vector). Los coeficientes de un punto (dato) en la nueva base se denominan *Componentes Principales* (CPs).

Dada la muestra (2.23), se puede estimar la matriz de varianzas y covarianzas $C_{\mathbf{x}}$, donde cada elemento σ_{ij} representa la covarianza de los datos entre la variable i y la j del espacio original:

$$\sigma_{ij} = \langle (x_{ki} - \mu_i)(x_{kj} - \mu_j) \rangle_k; \mu_i = \langle x_{ki} \rangle_k, \mu_j = \langle x_{kj} \rangle_k, \quad (2.25)$$

donde $\langle \rangle$ denota el promedio aritmético. Esta matriz de varianzas y covarianzas es cuadrada y simétrica por lo que se puede calcular una nueva base ortogonal encontrando sus autovalores λ_i (que serán reales) y los correspondientes autovectores \mathbf{f}_i :

$$C_{\mathbf{x}} \mathbf{f}_i = \lambda_i \mathbf{f}_i, i = 1, \dots, n. \quad (2.26)$$

Es fácil resolver este problema cuando n es pequeño pero, a medida que aumenta la dimensión, el problema se complica debido al posible mal condicionamiento de la matriz. En estos casos es necesario aplicar métodos numéricos eficientes como la Descomposición en Valores Singulares (SVD), que proporciona una factorización de la matriz $C_{\mathbf{x}}$ de la forma (ver Press et al. (1992) para más detalles):

$$C_{\mathbf{x}} = F \Lambda F^T, \quad (2.27)$$

donde Λ es una matriz diagonal que contiene los autovalores λ_i (ordenados de forma decreciente) de $C_{\mathbf{x}}$, y las columnas de F son los correspondientes

autovectores \mathbf{f}_i . Además F es una matriz ortogonal y F^T es su inversa. De esta manera, si hacemos la proyección:

$$\bar{\mathbf{x}}_k = F^T \mathbf{x}_k = \begin{pmatrix} f_{11} & \dots & f_{1n} \\ \vdots & & \vdots \\ f_{n1} & \dots & f_{nn} \end{pmatrix} \begin{pmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{pmatrix} \quad (2.28)$$

se tendrá el elemento de la muestra proyectado sobre la base de autovectores de $C_{\mathbf{x}}$, mientras que la proyección inversa se obtendrá mediante $\mathbf{x}_k = F \bar{\mathbf{x}}_k$. Esta proyección tiene las siguientes propiedades:

- Componentes incorrelacionadas: $\langle \bar{x}_{ki} \bar{x}_{kj} \rangle_k = 0, i \neq j$.
- $Var(\bar{x}_{ki}) = \lambda_i, i = 1, \dots, n$.
- $\sum_{i=1}^n Var(x_{ki}) = \sum_{i=1}^n Var(\bar{x}_{ki}) = \sum_{i=1}^n \lambda_i$.

Dado que los vectores se eligen en orden decreciente de varianza, es posible hacer un recorte de dimensión reteniendo la máxima varianza posible (obviamente, si se quiere conservar toda la varianza habrá que tomar $d = n$). Si se toman sólo las d primeras EOFs, cada elemento de la muestra se podrá expresar aproximadamente como:

$$\mathbf{x}_k \approx \tilde{F} \tilde{F}^T \mathbf{x}_k = \begin{pmatrix} f_{11} & \dots & f_{d1} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ f_{1n} & \dots & f_{dn} \end{pmatrix} \begin{pmatrix} f_{11} & \dots & f_{1n} \\ \vdots & & \vdots \\ f_{d1} & \dots & f_{dn} \end{pmatrix} \mathbf{x}_k, \quad (2.29)$$

donde \tilde{F} representa a la matriz F truncada a los d primeros autovectores. El vector $\bar{\mathbf{x}}_k = \tilde{F}^T \mathbf{x}_k$ de dimensión $d \times 1$ contendrá las CPs del patrón \mathbf{x}_k , es decir, las componentes del vector en el nuevo espacio de dimensión d . Para recuperar la dimensión original, el vector de CPs se proyectará mediante $\tilde{F} \bar{\mathbf{x}}_k$, obteniendo una aproximación del vector original (mejor cuanto mayor sea la dimensión d del espacio proyector).

Maximizar la varianza es equivalente a minimizar la norma cuadrática de los residuos $\bar{\mathbf{x}}_k - \mathbf{x}_k$.

$$Var(\mathbf{x}_k) = Var(\mathbf{x}_k + \tilde{F} \bar{\mathbf{x}}_k - \tilde{F} \bar{\mathbf{x}}_k) = Var(\mathbf{x}_k - \tilde{F} \bar{\mathbf{x}}_k) + Var(\tilde{F} \bar{\mathbf{x}}_k) \quad (2.30)$$

Por tanto, la técnica de componentes principales obtiene la proyección lineal óptima en sentido de máxima varianza explicada y de mínimo error cuadrático de reconstrucción.

Para eliminar el efecto de las distintas escalas de cada una de las componentes del vector, es conveniente estandarizar los datos (componente a componente) como paso previo a realizar el análisis. De esta forma se evita que las variables de mayor varianza se hagan dominantes en el análisis. En el caso de

datos atmosféricos se han de estandarizar por separado los valores correspondientes a cada punto de rejilla, de forma que la variabilidad del patrón en toda la extensión espacial sea homogénea. Otro procedimiento para tener en cuenta este problema consiste en utilizar la matriz de correlaciones en lugar de la de varianza-covarianza para realizar el análisis (Noguer, 1994).

Recientemente se han descrito en la literatura extensiones no lineales de esta técnica que proyectan los datos mediante combinaciones no lineales de las variables originales maximizando la varianza explicada o minimizando el error cuadrático. Uno de los métodos no lineales más populares son las redes neuronales de cuello de botella (Kramer, 1991). Estos modelos tienen mayor flexibilidad que las técnicas lineales, pero en contraprestación tienen algunas deficiencias, como la pérdida de ortogonalidad en los vectores de la base y el coste computacional necesario para entrenar los modelos resultantes (en el Cap. 5 se analizan en más detalle estos modelos). Otra técnica reciente aplicada a este problema son los métodos basados en núcleos (kernel methods), que también dan lugar a modelos no lineales pero con un menor coste computacional en su entrenamiento (ver Schölkopf et al., 1998, para más detalles)

Ejemplo 2.10 (EOFs). En este ejemplo se realiza un análisis de componentes principales de dos muestras de datos $\mathbf{x} = (x, y, z)$ y $\mathbf{u} = (u, v)$, donde x , y y u son variables aleatorias uniformes, $z = y + x/2 + \epsilon_1$ y $v = u + \epsilon_2$ donde ϵ_1 y ϵ_2 son variables aleatorias gaussianas. En los ejemplos anteriores, una de las variables es dependiente de las restantes. En Matlab, se genera fácilmente una muestra de tamaño $N = 100$ para este problema

```
>> N=100; x=rand(N,1); y=rand(N,1).*2;
>> z=y+x./2+randn(N,1).*(1/8);
>> datos1=[x y z];

>> u=rand(N,1);
>> v=2*u+randn(N,1).*(1/4);
>> datos2=[u v];
```

Una vez generados los datos, se procede a calcular los valores singulares de su matriz de covarianzas. La función `svd` de Matlab realiza la descomposición en valores singulares de una matriz X , dando como salida tres matrices, U , S y V de modo que: $X = U S V^T$, donde S es una matriz diagonal que contiene los autovalores y cada columna de V es el autovector correspondiente ($U = V$ para matrices simétricas):

```
>> [res1,valsing1,eof1]=svd(cov(datos1));
>> [res2,valsing2,eof2]=svd(cov(datos2));
```

Para la primera muestra se obtienen los autovectores (o EOFs), $eof_1 = (0.04, 0.68, 0.73)$, $eof_2 = (-0.93, 0.29, -0.21)$ y $eof_3 = (0.35, 0.67, -0.64)$ (ordenados de mayor a menor contribución de la explicación de la varianza).

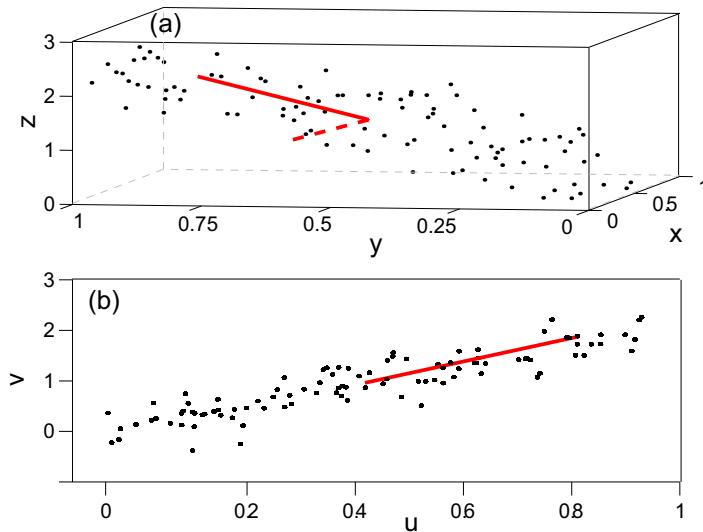


Figura 2.11: Muestras de datos (a) tridimensionales y (b) bidimensionales generados a partir de números aleatorios, con sus correspondientes EOFs.

<i>EOF</i>	% varianza explicada
1	88.93
2	10.05
3	1.01

En este caso, las primeras dos EOFs explican prácticamente toda la varianza de los datos. Para la segunda muestra se tienen los vectores $\text{eof}_1 = (0.39, 0.92)$ y $\text{eof}_2 = (-0.92, 0.39)$ con la contribución:

<i>EOF</i>	% varianza explicada
1	97.30
2	2.69

En las Figs. 2.11(a) y (b) están representadas las EOF para las dos muestras de datos. En la Fig. 2.11(a) se ve que las dos primeras EOF representan las dos direcciones principales que siguen los datos mientras que la tercera dirección sería perpendicular a ellas, y viene dada por el ruido que se ha introducido en la tercera variable. De la misma forma, en la Fig. 2.11(b) se aprecia que la primera EOF representa la dirección principal, mientras que la segunda representaría el ruido de la segunda variable de la muestra. ■

Ejemplo 2.11 (EOFs y CPs de Patrones Atmosféricos). En este ejemplo se aplica la técnica de componentes principales para hallar los patrones de presión a nivel del mar (Sea Level Pressure, SLP) dominantes en tres zonas del globo (ver Fig. 2.12), con características diferentes de circulación atmosférica y oceánica. Para ello, se han considerado medias decenales de patrones diarios de SLP en tres rejillas distintas que definen la región del Atlántico Norte (AN),

la región de El Niño (EN), y la zona de América Austral (Austral), respectivamente. Los datos disponibles cubren el período 1959-1999 correspondiente al reanálisis ERA40. La siguiente tabla muestra la varianza explicada por las cuatro primeras EOF en cada caso, ilustrando las diferencias entre las distintas zonas:

	AN	EN	Austral
EOF_1	32.91	59.90	25.90
EOF_2	19.14	12.71	22.62
EOF_3	14.40	7.69	20.52
EOF_4	8.86	5.26	9.86
Acumulado	75.31	85.66	78.90

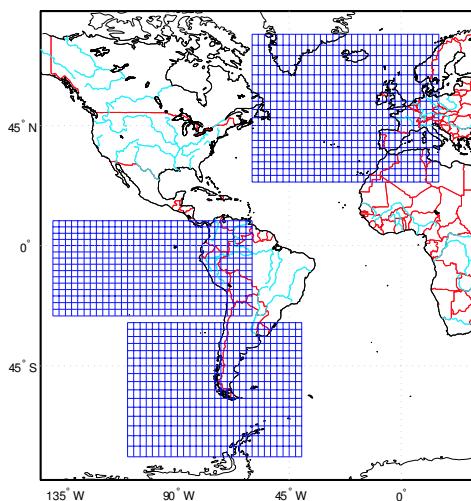


Figura 2.12: Zonas geográficas correspondientes al Atlántico Norte, El Niño, y América Austral. Las rejillas ($2.5^\circ \times 2.5^\circ$) muestran los puntos de grid utilizados para caracterizar los patrones de presión a nivel del mar.

En los tres casos existe una enorme redundancia en los datos, y una proporción muy pequeña de las variables permite explicar una alto porcentaje de la varianza. La zona EN es la que mayor redundancia muestra (la primera EOF explica cerca del 60 % de la varianza), mientras que las zonas AN y Austral presentan una varianza acumulada similar; sin embargo, esta varianza está igualmente distribuida entre las tres primeras EOFs en el caso Austral, mientras que decrece uniformemente en el AN. Esto nos muestra que la presión en los trópicos tiene mucha más correlación espacial que en latitudes medias y, a su vez, la correlación en latitudes medias se expresa de forma distinta en distintas regiones del globo.

La evolución temporal de las CPs nos da un idea de la frecuencia de variación temporal de los fenómenos caracterizados por la correspondiente EOF. Por ejemplo la Figura 2.13 muestra la evolución de la CPs para la región del AN. Puede observarse la frecuencia anual de la primera EOF, mientras que las

restantes presentan variabilidades temporales más complejas, relacionadas con las oscilaciones de los correspondientes patrones.

La Fig. 2.14 muestra las cuatro primeras EOFs para la zona NAO (North Atlantic Oscillation) y la zona Austral. Estos patrones maximizan la varianza proyectada de la muestra y , por tanto, definen los fenómenos sinópticos más relevantes que explican la variabilidad climática. Por ejemplo, la primera EOF de la zona AN corresponde al patrón de variabilidad anual de la presión, mientras que la segunda EOF corresponde al patrón de la NAO, y las siguientes están relacionadas con patrones como la EA (East Atlantic Index), AO (Artic Oscillation), etc. (ver, por ejemplo, Corte-Real et al., 1999; Rodríguez-Fonseca and Serrano, 2002, para una detallada descripción de estos patrones de teleconexión). ■

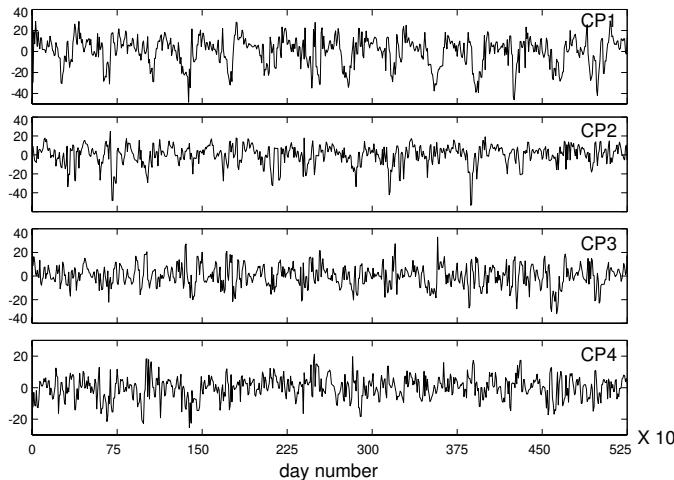


Figura 2.13: Evolución temporal de las cuatro primeras CPs en la zona del AN.

2.5.1. MeteoLab: Componentes Principales y EOF

La forma de trabajar con patrones atmosféricos en *MeteoLab* es mediante sus componentes principales. De esta forma, los datos son manejables y se mejora notablemente la eficiencia de los algoritmos. Para ello, se han implementado distintas funciones para obtener las componentes principales a partir de una serie de patrones dados, y también para reconstruir los campos originales a partir de las CPs almacenadas (ver Sec. 1.9.5). En este ejemplo, se considera la presión a nivel del mar (SLP) en la zona NAO. Cada dato es un patrón diario de presión a nivel del mar en el período 1958-2001 (período ERA40).

La función `computeEOF` permite generar el número de EOFs deseadas (por defecto genera todas las posibles) y las correspondientes CPs para el conjunto de datos; también se obtienen la media y la desviación típica de cada dimensión del campo (utilizadas para estandarizar los datos, punto a punto del grid) y los

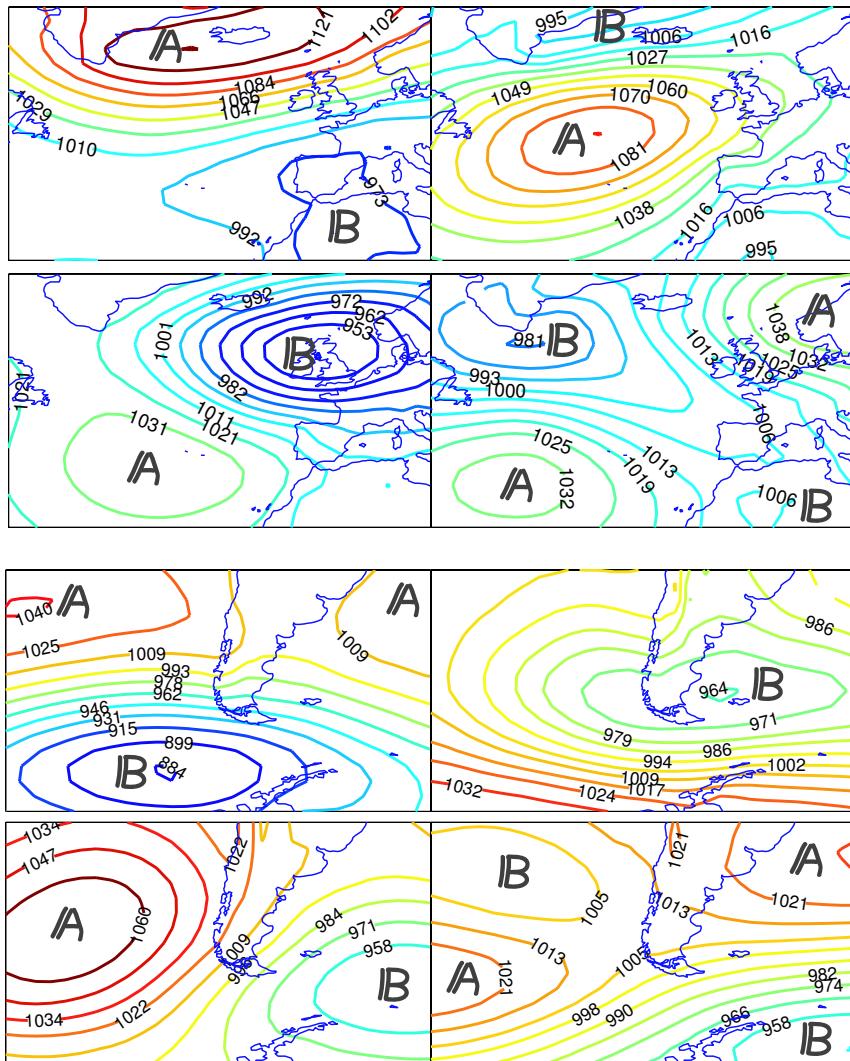


Figura 2.14: Cuatro primeras EOF correspondientes al área geográfica del Atlántico Norte (AN) y de América Austral (Austral).

valores propios asociados a las EOF. Toda esta información puede ser almacenada para utilizarse posteriormente según se ha descrito en la Sec.1.9.5; para ello, basta con tener cargados los campos (matriz `fields` en el ejemplo siguiente), e indicar el camino en el que se desea que se deseja guardar la información en la opción `path`.

```
>> dmn=readDomain('Nao');
>> [EOF,CP,MN,DV,OP]=computeEOF(X,'ncp',50,'path',dmn.path);
```

Si las EOFs ya hubiesen sido generadas con anterioridad, entonces se puede proceder como se indicó en la Sec. 1.9.5 para cargar un campo meteorológico dado a partir de las CPs almacenadas por medio de la función `getFieldfromEOF`. Sin embargo, en este caso, las CPs sólo se utilizan para reconstruir el campo sin cargarse en *Matlab* para poder explorarlas y manipularlas. La función `getEOF` cumple este cometido, cargando en memoria tanto las EOF, como las CPs y las medias y varianzas utilizadas en el proceso de standarización:

```
>> [EOF,CP,MN,DV]=getEOF(dmn,'ncp',50);
```

Con los datos cargados se pueden dibujar las primeras EOF (y las últimas) y las CPs de cada uno de los días disponibles (1958-2001 en este caso):

```
>> drawGrid(EOF(:,1:4)',dmn,'iscontourf',1);
>> drawGrid(EOF(:,[25 50])',dmn,'iscontourf',1);
>> figure
>> subplot(2,1,1); plot(CP(:,1));
>> subplot(2,1,2); plot(CP(:,50));
```

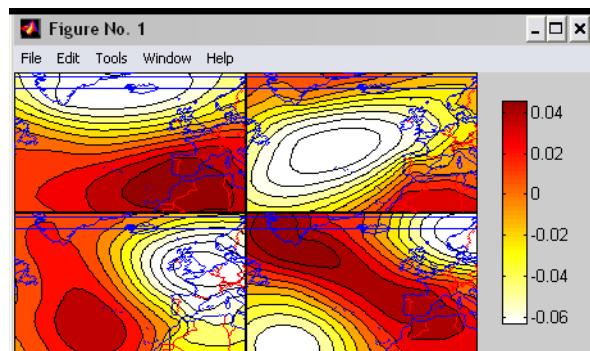


Figura 2.15: Patrones de las cuatro primeras EOF. La primera EOF representa el patrón bimodal de la NAO.

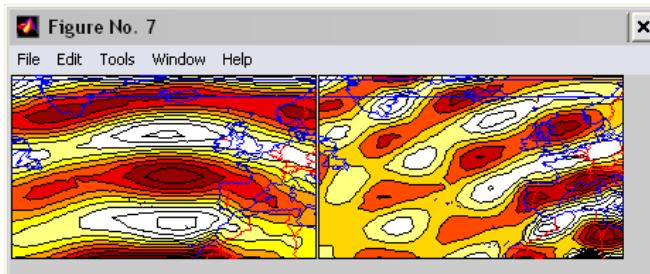


Figura 2.16: Patrones de las EOF número 25 y 50. Los patrones se van volviendo más ruidosos a medida que aumenta la posición de la EOF.

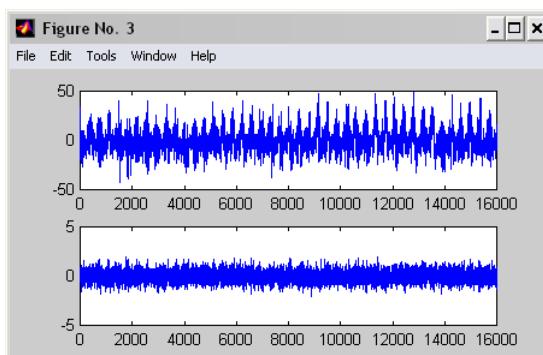


Figura 2.17: Coeficiente de la EOF (o CPs) números 1 y 50 para la serie de patrones diarios correspondientes al período 1958-2001. La primera CP ocupa la posición superior y se ve que posee más estructura que la otra.

En las figuras anteriores puede observarse que la primera EOF corresponde al patrón de oscilación del Atlántico Norte (NAO) descrito en la Sec. 1.4. Para analizar en mayor detalle la estructura periódica de las CPs se puede considerar el espectro de potencias (transformada de Fourier); por otra parte para estudiar variaciones interanuales de este patrón se puede considerar la media anual de estos valores:

```

>> figure
>> subplot(2,1,1)
>> f=fft(CP(:,1)); loglog(abs(f(2:end/2,:)))
>> subplot(2,1,2)
>> f=fft(CP(:,50)); loglog(abs(f(2:end/2,:)))
>> figure
>> period=365;
>> a=aveknt(CP(:,1),period);
>> plot(a(1,1:period:end))

```

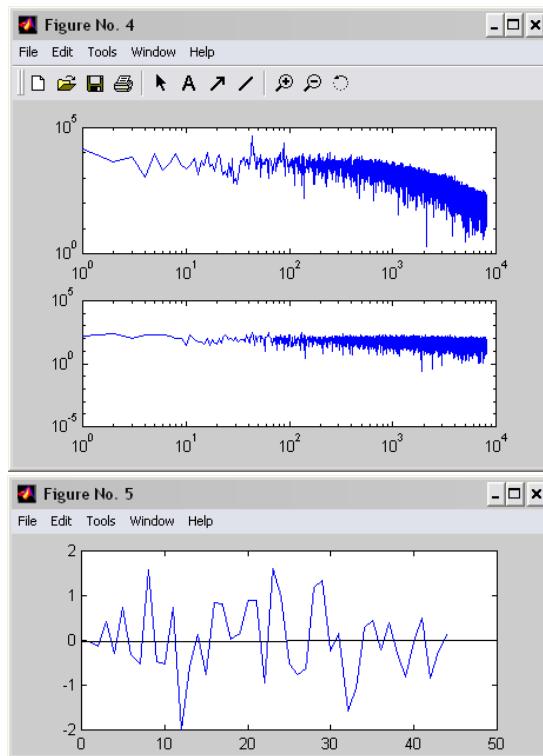


Figura 2.18: (arriba) Espectro de potencias en coordenadas logarítmicas para las CPs 1 y 50; en el primer caso se observa un régimen de crecimiento potencial, mientras que la serie de coeficientes de la última CP es similar a un ruido blanco. (abajo) Promedio anual del coeficiente de la primera EOF entre los años 1958 y 2001.

Por último, se pueden comparar los campos originales reconstruidos con distintos número de CPs. En el siguiente ejemplo, se consideran 50 y 4 CPs, respectivamente, para reconstruir los campos originales de presión a nivel del mar.

```
>> days=[size(CP,1), 1];
>> pat=((CP*EOF').*repmat(DV,[days 1]))+repmat(MN,days);
>> patAprox=((CP(:,1:4)*EOF(:,1:4)').*repmat(DV,days))+repmat(MN,days);
>> drawGrid(pat(1:9,:),dmn);
>> drawGrid(patAprox(1:9,:),dmn);
```

Se puede ver que los campos reconstruidos a partir de sólo cuatro CPs están muy suavizados.

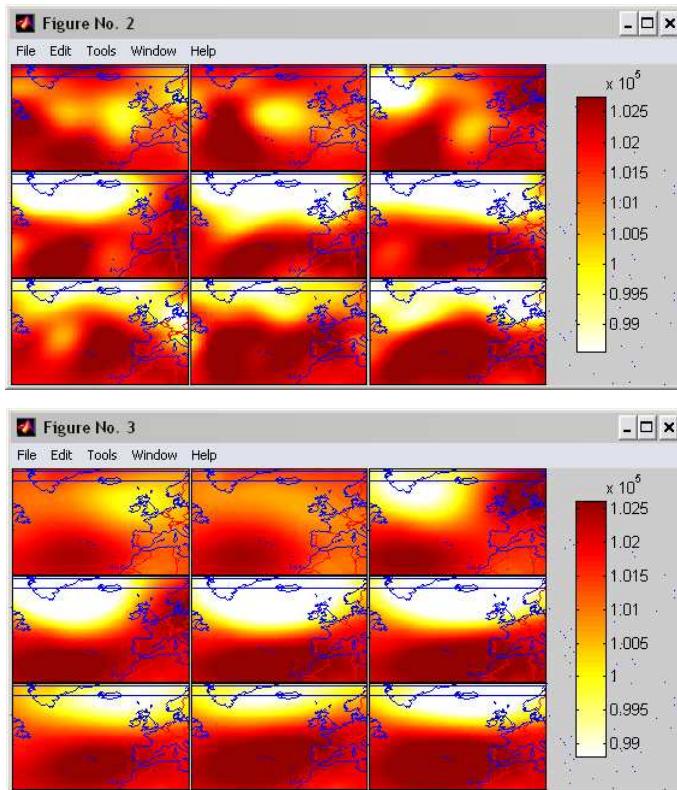


Figura 2.19: Reconstrucción de los nueve primeros días del año 1958 utilizando 50 (arriba) y 4 (abajo) componentes principales.

2.5.2. Elección del Número de Componentes

Una cuestión importante en la práctica es determinar el número de CPs que deben tomarse para un determinado problema, de forma que haya un equilibrio entre la reducción de información deseada y la calidad de la aproximación resultante. Una forma objetiva de seleccionar el número necesario de CPs es imponer un umbral para el error de reconstrucción obtenido (el error residual). Por ejemplo, la Figura 2.20 muestra el error de reconstrucción (Root Mean Square Error, RMSE) frente al porcentaje de CPs utilizadas para el patrón atmosférico *Modelo 2* de la Sección 1.9.4 (estandarizando las variables antes de aplicar el algoritmo). En la figura se muestran los errores separadamente para cada variable. A pesar de que las CPs se han obtenido globalmente combinando todas las variables, los errores de reconstrucción son similares para todas ellas. Sólo en el caso de considerar un número pequeño de CPs se pueden apreciar diferencias en los errores de reconstrucción, siendo éstos inferiores en patrones más suaves (por ejemplo *Z* o *T*).

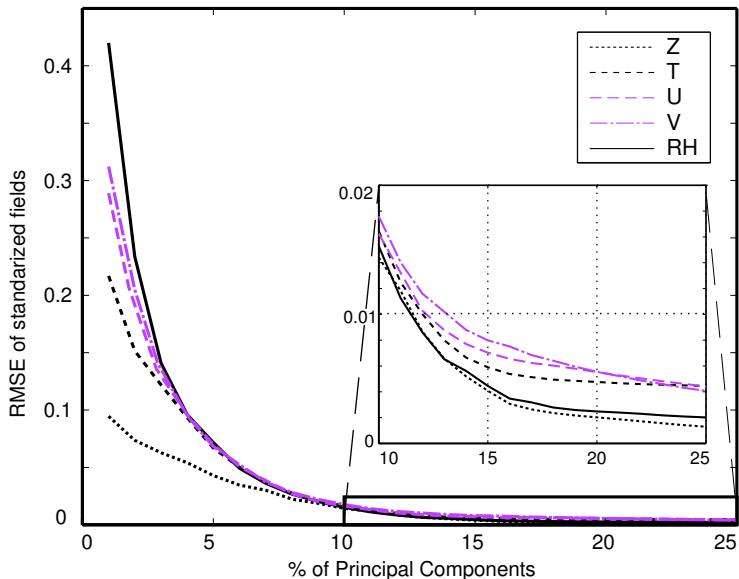


Figura 2.20: Error RMSE de reconstrucción para cada una de las cinco variables en el Modelo 1 (el error es calculado para los campos 3D normalizados) frente al número de CPs (variando desde 1 % de la dimensión del vector original, al 25 %).

Por ejemplo, se puede adoptar como criterio que los errores de reconstrucción sean inferiores a los errores de asimilación habituales en los modelos numéricos. En la Fig. 2.20 se observa que utilizando tan sólo un 10 % de las variables originales se tiene un error de reconstrucción menor del 2 % de la desviación estándar de los campos 3D, cifra inferior a los errores promedio de asimilación. Un criterio alternativo para seleccionar el número apropiado de componentes principales sería utilizar la distancia promedio entre los patrones vecinos en una base de datos de re-análisis.

Otro criterio más práctico para seleccionar el número óptimo de CPs sería elegir el que proporcione mejores resultados de validación cuando se aplique un método concreto. Por ejemplo, en algunos ejemplos de este libro, se aborda el problema de la predicción meteorológica probabilística local. Por tanto, un criterio a seguir para elegir el número óptimo de CPs sería en base al menor error de validación. Por ejemplo, la Fig. 2.21 muestra la evolución del índice de pericia de Brier (Brier Skill Score, BSS; ver Cap. 7) en función del número de CPs consideradas al aplicar un método estándar de predicción local denominado k -NN (ver Cap. 3.2). Esta figura indica que el número de componentes relevantes para el método es sustancialmente bajo (menor de 25 para umbrales bajos de precipitación). Así mismo, se observa que a medida que el evento es más raro (por ejemplo $Precip > 20mm$), el número de componentes óptimo se incrementa sustancialmente. Este ejemplo ilustra que el número de componentes principales relevantes depende sustancialmente del problema que se desea resolver y del método utilizado para su resolución.

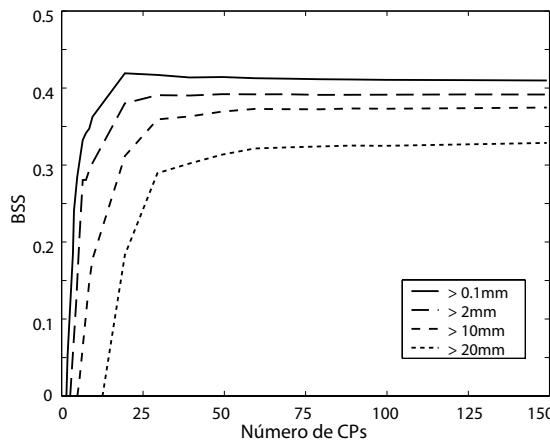


Figura 2.21: Evolución del Brier Skill Score (BSS) para la predicción de los eventos $Precip > 0.1mm$, 2 , 10 , y $20mm$.

2.5.3. Efectos de la Escala Temporal

La escala temporal de los datos viene dada por el tipo de estudio que se deseé realizar. En estudios de tipo climático, que sólo analizan patrones sinópticos promedio semanales o mensuales de gran escala, el número de componentes tomadas suele ser reducido, y cada una de las EOFs resultantes se analiza en el contexto de los distintos patrones de teleconexión, buscando una interpretación del patrón resultante, como se mostró en el Ejemplo 2.11. En cambio, cuando se llevan a cabo estudios sobre regiones más reducidas y con patrones de mayor variabilidad temporal (diarios, horarios, etc.) el número de CPs crece de forma considerable, ya que la correlación espacial de los patrones disminuye. En el siguiente ejemplo, se ilustra este hecho.

Ejemplo 2.12 (Componentes Principales y Escala Temporal). *Un ejemplo notorio lo constituyen los datos de observaciones en una red de estaciones sobre una zona de interés. Se consideran las 100 estaciones de la red principal del INM mostradas en la Fig. 1.17, tomando patrones diarios formados por los correspondientes 100 valores diarios de temperatura, precipitación, o de racha máxima de viento. En la Figura 2.22 se muestra el porcentaje de varianza explicada en función del número de CPs tomadas para el patrón atmosférico de la Sec. 1.9.4 (Modelo 2), y los patrones de observaciones de temperatura, precipitación y racha máxima.*

En esta figura puede verse que el patrón atmosférico está altamente correlacionado y, por tanto, el porcentaje de varianza explicado con unas pocas CPs es muy elevado; por el contrario, la precipitación y la racha máxima presentan una menor correlación espacial y requieren de un número mayor de EOFs para ser representados adecuadamente. La temperatura muestra un comportamiento intermedio.

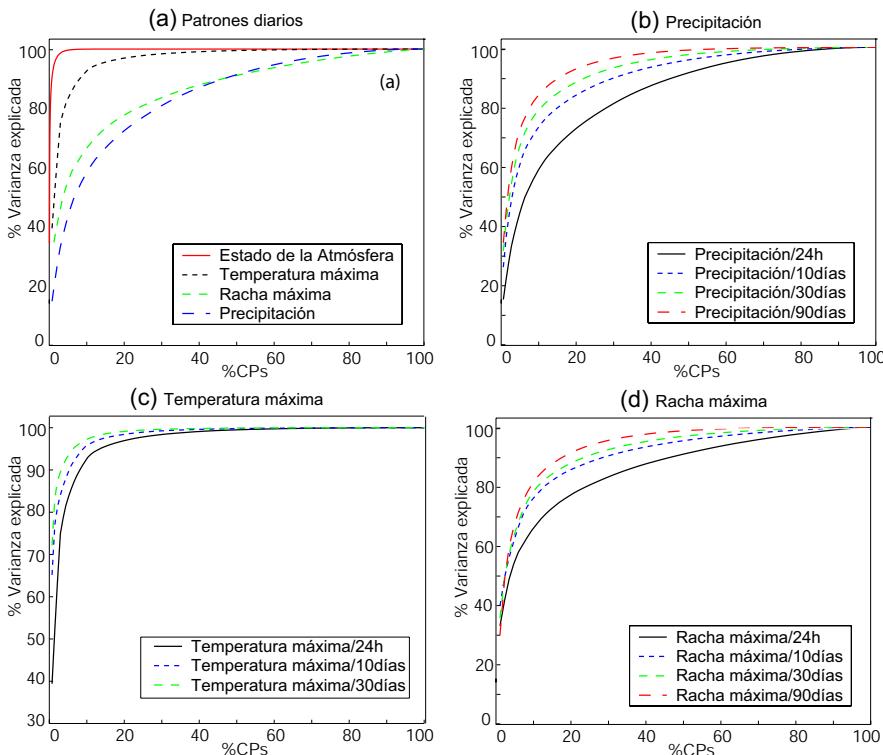


Figura 2.22: Porcentaje de varianza explicada en función del número de EOFs consideradas para patrones atmosféricos, temperatura, precipitación y racha máxima.

Por otra parte, las Figuras 2.22(b)-(d) muestran el aumento de la correlación espacial al considerar promedios temporales (medias diarias, decenal, mensual y estacional) de esas mismas variables. Por tanto, a medida que crece la escala temporal donde el patrón está promediado, decrece el número de CPs necesario para alcanzar un umbral requerido. Por tanto, estudios de escala estacional, o de cambio climático, que trabajan con promedios mensuales de las variables requerirán un número menor de CPs que estudios que requieran el uso de patrones diarios. ■

2.6. Técnicas de Agrupamiento

En esta sección se describen brevemente las técnicas de agrupamiento que son utilizadas en numerosas disciplinas para dividir o segmentar un conjunto de datos en subconjuntos homogéneos siguiendo algún criterio de similitud (ver Anderberg, 1973, para una descripción detallada de estos métodos). Una primera división de estas técnicas se puede establecer en base a su carácter *jerárquico* o *particional*, según las características del proceso seguido para construir los grupos.

2.6.1. Técnicas Jerárquicas

Las técnicas de agrupamiento jerárquico son iterativas y proceden uniendo grupos pequeños (técnicas *aglomerativas*), o dividiendo grupos grandes (técnicas *divisivas*), donde el concepto de tamaño viene dado por la medida de similitud utilizada (correlación, distancia, información mutua, etc.). Dentro de estos métodos destacan los llamados SHAN, que comparten las siguientes características:

- Secuencial (Sequential): el mismo algoritmo es aplicado iterativamente a los grupos disponibles.
- Jerárquico (Hierarchical): la secuencia de uniones de grupos se representa mediante una estructura de árbol.
- Aglomerativa (Agglomerative): inicialmente cada punto del conjunto de datos es asignado a un grupo distinto; y el algoritmo procede uniendo los grupos más similares hasta que el criterio de parada es alcanzado.
- Sin-solapamiento (Non-overalpping): ningún elemento puede pertenecer simultáneamente a dos grupos diferentes.

Varias alternativas son posibles, dependiendo de la métrica utilizada para definir la similitud entre grupos. Por un lado, el método conocido como “enlace promedio” define una distancia inter-grupo como la distancia promedio entre todos los posibles pares de elementos en los dos grupos comparados. Este método tiende a formar grupos con varianzas similares. Por otro lado, el “enlace de Ward” mezcla aquellos pares de grupos que minimizan la dispersión del grupo resultante. En este caso, el cuadrado de la distancia Euclídea es tomada como medida de diferencia (para más detalles sobre la aplicación de estos métodos en meteorología, consultar Kalkstein et al. (1987)). A continuación, se describe el método de Ward como ejemplo de estas técnicas.

Dado un conjunto de datos $\{v_1, v_2, \dots, v_n\}$, este método descompone la varianza total V en varianzas intragrupo, para los grupos actuales C_i con centroides c_i y peso, o masa, m_i (en un paso de iteración dado) y la varianza entre grupos:

$$V = \sum_q m_q \|c_q - c\|^2 + \sum_q \sum_{i \in C_q} m_i \|v_i - c_q\|^2, \quad (2.31)$$

donde c es el centroide global (media de los datos).

Si dos grupos C_i y C_j , con masas m_i y m_j respectivamente, son unidos en un solo grupo, D , con masa $m_i + m_j$ y centroide

$$d = \frac{m_i c_i + m_j c_j}{m_i + m_j}, \quad (2.32)$$

entonces la varianza V_{ij} de C_i y C_j respecto a D pueden ser descompuestas por la ecuación

$$V_{ij} = m_i \|c_i - d\|^2 + m_j \|c_j - d\|^2 + m_i \|d - c\|^2. \quad (2.33)$$

El último término es el único que permanece constante si cambiamos C_i y C_j por el centro de gravedad D . Entonces, la reducción de la varianza será:

$$\Delta V_{ij} = m_i \|c_i - d\|^2 + m_j \|c_j - d\|^2. \quad (2.34)$$

Usando (2.32), se tiene:

$$\begin{aligned}\Delta V_{ij} &= m_i \|c_i - \frac{m_i c_i + m_j c_j}{m_i + m_j}\|^2 + m_j \|c_j - \frac{m_i c_i + m_j c_j}{m_i + m_j}\|^2 \\ &= \frac{m_i m_j}{m_i + m_j} \|c_i - c_j\|^2.\end{aligned}\quad (2.35)$$

Luego la estrategia seguida por este método es la unión, en cada paso, de grupos C_i y C_j que minimiza ΔV_{ij} (initialmente cada punto es considerado como un solo grupo). Así que se puede considerar ΔV_{ij} como la medida de disimilitud. Los elementos con menos peso son los primeros en unirse.

El algoritmo de agrupamiento puede ser representado gráficamente mediante un “dendrograma” (un árbol representando en diferentes niveles la jerarquía de uniones de los grupos individuales o grupos en diferentes pasos).

Ejemplo 2.13 (Regionalización Automática). Una de las primeras clasificaciones climatológicas basadas en criterios estadísticos es la debida a Köppen (1918), que definió un conjunto de climas basado en combinaciones de umbrales para la precipitación y temperatura en las distintas estaciones del año (ver Oliver, 1991, para una descripción histórica de las técnicas de clasificación climática). Esta clasificación aún continúa vigente y es el punto de partida de estudios más sistemáticos de regionalización. En fechas más recientes, las técnicas basadas en métodos de agrupamiento han mostrado ser simples y eficientes para este problema (ver Fovell and Fovell, 1993).

A la hora de aplicar técnicas de agrupamiento para el problema de la regionalización, primero hay que decidir qué variables van a considerarse para definir la climatología local de las distintas estaciones. Existen numerosas fuentes de información que permiten discriminar adecuadamente distintas regiones con climatologías homogéneas: geográficas (como longitud, latitud, elevación, pertenencia a cuencas hidrográficas), estadísticas (temperaturas extremas, medias mensuales, precipitación acumulada, humedad relativa, etc.). Las técnicas de agrupamiento permiten realizar de forma automática distintos experimentos combinando estas variables. En el siguiente ejemplo se muestra una sencilla aplicación considerando sólamente información relativa a la precipitación (ésta es la variable esencial en problemas hidrológicos, agrícolas y ecológicos). Cada estación es representada por un vector que caracteriza su climatología local. En este ejemplo, cada estación está caracterizada por un vector $v = (mp_v, mpi)$, donde mp_v y mpi son las medias estacionales de precipitación para verano e invierno, respectivamente. Se consideran datos de 30 años para un conjunto de 54 estaciones de la península Ibérica (ver Fig. 2.23(a)).

En todos los casos, las observaciones de precipitación diaria están disponibles de 1970 a 2000, sin falta de datos ni lagunas. La Fig. 2.24 muestra los

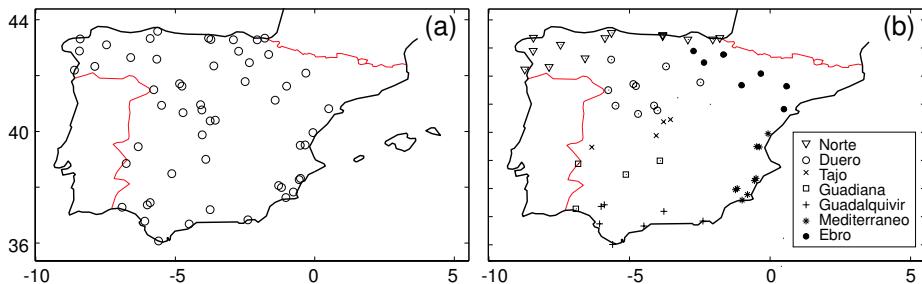


Figura 2.23: (a) Red de 54 estaciones automáticas en España; (b) Estaciones correspondientes a cada una de las siete cuencas hidrográficas principales consideradas en este ejemplo: Norte, Duero, Tajo, Guadiana, Guadalquivir-Sur, Mediterraneo, y Ebro.

grupos obtenidos aplicando el algoritmo de Ward considerando un máximo de 7 grupos. La razón para este criterio de parada es la coincidencia con las siete cuencas hidrográficas de la Península consideradas en este ejemplo (ver Fig. 2.23(b)).

La Figura 2.24 ilustra la capacidad de discriminación de la variable de precipitación promediada. Los símbolos en las figuras corresponden a los diferentes grupos obtenidos, y el dendrograma muestra el proceso aglomerativo. La lista de estaciones correspondientes a un grupo es dada bajo el correspondiente símbolo en el dendrograma; el número precedente a los nombres de las estaciones corresponde a la cuenca hidrográfica a la que pertenece la estación (Norte, Duero, Tajo, Guadiana, Guadalquivir-Sur, Mediterraneo, Ebro). En esta figura se puede ver como la cuenca Norte es claramente separada del resto de cuencas (esta es la principal separación climatológica en la Península Iberica, que corresponde con un clima “oceánico-húmedo” de acuerdo con la clasificación de Köppen). La única excepción corresponde a “Navacerrada” la cual se encuentra en una zona montañosa que presenta condiciones climatológicas parecidas a las estaciones de la cuenca Norte (al menos en promedio).

Por otra parte, el grupo etiquetado con una estrella corresponde a una región con un clima “semiárido”. Finalmente los dos grupos restantes corresponden a regiones con clima “verano seco subtropical”. El grupo etiquetado por “+” se encuentra básicamente en la cuenca del Guadalquivir y el grupo etiquetado con un cuadrado está disperso sobre toda la Península.

Este ilustrativo ejemplo del procedimiento de clasificación por agrupamiento automático no tiene valor climático alguno, y sólo sirve como ejemplo de las posibles aplicaciones de esta técnica. ■

2.6.2. Técnicas Particionales

Los métodos de agrupamiento más convenientes para un gran número de patrones en un espacio alto-dimensional son los métodos de ajuste de centroides iterativos. El método más común es el algoritmo de las k -medias (ver Hastie et al., 2001). Dado un grupo de vectores reales d -dimensionales $X =$

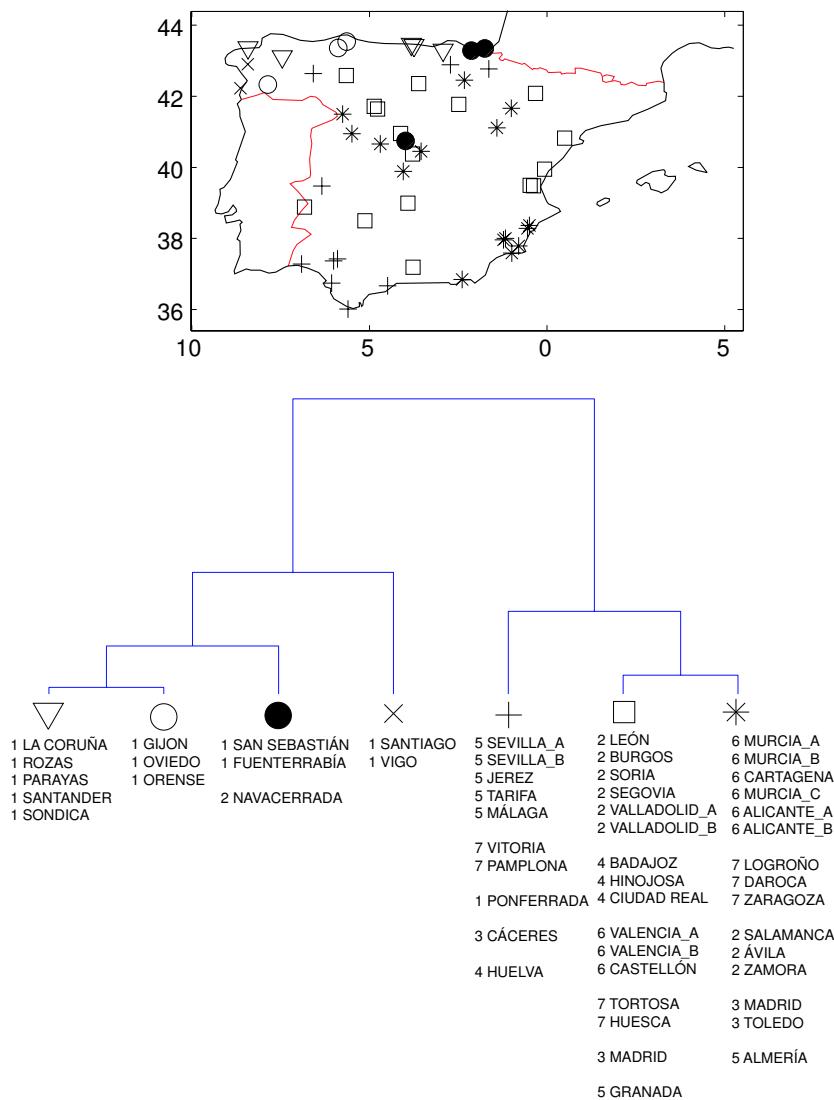


Figura 2.24: Análisis de agrupamiento mediante el enlace de Ward de 54 estaciones en la Península Ibérica caracterizado por el promedio de precipitación en Invierno y Verano. El dendrograma representa los diferentes grupos a un cierto nivel de profundidad, en el eje horizontal, y las distancias donde los elementos son unidos en diferentes niveles jerárquicos, en el eje vertical.

$\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, y un número determinado de grupos m , el algoritmo de las k -medias calcula un conjunto de prototipos d -dimensionales, $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$, o centroides, cada uno de ellos caracterizando a un grupo de datos $C_i \subset X$ formado por los vectores para los cuales \mathbf{v}_i es el prototipo más cercano. Esta tarea es realizada siguiendo un procedimiento iterativo, el cual comienza con un conjunto inicial de centroides $\mathbf{v}_1^0, \dots, \mathbf{v}_m^0$, elegidos aleatoriamente (ver Peña et al., 1999, para una descripción y comparación de diferentes procedimientos de inicialización). El objetivo del algoritmo es minimizar globalmente la distancia intra-grupos:

$$\sum_{i=1, \dots, m} \sum_{\mathbf{x}_j \in C_i} \|\mathbf{x}_j - \mathbf{v}_i\|^2 \quad (2.36)$$

Ya que una búsqueda exhaustiva del mínimo es prohibitiva, se calcula un mínimo local mediante un ajuste iterativo de los centroides de los grupos, y reasignando cada patrón al centroide más cercano. En la iteración $(r+1)$ -ésima, cada uno de los vectores \mathbf{x}_j es asignado al grupo i -ésimo, donde $i = \operatorname{argmin}_c \|\mathbf{x}_j - \mathbf{v}_i^r\|$, y los prototipos son actualizados por medio de los correspondientes patrones:

$$\mathbf{v}_i^{r+1} = \sum_{\mathbf{x}_j \in C_i} x_j / \#C_i,$$

donde $\#C_i$ denota el número de elementos en C_i . Bajo ciertas condiciones, el proceso iterativo anterior converge después de R iteraciones, y los centros finales \mathbf{v}_i^R son los prototipos (centroides). Cada uno de los centroides \mathbf{v}_i representa un grupo C_i formado por los patrones más cercanos a \mathbf{v}_i^R que a cualquier otro centroide. El algoritmo de k -medias consiste en los siguientes pasos:

1. Seleccionar el número de grupos deseados m .
2. Inicializar los centroides de los grupos (p.e. aleatoriamente).
3. Repetir:
 - a) Asignar cada vector (patrón atmosférico) al grupo más cercano.
 - b) Re-calcular los centroides de cada grupo, para que sean la media de los patrones asignados a ese grupo.

Ejemplo 2.14 (Clasificación de Patrones Atmosféricos). Se consideran los 5500 patrones atmosféricos del re-análisis ERA-15 definidos en la Sec. 1.9.4 en distintas rejillas (Modelos 1 y 2); para reducir la dimensionalidad del espacio, se toman las primeras 100 componentes principales (ver Ejemplo 2.11) y se consideran distintos números de grupos $m = 100, 200$, y 400 , que corresponden a distintos tamaños promedio de grupo: aproximadamente 50, 25, y 15, respectivamente. Por ejemplo, las Figs. 2.25 (a) y (b) muestran los prototipos obtenidos al aplicar el algoritmo con $m = 100$ para los Modelos 1 y 3, respectivamente. ■

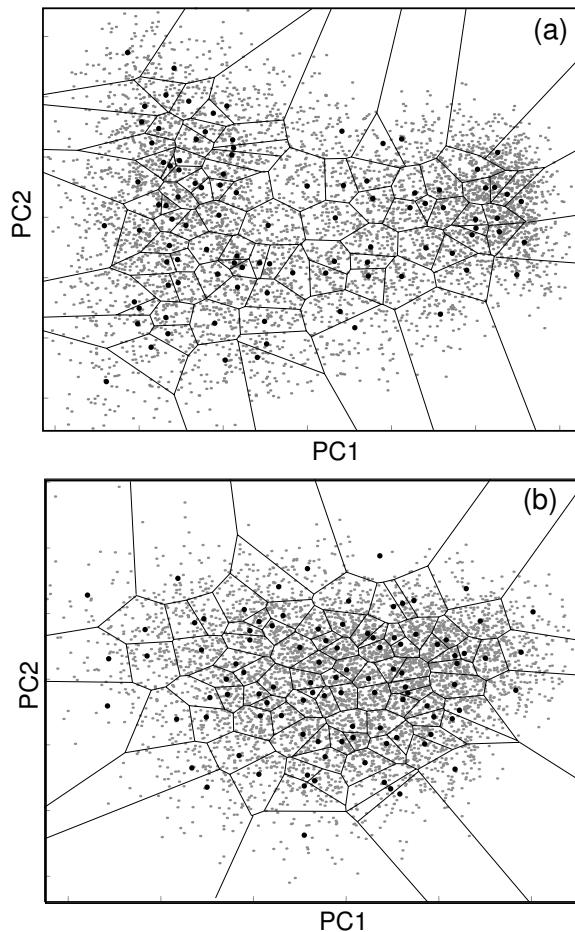


Figura 2.25: Agrupamiento del re-análisis ERA-15 con el algoritmo de k -medias considerando 100 grupos para (a) Modelo 1, (b) Modelo 3. El grafo muestra los patrones diarios y los centroides proyectados en el espacio de las dos primeras componentes principales. También se muestran las líneas de separación entre grupos (estas líneas corresponden al diagrama de Voronoi asociado a los centroides).

2.6.3. MeteoLab: *Técnicas de Agrupamiento*

Para aplicar las funciones de agrupamiento, se cargan primero los patrones atmosféricos relativos a un dominio (por ejemplo el dominio de la NAO) como se mostró en el Ej. 2.5.1

```
>> dmn=readDomain('Nao');
>> [EOF,CP,MN,DV]=getEOF(dmn,'ncp',50);
```

la función `makeClustering` permite aplicar los algoritmos descritos en las secciones anteriores (junto con una técnica neuronal que se describirá en la Sec. 5.8). Esta función requiere el número de grupos que se desean calcular, así como el tipo de agrupamiento deseado (en este caso *k*-medias, '`kmeans`')

```
>> Clustering=makeClustering(CP,'Kmeans',50)

Clustering = 

    NumberCenters: 50
        Type: 'KMeans'
        Centers: [50x50 double]
        PatternsGroup: [16436x1 double]
        PatternDistanceGroupCenter: [16436x1 double]
        SizeGroup: {50x1 cell}
        Group: {50x1 cell}
```

Esta función proporciona como salida una estructura (en este caso bajo el nombre de `Clustering`) con campos relativos al número de clases o grupos (`Clustering.NumberCenters`), tipo de agrupamiento, prototipos de cada uno de los grupos (`Clustering.Centers`), índice del grupo al que pertenece cada uno de los patrones (`Clustering.PatternsGroup`), distancia de cada patrón al prototipo de su grupo, tamaño de cada grupo y, finalmente, índices de los patrones que pertenecen a cada grupo.

A continuación se dibujan los patrones y los centros obtenidos por el agrupamiento proyectados sobre las dos primeras componentes principales.

```
>>figure
>>plot(CP(:,1),CP(:,2),'b.')
>>hold on
>>plot(Clustering.Centers(:,1),Clustering.Centers(:,2),...
    'ko',...
    'MarkerSize',6,...
    'MarkerEdgeColor',[0 0 0],...
    'MarkerFaceColor',[1 0 0])
```

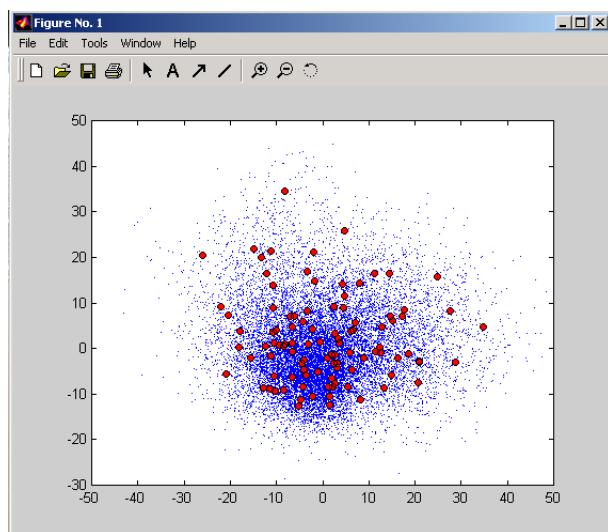


Figura 2.26: Agrupamiento del re-análisis ERA40 con el algoritmo de k -medias considerando 50 grupos para la zona NAO. El grafo muestra los patrones diarios y los centros proyectados en el espacio de las dos primeras componentes principales.

CAPÍTULO 3

Técnicas Estadísticas. Modelización y Predicción

3.1. *Introducción*

En el capítulo anterior se han descrito algunos métodos estadísticos para la exploración de datos. Estas técnicas permiten analizar la información disponible en un problema dado, filtrando los datos redundantes y obteniendo sólo el conocimiento útil para el mismo, expresado de la forma más compacta posible. En este capítulo se presentan las técnicas estadísticas estándar para la modelización y predicción que operan a partir de esos datos filtrados. Estas técnicas permiten obtener modelos lineales que explican la variabilidad y tendencias de un cierto sistema, y permiten predecir su estado futuro hasta un cierto horizonte. Se describen algunas de las técnicas de mayor aplicación en el ámbito de las ciencias atmosféricas, prestando especial atención a las aplicaciones prácticas en este campo.

La Sec. 3.2 analiza un ejemplo de gran importancia práctica que será utilizado en el resto del capítulo para ilustrar la aplicación de los métodos descritos: el downscaling o mejora de resolución estadística. En la Sec. 3.3 se analizan las técnicas de series temporales y se describen algunas aplicaciones en el ámbito de la meteorología. A continuación, la Sec. 3.4 describe el problema de la regresión, para relacionar una variable con otras a través de una muestra de datos. Finalmente, la Sec. 3.5 describe la técnica de correlación canónica que establece relaciones entre dos conjuntos distintos de variables. Otras técnicas estadísticas no lineales más avanzadas serán descritas en otros capítulos a lo largo del libro.

3.2. *Downscaling Estadístico*

El problema denominado downscaling (o aumento de resolución) consiste en “interpolar” las predicciones de un modelo numérico dadas en una cierta rejilla a puntos sub-rejilla (una red de estaciones distribuidas irregularmente, una rejilla de mayor resolución, etc.). En la práctica, este problema aparece cuando la resolución espacial del modelo numérico de circulación atmosférica (ACM) es insuficiente para una aplicación dada, o cuando se desean obtener predicciones locales en ciudades, o localidades de interés. El problema es aún más complicado cuando se desean obtener predicciones de variables en superficie cuya salida no proporcione directamente el modelo (temperatura máxima diaria, etc), sino que han de estimarse a partir de otras variables disponibles. La Fig. 3.1 ilustra este problema en un ejemplo en el que se desea obtener una predicción local de precipitación en distintos puntos de la red secundaria de estaciones del INM, a partir de la predicción operativa del modelo del ECMWF (en la actualidad, la resolución de este modelo es el doble de la mostrada en la figura).

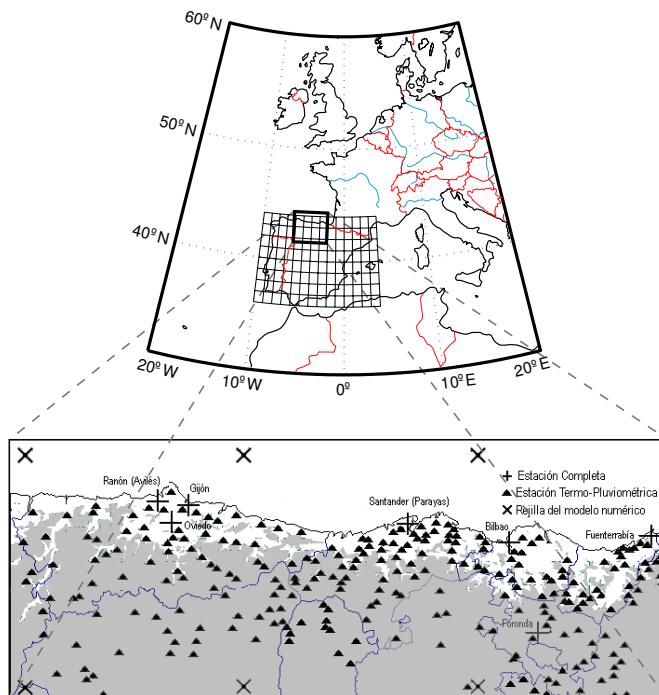


Figura 3.1: El problema de downscaling. La figura superior muestra la rejilla de un modelo numérico de 1° de resolución horizontal (aprox. 100 km) sobre la Península Ibérica; la figura inferior muestra una ampliación de la cornisa Cantábrica detallando los puntos de rejilla del modelo numérico ('x'), y los observatorios de la red principal ('+') y secundaria (triángulos) del INM donde se desea obtener una predicción local, y donde se dispone de observaciones históricas.

Se han propuesto diferentes métodos para abordar este problema desde distintas perspectivas, utilizando no sólo las predicciones de los modelos numéricos, sino también las series temporales de observaciones disponibles en la región o localidad de interés (cuando éstas están disponibles). Estos métodos podrían clasificarse en cuatro grandes grupos:

- *Técnicas dinámicas de downscaling* (o incremento dinámico de resolución), que incrementan la resolución de la rejilla del modelo atmosférico global, anidando a éste un modelo atmosférico regional o mesoescalar limitado a la zona de interés (ver Sec. 1.5). De esta forma se puede mejorar la orografía y la parametrización de algunos procesos físicos locales de importancia en el área de interés (radiación, capa límite, cúmulos, etc.). El principal problema de estos modelos es que las parametrizaciones han de elegirse cuidadosamente para la zona de estudio, pues distintas parametrizaciones pueden dar lugar a resultados muy diferentes (ver Sec. 1.3.2). Por ejemplo, la Fig. 3.2 muestra dos salidas de precipitación del modelo mesoscalar MM5 integrado para dos configuraciones distintas de las parametrizaciones. Otros trabajos muestran cómo la variación de las parametrizaciones de radiación influyen en la predicción del viento en superficie, etc.

En ocasiones, estos métodos numéricos se acoplan con técnicas estadísticas, como los filtros de Kalman (Bergman and Delleur, 1985), para eliminar la influencia de los errores sistemáticos del modelo global, adaptando dinámicamente las salidas de los modelos numéricos en base a las observaciones disponibles.

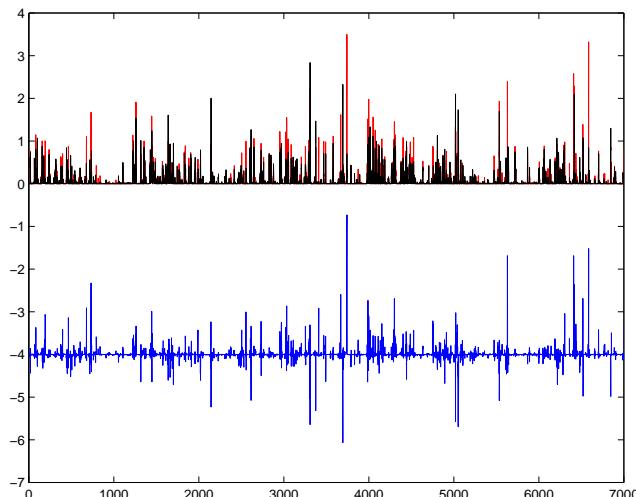


Figura 3.2: Precipitación en un punto de rejilla de un área de 35×35 km simulada cada 6 horas con el modelo MM5 utilizando dos parametrizaciones distintas (códigos 4.3.5.2.1.0 en rojo y 5.3.2.4.1.0 en negro). La diferencia entre ambas predicciones (figura inferior) tiene el eje vertical trasladado para claridad de visualización.

Ejemplo proporcionado por Álvaro Pascual, del CIEMAT (Madrid).

- *Técnicas físicas de downscaling*, que aplican relaciones físicas entre las distintas variables del modelo y otras variables auxiliares (orografía, etc.) para estimar el valor de los campos físicos que se desean predecir en una rejilla de alta resolución a partir de las salidas de baja resolución de los modelos. Estas técnicas suelen estar incorporadas en el post-proceso de las salidas de los modelos dinámicos anteriores. Por ejemplo, el sistema de diagnóstico CALMET (www.src.com/calpuff/calpuff1.htm) puede trabajar directamente con los campos de salida del modelo de mesoscala MM5.
- *Técnicas estadísticas de downscaling*. Estos métodos se aplican cuando se dispone de registros históricos de la variable que se desea predecir en la rejilla de alta resolución, o en los puntos locales de interés (por ejemplo, como se muestra en la Fig. 3.1). En este caso, distintos métodos estadísticos permiten combinar las salidas de los modelos numéricos con esta información, para desarrollar modelos de predicción que relacionen las salidas de los modelos numéricos con las observaciones locales. Para ello, es necesario disponer de un conjunto suficiente de datos simultáneos de registros históricos y de campos atmosféricos simulados por algún proyecto de re-análisis (ver Sec. 1.9 para más detalles sobre los datos disponibles en la actualidad).

En este capítulo se analizan las técnicas estadísticas de series temporales, regresión y correlación canónica que se aplican en la práctica en este y otros problemas para predecir el valor de unas variables (predictandos) en base a otras (predictores).

3.3. *Técnicas Estadísticas para Series Temporales*

En la literatura se han propuesto diferentes métodos para abordar el problema de la predicción meteorológica local a partir de series temporales de observaciones disponibles en una región o localidad de interés. Con esta información se pueden estimar modelos estadísticos de predicción (modelos autoregresivos, de Markov, etc.) que pueden ser posteriormente utilizados para predecir futuros valores y tendencias. Estos métodos son puramente estadísticos y no utilizan el conocimiento físico sobre la dinámica de la atmósfera ni las salidas numéricas de los modelos de circulación atmosférica. Planteado de forma genérica, un problema de series temporales es de la siguiente forma:

Dada una serie de observaciones x_1, \dots, x_N de una cierta variable (por ejemplo, la precipitación diaria en una localidad), se desea obtener un modelo que describa su evolución y permita predecir el estado futuro de la misma x_{N+1}, \dots .

En la actualidad se dispone de series históricas homogéneas y suficientemente largas en numerosos puntos geográficos (ver Sec. 1.9), por lo que es posible aplicar técnicas estándar de series temporales a este problema.

3.3.1. Modelos Autoregresivos

Las técnicas lineales autoregresivas tratan de explicar la estructura de la serie temporal de forma que el valor actual de la serie dependa de los valores inmediatamente anteriores y de una componente aleatoria ϵ_n de media cero y varianza σ^2 (ver, por ejemplo, Box and Jenkins, 1976). Así, un proceso x_n se dice que es un *modelo autoregresivo* de orden p (denotado por $AR(p)$) si

$$x_n = \alpha_1 x_{n-1} + \alpha_2 x_{n-2} + \dots + \alpha_p x_{n-p} + \epsilon_n, \quad (3.1)$$

donde $\alpha_1, \dots, \alpha_p$ son los parámetros del modelo. Suponiendo que el proceso es estacionario y de varianza finita, estos parámetros pueden obtenerse a partir las ecuaciones de Yule-Walker, que relacionan la función de autocorrelación $\rho(k)$ con los parámetros del modelo:

$$\rho(k) = \sum_{m=1}^p \alpha_m \rho(k-m). \quad (3.2)$$

Estas relaciones conducen a la ecuación matricial:

$$\mathbf{R} \hat{\boldsymbol{\alpha}} = \mathbf{r}, \text{ con } \mathbf{R} = \begin{pmatrix} 1 & r_1 & r_2 & \dots & r_{p-1} \\ r_1 & 1 & r_1 & \dots & r_{p-2} \\ r_2 & r_1 & 1 & \dots & r_{p-3} \\ \vdots & & & \ddots & \\ r_{p-1} & r_{p-2} & r_{p-3} & \dots & 1 \end{pmatrix}, \quad (3.3)$$

donde r_i es la estimación del coeficiente de autocorrelación $\rho(i)$ obtenida de la serie temporal.

En la práctica no es sencillo identificar el orden de un proceso autoregresivo a partir de la función de autocorrelación; una forma de obtener el valor óptimo es ir ajustando progresivamente modelos hasta que se stabilice el error, o hasta que la mejora del error no compense el incremento de complejidad del modelo resultante (distintos métodos de selección de modelos se describen en Hastie et al., 2001).

A parte de los modelos autoregresivos, han sido propuestas numerosas variantes de modelos de series temporales en la literatura, cada una de las cuales incorpora un nuevo elemento para modelizar el comportamiento de la serie. Por ejemplo, los *modelos autoregresivos de media móvil* incluyen la dependencia de los términos aleatorios en tiempos pasados. Un proceso $ARMA(p, q)$ es de la forma:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \dots + \beta_q \epsilon_{t-q} \quad (3.4)$$

Estos modelos también dan lugar a procesos estacionarios (sus parámetros estadísticos, como la media o varianza, permanecen constantes en el tiempo). Sin embargo, en la práctica, la mayoría de las series son no estacionarias y, por tanto, antes de ajustar un modelo AR o $ARMA$ es necesario eliminar las fuentes de variación. Otros modelos permiten trabajar con la serie no estacionaria

original; por ejemplo, los modelos ARIMA permiten tratar series no estacionarias en la media; por otra parte, los modelos SARIMA incluyen componentes periódicas estacionales y permiten tratar series con este tipo de no estacionariedad (ver Chatfield, 2003, para una descripción actualizada de estos métodos).

Estos modelos han sido aplicados en meteorología con cierto éxito en escalas de tiempo mensual o estacional, donde los promedios de las variables son aproximadamente gaussianos. Por ejemplo, Zwiers and von Storch (1990) emplean un modelo AR para la modelización del patrón El Niño/Oscilación del Sur (ENSO); por otra parte Verma et al. (2002) utilizan un modelo ARIMA para caracterizar medias mensuales de precipitación. Sin embargo, dado que la mayoría de las variables meteorológicas son no gaussianas en el corto plazo, estas técnicas son inadecuadas para la predicción local aplicada a esta escala temporal. Para solventar este problema, se han introducido distintas extensiones no lineales de los modelos autoregresivos; por ejemplo, Buhamra et al. (2003) analizan en detalle las propiedades de distintas extensiones basadas en redes neuronales (estos modelos se describen con detalle en el Cap. 5).

3.3.2. MeteoLab: *Modelos Autoregresivos.*

Existen distintos paquetes de Matlab para ajustar modelos autoregresivos a partir de un conjunto de datos. Uno de estos paquetes es *ARFit* (ver www.gps.caltech.edu/~tapio/arfit). En este ejemplo se consideran datos horarios de temperatura de Oviedo y se analiza la pericia de los modelos autoregresivos para predecir estos datos a distintos alcances (desde horario a diario):

```

Example.Network={'MyStations'};
Example.Stations={'hourlyData.stn'}; %Hourly temp in Oviedo
Example.Variable={'Common'};
[dat,Example]=loadStations(Example,'ascfile',1);
lt=5; data=dat(1:lt:end); % Choosing the leadtime: 1 is an hour

[w, A]=arfit(data,2,7);
% The order is automatically selected from 2 to 7.

k=size(A,2); % Order of the model
lag=[];
for i=1:k+1
    lag=[lag; data(i:end-(k+1)+i,1)'];
end

X=lag(1:k,:); % Lagged input variables
Y=lag(k+1,:); % Predicted values

pred=w*X*flipud(A');
figure; plot(pred,Y,'.k')
error=mean(abs(Y-pred))

```

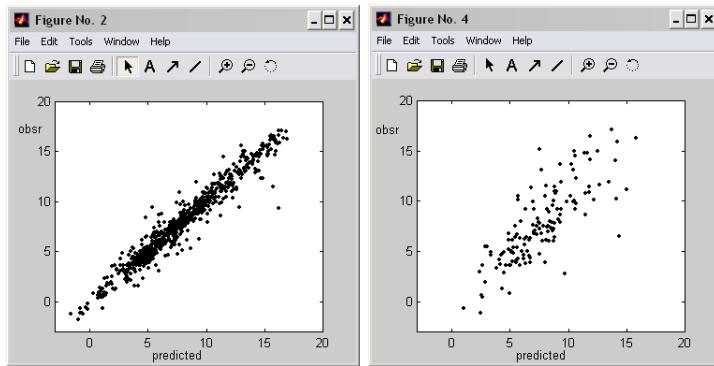


Figura 3.3: Valor previsto frente al observado para un horizonte de predicción de (izquierda) una hora y (derecha) cinco horas. Una predicción perfecta correspondería a la diagonal.

A medida que aumenta el horizonte de predicción del modelo (considerando datos cada dos horas, etc.) el error del modelo se incrementa rápidamente, como se muestra en la Fig. 3.4 (ver los ejemplos del paquete MeteoLab en `Meteolab/Statistics/TimeSeries` para más detalles). Puede observarse que más allá de las diez horas, el error del modelo se satura y éste pierde su capacidad predictiva.

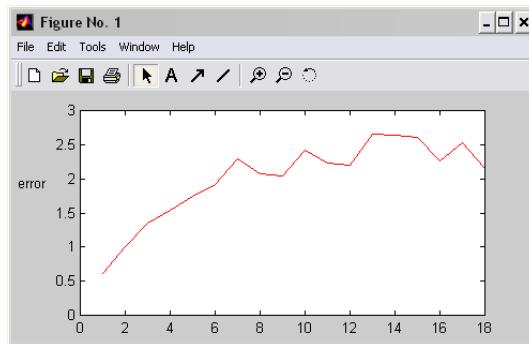


Figura 3.4: Error absoluto medio del modelo autoregresivo en función del horizonte de predicción (en horas).

3.3.3. Series Caóticas. Técnicas de Inmersión

En los últimos años, el análisis de series temporales no lineales ha cobrado un fuerte interés debido a sus múltiples aplicaciones (ver, por ejemplo, Abarbanel, 1995). Las técnicas clásicas autoregresivas de análisis de series temporales descritas en la sección anterior han sido generalizadas teniendo en cuenta las propiedades dinámicas de los sistemas no lineales (ver Sec. 1.6). Desde un punto

de vista práctico, el avance más importante en este campo ha sido la llamada técnica de *inmersión o delay-embedding* (Sauer, 1994), que supone que la serie temporal proviene de un sistema dinámico determinista (por ejemplo, una ecuación diferencial) definido en un conjunto de variables que componen su espacio de fases:

$$\frac{d\mathbf{x}(t)}{dt} = F(\mathbf{x}(t)), \quad (3.5)$$

Una serie temporal $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ tomada del sistema no lineal responderá a la relación funcional del sistema (3.5). Por tanto, con técnicas apropiadas se podrá obtener un modelo aproximado del mismo. Sin embargo, en la práctica no suele ser posible medir todas las variables que componen el espacio de fases de un sistema, sino que sólo se tiene acceso a observables derivados de estas variables $o(t) = G(\mathbf{x}(t))$; por ejemplo, no se pueden medir todos los grados de libertad que componen la atmósfera, pero se dispone de la medición de algunas variables derivadas (precipitación, temperatura, etc.) en puntos concretos. En estos casos, la técnica de inmersión ofrece una forma de reconstruir un espacio de fases topológicamente equivalente al original, tomando vectores de retrasos temporales del observable de la forma: $(o(t), o(t-\tau), \dots, o(t-(m-1)\tau))$, donde $o(t)$ es el valor del observable en el tiempo t y los parámetros τ y m son el *retraso* y la *dimensión de embedding*, respectivamente. Bajo ciertas condiciones (Sauer, 1994), existe un difeomorfismo (aplicación biyectiva, diferenciable y con inversa diferenciable) entre el espacio de fases original y el espacio de embedding resultante, de forma que la evolución del sistema en el espacio de fases puede ser determinada a partir de una función en el espacio de inmersión. De este modo se tiene una relación funcional de la forma:

$$o(t) = G(\mathbf{x}(t)) = H(o(t-\tau), \dots, o(t-(m-1)\tau)), \quad (3.6)$$

donde H será la función compuesta dada por el difeomorfismo G , la función de evolución del sistema dinámico y la función característica del observable. Los parámetros τ y m han de elegirse adecuadamente para que el espacio resultante sea topológicamente equivalente al espacio de fases original. La dimensión de este espacio (dimensión de embedding) está acotada teóricamente por el valor $m < 2D + 1$, donde D es la dimensión del atractor del sistema original. El retraso τ es arbitrario desde el punto de vista teórico; un retraso óptimo proporciona la cota anterior, pero otro retraso cualquiera sólo incrementaría la dimensión del espacio de inmersión resultante. Una mala elección obligará a tomar más retrasos de los necesarios, incrementando la complejidad del modelo resultante. En la práctica, una forma simple de seleccionar τ es garantizar la independencia lineal entre sucesivos retrasos; para ello se puede utilizar la función de autocorrelación o la información mutua de la serie (ver Abarbanel, 1995, para más detalles).

Para poder aplicar este tipo de técnicas a datos meteorológicos es necesario considerar la atmósfera como un sistema caótico de baja dimensión (Tsonis and Elsner, 1988). Un observable puede ser la temperatura medida en una estación meteorológica (por ejemplo, en Oviedo). La Fig. 3.5(a) muestra la serie temporal de la temperatura media horaria registrada en Oviedo entre los

años 1979 y 1993. En la figura sólo se han representado 3 años para observar con claridad la onda anual, con los correspondiente máximos de temperatura alcanzados en la época estival. En la gráfica superior se ha ampliado una zona de la serie temporal para mostrar la onda diurna.

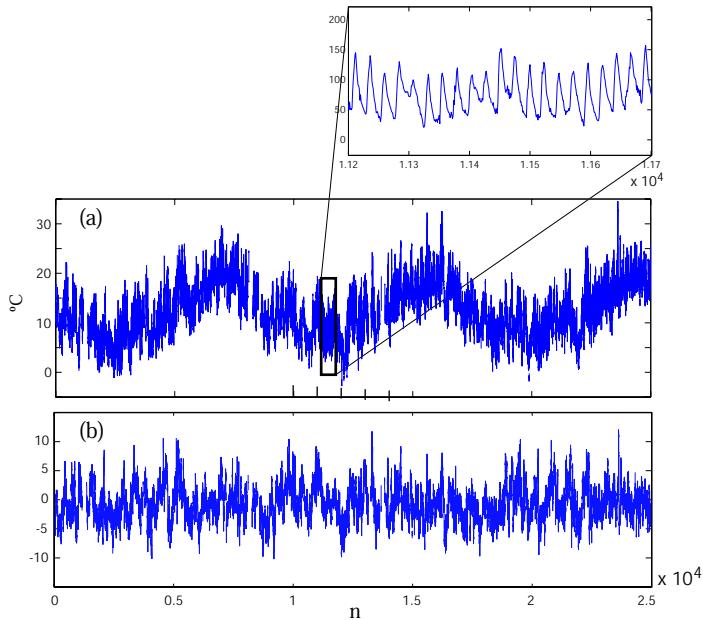


Figura 3.5: Temperatura media horaria tomada en el observatorio de Oviedo entre los años 1979 y 1981.

Como puede observarse en este ejemplo, la situación real cuando se trabaja con datos meteorológicos no es sencilla, pues la serie contiene varias componentes periódicas (onda anual, diurna, etc.) y ruido, además de la componente no lineal. Las componentes periódicas pueden separarse fácilmente (ver Fig. 3.5(b)), pero el efecto del ruido complica la aplicación de la técnica de embedding. Por ejemplo, la Fig. 3.6 muestra el espacio de inmersión para el ejemplo de la temperatura en Oviedo, considerando tanto la señal original, como la señal obtenida eliminando la onda anual. En esta última figura resulta difícil observar alguna dinámica determinista. Este hecho se puede comprobar de forma práctica comparando un modelo lineal y uno no lineal (por ejemplo, una red neuronal, como se muestra en el Cap. 5) entrenados con los datos. Como se verá, es difícil observar ninguna diferencia significativa entre ambos modelos.

Por otra parte, la hipótesis de la baja dimensionalidad de la atmósfera es bastante cuestionable (Lorenz, 1991) y los modelos desarrollados no se han mostrado competitivos con otras técnicas. No obstante, en algunas aplicaciones, estas técnicas han mostrado una gran utilidad en combinación con modelos numéricos. Por ejemplo, Pérez-Muñozuri and Gelpí (2000) presentan un modelo de predicción basado en esta técnica para la cubierta nubosa, y muestran

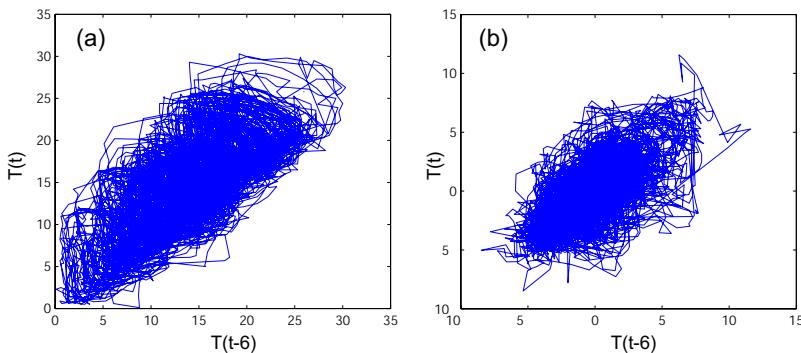


Figura 3.6: (a) Espacio de inmersión $(T(t), T(t - 6))$ para la serie de temperaturas de Oviedo. (b) Espacio de inmersión para la serie anterior restando una media móvil $(T(i - 24), T(i), T(i + 24))$ para eliminar la componente anual.

como estas predicciones pueden ser utilizadas por un modelo numérico mesoscalar (ARPS) para mejorar la capacidad de predicción del mismo.

Los recientes avances en el campo del caos espacio-temporal suponen una prometedora línea de trabajo para poder llegar a modelizar la dinámica atmosférica de forma más sólida y global (Gollub and Cross, 2000), sin tener que recurrir a hipótesis simplistas como su baja dimensionalidad.

3.3.4. Modelos de Markov

En el ámbito de la predicción probabilística también han sido aplicadas distintas técnicas estocásticas que operan a partir de las series temporales para pronosticar la probabilidad de un cierto evento meteorológico (por ejemplo la ocurrencia de lluvia “*Precip > 0.5mm*”, de vientos fuertes “*Racha > 80km/h*”, etc.). Las cadenas de Markov se aplican a variables discretas, y han sido utilizadas profusamente en este campo, comenzando con Gabriel and Neumann (1962) que desarrollaron un modelo probabilístico de precipitación para Tel Aviv, como se mostró en la Sec. 2.3. Estos modelos generalizan la idea de la persistencia estableciendo probabilidades de transición entre los distintos estados de la variable en sucesivos valores del tiempo. Suponiendo que el estado actual sólo depende de los k estados anteriores, un modelo de Markov de orden k permitiría cuantificar probabilísticamente la dinámica de transiciones del sistema.

El uso de estos modelos se ha extendido a la predicción de eventos más complejos, como el día de comienzo de lluvias (el primer día que lloverá a partir de la fecha actual), o el primer día que se superará un cierto umbral de precipitación (Stern, 1982). Además de las cadenas de Markov, también han sido aplicados otros métodos estocásticos a estos problemas: procesos continuos puntuales (point process) (Rodríguez-Iturbe et al., 1987), procesos de renovación (Markov renewal process) (Faoufoula-Georgiou and Lettenmaier, 1987), etc.

3.4. Regresión

La regresión es la técnica estadística más utilizada en la práctica, pues permite inferir relaciones funcionales entre distintas variables a partir de una muestra de las mismas, y hacer predicciones en base a ellas (ver Chatterjee et al., 2000, para una introducción a este tema). El planteamiento general del problema es el siguiente:

Dada una muestra simultánea de variables X_1, \dots, X_p (predictores) e Y (predictando), se desea obtener una relación funcional que permita estimar el valor del predictando a partir de los valores de los predictores.

Para ello, se pueden usar modelos de regresión de la forma

$$\hat{y} = f(x_1, \dots, x_n) = f(\mathbf{x}), \quad (3.7)$$

y obtener f a partir de una muestra $\{(\mathbf{x}_k, y_k), k = 1, \dots, n\}$, de forma que los residuos $\epsilon_k = y_k - \hat{y}_k$, con $\hat{y}_k = f(\mathbf{x}_k)$, sean mínimos (utilizando alguna métrica).

3.4.1. Regresión Lineal

El más simple de estos modelos es el de regresión lineal, que supone una relación funcional lineal entre las variables de la forma:

$$y = a_0 + a_1 x_1 + \dots + a_p x_p + \epsilon, \quad (3.8)$$

donde a_0, \dots, a_p son los parámetros del modelo y ϵ es una variable aleatoria gaussiana. En ocasiones es habitual considerar variables transformadas para tener en cuenta, por ejemplo, fluctuaciones periódicas (como el ciclo diario o anual) como en el siguiente modelo:

$$y = a_0 + a_1 \sin(x_1) + a_2 x_2 + \dots + a_p x_p + \epsilon. \quad (3.9)$$

Obsérvese que estos modelos siguen siendo lineales (en las variables transformadas).

En la práctica, se está interesado en estimar los parámetros del modelo a partir de una muestra $\{(y_k, x_{1k}, \dots, x_{pk}), k = 1, \dots, n\}$. Por ejemplo, en el caso simple de una regresión lineal unidimensional: $y_k = a x_k + b$, los parámetros a y b se pueden estimar aplicando el método de mínimos cuadrados, que consiste en minimizar la suma de los cuadrados de los errores:

$$E = \sum_1^n (\epsilon_k)^2 = \sum_{k=1}^n (y_k - \hat{y}_k)^2 = \sum_{k=1}^n (y_k - b - ax_k)^2 \quad (3.10)$$

Por tanto, se tienen las denominadas ecuaciones normales:

$$\frac{\partial M}{\partial b} = 0 \Rightarrow \sum_{k=1}^n y_k = nb + a \sum_{k=1}^n x_k, \quad (3.11)$$

$$\frac{\partial M}{\partial a} = 0 \Rightarrow \sum_{k=1}^n x_k y_k = b \sum_{k=1}^n x_k + a \sum_{k=1}^n x_k^2, \quad (3.12)$$

de las que se obtiene la solución final:

$$b = \frac{\sum_{k=1}^n y_k}{n} - a \frac{\sum_{k=1}^n x_k}{n} = \bar{y} - a \bar{x} \quad (3.13)$$

$$a = \frac{n \sum_{k=1}^n x_k y_k - \sum_{k=1}^n x_k \sum_{k=1}^n y_k}{n \sum_{k=1}^n x_k^2 - (\sum_{k=1}^n x_k)^2} = \frac{\sigma_{xy}}{\sigma_x^2} \quad (3.14)$$

Por tanto, la recta de regresión lineal tiene la siguiente expresión directa:

$$\hat{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x}) + \bar{y} \quad (3.15)$$

Antes de utilizar este modelo es importante entender las hipótesis implícadas, y comprobar que los datos a analizar no las contradicen. El modelo de regresión lineal implica que existe una relación lineal entre el valor y_k del predictando y los valores x_{1k}, \dots, x_{pk} de los predictores. Por otra parte, las observaciones x_{1k}, \dots, x_{pk} no contienen errores, mientras que las observaciones de y_k se miden con un error aleatorio ϵ_k de media cero y varianza σ^2 constante e incorrelacionado.

Sin embargo, en la práctica, estas condiciones no se cumplen de forma exacta y es necesario medir la calidad del ajuste entre los puntos observados y_k y la recta de regresión. Para ello, Pearson introdujo el “coeficiente de correlación lineal”, que definió como:

$$R^2 = \frac{\sum_{k=1}^n (\hat{y}_k - \bar{y})^2}{\sum_{k=1}^n (y_k - \bar{y})^2}. \quad (3.16)$$

Si $R^2 = 1$, entonces los valores observados y_k y los valores estimados \hat{y}_k coinciden; es decir, todos los valores están sobre la recta de regresión. Teniendo en cuenta la definición de la recta de regresión (3.15), y la definición de varianza, se tiene que:

$$R^2 = \frac{\frac{\sigma_{xy}^2}{\sigma_x^4} n \sigma_x^2}{n \sigma_y^2} = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2}. \quad (3.17)$$

Esto permite definir la ecuación de regresión en función de R

$$\hat{y} - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} (x - \bar{x}) = \frac{R \sigma_y}{\sigma_x} (x - \bar{x}).$$

Obsérvese que R no es más que el coeficiente de correlación definido en el capítulo anterior.

Ejemplo 3.1 (El Niño. Regresión). *En este ejemplo se utilizan de nuevo los datos de precipitación, caudal del río y temperatura de la superficie del mar, descritos en el Ejemplo 2.8. Se trata de establecer una relación lineal entre la precipitación en Piura (predictando), y la temperatura del mar en Chicama (predictor). En este caso, operando de acuerdo a lo visto anteriormente, se obtiene la recta de regresión*

$$\hat{y} = 15.986x - 261.438,$$

donde \hat{y}_k es el valor estimado para la precipitación en Piura para cada valor x_k de la temperatura del mar en Chicama (ver Fig. 3.7(a)). El parámetro de ajuste es $R^2 = 0.34$ ($R = 0.58$, que coincide con el coeficiente de correlación obtenido en el Ejemplo 2.8). Este valor indica que el modelo obtenido no es apropiado para representar los datos.

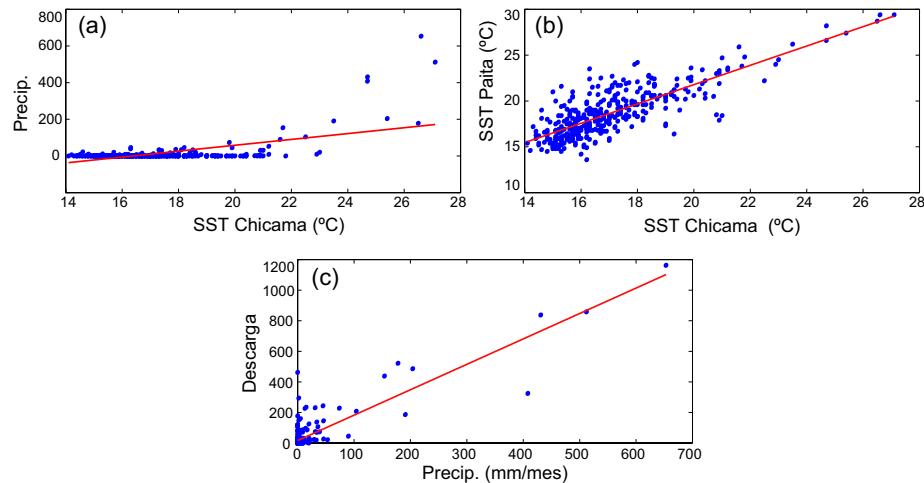


Figura 3.7: (a) Temperaturas del agua del mar en Chicama, frente a la precipitación. (b) Temperaturas del agua del mar en Chicama y Paita. (c) precipitación frente al caudal del río.

Por otra parte, interesa la relación que se establece entre la SST de Paita y de Chicama; en este caso, se toma como predictando la temperatura de Paita, y como predictor la de Chicama, obteniéndose la recta de regresión estimada $\hat{y} = 1.053x + 0.717$, con $R^2 = 0.649$ (ver Fig. 3.7(b)). Por último, para el caudal del río Piura y la precipitación en Piura se obtiene $\hat{y}_k = 15.969x_k + 1.663$, con $R^2 = 0.781$ (ver Fig. 3.7(c)).

A partir de estos cálculos y figuras puede observarse que la hipótesis lineal es razonable en (c) y parcialmente razonable en (b). Para el ejemplo (a) sería más razonable considerar una variable transformada. Dado que los datos sugieren una relación exponencial entre las variables, se considera un modelo de regresión de la forma $y = a + b e^x$ (donde la variable x ha sido sustituida por $z = e^x$, para trabajar con el modelo lineal $y = a + bz$). En este caso, debido a las unidades en que están tomados los datos, y a que el método de optimización que se utiliza es el de mínimos cuadrados, es necesario normalizar los datos x antes de calcular su exponencial, de otro modo las diferencias de errores se harían enormes. Así, se obtiene que $y = -1.863 + 4.482 e^{\bar{x}}$ con $R^2 = 0.7551$, considerablemente mejor que el que se había obtenido en el caso anterior (Fig. 3.8). ■

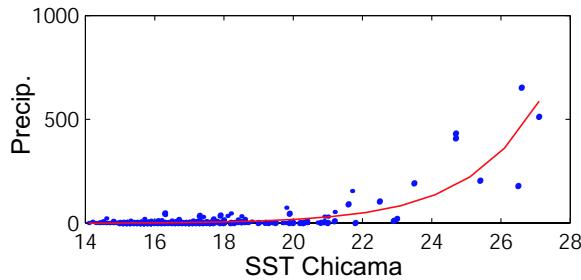


Figura 3.8: Modelo exponencial de temperaturas del agua del mar frente a la precipitación en Piura.

3.4.2. Regresión Lineal Múltiple

En el caso de la regresión lineal múltiple, es conveniente escribir el modelo (3.8) en forma matricial:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (3.18)$$

con

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{p1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{pmatrix}; \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}; \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad (3.19)$$

donde \mathbf{X} es la *matriz de diseño*. De forma análoga al caso lineal unidimensional, el método de mínimos cuadrados proporciona una estimación de los parámetros del modelo:

$$\begin{aligned} \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}^T \mathbf{Y} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \\ &= \mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}. \end{aligned} \quad (3.20)$$

Para obtener el valor mínimo, se deriva (3.20) con respecto a $\boldsymbol{\beta}$ y se iguala la expresión resultante a cero, obteniendo las *ecuaciones normales*

$$(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y}, \quad (3.21)$$

donde $\hat{\boldsymbol{\beta}}$ es el *estimador de mínimos cuadrados* de $\boldsymbol{\beta}$. Suponiendo que X es de rango total, (3.21) conduce al estimador

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (3.22)$$

que demuestra que $\boldsymbol{\beta}$ es una función lineal de \mathbf{Y} .

La matriz de covarianzas de $\hat{\boldsymbol{\beta}}$ puede obtenerse mediante el siguiente resultado bien conocido: La matriz de covarianzas Σ_V de un conjunto de variables \mathbf{V} que son combinaciones lineales de otro conjunto \mathbf{U} (es decir, $\mathbf{V} = \mathbf{C}\mathbf{U}$) es

$$\Sigma_V = \mathbf{C}\Sigma_U\mathbf{C}^T. \quad (3.23)$$

Entonces, como $\hat{\beta}$ es una combinación lineal de \mathbf{Y} , se tiene

$$\text{Var}(\hat{\beta}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{Y}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma_y^2 (\mathbf{X}^T \mathbf{X})^{-1}, \quad (3.24)$$

dado que $\text{Var}(\mathbf{Y}) = \sigma_y^2 \mathbf{I}$. En este caso:

$$r^2 = \frac{\beta^T \mathbf{X}^T \mathbf{Y} - n \bar{Y}^2}{\mathbf{Y}^T \mathbf{Y} - n \bar{Y}^2} \quad (3.25)$$

Ejemplo 3.2 (Precipitación de Santander). En este ejemplo se relacionan los datos de precipitación en Santander con los datos de precipitación en otras ciudades españolas: Bilbao (B), Oviedo (O), Santiago (S), Madrid (M) y Almería (A). En un primer ejemplo, se consideran variables predictoras $\mathbf{x} = (1, b, o, s, m, a)$ los datos de precipitación de todas las ciudades, a excepción de Santander, y como variable predictando y la precipitación en Santander. Aplicando (3.22) se tiene:

$$\beta^T = (5.2052, 0.5136, 0.4080, -0.0045, -0.0099, 0.0158); \quad (3.26)$$

con un $r^2 = 0.5576$. Bilbao y Oviedo son las ciudades que más aportan; el resto prácticamente no influyen en la estimación de los datos de Santander. En un segundo ejemplo se consideran únicamente las variables predictoras $\mathbf{x} = (1, b, o)$; en este caso se tiene:

$$\beta^T = (5.0224, 0.5139, 0.4054); \quad (3.27)$$

con $r^2 = 0.5575$. Por tanto, el coeficiente de ajuste es similar en los dos casos, lo que significa que añadir variables predictoras, una vez se tiene la información de Bilbao y Oviedo es superfluo. En este caso, de acuerdo a las variables disponibles, se ha obtenido el modelo lineal de regresión óptimo. Gráficamente, si se dibujan los datos originales de precipitación en Santander, frente a los datos simulados, se ve que utilizar como regresores los datos de Oviedo y Bilbao subestima los datos reales, aunque la aproximación puede considerarse buena.

En este mismo ejemplo, se podría considerar la predicción simultánea de la precipitación en Santander y Oviedo, utilizando el resto de las estaciones; sin embargo, en este caso, la relación existente entre Santander y Oviedo no sería utilizada en el modelo y se perdería información valiosa. El problema de la regresión multivariada es que los predictandos se consideran independientes en el modelo, y los parámetros obtenidos serían los mismos que si se ajustasen modelos independientes para cada variable. El método de correlación canónica descrito más adelante resuelve parcialmente este problema. ■

Los métodos iterativos de regresión múltiple trabajan por etapas, añadiendo los regresores (de entre todas las posibilidades) que maximizan la varianza explicada en cada etapa (ver, por ejemplo, Chatterjee et al., 2000).

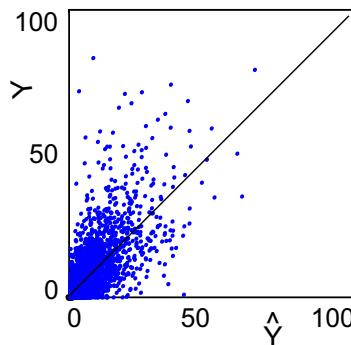


Figura 3.9: Datos de precipitación obtenidos con la recta de regresión frente a los datos originales de Santander (las unidades utilizadas son mm).

3.4.3. Regresión No Lineal

Existen numerosos problemas prácticos donde las hipótesis de los modelos lineales no se cumplen y, por tanto, es necesario considerar formas funcionales más complejas que permitan modelizar este tipo de relaciones. El modelo general de regresión no lineal es de la forma

$$y = f(a_0, \dots, a_q; x_1, \dots, x_p) + \epsilon, \quad (3.28)$$

donde f es una función conocida y a_0, \dots, a_q son los parámetros del modelo. Por ejemplo el conocido modelo de regresión logística es de la forma:

$$y = f(a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n), \text{ con } f(z) = \frac{1}{1 + e^{-z}}. \quad (3.29)$$

Estos parámetros pueden estimarse por el método de los mínimos cuadrados, es decir, minimizando

$$E = \sum_{j=1}^n (y_j - f(a_0, \dots, a_p; x_{1j}, \dots, x_{qj}))^2. \quad (3.30)$$

Desgraciadamente, este proceso conduce a sistemas de ecuaciones no lineales, y es necesario utilizar métodos costosos de optimización (descenso de gradiente, etc.) para su resolución. Además, dada la existencia de mínimos locales en la función de error, estos métodos no garantizan la obtención del óptimo global.

Sin embargo, aparte de modelos no lineales específicos para aplicaciones concretas, como el modelo de regresión logística, en la práctica no suele ser posible prefijar una cierta relación funcional para las variables involucradas en un problema. Por ello, el mayor esfuerzo investigador en este ámbito se ha centrado en desarrollar técnicas de estimación no paramétrica, que tienen una forma funcional flexible en base a un conjunto de parámetros transparentes para el usuario que permiten ajustar el modelo a una clase amplia de funciones (funciones continuas, etc.). Las redes neuronales multicapa son uno de estos modelos de regresión lineal no paramétrica (ver Cap. 5).

3.4.4. MeteoLab: Regresión

En el siguiente ejemplo se muestra el uso de las técnicas de regresión para la predicción de dos variables diferentes (medias diarias de temperatura y presión) en una estación del GSN (en este ejemplo Navacerrada, Madrid), a partir de las componentes principales del patrón atmosférico de la NAO (ver Sec. 2.5.1). Es de esperar que con este patrón atmosférico tan específico (SLP), la predicción de la presión sea mejor que la de la temperatura. En este primer ejemplo simplemente se trata de ilustrar la técnica, aplicada a este problema de downscaling, sin pretender obtener el mejor modelo de predicción. En el siguiente ejemplo se utilizará un patrón atmosférico más apropiado para mejorar estos resultados.

El siguiente código es un script de *MeteoLab*, que carga los datos necesario (como se describe en secciones anteriores) y aplica el comando **regress** para ajustar un modelo de regresión.

```
% Reading CPs from NAO domain
dmn=readDomain('Nao');
[EOF,CP]=getEOF(dmn,'ncp',10);

% Reading temperature from Spanish GSN stations
Example.Network={'GSN'};
Example.Stations={'Spain.stn'};
Example.Variable={'Temp'};
%Example.Variable={'Press'};
[data,Example]=loadStations(Example,'dates',...
    {dmn.startDate,dmn.endDate},'ascfile',1);
Example.Info.Name(4,:)
data=data(:,4); % Taking data for Navacerrada station

j=find(isnan(data)==0); % Selecting days with no missing data
X=[ones(size(j),1), CP(j,:)];
y=data(j,:);

c=regression(y,X);

%Forecasting using the regression
yhat=X*c;
plot([y,yhat])
figure; plot(y,yhat,'.k')
```

Las Figs. 3.10 y 3.11 muestran los resultados obtenidos para la temperatura y la presión, respectivamente ajustando los modelos con los datos del período ERA-40 (1959-1999); en las figuras de la izquierda se muestra los valores reales (en negro) y predichos (en rojo) para los tres primeros años. Estas figuras muestran que mientras la presión se predice con cierta pericia, la predicción de la temperatura es muy deficiente. Modificando ligeramente el código anterior se puede cambiar de estación, número de CPs utilizadas en la regresión, etc.

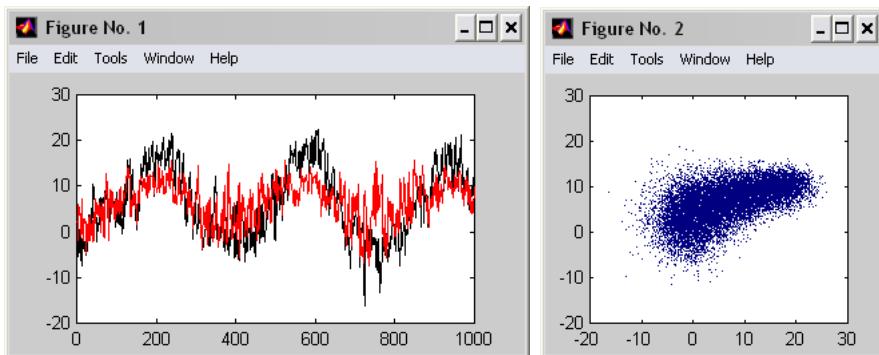


Figura 3.10: Predicción de la temperatura en la estación de Navacerrada a partir de 10 CPs de la NAO, con un modelo regresivo.

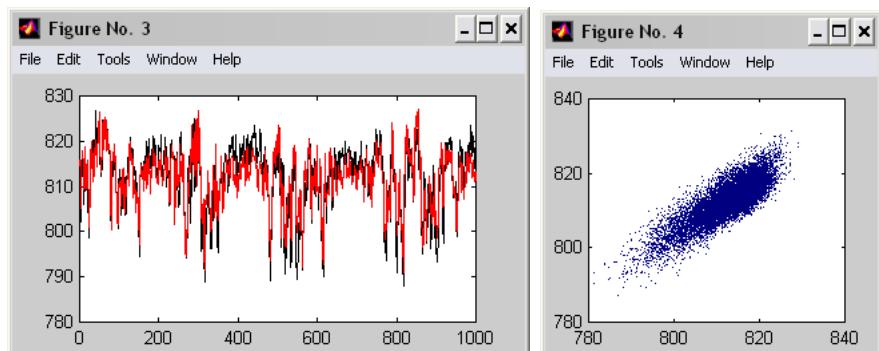


Figura 3.11: Predicción de la presión atmosférica en la estación de Navacerrada a partir de 10 CPs de la NAO, con un modelo regresivo.

Para mejorar la predicción se consideran las CPs asociadas a un patrón más relacionado con la circulación atmosférica sobre la península (obsérvese que la estación estudiada es Navacerrada). Para ello, se considera el Modelo 2 de patrón atmosférico descrito en la Sec. 1.9.4, en lugar del Modelo 1 (la zona NAO). Para ello, basta sustituir en el código anterior la primera línea por:

```
% Reading CPs from Iberia domain
dmn=readDomain('Iberia');
```

En este caso, como muestran las Figs. 3.12 y 3.13, los resultados son claramente mejores, tanto para la temperatura, como para la presión, y pueden considerarse ya operativos.

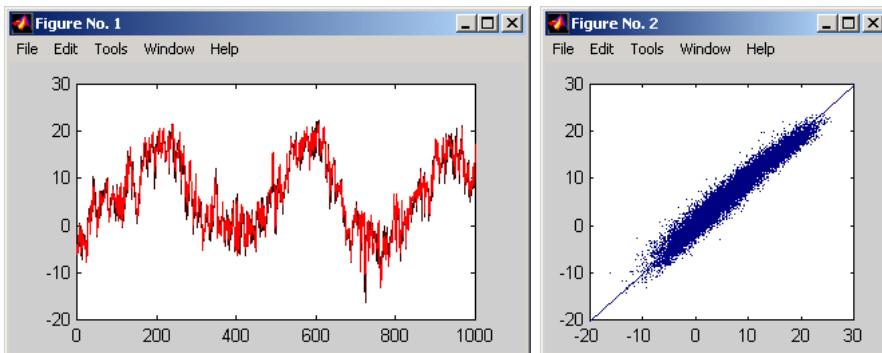


Figura 3.12: Predicción de la temperatura en la estación de Navacerrada a partir de 10 CPs de un patrón atmosférico definido sobre la península ibérica (Modelo 2).

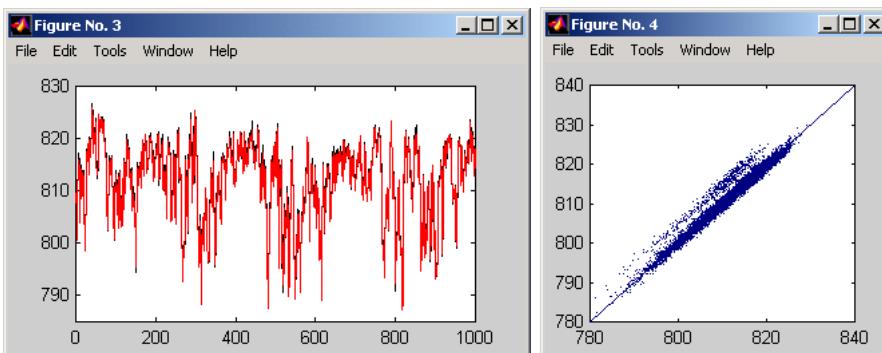


Figura 3.13: Predicción de la presión atmosférica en Navacerrada a partir de 10 CPs de un patrón atmosférico definido sobre la península ibérica (Modelo 2).

3.5. Correlación Canónica

La correlación canónica es una generalización de la regresión lineal multivariada descrita en la sección anterior, que tiene en cuenta las relaciones “espaciales” entre las variables que se desean predecir (Hotelling, 1936). El planteamiento general del problema es el siguiente:

Dada una muestra simultánea de predictores $\mathbf{X} = (X_1, \dots, X_n)$ y predictandos $\mathbf{Y} = (Y_1, \dots, Y_m)$, se desea estimar el valor del vector predictando de forma conjunta a partir de los patrones espaciales que mayor correlación presenten con el vector predictor.

Esta técnica trata de suplir las deficiencias de la regresión multivariada cuando se aplica a un conjunto de predictandos correlacionados (por ejemplo, observaciones en un conjunto de estaciones). En ese caso, la regresión obtiene un

modelo independiente para cada predictando (para cada estación), sin tener en cuenta las relaciones espaciales existentes. El Análisis de Correlación Canónica (CCA) consiste en obtener sendos espacios de proyección para predictandos y predictores de forma que los vectores de una base estén correlacionados de forma óptima con los de la otra. Así, conocido el predictor se podrá inferir un valor para el predictando utilizando las correlaciones obtenidas. De esta forma, se relacionarán los predictores con el vector de predictandos de forma compacta (como un vector) y no de forma independiente para cada componente. Por tanto, la idea es obtener parejas de vectores (en los espacios predictor y predictando, respectivamente), de forma que los datos proyectados en los respectivos vectores muestren la mayor correlación posible. La idea del método es similar al análisis de componentes principales, pero en este caso los vectores canónicos no maximizan la varianza sino la correlación entre dos muestras de datos. Algunas de las propiedades como la ortogonalidad, etc. no se cumplen en este caso.

El procedimiento de cálculo es sencillo. Dadas dos muestras simultáneas de tamaño N :

$$\mathbf{x}_k = (x_{k1}, \dots, x_{kn})^T, \quad \mathbf{y}_k = (y_{k1}, \dots, y_{km})^T, \quad k = 1, \dots, N,$$

se trata de hallar dos subespacios proyectores ($\{f_1, \dots, f_d\}$ y $\{g_1, \dots, g_d\}$, respectivamente, con $d \leq \min\{n, m\}$) para los dos espacios de datos:

$$\bar{\mathbf{x}}_k = F^T \mathbf{x}_k = \begin{pmatrix} f_{11} & \dots & f_{1n} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ f_{d1} & \dots & f_{dn} \end{pmatrix} \begin{pmatrix} x_{k1} \\ \vdots \\ x_{kn} \end{pmatrix}, \quad \bar{\mathbf{y}}_k = G^T \mathbf{y}_k, \quad (3.31)$$

de forma que las proyecciones \bar{x}_{ki} y \bar{y}_{kj} tengan correlación máxima para cada pareja $i = j$ de vectores, y estén incorrelacionadas para $i \neq j$.

El cálculo de estos vectores se realiza a partir de la matriz de varianza-covarianza de los datos:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix},$$

donde C_{xx} y C_{yy} son las matrices de covarianza de cada vector y $C_{xy} = C_{yx}^T$ es la matriz de covarianza. Los vectores de correlación canónica se obtienen resolviendo las siguientes ecuaciones de autovalores:

$$(\mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \mathbf{C}_{yy}^{-1} \mathbf{C}_{yx}) \mathbf{f}_i = \rho_i^2 \mathbf{f}_i \quad (3.32)$$

$$(\mathbf{C}_{yy}^{-1} \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}) \mathbf{g}_i = \rho_i^2 \mathbf{g}_i, \quad (3.33)$$

donde los autovalores ρ_i^2 son los cuadrados de las correlaciones canónicas. El número de soluciones no nulas de estas ecuaciones está limitado por el mínimo de las dimensiones de las variables multidimensionales ($d \leq \min\{n, m\}$). Además, las ecuaciones anteriores están relacionadas por:

$$\mathbf{C}_{xy} \mathbf{g}_i = \rho_i \lambda_x \mathbf{C}_{xx} \mathbf{f}_i \quad (3.34)$$

$$\mathbf{C}_{yx}\mathbf{f}_i = \rho_i \lambda_y \mathbf{C}_{yy} \mathbf{g}_i \quad (3.35)$$

donde

$$\lambda_x = \lambda_y^{-1} = \sqrt{\frac{\mathbf{g}_i^T \mathbf{C}_{yy} \mathbf{g}_i}{\mathbf{f}_i^T \mathbf{C}_{xx} \mathbf{f}_i}}, \quad (3.36)$$

por lo que sólo hace falta resolver una de las ecuaciones de autovalores.

3.5.1. Predicción con Correlación Canónica

Las correlaciones existentes entre las componentes de los vectores proyectados pueden ser utilizadas para realizar predicciones de una de ellas dada la otra. Una vez obtenidos los vectores canónicos en cada espacio y dado un valor \mathbf{x} para el vector de predictores, se puede obtener una estimación de \mathbf{y} de la siguiente forma: Se proyecta \mathbf{x} sobre el espacio de vectores canónicos según (3.31) obteniéndose $\bar{\mathbf{x}}$. Cada componente del vector proyectado está linealmente relacionada con las componentes de \mathbf{y} . Por tanto, se estiman las regresiones entre estos parámetros en base a las muestras disponibles:

$$\bar{y}_i = a_i \bar{x}_i + b_i, \quad i = 1, \dots, d. \quad (3.37)$$

De esta forma, se obtiene el valor estimado de $\bar{\mathbf{y}}$ y se invierte la proyección para recuperar el vector estimado en el espacio original:

$$\hat{\mathbf{y}} = (G^T)^{-1} \bar{\mathbf{y}}. \quad (3.38)$$

Obsérvese que la inversa anterior sólo puede calcularse cuando $d = m$ (es decir, el número de predictandos es menor que el de predictores); en caso contrario habría que utilizar pseudoinversas para resolver el problema de forma similar.

Hay que tener presente que la habilidad de la Correlación Canónica para predecir está limitada por la capacidad de los predictores para explicar de forma lineal los predictandos; es por esto que la CCA se suele aplicar sobre campos de naturaleza fundamentalmente gaussiana como son las temperaturas en predicción a medio plazo, o bien sobre campos de variables promediadas en el tiempo como son valores medios decenales, mensuales y anuales en predicción estacional y climática. También se han descrito en la literatura extensiones no lineales de este método (Hsieh, 2000).

3.5.2. MeteoLab: Correlación Canónica

En esta sección se describe el uso de la toolbox *MeteoLab* en problemas de cálculo de CCA. Como ejemplo, se analiza la relación entre los campos de presión a nivel del mar en la zona de la NAO e Iberia y los datos de racha máxima de viento, temperatura máxima y precipitación sobre una red de estaciones meteorológicas europeas. Siguiendo los pasos del ejemplo de regresión (Sec. 3.4.4), se cargan las CPs de los campos de la NAO (también podrían utilizarse los campos originales de presión a nivel del mar, pero el uso de CPs permite adecuar el número de predictandos al de predictores) y las

observaciones de uno de los predictandos (racha máxima de viento). En este ejemplo se utiliza la red de observatorios del INM, que tiene mayor resolución espacial que el GSN (estos datos no se distribuyen con *MeteoLab*):

```

dates={'1-Jan-1960','31-Dec-1999'};
dmn=readDomain('Nao');
%dmn=readDomain('Iberia');
[EOF,CP,MN,DV]=getEOF(dmn,'dates',dates);
X=CP';

Obsr.Network={'INM'};
Obsr.Stations={'completas.stn'};
Obsr.Variable={'Rellenos/Windx'};
[data,Obsr]=loadStations(Obsr,'dates',dates,'zipfile',1);
Y=data';

```

Para llevar a cabo distintos estudios variando los valores de agregación temporal (diaria, semanal, mensual), se calculan las medias correspondientes tanto de predictandos como de predictores:

```

period=30; % days of the aggregation period
XX=[];YY=[];
for i=1:size(X,2);
    XX(i,:)=movingAverage(X(:,i),period);
end
for i=1:size(Y,2);
    YY(i,:)=movingAverage(Y(:,i),period);
end
XX=XX(:,1:period:end);
YY=YY(:,1:period:end);

%testing with the last year
ntest=fix(365/period);
train=[1:size(XX,2)-ntest];
test=[size(XX,2)-ntest+1:size(XX,2)];

```

La función **computeCCA** calcula la correlación canónica entre dos matrices X e Y, cuyas columnas contienen elementos simultáneos de la muestra de datos (es necesario que ambas matrices tengan el mismo número de columnas).

```
[F, G, r] = computeCCA(XX(:,train),YY(:,train));
```

F y G contienen los vectores canónicos correspondientes a cada una de las muestras y r es el vector que contiene los coeficientes de correlación entre cada par de variables en los espacios proyectados sobre los vectores canónicos. Las

correlaciones vienen ordenadas de forma descendente. Obsérvese que se ha seleccionado un conjunto de entrenamiento para calcular la correlación canónica, reservando un conjunto de test sobre el que se hará la predicción y se validará el resultado.

A continuación se calculan los espacios proyectados, y los coeficientes de regresión entre las componentes de dichos espacios. Como se ha visto en la Sec. 3.5.1, a partir de un valor estimado de los predictandos (en este caso los valores del conjunto de test), se puede calcular una estimación de los predictores:

```
Xp=F'*XX;
Yp=G'*YY;
Ype=[];
for j=1:min(size(YY,1),size(XX,1))
    b=regression(Yp(j,train)',[ones(length(train),1) Xp(j,train)'']);
    Ype=[Ype; ([ones(length(test),1) Xp(j,test)']*b)'];
end
```

por último se vuelve al espacio original de los predictores:

```
Ye=inv(G(:,1:size(G,1))')*Ype;
```

Ahora se pueden comparar los datos originales de la racha máxima mensual con los datos reconstruidos utilizando la correlación canónica para un mes cualquiera (el primero, en este caso):

```
show=1;
drawIrregularGrid([Ye(:,show)'; YY(:,test(show))'],...
    Observation.Info.Location)
```

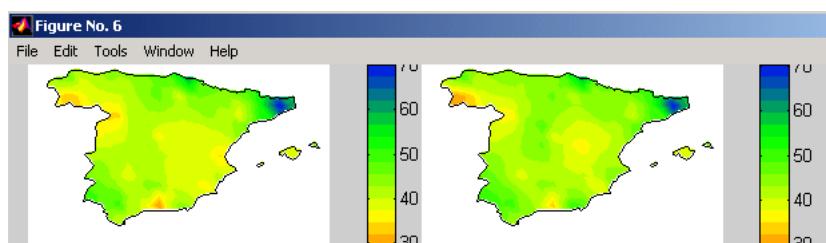


Figura 3.14: Media mensual de racha máxima diaria de viento observada (derecha) y predicha con CCA (izquierda), utilizando como predictor el patrón de la NAO.

La función `drawIrregularGrid` es análoga a `drawGrid` (vista en la Sec. 1.9.5), pero no requiere una rejilla regular como dato de entrada, sino que la crea automáticamente a partir de una rejilla irregular.

En los ejemplos anteriores, el parecido entre predicciones y observaciones se puede analizar visualmente. Sin embargo, se puede utilizar la correlación espacial como medida global de la pericia del modelo de predicción. El siguiente código muestra la forma de calcular la correlación espacial para los doce meses del período de prueba (el año 1999). La figura resultante muestra valores de correlación superiores a 0.85 en todos los casos.

```
cor=[]; for i=1:size(Ye,2),
    a=corrcoef(Ye(:,i),YY(:,test(i)));
    cor(i)=a(2,1);
end
figure;
plot(cor)
```

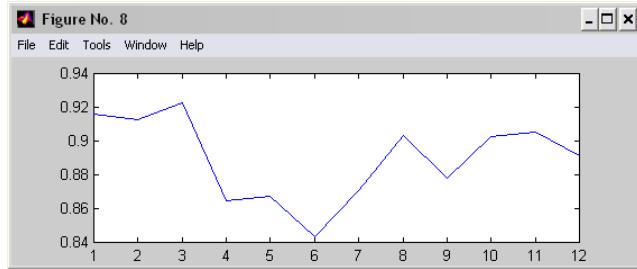


Figura 3.15: Correlación espacial entre patrones previstos y observados de media mensual de racha máxima diaria de viento.

El análisis anterior también puede realizarse a escala diaria, sin más que cambiar la variable `period=1` en el código. Sin embargo, en este caso, los patrones diarios previstos no son tan parecidos a los reales como en el caso mensual. Por ejemplo, la Fig. 3.16 muestra un patrón diario previsto, junto al observado.

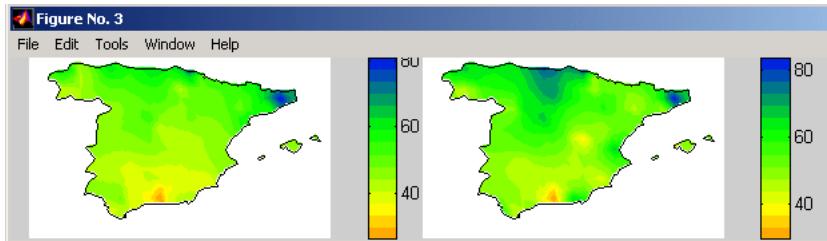


Figura 3.16: Racha máxima diaria de viento observada (derecha) y predicha (izquierda), utilizando como predictor el patrón de la NAO.

El código completo de estos ejemplos se adjunta en el fichero `Examples.m` del directorio `MeteoLab/Statistics/CCA`.

3.5.3. Efectos de la Escala Temporal

En la Sec. 2.5.3 se analizó el efecto de la escala temporal (diaria, semanal, mensual) en el método de Componentes Principales. Se mostró que los resultados mejoran al incrementar el período sobre el que se promedian los campos atmosféricos, ya que éstos se normalizan. En el caso del análisis de CCA, el efecto de la escala temporal es aún más problemático, pues la aproximación normal no sirve a escala diaria y semanal para algunos predictandos, como la precipitación (ver Fig. 2.22).

Para estudiar este problema y hallar los límites operativos del método de CCA para distintas variables y escalas temporales, se han considerado datos de temperatura máxima y precipitación en la red de estaciones del INM, y se han realizado distintos experimentos de predicción aplicando CCA con distintos patrones atmosféricos (la región NAO y la región Ibérica mostradas en la Sec. 1.9.4) y distintas escalas temporales (de diaria a mensual). La Fig. 3.17 muestra las correlaciones espaciales obtenidas entre los patrones observados y los previstos utilizando la correlación canónica para la temperatura máxima y la precipitación. Se muestran los resultados obtenidos con los dos patrones atmosféricos considerados: la zona de la NAO y la zona Ibérica.

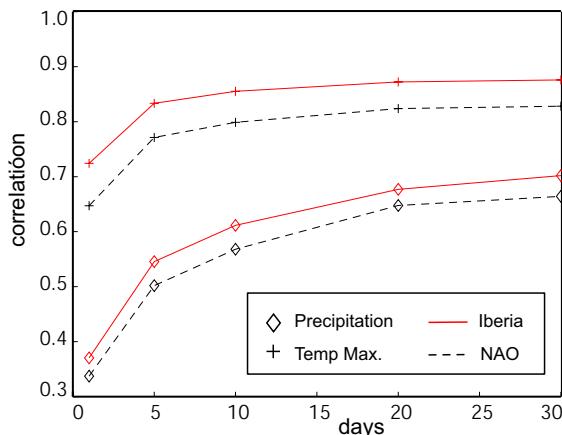


Figura 3.17: Correlación espacial entre los patrones observados y previstos para distintos valores de agregación temporal (de diaria a mensual) para la temperatura y precipitación, utilizando dos patrones atmosféricos distintos.

Una primera conclusión es que el patrón Ibérico mejora los resultados del patrón NAO en todos los casos. Por otra parte, los resultados obtenidos para la temperatura muestran pericia desde la escala diaria, manteniéndose constante aproximadamente desde la escala semanal. Sin embargo, el caso de la precipitación es distinto, pues los resultados sólo empiezan a ser buenos a partir de la quincena, no observándose saturación en la pericia al incrementar la escala temporal hasta un mes. Esto es debido a que la temperatura es un campo aproximadamente normal y con suaves fluctuaciones espaciales y temporales,

mientras que la precipitación tiene grandes fluctuaciones en el espacio y en el tiempo. Estos resultados indican que el método de CCA es inapropiado para predecir este último tipo de variables a escala diaria.

La Fig. 3.18 muestra un ejemplo de predicción diaria y mensual de temperatura máxima, mientras que la Fig. 3.19 muestra un ejemplo ilustrativo de la precipitación.

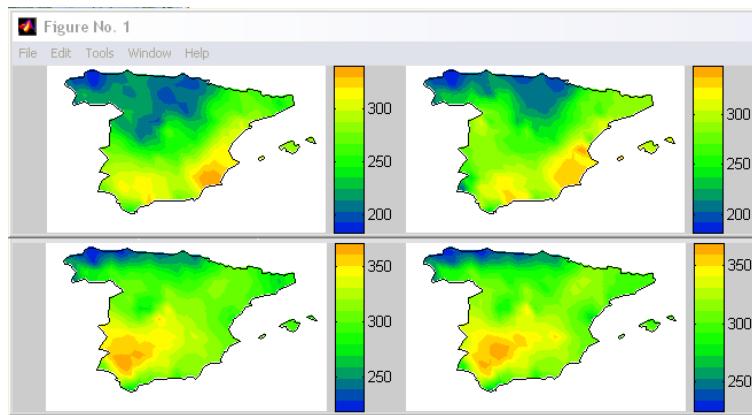


Figura 3.18: Ejemplo de patrones observados (derecha) y previstos con CCA (izquierda) para la temperatura máxima. La figura superior muestra un patrón diario y la inferior uno mensual.

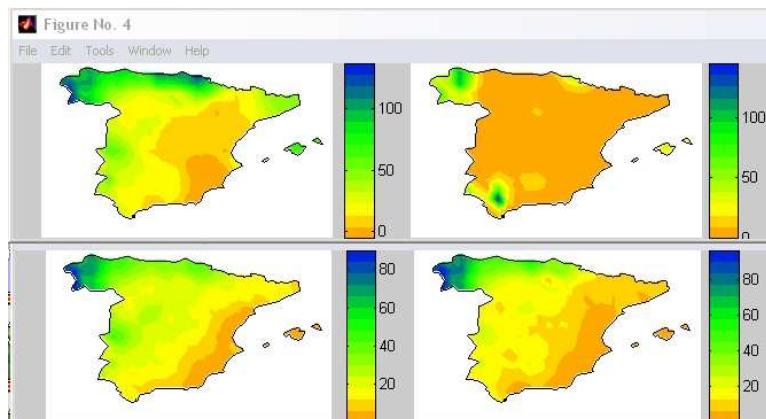


Figura 3.19: Ejemplo de patrones observados (derecha) y previstos con CCA (izquierda) para la precipitación. La figura superior muestra un patrón diario y la inferior uno mensual.

3.6. Aplicación al Downscaling. MOS y Perfect Prog

Los métodos lineales de series temporales, regresión y correlación canónica, descritos en las secciones anteriores, han sido aplicados de distintas formas al problema de la predicción local descrito en la Sec. 3.2. Las distintas aplicaciones se diferencian en la forma de considerar la relación (local o globalmente) que existe entre el estado de la atmósfera y los fenómenos meteorológicos observados en superficie que se quieren predecir. El estado de la atmósfera se puede aproximar utilizando un modelo numérico de circulación, que proporciona los campos atmosféricos estimados para un tiempo dado. Por tanto, una aproximación global a este problema sería analizar estadísticamente la relación entre campos y observaciones utilizando todos los registros históricos disponibles en un punto de interés, y los correspondientes campos atmosféricos simulados (por un modelo numérico de reanálisis). Este esquema se denomina *Perfect-Prog* puesto que supone que el modelo atmosférico proporciona una representación fidedigna del estado de la atmósfera que no varía en el tiempo, al menos en la periodo del reanálisis (ver Wilks, 1995, para una descripción más detallada). De esta forma, se puede entrenar un modelo global de regresión para estimar una relación lineal entre los datos (ver, por ejemplo, Enke and Spekat, 1997; Billet et al., 1997). También han sido utilizados recientemente modelos no-lineales más generales, que son ajustados a los datos usando técnicas no paramétricas modernas como las redes neuronales que se describen en el Cap. 5 (Gardner and Dorling, 1998; McGinnis, 1994). Otra técnica global que ha sido aplicada a este problema es el Análisis de Correlación Canónica (CCA), que tiene en cuenta las dependencias espaciales existentes entre los distintos observatorios, tanto lineales (Bergman and Delleur, 1985), como no lineales (Hsieh, 2001). El denominador común de todos estos métodos es su carácter global; es decir, se considera un único modelo para analizar todas las situaciones que puedan presentarse.

Una aproximación distinta a este problema consiste en suponer que los patrones atmosféricos disponibles no son una representación perfecta de la atmósfera, sino que dependen del modelo numérico utilizado (por ejemplo, cada cambio realizado en el modelo atmosférico operativo con nuevos esquemas de asimilación o parametrizaciones tienen influencia directa en los campos previstos). La técnica denominada *Model Output Statistics (MOS)* (Klein and Glahn, 1974) es una técnica local que ajusta de forma dinámica un modelo de regresión entre los campos previstos por el modelo y las observaciones ocurridas, de forma que este modelo está continuamente actualizándose con las nuevas observaciones y predicciones disponibles. Este proceso se realiza de distintas formas, incluyendo filtros de Kalman (ver Bergman and Delleur, 1985, para más detalles). También han sido propuestas distintas extensiones no lineales de estos modelos utilizando, por ejemplo, redes neuronales (Yuval and Hsieh, 2003; Marzban, 2003). En este caso, el problema es la necesidad de disponer de observaciones y salidas del modelo recientes, que sean representativas de la configuración actual del mismo. Esta imposición limita el número de estaciones sobre el que es posible aplicar las técnicas, ya que no se dispone de información

en tiempo real para la mayoría de las estaciones.

Otra aproximación a este problema, a medio camino entre las dos metodologías anteriores, es la técnica local de análogos, que consiste en identificar conjuntos de patrones atmosféricos homogéneos (de un mismo modelo, o asociados a una misma situación atmosférica) y entrenar un modelo local para cada uno de ellos. La necesidad de considerar modelos locales (por ejemplo, en Perfect Prog) se justifica en la siguiente sección.

3.6.1. Modelos Globales y Locales. Dependencia Espacial

En el Ejemplo 2.9 se analizó la dependencia existente entre fenómenos ocurridos (precipitación, etc.) en distintas estaciones distribuidas espacialmente en la geografía española. Se comprobó que estas relaciones no son estáticas, sino que varían dependiendo de la información disponible sobre la ocurrencia del fenómeno en alguna estación o región. En esta sección se muestra que estas relaciones también están condicionadas al estado de la atmósfera; es decir, las dependencias entre estaciones varían dependiendo de los distintos modos atmosféricos relevantes en la región dada. Por tanto, el sistema se describirá mejor con modelos locales, condicionados al estado de la atmósfera, que es la esencia del método de análogos.

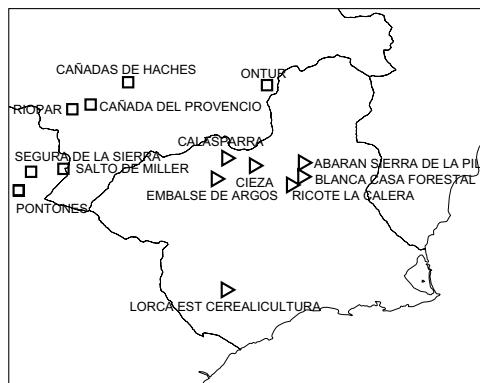


Figura 3.20: (a) Cuenca del Segura con la ubicación de los observatorios considerados a ambos lados de la cordillera. Las estaciones de Jaén y Albacete se agrupan en la zona NW (marcadas con un cuadrado) y las de la provincia de Murcia en la zona SE (marcadas con un triángulo).

Como ejemplo considérese la cuenca del Segura, región muy seca donde la precipitación es escasa pero intensa. En un trabajo previo Ribalaygua and Borén (1995) consideran un único patrón de precipitaciones en el Sudeste Peninsular, y una sola configuración atmosférica para explicarlo, lo que conduce a un modelo global. Sin embargo, para analizar este problema en detalle se ha rea-lizado un experimento similar al del Ejemplo 2.9, pero esta vez añadiendo el estado de la atmósfera al sistema. El estado de la atmósfera se reduce a un

escalar discreto aplicando la técnica de agrupamiento k -medias descrita en la Sec. 2.6.2, de forma que se dispone de 25 modos atmosféricos distintos. De entre estos modos, existen al menos 4 ó 5 claramente diferentes, que pueden producir precipitaciones fuertes (generalizadas o locales) en la cuenca del Segura. Se han seleccionado dos modos distintos para que no tenga ningún sentido una estructura promedio (ver Fig. 3.21). Para el cálculo de las matrices de correlación, se han seleccionado dos zonas mostradas en la Fig. 3.20, una a cada lado del Sistema Bético en la cuenca del río Segura.

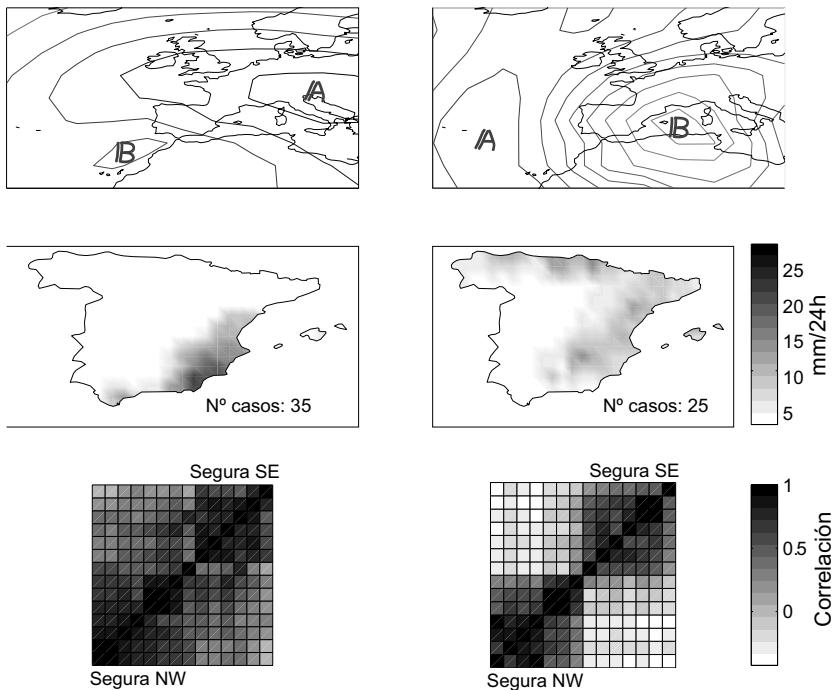


Figura 3.21: Correlación entre las estaciones de las dos zonas de Levante mostradas en la Fig. 3.20 para dos configuraciones atmosféricas del geopotencial en 1000 mb.

Los resultados se muestran en la Fig. 3.21. Las dos figuras superiores corresponden al geopotencial de 1000 mb para dos tipos diferentes de situaciones de precipitaciones intensas en la cuenca del Segura; la izquierda corresponde a flujos del Este en niveles bajos, asociados con una depresión fría entre el golfo de Cádiz y Canarias que provoca precipitaciones muy intensas en el litoral, afectando a la casi totalidad de las provincias de Murcia y Almería; la superior derecha corresponde a flujos del Noroeste impulsados por una profunda depresión cálida mediterránea fuertemente baroclíntica (con SW en 500 mb) y que afecta principalmente a la zona de barlovento del Sistema Bético y a otras regiones como se puede ver en la figura centro-derecha. Lo verdaderamente interesante de esta figura, aparte de la lógica diferencia en la distribución de la precipitación sobre la cuenca del Segura, es la clara diferencia entre las matrices

de correlación de las dos zonas definidas en la Fig. 3.20 y que se muestra en la parte inferior de la figura; en el caso de la borrasca fría del golfo de Cádiz, la correlación es muy uniforme en toda la cuenca, en cambio en el caso de la baja dinámica mediterránea se observa una clara independencia entre las zonas a ambos lados de la cordillera.

Estos resultados se pueden generalizar en mayor o menor medida a casi todas las cuencas de la geografía española, por lo que se puede afirmar que los modelos estadísticos globales no son apropiados cuando el patrón atmosférico utilizado en la predicción abarca una zona amplia de estudio, para un conjunto de estaciones dado.

3.7. Técnicas Locales de Análogos

El método de análogos introducido por Lorenz (1969) en el marco de la predicción de series temporales es una versión particular de una metodología más general llamada técnica de vecinos próximos (Nearest Neighbors, NN) (Hastie et al., 2001). Esta técnica utiliza el “entorno” del patrón atmosférico previsto para entrenar un modelo local y obtener una predicción. En la práctica, el espacio de patrones atmosféricos se aproxima por la muestra aleatoria de patrones proporcionada por un reanálisis suficientemente amplio (ver Sec. 1.9.4). Por ejemplo, utilizando ERA40 se dispone de casi 15000 patrones atmosféricos diarios para aproximar el espacio de configuraciones atmosféricas. Por tanto, los modelos locales se entrena en base a los patrones atmosféricos de un *conjunto de análogos*, que está formado por los k días más próximos al patrón previsto en la base de datos del reanálisis. (ver Cofiño, 2004, para una revisión de estos métodos).

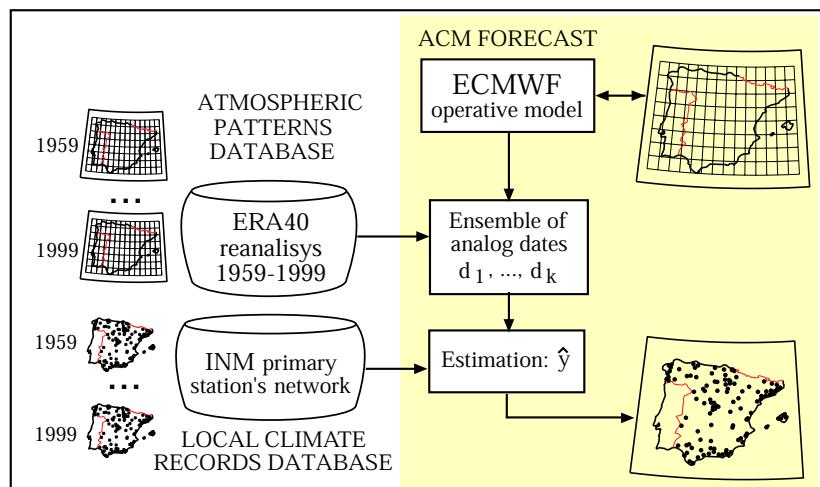


Figura 3.22: Esquema del algoritmo estándar de downscaling por análogos (k -NN).

La Fig. 3.22 muestra un esquema con las componentes fundamentales de este tipo de métodos. En esta figura puede observarse que es necesaria una base de datos con los patrones atmosféricos y otra con los registros históricos de la climatología local. El algoritmo parte de una predicción numérica del patrón atmosférico previsto, y obtiene una predicción local para los observatorios disponibles. Las zonas sombreadas representan las componentes del algoritmo implicadas en la fase operativa de obtención de predicciones locales. En la literatura han sido descritas distintas implementaciones de este tipo de métodos, tanto para la predicción de anomalías climáticas (Zorita and von Storch, 1999; Wilby and Wigley, 1997), como para la predicción a corto plazo (van den Dool, 1989). En general, se ha demostrado que el método de análogos funciona tan bien como otras técnicas más complicadas de downscaling (ver, por ejemplo, Zorita and von Storch, 1999), indicando que estos métodos de “hombre pobre” son alternativas eficientes para diversos problemas de downscaling.

3.7.1. Influencia del Período de Reanálisis Utilizado

La pericia de la técnica de análogos descrita en la sección anterior depende, en gran medida, de la longitud de las bases de datos utilizadas en el proceso.

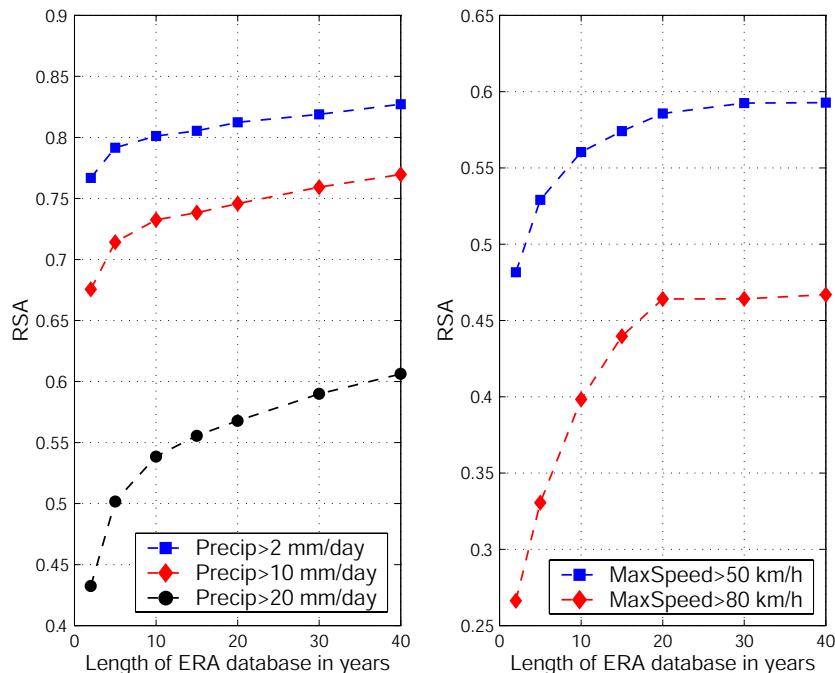


Figura 3.23: RSA (ROC Skill Area) de la predicción con el método de análogos para distintos umbrales de la precipitación (izquierda) y racha máxima de viento (derecha) en la red de estaciones completas del INM.

En la Fig. 3.23 se muestra la variación de la pericia en función del número de años considerados para el análisis (hasta un máximo de 40 años, correspondientes al reanálisis ERA40 completo). Se ha aplicado el método de análogos con $k = 50$ vecinos considerando 50 CPs del estado de la atmósfera definido sobre la rejilla que cubre la península Ibérica utilizada en ejemplos anteriores. Se muestran resultados para distintos umbrales de la precipitación (izquierda) y racha máxima de viento (derecha) en la red de estaciones completas del INM.

A partir de los resultados mostrados en esta figura, se observa un crecimiento exponencial de la pericia que tiende a saturar en períodos distintos para diferentes eventos. Además, se observa que los eventos extremos requieren de un período más largo para la saturación, en consonancia con la hipótesis de similitud del método de análogos.

3.7.2. MeteoLab: *Técnicas Locales de Análogos*

En esta sección se ilustra la aplicación de *MeteoLab* a la predicción local mediante la técnica de análogos. Para ello, se considera el mismo ejemplo utilizado en ocasiones anteriores, en el que se trata de dar predicciones locales en estaciones del GSN (en este caso, la precipitación), a partir de los patrones atmosféricos representativos de la zona (en este caso, se considera la zona Iberia definida en *Patterns/Zones.txt* definida por una rejilla sobre la península ibérica). El espacio de búsqueda de análogos será el espacio de las cincuenta primeras componentes principales de los patrones atmosféricos diarios del período ERA40. En este ejemplo se obtendrá tanto una predicción numérica (utilizando la media de las observaciones del conjunto de análogos), como una probabilística (utilizando la frecuencia del evento en el conjunto de análogos).

Se ha definido un conjunto de entrenamiento, sobre el que se buscarán los análogos (datos correspondientes al período 1960-1998), y un conjunto de test (datos correspondientes al año 1999) en el que se realizará la predicción utilizando los patrones diarios del reanálisis como patrones previstos.

```

dmn=readDomain('Iberia');
Stations.Network={'GSN'};
Stations.Stations={'Spain.stn'};
Stations.Variable={'Precip'};

%Training data
dates={'1-Jan-1960','31-Dec-1998'};
[EOF,CP]=getEOF(dmn,'ncp',50,'dates',dates);
[dataE,Stations]=loadStations(Stations,'dates',dates,'ascfile',1);

%Test data
dates={'1-Jan-1999','31-Dec-1999'};
[EOF,CPT]=getEOF(dmn,'ncp',50,'dates',dates);
[dataT,Stations]=loadStations(Stations,'dates',dates,'ascfile',1);

```

Una vez cargados los datos, para cada uno de los patrones previstos, se buscan los $k = 25$ vecinos más cercanos. La función `getAnalogous(pat,rean,k)` permite obtener los k vecinos más cercanos para un patrón `pat` dado, de entre los campos del reanálisis `rean`. Como parámetros de salida, la función devuelve los k patrones análogos, `AnalogPat`, el índice de cada uno de los días análogos, `Neig` y la distancia de cada uno de los patrones análogos al patrón problema `NeigDist`. Una vez que se han seleccionado los análogos, se puede obtener una predicción numérica utilizando, por ejemplo, la media de las observaciones de precipitación de los días análogos al día problema.

```
O=[];P=[]; for j=1:1:180
    [AnalogPat,Neig,NeigDist]=getAnalogous(CPT(j,:),CP,25,'knn',[]);
    O=[O;dataT(j,:)];
    P=[P;nanmean(dataE(Neig,:))];
end
plot([O(:,1),P(:,1)])
```

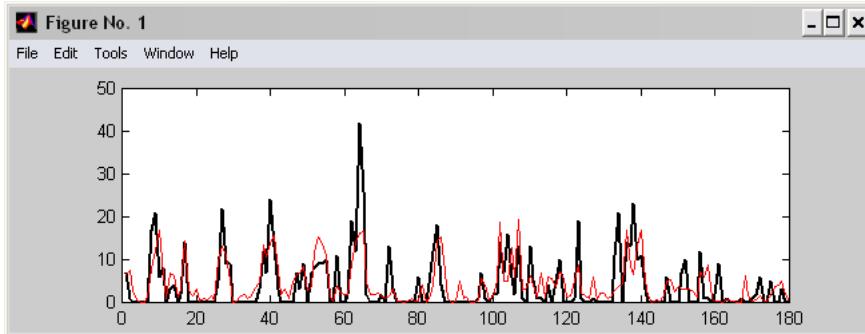


Figura 3.24: Predicción numérica de la precipitación para Iguelo (San Sebastián) para los seis primeros meses del año 1999.

De forma similar, se puede obtener una predicción probabilística para un evento concreto; por ejemplo que la precipitación supere 5 mm (`Precip > 5mm`). En este caso, la validación ha de hacerse utilizando los índices de calidad descritos en el Cap. 7.

```
umbral=5;
i=find(~isnan(dataE));
dataE(i)=dataE(i)>umbral; dataT=dataT>umbral;

O=[];P=[];
for j=1:1:180
    [AnalogPat,Neig,NeigDist]=getAnalogous(CPT(j,:),CP,25,'knn',[]);
    O=[O;dataT(j,:)];
    P=[P;nanmean(dataE(Neig,:))];
end
plot([O(:,1),P(:,1)])
```

```

O=[0;dataT(j,:)];
P=[P;nanmean(dataE( Neig,:))];
end
%validation of the probabilistic forecast
makeValidation(O(:,1),P(:,1))

```

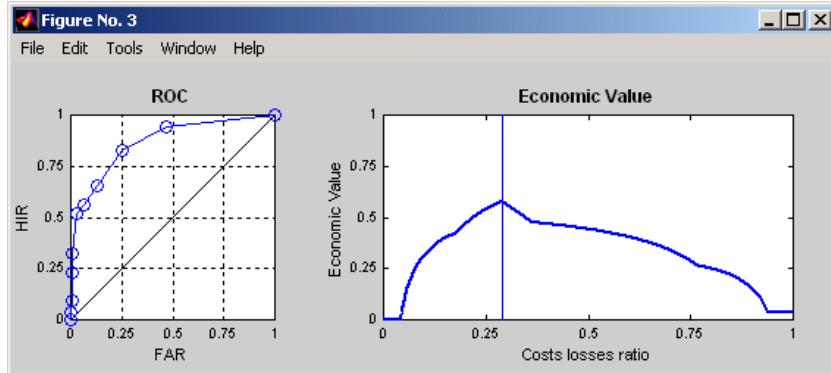


Figura 3.25: Validación (curva ROC y de valor económico) para predicción probabilística del evento *Precip > 5mm* en Igueldo (San Sebastián) para los seis primeros meses del año 1999.

3.8. Comparación de Técnicas en el Corto Plazo

En esta sección se presentan los resultados de un experimento llevado a cabo para comparar entre sí los métodos descritos en las secciones anteriores aplicados a la predicción probabilística a corto plazo. Dado que uno de los modelos a comparar es la regresión lineal se ha decidido seleccionar la variable más “normal” de las que se dispone: la racha máxima diaria de viento (*Wind*). Los métodos son aplicados para obtener un pronóstico probabilístico del evento “*Wind > 50km/h*”.

Se han considerado las salidas del modelo operativo del ECMWF con un alcance de 1 día para los meses Diciembre, Enero y Febrero del año 2000. Obsérvese que no existe solapamiento con los 40 años de re-análisis de ERA40 (1959-1999) que se utilizan conjuntamente con observaciones históricas de ese período para entrenar los modelos. Se han probado diferentes estaciones de la red principal de estaciones del INM (ver Fig. 1.17(c)) obteniéndose resultados similares; en esta sección se ilustran los resultados obtenidos para Santander. Una metodología estándar de validación para predicciones probabilísticas es el uso de curvas ROC (ver Cap. 7). El cálculo de las curvas ROC se basa en la tabla de contingencia que viene dada por las ocurrencias y no ocurrencias reales de un evento en función de las predichas. Con objeto de comprobar la importancia de la resolución espacial y temporal del patrón atmosférico, se

realizaron pruebas considerando una rejilla global y una local (mostradas en las Fig. 1.22(a) y (c), respectivamente). En ambos casos se consideran las diez primeras componentes principales para representar cada uno de los patrones. En las Fig. 3.26(a) y (b) se muestran las curvas ROC y los Brier Scores (BS) para cuatro algoritmos distintos considerando los patrones local y global, respectivamente. Los métodos utilizados son:

- Un modelo de regresión lineal, ajustado para predecir la probabilidad del evento $Wind > 50km/h$ en función de las componentes principales de los patrones atmosféricos.
- Un modelo de red neuronal de la forma 10:m:1, donde las entradas son las componentes principales y la salida es la probabilidad del evento. En este caso se probaron distintos valores de m y los mejores resultados se obtuvieron con $m = 5$ (esto método se describe en detalle en el Cap. 5).
- La técnica k-NN de análogos, utilizando como predictor de la probabilidad del evento la distribución empírica obtenida del conjunto de análogos. Se han considerado cuatro valores diferentes para k (100, 50, 25 y 15), obteniéndose los mejores resultados con $k = 50$.

Estos resultados indican que un modelo lineal es fácilmente mejorable con un modelo neuronal. Por otra parte, una técnica local simple de vecinos cercanos (k-NN) permite obtener resultados similares a una técnica complicada como una red neuronal. En lo referente a la influencia de la escala del patrón en los resultados, se observa que los métodos globales mejoran su pericia más que el método local cuando se aplican sobre patrones definidos en rejillas locales, especialmente la red neuronal. Este resultado es lógico puesto que la reducción de escala en el patrón atmosférico tiene un efecto similar a una localización en los patrones que se utilizan para la predicción. Una comparación de estas técnicas en una escala temporal distinta se puede encontrar en Schoof and Pryor (2001).

Como conclusión se obtiene que los métodos locales basadas en análogos son las técnicas de downscaling estadístico más apropiadas para trabajar en esta escala temporal (ver Gutiérrez et al., 2004).

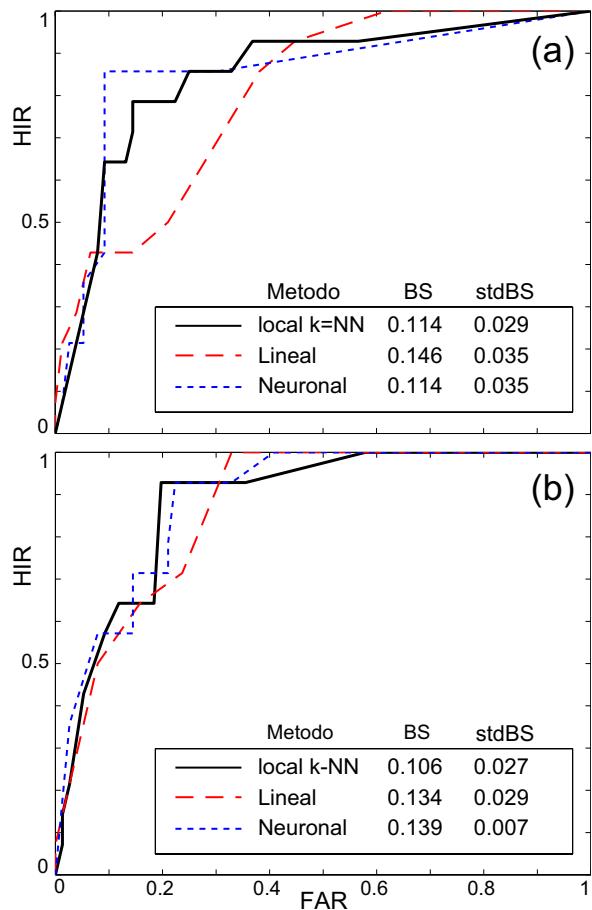


Figura 3.26: Curvas ROC y Brier Scores obtenidos para predecir $p(\text{Wind} > 50\text{km}/\text{h})$ en Santander aplicando distintos métodos a los patrones dados por una rejilla (a) local (Fig. 1.22(c)) y (b) global (Fig. 1.22(a)).

Parte III

Técnicas de Aprendizaje Automático

CAPÍTULO 4

Redes Probabilísticas (Bayesianas)

4.1. Introducción

En los capítulos anteriores se han analizado las técnicas estándar de modelización estadística. Sin embargo, la creciente disponibilidad de información meteorológica y climática (registros históricos, simulaciones de modelos numéricos, imágenes de satélites, etc.) hace necesario el uso de nuevas técnicas más eficientes y automáticas. En las últimas décadas, se han desarrollado numerosas técnicas en distintas áreas de la *inteligencia artificial* para analizar y modelizar la gran cantidad de información disponible en numerosos problemas prácticos. El objetivo de estas técnicas de aprendizaje automático es preprocesar de forma rápida y fiable la información, capturando distintos patrones de conocimiento (reglas, grafos de dependencias, funciones, etc.) que sean apropiados para resolver un problema dado, y que resuman la información disponible haciéndola manejable. Se trata de que estos sistemas operen de forma objetiva y automática, precisando de la mínima intervención humana en los procesos de modelización e inferencia; la liberación de este tipo de tareas permite incrementar esfuerzos en trabajos más creativos como el diseño, supervisión e interpretación del sistema.

Durante los últimos años se ha producido un notable desarrollo de este tipo de herramientas en distintas áreas de conocimiento: aprendizaje automático (machine learning), computación neuronal, estadística, computación paralela, bases de datos, etc. En la última década se ha acuñado el término *Minería de Datos* (Data Mining) para referirse a este área interdisciplinar que extrae conocimiento de forma automática a partir de un conjunto de datos (ver Cofiño et al., 2003, para más detalles). Los sistemas resultantes se denominan *sistemas inteligentes* (o, más genéricamente, *sistemas expertos*) pues permiten

modelizar un problema, obtener conclusiones, e incluso tomar decisiones de forma casi automática.

En este capítulo se describe una de estas técnicas, las redes probabilísticas, que permiten caracterizar probabilísticamente un conjunto de datos de forma automática. En la Sec. 4.2 se comienza describiendo las características y componentes de los sistemas inteligentes. La Sec. 4.3 analiza la modelización probabilística, y describe los problemas que conlleva trabajar con un número elevado de variables. La Sec. 4.4 muestra la aplicación de los grafos para resolver estos problemas, representando únicamente las dependencias importantes para cada problema (los modelos resultantes se denominan redes probabilísticas). Las Sec. 4.5 introduce las redes Bayesianas, modelos probabilísticos basados en grafos dirigido. La Sec. 4.6 describe el problema de la inferencia en este contexto. La Sec. 4.7 analiza el problema más importante desde un punto de vista práctico: el aprendizaje automático de estos modelos a partir de datos. Finalmente, la Sec. 4.8 describe numerosas aplicaciones de estas técnicas.

4.2. Sistemas Inteligentes

La Fig. 4.1 muestra el esquema general de un sistema inteligente con sus dos componentes principales: la *base de conocimiento* y el *módulo de razonamiento*. La base de conocimiento almacena de forma estructurada (utilizando algún patrón de conocimiento adecuado, como reglas, grafos, etc.) la información relevante para un problema dado. En los antiguos sistemas expertos, este conocimiento era obtenido de un experto humano. En los modernos sistemas inteligentes, este conocimiento se extrae directamente de los datos utilizando algoritmos de aprendizaje automático (algoritmos de extracción de reglas Agrawal et al. (1993), etc.).

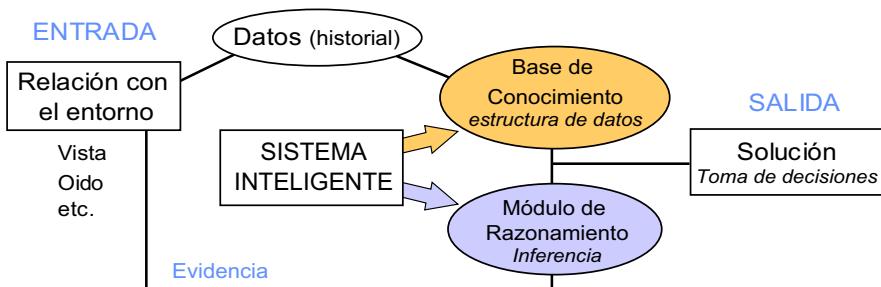


Figura 4.1: Esquema de un sistema inteligente, con sus componentes principales.

Por otra parte, una vez adquirido el conocimiento, el módulo de razonamiento ha de ser capaz de extraer las conclusiones apropiadas a partir de una evidencia concreta para un problema particular. Por ejemplo, si la base de conocimiento contiene la información relevante sobre la precipitación en un conjunto de estaciones, entonces el módulo de razonamiento tendrá que obtener conclusiones sobre el valor o estado de la precipitación en un estación o conjunto de

estaciones cuando se tiene una evidencia concreta (por ejemplo, se sabe que ha llovido en una cierta estación). El módulo de razonamiento está implementado en función del tipo de patrón de conocimiento empleado en el sistema. Por tanto, la forma de obtener conclusiones será distinta en cada caso. Por ejemplo, un sistema basado en reglas tiene un módulo de razonamiento que consiste en “disparar” aquellas reglas cuya premisa se conozca (formen parte de la evidencia), obteniendo las conclusiones indicadas en la regla, que a su vez pasarán a formar parte de la evidencia para tratar de seguir disparando reglas. En otros sistemas inteligentes más modernos, que utilizan patrones de conocimiento más abstractos, el proceso de razonamiento es más complicado y menos intuitivo. Dos ejemplos notables de estas nuevas metodologías son los algoritmos inspirados en la Biología, como las redes neuronales, y los algoritmos estadísticos de aprendizaje estadístico, como las redes probabilísticas:

- Las *redes probabilísticas* utilizan la expresividad de los grafos para automatizar el proceso de modelización probabilística (incluso para un número elevado de variables). Los modelos resultantes combinan resultados de la teoría de grafos (para representar las relaciones de dependencia e independencia del conjunto de variables) y de la probabilidad (para cuantificar estas relaciones). Esta unión permite realizar de forma eficiente tanto el aprendizaje automático del modelo, como la inferencia a partir de la evidencia disponible. La base de conocimiento de estos sistemas es una función de probabilidad conjunta de todas las variables del modelo, mientras que el módulo de razonamiento implementa técnicas eficientes de cálculo de probabilidades condicionadas. Por ejemplo, las redes probabilísticas permiten definir un modelo conjunto con toda la información relevante sobre la precipitación en las estaciones de la red secundaria (miles de variables), y también propagar cualquier evidencia en el modelo (por ejemplo, “ha llovido en Santander”), obteniendo las probabilidades condicionadas de precipitación en el resto de estaciones.
- Las *redes neuronales* se inspiran en la Biología para definir un patrón de conocimiento que “simule” la estructura de un cerebro, considerando un gran número de unidades de proceso simples (o “neuronas”) conectadas entre sí. En este caso, la base de conocimiento es abstracta, pues viene dada por los pesos de las conexiones que se establecen entre las neuronas. La forma de aprender estos pesos a partir de un conjunto de datos sigue diversas analogías con la Biología, como las técnicas de aprendizaje no supervisado basadas en el córtex visual. Por otra parte, en este caso el módulo de razonamiento es simple y se basa en la propagación por las distintas capas de neuronas de un conjunto de datos de entrada, hasta una capa final que proporciona la salida.

Estas dos metodologías se describen en detalle en los dos siguientes capítulos. Por una lado, este capítulo describe las redes probabilísticas, que utilizan estructuras de datos y algoritmos de inferencia heredados del razonamiento humano (los grafos y la probabilidad); por otra parte, el siguiente capítulo describe las redes neuronales, que están inspiradas en la estructura del cerebro y

corresponden a un nuevo paradigma de modelos inspirados en la Biología. El estudio de estas dos técnicas proporciona una visión global del problema del aprendizaje automático, y permite entender mejor otras técnicas alternativas que se describen, por ejemplo, en Hastie et al. (2001) y en Duda et al. (2001).

4.3. Sistemas Inteligentes Probabilísticos

Las redes probabilísticas son sistemas inteligentes basados en probabilidad, fruto de un largo esfuerzo investigador para desarrollar modelos que permitan tratar problemas con incertidumbre que involucren un número elevado de variables. Estos modelos codifican de forma apropiada la probabilidad conjunta y las relaciones de dependencia entre las variables del problema, de forma que permiten calcular probabilidades marginales y actualizar éstas cuando se dispone de nueva información (evidencia). El Ejemplo 2.1 del Capítulo 2 muestra un modelo probabilístico simple para tres variables: *estación del año*, *viento* y *lluvia* en Parayas (Santander), con 4, 4 y 2 estados, respectivamente. El modelo está definido por una Función de Probabilidad Conjunta (FPC) de todas las variables (la Tabla 2.1 muestra los $4 \times 4 \times 2 = 32$ parámetros necesarios para definirla; en realidad sólo son necesarios 31, porque la suma de todos ha de ser uno). Una vez aprendidos, o ajustados, los parámetros del modelo (utilizando, por ejemplo, las frecuencias de un registro histórico), las fórmulas de la probabilidad permiten obtener información “a priori” sobre las variables: $P(Lluvia) = 0.564$, y actualizar esta información cuando se dispone de nueva evidencia (por ejemplo, “el viento sopla del SW”): $P(Lluvia|SW) = 0.746$. De esta forma se pueden obtener conclusiones para un problema dado: “cuando el viento sopla del SW es más probable que llueva”. Ésta es la base de los sistemas inteligentes probabilísticos (ver Castillo et al., 1997, Cap. 1, para más detalles):

- Por una parte, la base de conocimiento está dada en términos de probabilidades, a través de la FPC que cuantifica las relaciones inciertas entre las variables del problema.
- Por otra parte, el módulo de razonamiento lo constituyen los algoritmos de cálculo de probabilidades condicionadas, que permite obtener conclusiones a partir de información particular (o evidencia) conocida en un instante concreto.

El problema de estos modelos se origina cuando crece el número de variables. Por ejemplo, el número de parámetros necesario para definir una FPC de 50 variables binarias es $2^{50} > 10^{16}$, lo cual es un número prohibitivo. Por tanto, es necesario disponer de una forma más flexible para definir una FPC a partir de cierta información que permita simplificar el número de parámetros, sin tener que especificar la probabilidad de todos los estados posibles del sistema. Un modelo global contempla todas las dependencias posibles que se podrían dar entre las variables; son los valores de los parámetros quienes determinan las relaciones se dan o no en cada caso. Es decir, la estructura paramétrica de una probabilidad conjunta global es demasiado general para que tenga utilidad

práctica. Por ello, es necesario especificar qué relaciones de dependencia e independencia se quieren mantener en un modelo y tratar de simplificar la forma de definir la FPC.

Para ilustrar este problema considérese el siguiente ejemplo: se desea construir un modelo para explicar las relaciones entre meteoros en el observatorio de Parayas (Santander). Para ello se consideran las siguientes variables: *tormenta*, *neblina*, *viento*, *granizo*, *nieve*, *escarcha*, *rocío*, *nieve en el suelo*, *precipitación*, y *niebla* (ver Fig. 4.2(a)). Se dispone de una base de datos con las ocurrencias diarias de estos fenómenos en un período de 10 años, entre 1986 y 1995 (Fig. 4.2(b)). El número de estados posibles de este sistema es $2^{10} = 1024$, demasiado elevado para el número de observaciones disponible, por lo que no cabe plantearse un modelo global que contenga todos los parámetros. Como se verá más adelante, el grafo de la Fig. 4.2(c) constituye la base de una red Bayesiana y es una forma de especificar un modelo más simple, con menos parámetros, que sólo contiene las relaciones de dependencia relevantes para el problema, expresadas mediante el grafo.

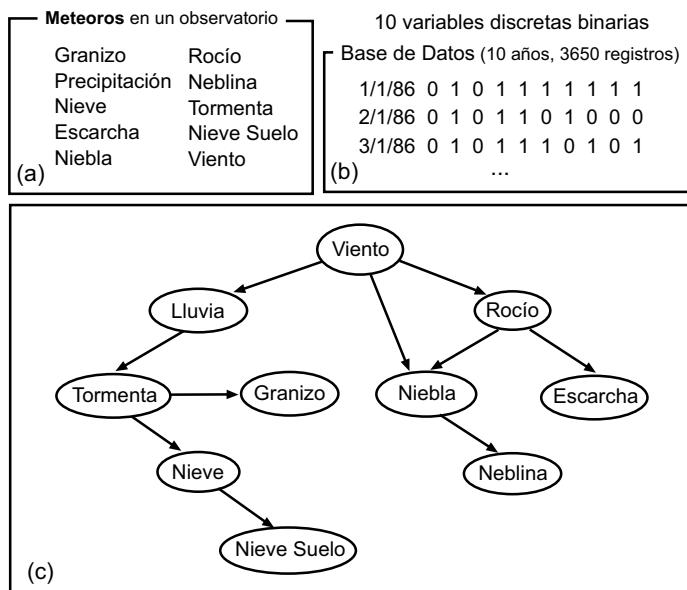


Figura 4.2: Ejemplo de aplicación de redes Bayesiana. (a) 10 meteoros registrados en un observatorio; (b) base de datos con las observaciones de estas variables durante un período de 30 años; (c) grafo dirigido acíclico que representa las relaciones existentes entre las variables.

Los primeros sistemas expertos basados en probabilidad optaron por distintas soluciones para resolver este problema. Cada uno de estos modelos asumía distintas hipótesis de independencia entre las variables para simplificar la FPC y, así, el número de parámetros. Por ejemplo, el modelo más simple se obtendría suponiendo que todas las variables son independientes entre sí. Por tanto,

se tendría que la FPC se podría expresar como: $p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i)$; en este caso, el número de parámetros necesarios en un problema con 10 variables binarias (u once binarias y una cuaternaria, como en el ejemplo anterior) se reduciría a 10, en contraposición a los más de 1000 necesarios en el caso general, que incluye todas las posibles dependencias entre las variables. Sin embargo, el modelo global de independencia es de muy poca utilidad, pues no permite codificar ninguna relación de dependencia entre las variables. En la práctica se utilizaron modelos intermedios que suponían distintas hipótesis de independencia entre las variables, por ejemplo los modelos de enfermedades y síntomas para diagnóstico médico, donde se suponía que los síntomas eran independientes y las enfermedades condicionalmente independientes, dados los síntomas (ver Castillo et al., 1997, Cap. 3). Sin embargo, estos modelos son sólo válidos en problemas concretos que se adecúan a las hipótesis impuestas, pero no permiten construir modelos probabilísticos para un problema general.

En el siguiente ejemplo se ilustra la forma de construir un modelo probabilístico simplificado (factorizado) que contenga un conjunto de relaciones de independencia arbitrario.

Ejemplo 4.1 (Modelo Probabilístico Factorizado) *Considérese un problema definido con cuatro variables binarias $\{X_1, X_2, X_3, X_4\}$, con un total de 16 estados distintos. Si no se conoce ninguna relación de independencia, tendrá 15 parámetros libres. Supóngase que se sabe que las variables del problema cumplen:*

$$I(X_3, X_1|X_2) \text{ y } I(X_4, \{X_1, X_3\}|X_2).$$

Tal como se vio en la Sec. 2.4, estas relaciones de dependencia implican:

$$p(x_3|x_1, x_2) = p(x_3|x_2), \quad p(x_4|x_1, x_2, x_3) = p(x_4|x_2);$$

por tanto, los parámetros de la FPC del modelo resultante no son libres, sino que están sujetos a las restricciones anteriores. Sería deseable encontrar una forma de expresar la FPC con un número reducido de parámetros que fuesen libres para este problema (de esta manera, además de reducir el número de parámetros, se eliminaría redundancia y posibles fuentes de error). En este ejemplo, la regla de la cadena proporciona una solución simple. La regla de la cadena indica que una FPC siempre se puede factorizar de la siguiente forma

$$p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)p(x_4|x_1, x_2, x_3).$$

En este caso concreto, las condiciones anteriores permiten sustituir alguna de estas funciones, resultando:

$$p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_2),$$

que es la expresión general de las FPC que cumplen las relaciones de independencia anteriores (los parámetros libres se reducen de 15 a 7). ■

Sin embargo, en general no es tan simple hallar una factorización apropiada de la FPC que contenga las condiciones de independencia requeridas para un

problema dado. Es decir, dado un conjunto de variables y un conjunto de relaciones de independencia (ver Sec. 2.4) definidas sobre ellas, no es trivial hallar un modelo probabilístico (una factorización de la FPC) que incluya estas relaciones y sólo requiera especificar los parámetros libres del modelo.

Las redes probabilísticas fueron introducidas por Pearl (1988) para dar solución a este problema, combinando grafos y probabilidades.

4.4. Redes Probabilísticas

Las redes probabilísticas combinan grafos y probabilidad para definir modelos que satisfagan un cierto conjunto de relaciones de dependencia. Las relaciones de dependencia entre las variables se expresan mediante un grafo, que es una representación visual e intuitiva; los nodos de grafo representan a las variables y las aristas del grafo las relaciones entre ellas. El grafo proporciona un criterio para factorizar la FPC acorde a las relaciones indicadas en el mismo, incluyendo el mínimos de parámetros necesario. De forma muy general se podrían definir de la siguiente forma:

Las redes probabilísticas combinan grafos y funciones de probabilidad para definir de forma eficiente modelos probabilísticos que contengan las relaciones de dependencia deseadas para un problema y que sean tratables computacionalmente.

Una ventaja de estos modelos es que tanto la estructura gráfica como las probabilidades (los parámetros del modelo factorizado resultante) pueden ser aprendidas directamente de los datos, con algoritmos de aprendizaje apropiados, y sin necesidad de supervisión humana. Por tanto, pese a su complejidad teórica, estos modelos son bastante intuitivos y fáciles de aplicar en situaciones prácticas.

Dependiendo del tipo de grafo utilizado (dirigido, no dirigido, mixto) y de la forma de codificar las relaciones de dependencia en el mismo, se tienen distintos tipos de redes probabilísticas. Las más famosas son las redes Bayesianas (grafos dirigidos acíclicos) y las redes de Markov (grafos no dirigidos); este capítulo se centra básicamente en el primer tipo de modelos (ver Castillo et al., 1997, para más detalles). Por ejemplo, el grafo de la Fig. 4.2(c) define una red Bayesiana para el problema tratado. En primer lugar es necesario conocer cómo se codifican las relaciones de dependencia en el grafo.

4.4.1. Codificación de Dependencias. Criterios de Separación

Los criterios que permiten conocer si una cierta sentencia de dependencia/independencia condicional se cumple en un grafo son conocidos como criterios de separación. El caso más simple es el de los grafos no dirigidos, donde las aristas entre nodos denotan dependencias y, por tanto, el criterio de separación más simple consiste en que dos nodos (o conjuntos de nodos) son independientes si no existe ningún camino que los une en el grafo, y son independientes dado un tercero si todos los caminos que los unen pasan por el tercer nodo. Por ejemplo,

en el grafo de la Fig. 4.3(a) se dan todas las posibles dependencias entre nodos. Sin embargo, en el grafo de la Fig. 4.3(b) se tiene que $D(\text{humedad}, \text{viento})$, pero $I(\text{humedad}, \text{viento}|\text{niebla})$; es decir, la *humedad* y el *viento* son dependientes, pero son condicionalmente independientes dada la *niebla*. Estas relaciones no se cumplen en la práctica, sino más bien las contrarias, pero éstas no pueden especificarse en un grafo no dirigido. Es decir, la capacidad expresiva de los grafos no dirigidos es limitada y no permite expresar este tipo de relaciones.

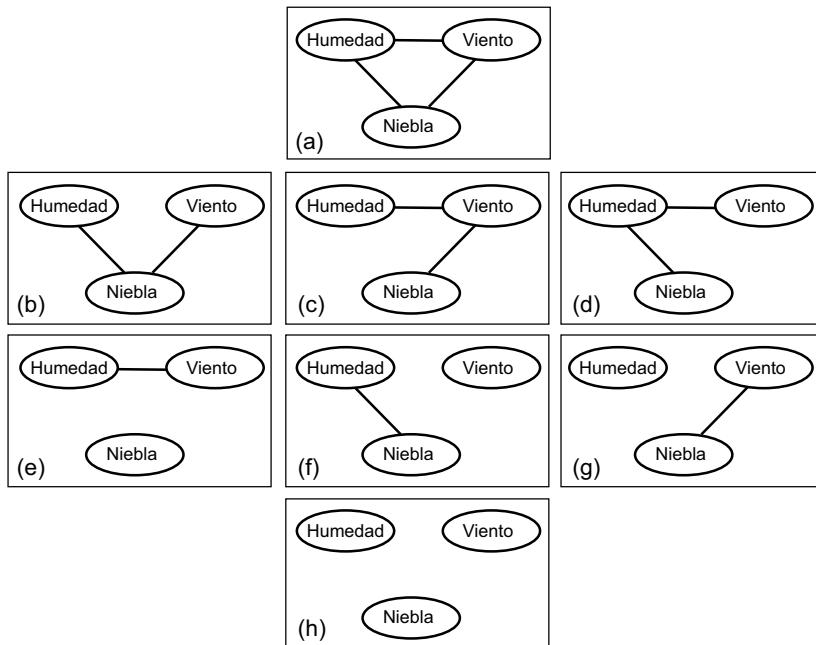


Figura 4.3: Posibles grafos no dirigidos para tres variables.

Ejemplo sugerido por José Luis Navarro, del observatorio del INM en Oviedo.

El criterio de separación para grafos dirigidos es más complejo, pero a la vez más potente, dando cabida a relaciones de dependencia más generales (como las que se cumplirían en la realidad en el ejemplo anterior). El sentido de las aristas proporciona un grado de libertad más para definir el criterio de separación; por ejemplo, los vecinos de un nodo se dividen en *padres* (de los que sale una arista que confluye en el nodo) e *hijos* (a los que llega una arista desde el nodo). Los nodos que se alcanzan ‘corriente arriba’ de uno dado son sus ancestros: padres, abuelos, etc; y todos los nodos que se encuentran ‘corriente abajo’ son sus descendientes: hijos, nietos, etc.

En un grafo dirigido, dos nodos son dependientes si existe un camino que los une en el subgrafo ancestral (que sólo contiene nodos ancestrales de ellos, como padres, abuelos, etc.). Por otra parte, dos nodos son condicionalmente dependientes dado un tercero si existe algún camino que los una y que no contenga al tercero, o que lo contenga en el vértice de una v-estructura (una *v*-

estructura es un camino de la forma $A \rightarrow C \leftarrow B$, donde C es el vértice). Dicho de otra forma, las v-estructuras cumplen la misión de codificar relaciones de la forma: $I(A, B|\emptyset)$, $D(A, B|C)$, donde C es el vértice de una v-estructura entre A y B . Este tipo de relaciones se denominan transitivas y son importantes en la práctica. el criterio de separación en grafos dirigidos (denominado d-separación) fue definido de esa forma para incluir este tipo de relaciones (ver Castillo et al., 1997, para más detalles).

El número de grafos dirigidos se incrementa con respecto al de no dirigidos. Por ejemplo, la Fig. 4.4 muestra tres grafos dirigidos distintos asociados al mismo grafo no dirigido. En el grafo (a) se tiene que $I(\text{humedad}, \text{viento})$ (pues no existen ancestros de ambos), pero $D(\text{humedad}, \text{viento}|\text{niebla})$ (pues existe el camino $\text{humedad} \rightarrow \text{niebla} \leftarrow \text{viento}$). Éste es por tanto el modelo que mejor refleja la realidad. Los otros dos grafos (b)-(c) cumplen $D(\text{humedad}, \text{viento})$, $I(\text{humedad}, \text{viento}|\text{niebla})$; en realidad, ambos grafos son equivalentes, es decir, definen el mismo conjunto de relaciones de dependencia (en general, dos grafos son equivalentes si tienen asociado el mismo grafo no dirigido y poseen las mismas v-estructuras). En realidad, las v-estructuras son los elementos clave en un grafo dirigido que le confieren su capacidad para codificar modelos de dependencia con propiedades convenientes (ver Castillo et al., 1997, Cap. 4).

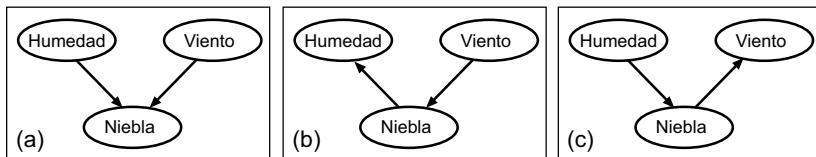


Figura 4.4: Tres grafos dirigidos definidos a partir del mismo grafo no dirigido.

4.4.2. Factorización de la Probabilidad

La complejidad de las redes probabilísticas consiste en hallar un forma de factorizar la FPC de manera que la probabilidad resultante contenga las mismas dependencias e independencias que el grafo con el criterio de separación correspondiente (en el Ejemplo 4.1, la factorización de la probabilidad y el conjunto de dependencias indicado tienen una correspondencia perfecta). Este problema no es trivial ya que el conjunto de modelos de dependencia asociados a una función de probabilidad, no tiene porqué ser el mismo que el que se puede representar mediante un grafo, dirigido o no dirigido. Por tanto, en general, es posible que el modelo de dependencia asociado a un grafo no tenga una correspondencia exacta con el derivado de la función de probabilidad (en caso afirmativo el grafo se denomina un mapa perfecto de la probabilidad); el problema consiste en hallar una factorización de la probabilidad que tenga todas las dependencias que contiene el grafo y quizás alguna adicional, lo que implica algún parámetro extra, pero no una pérdida de generalidad; estos modelos se denominan I-mapas. Las redes de Markov son los I-mapas de los modelos probabilísticos asociados a grafos no dirigidos, donde la probabilidad se factoriza

en base a funciones locales definidas sobre conjuntos de nodos que se obtienen manipulando el grafo. La factorización dada por un grafo dirigido es mucho más intuitiva y los modelos resultantes se denominan redes Bayesianas. En la siguiente sección se describen en detalle estos modelos.

4.5. Redes Bayesianas. Definición y Tipos

Una Red Bayesiana (RB) es un modelo probabilístico de una función de probabilidad conjunta (FPC) definido por un grafo dirigido acíclico (DAG) y un conjunto de funciones de probabilidad condicionada, de forma que la estructura de dependencia/independencia mostrada por el DAG puede ser expresada en términos de la función de probabilidad conjunta mediante el producto de varias distribuciones condicionadas, como sigue:

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i | \pi_i), \quad (4.1)$$

donde Π_i es el conjunto de los padres del nodo X_i en el grafo. De esta manera, las independencias del grafo son inmediatamente traducidas al modelo probabilístico de una manera muy práctica. Por ejemplo, la FPC de una red Bayesiana definida por el gráfico de la figura 4.2 requiere la especificación de 10 tablas de probabilidad condicional, una para cada variable, condicionada al conjunto de sus padres. La Tabla 4.5 muestra una de estas tablas de probabilidad, la correspondiente al nodo *nieve en suelo* dado su único parente *nieve*: $p(\text{nieve suelo} | \text{nieve})$. Los valores de la tabla han sido estimados a partir de los datos de 10 años.

		nieve en suelo	
		si	no
π	nieve=si	0.218	0.782
	nieve=no	0.001	0.999

Tabla 4.1: Tabla de probabilidad condicional para el nodo *nieve en suelo*, dados sus padres $\Pi = \{\text{nieve}\}$.

Dependiendo del carácter discreto o continuo de las variables, las probabilidades condicionales en (4.1) son familias paramétricas específicas. Fundamentalmente hay tres tipos de redes Bayesianas -en este libro se considerarán únicamente los dos primeros-:

- *Redes Bayesianas Multinomiales.* En una Red Bayesiana Multinomial se considera que todas las variables son discretas, es decir, cada variable tiene un número finito de posibles estados. Asimismo se considera que la probabilidad condicionada de cada variable dados sus padres es multinomial y por tanto, queda especificada por la tabla de probabilidades

correspondientes a las diferentes combinaciones de estados entre las variables implicadas. Nótese que el número de parámetros requeridos para especificar la FPC en el caso de dependencia total en el término de la izq. de (4.1) es m^n , siendo n el número de variables y m el número de estados, mientras que la red Bayesiana sólo requiere la especificación de n tablas de probabilidad condicional, una por cada variable condicionada a sus padres.

- *Redes Bayesianas Gaussianas.* En una red Bayesiana Gaussiana, se considera que las variables tienen una distribución normal multivariante, $N(\mu, \Sigma)$, cuya función de densidad conjunta viene dada por

$$f(x) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left\{ -1/2(x - \mu)^T \Sigma^{-1}(x - \mu) \right\}, \quad (4.2)$$

donde μ es el vector de medias n -dimensional, Σ es la matriz de covarianzas $n \times n$, $|\Sigma|$ es el determinante de Σ , y μ^T denota la traspuesta de μ . En una red Bayesiana Gaussiana esta FPC se especifica como en (4.1) por:

$$f(x_i | \pi_i) \sim N \left(\mu_i + \sum_{j=1}^{i-1} \beta_{ij} (x_j - \mu_j), v_i \right), \quad (4.3)$$

donde β_{ij} es el coeficiente de regresión de X_j en la regresión de X_i sobre los padres de X_i , Π_i , y $v_i = \Sigma_i - \Sigma_{i\Pi_i} \Sigma_{i\Pi_i}^{-1} \Sigma_{i\Pi_i}^T$ es la varianza condicional de X_i , dados $\Pi_i = \pi_i$, donde Σ_i es la varianza incondicional de X_i , $\Sigma_{i\Pi_i}$ es el vector de covarianzas entre X_i y las variables en Π_i , y Σ_{Π_i} es la matriz de covarianza de Π_i . Nótese que β_{ij} mide el grado de relación entre X_i y X_j . Si $\beta_{ij} = 0$, entonces X_j no es parente de X_i . Así, la red Bayesiana Gaussiana está definida por un conjunto de parámetros $\{\mu_1, \dots, \mu_n\}$, $\{v_1, \dots, v_n\}$, y $\{\beta_{ij} \mid j < i\}$, como se mostró en (4.3).

- *Redes Bayesianas Mixtas.* En general, el problema de definir modelos de redes Bayesianas que incluyan variables discretas y continuas al mismo tiempo es complicado. Un caso particular lo constituyen los modelos Gaussianos condicionados, donde se permite que una variable continua tenga padres discretos, pero no al contrario (Jordan, 1998). También existen otros modelos más particulares, que se han aplicado a ejemplos concretos. Por ejemplo, Castillo et al. (1998) describen una aplicación de modelos mixtos con variables discretas y Beta.

4.6. Razonamiento Probabilístico. Inferencia

Una vez definido el grafo y la estructura de la FPC resultante, es necesario asignar un valor a los parámetros del modelo (es decir, a las distintas funciones de probabilidad condicionada que definen la factorización). Una vez especificado el modelo completo (*grafo + FPC + parámetros*), la utilidad de las redes probabilísticas consiste en hallar de forma eficiente la probabilidad de cualquier nodo (o conjunto de nodos), dada una cierta información (evidencia). Este proceso

se denomina inferencia o razonamiento probabilístico y permite cuantificar la incertidumbre de las distintas variables y eventos del problema a medida que se va teniendo nueva información, o evidencia. Por ejemplo, el grafo de la Fig. 4.2 define las relaciones entre los meteoros asociados a una estación (Santander). Si la variable de interés es la *precipitación*, entonces la probabilidad $p(precip.)$ proporciona información sobre su estado inicial, y la probabilidad condicionada $p(precip.|e)$ proporciona información actualizada cuando se conoce que ha ocurrido cualquier evento e , por ejemplo $e = \{viento = no, tormenta = si\}$. En general, el cálculo de estas probabilidades es computacionalmente intratable para problemas con muchas variables; sin embargo, existen métodos eficientes de propagación de evidencia que utilizan la estructura de dependencia del grafo para realizar esta tarea (ver Castillo et al., 1997, Caps. 6 y 7).

4.6.1. MeteoLab: *Modelo de Meteoros (I)*

Antes de pasar a describir los métodos automáticos de generación de redes Bayesianas a partir de bases de datos, este ejemplo trata de ilustrar la dificultad de establecer, aún por personal experto, un conjunto apropiado de relaciones relevantes en un problema práctico dado. En concreto, en este ejemplo se analiza el problema descrito anteriormente de los meteoros observados en una estación. El código completo del ejemplo que se describe a continuación se incluye en el fichero `Meteoros.I.m` del directorio `MeteoLab/BayesNets` de *MLToolbox*.

Cuando el modelo es elaborado por un experto humano, se pone de manifiesto la dificultad de establecer relaciones de dependencia que no se basen en asociaciones directas causa-efecto concebidas fundamentalmente a través de la experiencia y/o el conocimiento de las leyes físicas que interrelacionan las diferentes variables. Esto dificulta, o anula, la posibilidad de tener en cuenta dependencias condicionadas, derivadas de la acción combinada de varias variables sobre otra u otras diferentes. Cuando se apliquen más adelante los métodos de aprendizaje automático a este mismo ejemplo, quedará de manifiesto que no necesariamente el grafo más ‘lógico’ es el que mejor explica y compatibiliza las interacciones combinadas que unas variables ejercen sobre otras, en el sentido de dar la máxima verosimilitud a la base de datos.

En este ejemplo se describe el proceso de construcción de una red Bayesiana utilizando la toolbox *BNT* (www.ai.mit.edu/~murphyk), que se incluye con en el paquete *MeteoLab*. Se consideran datos diarios de ocurrencia de meteoros en el aeropuerto de Parayas (a 7 kilómetros de la ciudad de Santander), considerando las variables, precipitación, nieve, granizo, tormenta, niebla, rocío, escarcha, nieve en el suelo, neblina y racha máxima de viento superior a 50 km/h. Las observaciones de cada meteoro para un período de diez años se encuentran en el fichero `dato.txt`, de forma que cada columna corresponde a un meteoro distinto, con un 1 o un 2 indicando la ausencia o presencia del fenómeno, respectivamente. Se trata de un ejemplo de sistema multinomial donde todas las variables son binarias.

Se comienza cargando los datos y definiendo el número de variables, el número de estados de cada una, y los nombres de cada una de ellas (para ello,

se debe estar en el directorio MeteoLab/BayesNets, donde se hallan los datos):

```
>> load('dato.txt');
>> novar=size(dato,2);
>> ns = max(dato);
>> label={'Lluvia', 'Nieva', 'Granizo', 'Tormenta', 'Niebla',...
    'Rocio', 'Escarcha', 'NievaSuelo', 'Neblina', 'V>50'}
```

la variable `novar` indica el número de variables con el que se trabaja, `ns` es el vector con el número de estados posibles de cada variable (dos en este caso, puesto que son todas variables binarias).

A continuación, a partir de la experiencia de un experto humano, se construye la matriz de adyacencia estableciendo manualmente las relaciones entre variables. Inicialmente se define el grafo, con todos los nodos independientes, definiendo una matriz cuadrada de tamaño `novar` rellena de ceros. A continuación, se van añadiendo los enlaces necesarios, de modo que el nodo i -ésimo es el padre del nodo j si el elemento de la fila i y columna j de la matriz es un uno:

```
>> dag=zeros(novar,novar);
>> dag(1,4)=1; dag(2,8)=1;
>> dag(4,2)=1; dag(4,3)=1; dag(5,9)=1; dag(6,5)=1;
>> dag(6,7)=1; dag(10,1)=1; dag(10,5)=1; dag(10,6)=1;
>> draw_graph(dag,label)
```

por ejemplo, el nodo 1 es padre del nodo 4 (la lluvia es padre de la tormenta), el nodo 2 es padre del nodo 8, etc. El grafo resultante se puede obtener con el comando `draw_dag`:

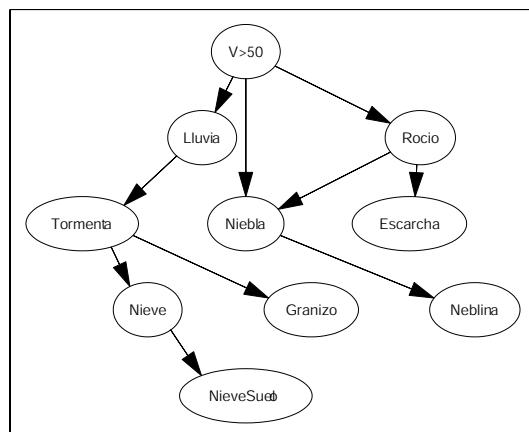


Figura 4.5: Grafo diseñado por un experto para relacionar diferentes fenómenos meteorológicos en el aeropuerto de Parayas.

La matriz de adyacencias resultante, es de la forma

```
>> dag =
0   0   0   1   0   0   0   0   0   0
0   0   0   0   0   0   0   1   0   0
0   0   0   0   0   0   0   0   0   0
0   1   1   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   1   0
0   0   0   0   1   0   1   0   0   0
0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0
0   0   0   0   0   0   0   0   0   0
1   0   0   0   1   1   0   0   0   0
```

Una cuestión muy importante a la hora de trabajar con la Toolbox *BNT*, es que los grafos deben estar siempre construidos y ordenados de forma topológica; esto es, los ancestros tienen que tener numeración menor que los descendientes (padres antes que hijos). Por ello, es necesario, reordenar los datos, las etiquetas y la propia matriz de adyacencias. La función `topological_order` hace esta tarea de forma automática a partir de la matriz de adyacencias, devolviendo un orden topológico:

```
>> ind= topological_sort(dag);
>> dag=dag(ind,ind);
>> dato=dato(:,ind);
>> label=label(ind);
```

Hay que tener en cuenta que el nuevo orden en el que se están definidas las variables es: $V > 50$, Rocío, Escarcha, Niebla, Neblina, Lluvia, Tormenta, Granizo, Nieve y NieveSuelo.

La función `mk_bnet` inicializa una estructura que actuará como recipiente donde se ubicarán todos los detalles necesarios para completar la red bayesiana.

```
>> bnet = mk_bnet(dag, ns);
>> bnet =
equiv_class: [1 2 3 4 5 6 7 8 9 10]
dnodes: [1 2 3 4 5 6 7 8 9 10]
observed: []
names: {}
hidden: [1 2 3 4 5 6 7 8 9 10]
hidden_bitv: [1 1 1 1 1 1 1 1 1]
dag: [10x10 double]
node_sizes: [2 2 2 2 2 2 2 2 2 2]
cnodes: []
```

```

    parents: {1x10 cell}
    members_of_equiv_class: {1x10 cell}
        CPD: {[[]] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ]}
    rep_of_eclass: [1 2 3 4 5 6 7 8 9 10]
        order: [1 6 7 9 10 8 2 4 5 3]

```

y contiene, entre otros, el grafo asociado `dag`, los nodos discretos `dnodes`, el número máximo de estados de cada nodo `node_sizes`, los padres de cada nodo `parents`, el orden en el que se consideran los nodos `order`, etc.

Para construir la red Bayesiana asociada al grafo, es necesario definir los valores de las probabilidades de cada nodo condicionado a sus padres, que factorizan la función de probabilidad conjunta según (4.1). En este caso, según el grafo de la Fig. 4.5, es necesario especificar las probabilidades: $p(V > 50)$, $p(Lluvia|V > 50)$, $p(Tormenta|Lluvia)$, etc. En la estructura anterior, falta dimensionar las tablas de probabilidad condicional para cada uno de los nodos, `bnet.CPD` (se puede observar que, por defecto, se inicializa vacía); para ello se utiliza la función `tabular_CPD`, que inicialmente crea una tabla con valores aleatorios para los parámetros necesarios:

```

>> for i=1:novar
>> bnet.CPD{i} = tabular_CPD(bnet, i);
>> end

```

Una vez estructurada la red Bayesiana por completo, está preparada para que podamos asignar un valor concreto a cada parámetro. En Este ejemplo, para mayor simplicidad, se estiman los parámetros a partir de los datos disponibles, lo que modificará las tablas de probabilidad condicional de acuerdo a las frecuencias observadas en los datos. Para ello se puede utilizar la función `learn_params` (ver Sec. 4.7 para más detalles):

```

>> bnet = learn_params(bnet, dato');

```

Las probabilidades condicionadas que definen el modelo están almacenadas en estructuras de manejo complicado; sin embargo, normalmente no es necesario manipular estos valores. A continuación se muestran algunos de los valores estimados, a modo ilustrativo. Por ejemplo, en el caso de la variable sin padres viento superior a 50 km/h, se tiene $p(V > 50)$:

```

>> s=struct(bnet.CPD{1});
>> dispct(s.CPT)
1 : 0.8558
2 : 0.1442

```

que contiene las probabilidades de ocurrencia de cada uno de los estados en que se ha discretizado dicha variable. Otro ejemplo con más parámetros es la probabilidad $p(Granizo|Tormenta)$:

```
>> s=struct(bnet.CPD{8});
>> dispcpt(s.CPT)
1 : 0.9903 0.0097
2 : 0.7463 0.2537
```

De esta forma se completa la definición de la red Bayesiana y se ha construido la base de conocimiento del sistema inteligente. A continuación se puede calcular la probabilidad de cualquier variable o conjunto de variables condicionadas a cualquier evidencia que se tenga disponible para un problema dado. Inicialmente, interesa calcular las probabilidades marginales (sin evidencia). Para ello, es necesario construir un motor de inferencia (el módulo de razonamiento del sistema). Existen distintos módulos de razonamiento incluidos en la toolbox; por ejemplo, el más simple de todos es el módulo global `engine = global_joint_inf_engine(bnet)`, que realiza los cálculos de probabilidades marginalizando directamente la función de probabilidad conjunta (en este caso, el proceso de razonamiento puede resultar extremadamente lento). También existen módulos de razonamiento que calculan las probabilidades de forma aproximada, utilizando algoritmos de simulación tipo Montecarlo, `engine=likelihood_weighting_inf_engine(bnet)`. Sin embargo, el módulo de razonamiento más adecuado para resolver problemas reales es un método de inferencia exacta que aprovecha la estructura del grafo para calcular eficientemente las probabilidades en base a cálculos locales:

```
>> engine = jtree_inf_engine(bnet);
```

Una vez elegido el módulo de razonamiento (o motor de inferencia) se pueden calcular probabilidades de forma sencilla, asignando la evidencia disponible (o dejándola vacía si no se dispone de evidencia), y calculando la probabilidad marginal deseada (por ejemplo, la probabilidad de $V > 50$, en este caso sin evidencia o “a priori”):

```
>> evidence = cell(1,novar);
>> engine = enter_evidence(engine,evidence);
>> v=marginal_nodes(engine,1);%viento
>> v.T
    0.856
    0.144
```

Procediendo de forma similar para el resto de las variables, definiendo las correspondientes probabilidades marginales `nv = marginal_nodes(engine,2)`,

etc., se tendría (el código en MeteoLab/BayesNets/Meteoros_I.m contiene todos los detalles):

```
>> ro=marginal_nodes(engine,2);%ocio
>> es=marginal_nodes(engine,3);%escarcha
>> nb=marginal_nodes(engine,4);%niebla
>> ne=marginal_nodes(engine,5);%neblina
>> Ll=marginal_nodes(engine,6);%lluvia
>> tor=marginal_nodes(engine,7,1);%tormenta
>> gr=marginal_nodes(engine,8);%granizo
>> nv=marginal_nodes(engine,9);%nieve
>> nvs=marginal_nodes(engine,10);%nieve suelo
>> [Ll.T nv.T gr.T tor.T nb.T ro.T es.T nvs.T ne.T v.T]
.444 .994 .975 .938 .931 .748 .981 .998 .453 .856
.556 .006 .025 .062 .069 .252 .019 .002 .547 .144
```

A continuación, si se dispone de evidencia para un problema dado, se pueden obtener conclusiones estudiando cómo varía la probabilidad de cada variable, dada la evidencia. Por ejemplo, si se conoce que ha ocurrido una tormenta; es decir, el estado del nodo 7 es 2 (ocurrencia), se pueden calcular de nuevo las probabilidades de las variables:

```
>> evidence{7}=2;
>> engine = enter_evidence(engine, evidence);
>> v=marginal_nodes(engine,1); % ...etc...
>> [Ll.T nv.T gr.T tor.T nb.T ro.T es.T nvs.T ne.T v.T]
.039 .956 .746 .000 .936 .764 .981 .990 .455 .783
.961 .044 .254 1.00 .064 .236 .019 .010 .545 .217
```

Como se puede ver, la presencia *tormenta*, afecta sobre todo al hecho de que haya *lluvia* (aumenta considerablemente la probabilidad de *lluvia*) y por supuesto también aumenta la probabilidad de *granizo*, mientras que permanecen indiferentes la *escarcha*, *niebla* y *neblina*. Por otra parte, también se nota un aumento de la probabilidad de rachas de viento superiores a 50 km/h.

De la misma manera, dando evidencia a la *niebla* y propagando se obtiene:

```
>> evidence = cell(1,novar); %Removing previous evidence
>> evidence{5}=2;
>> engine = enter_evidence(engine,evidence);
>> v = marginal_nodes(engine,1); % wind
>> [Ll.T nv.T gr.T tor.T nb.T ro.T es.T nvs.T ne.T v.T]
.490 .994 .976 .942 .000 .485 .975 .998 .106 .982
.510 .006 .024 .058 1.00 .515 .025 .002 .894 .018
```

Al contrario de lo que ocurría antes, ahora las variables más afectadas son el *ocio* y la *neblina*, al aumentar considerablemente su probabilidad de ocurrencia, mientras que otras variables, como son la *lluvia*, la *tormenta* o el *granizo*

son indiferentes a la presencia de *niebla*. En este caso, disminuye la probabilidad de *rachas de viento superiores a 50 km/h*. También se pueden dar evidencias a varios nodos simultáneamente.

4.7. Algoritmos de Aprendizaje Automático

El aprendizaje automático de Redes Bayesianas se puede dividir en dos tareas distintas:

- Aprendizaje estructural, para establecer las relaciones de dependencia entre las variables del modelo a partir de un conjunto de datos, obteniendo el grafo que mejor los represente.
- Aprendizaje paramétrico, para calcular los parámetros (tablas de probabilidad, etc.) de acuerdo con la estructura del grafo, el tipo de red Bayesiana elegida y los datos.

El aprendizaje estructural es el más complejo de los dos, ya que el paramétrico se reduce a estimar los parámetros estadísticos necesarios (tablas de probabilidad condicionada, medias, varianzas, etc.) a partir de los datos disponibles, para lo que existen distintos estimadores estadísticos. Sin embargo, la búsqueda de la estructura que optimiza el conjunto de datos es un problema NP-completo (intratable) y requiere algoritmos eficientes que, aunque no garanticen la optimidad, tengan un coste computacional acotado. En esta sección se describe el aprendizaje estructural. Para una descripción detallada de los métodos de aprendizaje puede consultarse Neapolitan (2003); Jordan (1998).

La complejidad del aprendizaje estructural se deriva de la enorme cantidad de grafos dirigidos acíclicos distintos para un determinado número de variables. Por ejemplo, para 3 variables existen 12, para 4 variables 543, y el número crece de forma combinatoria con el número de variables. Este problema es especialmente crítico en meteorología, ya que cualquier aplicación práctica involucra un gran número de variables, lo que dificulta la definición del modelo gráfico más conveniente. Por tanto, la aplicación de las redes Bayesianas en los problemas reales depende de la disponibilidad de procedimientos de aprendizaje automático para inferir, a partir de los datos, una estructura gráfica gráfica conveniente.

Dado que normalmente el aprendizaje estructural no puede ser exhaustivo, se han introducido distintos métodos de aprendizaje: técnicas de detección de dependencias, técnicas iterativas de búsqueda, etc; recientemente también se han planteado algoritmos que combinan distintas técnicas (de Campos et al., 2003). De entre estos métodos, los más sencillos son los algoritmos iterativos de búsqueda, que se componen de dos partes:

1. Una *medida de calidad*, usada para evaluar la calidad de cada red candidata. Esta medida es global ya que mide simultáneamente la calidad de la estructura gráfica y la calidad de los parámetros estimados.

2. Un *algoritmo de búsqueda*, utilizado para explorar el espacio de grafos posibles de manera eficiente, y así encontrar una red Bayesiana de buena calidad.

Entre las diferentes medidas de calidad propuestas en la literatura, las Bayesianas tienen una sólida fundamentación teórica. A cada red $B = (M, \theta)$, con estructura M y parámetros θ , se le asigna una medida de calidad que es función de su probabilidad a posteriori dados los datos D . Esta distribución de probabilidad a posteriori $p(B|D)$ se calcula como sigue:

$$p(B|D) = p(M, \theta|D) = \frac{p(M, \theta, D)}{p(D)} \propto p(M)p(\theta|M)p(D|M, \theta), \quad (4.4)$$

En el caso de redes multinomiales, asumiendo ciertas hipótesis acerca de las distribuciones a priori de los parámetros y una probabilidad inicial uniforme para todos los modelos, se puede obtener la siguiente medida de calidad (Geiger and Heckerman, 1995):

$$p(B|D) \propto \sum_{i=1}^n \left[\sum_{k=1}^{s_i} \left[\log \frac{\Gamma(\eta_{ik})}{\Gamma(\eta_{ik} + N_{ik})} + \sum_{j=0}^{r_i} \log \frac{\Gamma(\eta_{ijk} + N_{ijk})}{\Gamma(\eta_{ijk})} \right] \right], \quad (4.5)$$

donde n es el número de variables, r_i es el cardinal de la i -th variable, s_i el número de realizaciones del conjunto de padres Π_i , η_{ijk} son los hiper-parámetros de Dirichlet “a priori” para la distribución condicional del nodo i , N_{ijk} es el número de realizaciones en la base de datos que son consistentes con $y_i = j$ y $\pi_i = k$, y N_{ik} el número de aquellas consistentes con $\pi_i = k$. Todos los parámetros de (4.5) se pueden obtener fácilmente a partir de los datos.

En el caso de redes Gaussianas, se obtiene una medida de calidad muy similar considerando que la distribución de los parámetros es de tipo normal-Wishart (ver Geiger and Heckerman, 1994, para una descripción detallada). Existen otras medidas de calidad pero para no extender demasiado el tema, en este capítulo se considerarán únicamente las medidas anteriormente descritas.

4.7.1. Algoritmo K2

El algoritmo *K2* realiza una búsqueda voraz (greedy) eficiente para encontrar una red de calidad en un tiempo razonable (Cooper and Herskovitz, 1992). Este algoritmo ordena previamente los nodos para evitar la generación de ciclos. Se comienza con un grafo vacío. A continuación, siguiendo el orden establecido, para cada variable Y_i el algoritmo añade a su conjunto de padres Π_i aquel nodo cuyo orden es anterior al suyo propio y además produce un mayor incremento en la medida de calidad. Este proceso se repite para cada nodo, mientras el incremento supere un cierto umbral. También se puede limitar el número máximo de padres de cada variable en el proceso de búsqueda.

Considerando la descomponibilidad de la red, la contribución de la variable

Y_i con padres Π_i a la calidad de la red viene dada por

$$\sum_{k=1}^{s_i} \log \frac{\Gamma(\eta_{ik})}{\Gamma(\eta_{ik} + N_{ik})} + \sum_{j=0}^{r_i} \log \frac{\Gamma(\eta_{ijk} + N_{ijk})}{\Gamma(\eta_{ijk})}. \quad (4.6)$$

Esto lleva a un algoritmo de aprendizaje iterativo muy simple. Por ejemplo, el grafo de la Fig. 4.15 ha sido obtenido aplicando este algoritmo.

4.7.2. MeteoLab: *Modelo de Meteoros (II)*

En esta sección, se muestra la forma de construir automáticamente una red Bayesiana para el problema de la Sec. 4.6.1, utilizando el algoritmo de aprendizaje *K2*. En primer lugar, se carga la base de datos, y se define el número de variables con sus nombres y número de estados:

```
>> load('dato.txt');
>> novar=size(dato,2);
>> ns = max(dato);
>> label={'Lluvia', 'Nieve', 'Granizo', 'Tormenta', 'Niebla',
'Rocio', 'Escarcha', 'NieveSuelo', 'Neblina', 'V>50'};
```

Para ejecutar el algoritmo *K2*, es necesario definir una ordenación de las variables; en este caso se toman tal como se encuentran en la matriz de datos. También hay que especificar el número máximo de padres que se quieren considerar para cada variable en el proceso de búsqueda:

```
>> order=1:novar;
>> max_par=1;
```

Finalmente, se aplica el algoritmo *K2* para aprender la estructura y posteriormente se dibuja el grafo resultante (ver Fig. 4.6, izquierda):

```
>> dag= learn_struct_K2(dato', ns, order,max_par);
>> figure; draw_graph(dag,label);
```

Como se puede comprobar, el proceso es enormemente simple si se compara con los pasos necesarios para construir manualmente un modelo (Sec. 4.6.1). Sin embargo, estos ejemplos también ilustran un problema muy común en esta disciplina: El algoritmo automático de aprendizaje permite construir una solución de forma rápida y simple; sin embargo, muchos elementos importantes del modelo (factorización de la probabilidad, tablas condicionadas, etc.) permanecen ocultos durante el proceso y el usuario obtiene un modelo tipo “caja negra” como respuesta a su problema. Este hecho es muy común en otras técnicas de aprendizaje automático; sin embargo, en este caso es posible analizar el modelo

una vez obtenido si se conoce la teoría de las redes probabilísticas, pudiéndose encontrar la explicación física o estadística de la solución, una vez obtenido el modelo. Con otras técnicas (como las redes neuronales), esta interpretación a posteriori es más complicada (o incluso es imposible) pues los patrones de conocimiento que emplean son más abstractos.

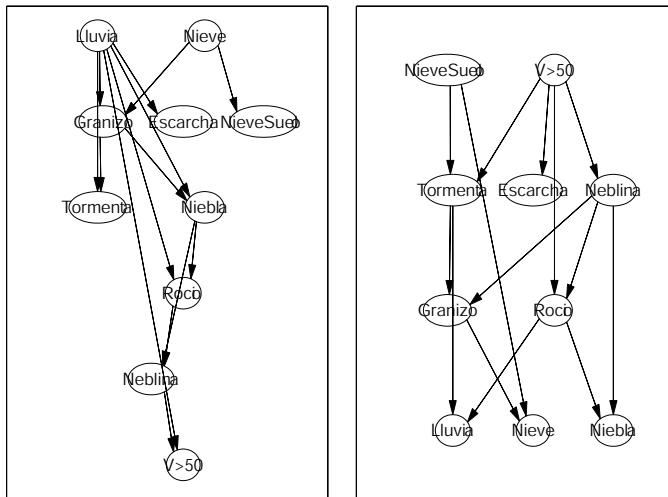


Figura 4.6: Ilustración del efecto de un cambio de ordenación de las variables en el algoritmo $K2$ (el orden está invertido en una figura con respecto a la otra).

A pesar de su eficiencia, el algoritmo $K2$ tiene una serie de características que limitan su aplicación práctica:

- *Efectos de la ordenación.* Se ha constatado la fuerte dependencia que presenta el grafo resultante respecto del orden previamente establecido en las variables cuando se aplica $K2$ a datos meteorológicos. Un efecto claramente apreciable es el hecho de que las variables colocadas en las primeras posiciones pueden tener más hijos que aquellas que eventualmente estén en las últimas posiciones. Esta ordenación, más o menos subjetiva, altera la dependencia espacial o física natural del sistema, creando estructuras cuya interpretación meteorológica es muy pobre. Esta dependencia de la ordenación previa provoca que se obtengan grafos diferentes para diferentes ordenaciones. En la Fig. 4.6 se ilustra el efecto de este algoritmo de aprendizaje sobre la dependencia física del problema. En la figura de la izquierda la lluvia ocupa el primer lugar y es padre de 6 variables mientras que en la figura de la derecha, al ocupar el último lugar no es padre de ninguna. Esto no afecta especialmente a la calidad del grafo para modelizar probabilísticamente el problema, pero es un inconveniente a la hora de tratar de obtener una interpretación del modelo.

Lo mismo se puede decir respecto de la dependencia espacial. Por ejemplo, en la Fig. 4.7 se muestra el gran número de enlaces creados desde el nodo que ocupa eventualmente el primer lugar cuando se aplica el algoritmo *K2* a la racha máxima de viento observada en 100 estaciones de la red de estaciones completas del INM.

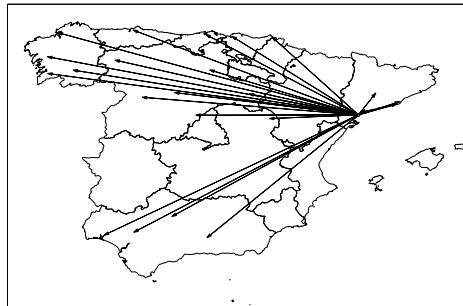


Figura 4.7: Ilustración del efecto sobre la dependencia espacial del algoritmo de aprendizaje *K2* para un ordenamiento concreto de las variables. El primer nodo es el padre de 22 de las 100 estaciones (se han omitido el resto de enlaces).

- *Complejidad del Algoritmo K2.* El tiempo de cómputo del algoritmo *K2* crece cuadráticamente con el número de variables $o(n^2)$. Concretamente, para un límite máximo de un parente por nodo, la complejidad del algoritmo es $o(n(n - 1)/2)$. A pesar de ser uno de los algoritmos más simples y eficientes, el número de variables implicadas en un problema meteorológico real suele ser tan grande que incluso éste se vuelve ineficiente.

Estos problemas tienen una solución parcial en meteorología siempre que exista una relación de dependencia espacial entre las variables que componen el modelo. En este caso, se puede realizar una búsqueda en un espacio más reducido, donde los padres de cada nodo se buscan localmente entre sus “vecinos”. Esto implica una importante reducción en el tiempo de cómputo, por lo que se puede utilizar un algoritmo más costoso que el *K2* pero que elimine los problemas de la ordenación (por ejemplo, el algoritmo *B*). Estas dos modificaciones se describen en las siguientes secciones.

4.7.3. *El Algoritmo de Aprendizaje B*

El algoritmo *B* elimina el problema de la dependencia de la ordenación previa de los nodos (Buntine, 1991). Al igual que el algoritmo *K2*, se inicia con conjuntos de padres vacíos y en cada etapa se añade aquel enlace que maximiza el incremento de calidad, eliminando aquellos enlaces que originen ciclos; sin embargo, en este caso, en cada iteración se busca entre todos los nodos el enlace que maximiza la calidad de la red; la independencia de la ordenación se consigue a cambio de controlar la generación de ciclos. Este proceso se itera hasta que

no se consigue un incremento de calidad, o bien se obtiene una red completa. A cambio de obtener redes de mayor calidad, el tiempo de cómputo del algoritmo B crece cúbicamente con el número de variables $o(n^3)$, por lo que su aplicación en problemas meteorológicos es muy limitada, quedando restringido a sistemas con muy pocas variables y estados.

4.7.4. Algoritmos de Aprendizaje Local. $LK2$ y LB

Para mejorar la eficiencia de la estrategia de búsqueda de los algoritmos anteriores, se puede modificar el conjunto de candidatos a padres para cada nodo incluyendo no todos los anteriores, sino únicamente aquellos cuyas características climatológicas sean similares a la del nodo Y_i . Cuando las variables corresponden a mediciones u ocurrencias en puntos geográficos distintos de una misma variable meteorológica, entonces una forma simple de elegir a los padres “potenciales” es considerar la correlación entre las series de observaciones simultáneas (en un contexto más general puede ser interesante emplear la información mutua o métodos de agrupamiento en lugar de la correlación). Esta modificación reduce notablemente la complejidad del proceso de búsqueda, pues mientras que el algoritmo $K2$ debe evaluar un promedio de $(n - 1)/2$ candidatos por enlace, el nuevo algoritmo (denominado $LK2$, Local $K2$) debe evaluar tan sólo $(m - 1)/2$, donde m es el número máximo de padres potenciales que se quieren considerar. Por tanto, este término puede considerarse constante cuando n es mucho mayor que m (es decir, para problemas con muchos nodos). Por tanto, la reducción en el tiempo de cómputo es de un orden de magnitud: $LK2$ es lineal, mientras que LB sería cuadrático (equivalente al $K2$ estándar). En la Fig. 4.8 se ve como evoluciona el tiempo de CPU del proceso de aprendizaje estructural en función del número de nodos para tres tipos diferentes de algoritmos de aprendizaje ($K2$ estándar y $LK2$ con 5 y 10 vecinos respectivamente).

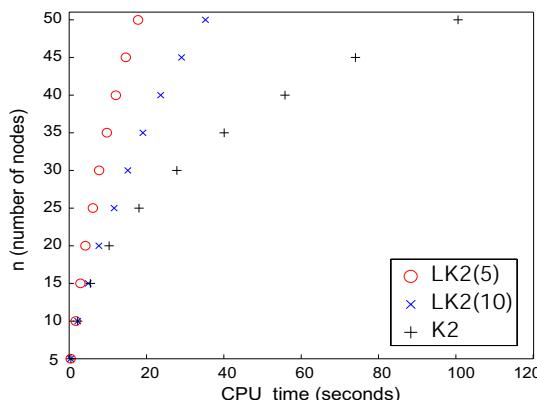


Figura 4.8: Número de nodos versus tiempo de CPU para tres algoritmos diferentes: $K2$ estándar y $LK2$ con 5 y 10 vecinos, respectivamente (Pentium IV 1.6 GHz).

La Figura 4.9 muestra dos grafos obtenidos aplicando los algoritmos $K2$ y $LK2(10)$ para modelizar las dependencias de la precipitación en sesenta estaciones de red de estaciones completas del INM. En ambos casos, el número máximo de padres permitido es dos, y el algoritmo $LK2$ considera sólo diez padres potenciales. La figura 4.9(a) muestra el grafo obtenido usando el algoritmo $LK2$ con 10 vecinos. La medida de calidad puede ser considerada como muy similar a la obtenida utilizando el algoritmo $K2$, mostrada en la Fig.4.9(b). Con el algoritmo $K2$ cada ordenación genera un grafo diferente, además de establecer enlaces de largo alcance que tienen difícil interpretación. En el caso de $LK2$ se ha constatado que las diferentes ordenaciones generan grafos muy similares, y aunque la calidad es ligeramente inferior a la del grafo obtenido con $K2$, la reducción de tiempo de cómputo por una parte y la mejor interpretación meteorológica del grafo por otra, justifican sobradamente su aplicación.

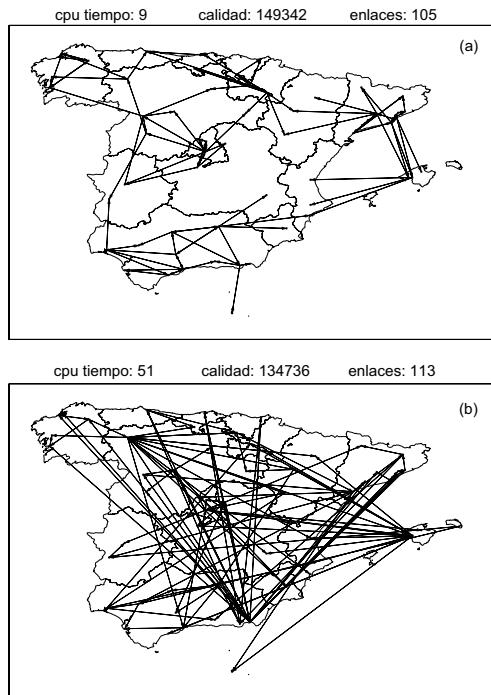


Figura 4.9: DAG para precipitación obtenidos aplicando el algoritmo (a) $LK2(10)$ y (b) $K2$. El número de nodos es 60 y el máximo número de padres es de dos.

Por otro lado, en la Figura 4.10 se compara la medida de calidad de $K2$ con $LK2(m)$ para diferente número de vecinos m , aplicado al caso anterior de la precipitación. En esta figura se puede ver la convergencia exponencial de la calidad de $LK2(m)$ hacia $K2$ para valores grandes de m . Esta figura indica que con un reducido número de padres potenciales se pueden tener resultados similares al algoritmos $K2$.

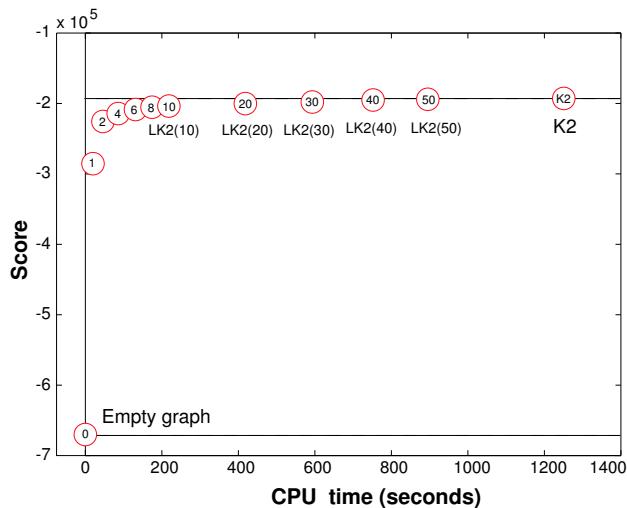


Figura 4.10: Comparación entre el algoritmo $K2$ y varios $LK2$ para diferentes valores de m (Pentium IV 1.6 GHz). En este ejemplo concreto se ve que tomando 10 vecinos prácticamente se alcanza la misma calidad que tomándolos todos. La línea horizontal indica la medida de calidad obtenida con el algoritmo $K2$.

Las mismas modificaciones se han aplicado al algoritmos B obteniendo resultados similares.

4.7.5. MeteoLab: Modelado de Meteoros (III)

Continuando con los ejemplos anteriores, ahora se trata de aprender el grafo automáticamente a partir de los datos utilizando el algoritmo LB . En este caso se obtendrá un grafo más representativo del problema (independiente del orden) en un tiempo similar al empleado por el algoritmo $K2$.

De nuevo, lo primero que se debe hacer es cargar los datos, establecer el número de variables, así como el número de estados de cada una de ellas, como se hizo en la primera parte de los ejemplos anteriores. Para cada nodo, el algoritmo LB busca padres únicamente entre los nodos que tengan una mayor correlación con él (los padres potenciales). Por tanto, lo primero que se hace es calcular la correlación entre los datos, y ordenarla en orden ascendente. También se fijan el número de padres potenciales y el máximo número de padres que puede tener cada nodo en el grafo final:

```
>> N=abs(corrcoef(datos));
>> [a,b]=sort(N);
>> pot_par=5;
>> max_par=2;
```

Como en este ejemplo hay pocas variables, podrían considerarse como padres potenciales todos los nodos, `pot_par=ncard-1`, lo que equivaldría a aplicar el algoritmo B directamente. Por último, antes de poder aprender el grafo es necesario dar la lista de posibles padres para cada nodo, que vendrá determinada por el orden de correlaciones antes calculado

```
>> lis = b(end-pot_par+1:end,:);
>> [dag,score] = L_B(dato',ns,lis,pot_par,max_par);
>> draw_graph(dag,label);
```

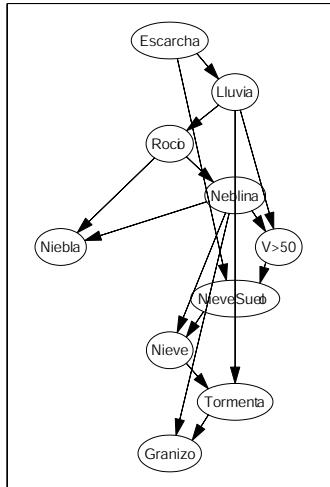


Figura 4.11: Grafo obtenido automáticamente a partir de los datos usando el algoritmo de aprendizaje Local B

De nuevo es necesario reorganizar el grafo y los datos, según el orden topológico antes definido, para poder aplicar el aprendizaje de parámetros y la inferencia. Según el grafo, ahora el orden queda, *Escarcha, Lluvia, Rocío, Neblina, V > 50, NieveSuelo, Nieve, Tormenta, Granizo y Niebla*. Dibujándolo, se tiene la figura 4.11. En este caso las probabilidades marginales (a priori) resultan:

```
>> [Ll.T nv.T gr.T tor.T nb.T ro.T es.T nvs.T ne.T v.T]
.444 .994 .975 .938 .931 .748 .982 .998 .452 .858
.556 .006 .025 .062 .069 .252 .018 .002 .548 .142
```

Estas probabilidades son idénticas a las obtenidas a partir del grafo definido por el meteorólogo en la primera parte de este ejemplo. Sin embargo, los resultados pueden ser distintos al calcular probabilidades condicionadas a una cierta evidencia. Por ejemplo, la Tabla 4.7.5 muestra las probabilidades de ocurrencia para algunas variables, dadas distintas evidencias. p_1 denota las probabilidades

del modelo ‘humano’ y p_2 las del ‘automático’. Se han resaltado en negrita aquellos valores sensiblemente discrepantes del modelo humano respecto del automático. Este ejemplo ilustra la limitación humana para poder definir un grafo que represente tanto las dependencias incondicionales, como las condicionales de un modelo.

	Ll	Nv	Gr	Ro	Es	NS	Ne	V>50
$p_1(x Tr)$.961	.044	.254	.236	.019	.010	.545	.217
$p_2(x Tr)$.956	.043	.251	.123	.005	.010	.495	.215
$p_1(x Nb)$.510	.006	.024	.515	.025	.002	.894	.018
$p_2(x Nb)$.446	.003	.011	.050	.022	.002	.894	.077
$p_1(x Tr, Nb)$.954	.044	.254	.515	.025	.010	.894	.029
$p_2(x Tr, Nb)$.943	.027	.128	.306	.005	.007	.858	.139

Tabla 4.2: Probabilidades obtenidas con dos redes Bayesianas distintas para el mismo problema.

4.8. Aplicaciones

A continuación se describen algunos problemas ilustrativos de meteorología que se adecúan al tratamiento probabilístico (una introducción de las redes Bayesianas desde la perspectiva de la meteorología, con múltiples ejemplos de aplicación se tiene en Cano et al. (2004)). Para ello, se consideran las dos fuentes de información más utilizadas en este ámbito (ver Sec. 1.9 para más detalles):

- *Bases de datos Climatológicos*, con información fenomenológica local (p.e., registros históricos de precipitación) distribuidos en redes de observación sobre áreas geográficas específicas. Por ejemplo, en la Fig. 4.12(b) muestra 100 estaciones en la cuenca Norte, donde se dispone de registros históricos de precipitación durante los últimos 50 años.
- *Bases de datos de Reanálisis*, con predicciones de determinados modelos numéricos de circulación atmosférica durante largos períodos de tiempo. Estos datos están organizados en rejillas 4D con valores de presión, temperatura, humedad, viento, campos derivados, etc., sobre todo el planeta en caso de ser modelos globales, o sobre determinadas áreas de interés en el caso de ser modelos de área limitada. Por ejemplo, la Fig. 4.12(c) muestra los puntos de rejilla del modelo numérico del centro europeo ECMWF sobre la cuenca norte.

Las bases de datos climatológicas contienen información de las propiedades estadísticas de los diferentes observables registrados en cada localidad, mientras que las bases de datos de reanálisis contienen información acerca del estado y evolución de la atmósfera. De esta manera, ésta información puede ser utilizada para analizar problemas relacionados con la dinámica atmosférica y sus posibles impactos locales o regionales.

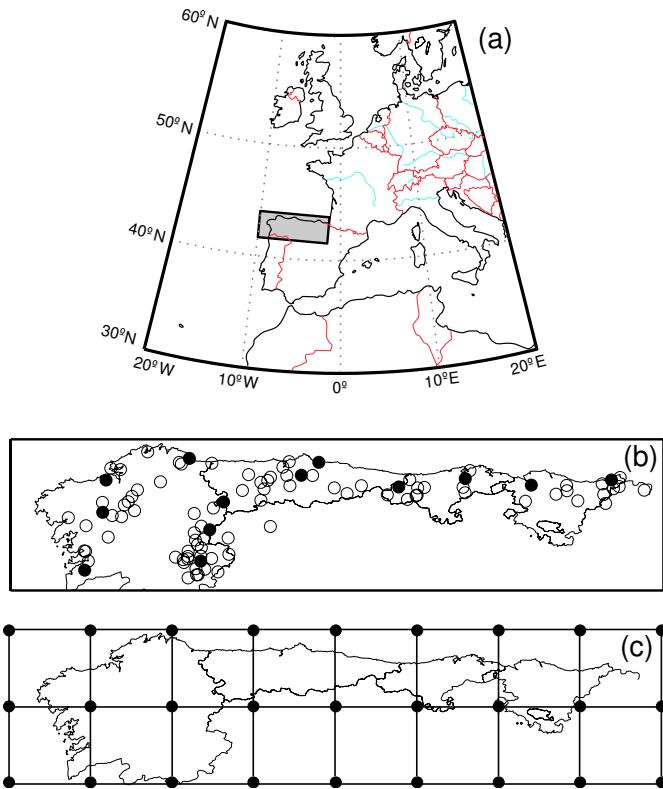


Figura 4.12: (a) Cuenca norte hidrográfica de la península Ibérica (región sombreada); (b) 100 estaciones de la red pluviométrica (los círculos negros indican estaciones de la red automática, donde se dispone de datos en tiempo real); (c) rejilla de $1^\circ \times 1^\circ$ de longitud y latitud utilizada para el reanálisis ERA40 del ECMWF.

4.8.1. Modelos Conceptuales

Un problema de las ciencias de la atmósfera en general, y de la meteorología en particular, es la síntesis de la información, ya que casi siempre se dispone de un gran número de variables y registros de donde hay que extraer la información relevante para caracterizar un problema dado. Por ejemplo, si se quiere desarrollar un modelo conceptual para explicar y predecir la precipitación a partir de un conjunto de variables simuladas por los modelos numéricos (humedad en 500mb, dirección del viento u y v en 1000mb, etc.), se puede obtener un modelo de dependencia que proporcione un modelo explicativo sencillo y que además tenga buenas propiedades predictivas.

En el ámbito de la meteorología, se entiende por *modelo conceptual* la descripción de los procesos característicos de las diferentes escalas y dinámicas de la atmósfera (una tormenta, un frente, etc.). De acuerdo con Chuck Doswell (Cooperative Institute for Mesoscale Meteorological Studies), los modelos con-

ceptuales casi nunca son definitivos; más bien al contrario, pues se encuentran sometidos a una revisión continua. La identificación y caracterización de un modelo conceptual es en buena parte un problema de clasificación, en el que los procesos físicos de la atmósfera no son idénticos sino similares; además, se producen en diferentes versiones: por ejemplo los frentes pueden ser fríos, cálidos u ocluidos, las supercélulas pueden ser secas, clásicas o torrenciales, los tornados igualmente se pueden producir en modalidad supercélula o no, etc. Estos argumentos ‘taxonómicos’ para caracterizar los modelos conceptuales son muy utilizados pero para ser realmente útiles han de ser capaces de manifestar diferencias reales en la dinámica de los procesos modelizados. En cierta medida los modelos conceptuales son referencias muy utilizadas para intercambiar conocimiento en el campo de las ciencias de la atmósfera. Por ello es muy importante que haya una definición muy clara de todos y cada uno de dichos modelos.

Este problema se puede abordar desde un punto de vista puramente teórico, recurriendo a expertos en la materia que con su experiencia y conocimiento establecen índices y relaciones físicas para explicar los datos y predecir el fenómeno considerado. Por otra parte, los modelos conceptuales también pueden ser obtenidos a partir de los datos, y esta es precisamente una de las tareas que pueden ser abordadas desde el punto de vista de las redes Bayesianas (si bien no todos los fenómenos a modelizar conceptualmente cuentan con una base de datos suficiente como para aplicar este método).

Una red Bayesiana, utilizada como modelo conceptual, tiene una doble utilidad puesto que puede ser utilizada para predecir, dando como evidencias los campos predictores, o bien para simular escenarios atmosféricos, dando como evidencia los campos predictandos. Así una vez que se conoce el modelo de dependencia del sistema es más fácil compatibilizar la interpretación física del fenómeno con los datos observados.

En el ejemplo que se muestra a continuación, se trata de encontrar las dependencias más relevantes entre los campos que definen el estado de la atmósfera y varios fenómenos meteorológicos en diferentes puntos. En este ejemplo únicamente se obtiene la estructura del modelo conceptual, queda por tanto pendiente la tarea de interpretar el modelo y evaluar su utilidad que dependerá de su capacidad para explicar y predecir la fenomenología.

Ejemplo 4.2 (Modelo Conceptual) *Para este ejemplo se utilizan 15 variables que definen el estado de la atmósfera: Z1000, Z850, Z500, T1000, T850, T500, U1000, U850, U500, V1000, V850, V500, H1000, H850, H500 y corresponden a los campos y niveles indicados en el nombre de las variables sobre el dominio de la cuenca norte mostrado en la Fig.4.12(c). En cuanto a la fenomenología meteorológica, para simplificar el análisis se considera una subcuenca, que comprende aproximadamente el área geográfica de la zona central comprendida entre el Cabo de Peñas y el Cabo de Ajo. En las estaciones de esta zona se analizan tres variables: la precipitación (77 observatorios), la temperatura mínima y la temperatura máxima (34 observatorios). La Fig. 4.13 muestra la red de observatorios utilizada.*

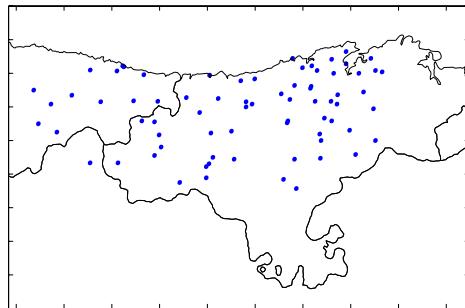


Figura 4.13: Red de observatorios de la segunda subcuenca de la cuenca norte.

Todas las variables están discretizadas teniendo en cuenta que no se debe definir una red con pocos datos y muchos parámetros, ya que se produciría un sobreajuste de los parámetros y se perdería la capacidad predictiva. Se han utilizado 10 estados por variable, obtenidos aplicando la técnica de agrupamiento k-medias a la serie diaria 1979-1993. A continuación se ha aplicado el algoritmo de aprendizaje B, limitando el número máximo de padres a tres para las variables fenomenológicas e imponiendo la independencia para las variables atmosféricas, ya que éstas son evidenciales en modo predictivo. El grafo resultante se muestra en la Fig.4.14. En el grafo se muestran únicamente aquellos aspectos del estado de la atmósfera que tienen influencia relevante en la determinación de la clase de distribución espacial de la precipitación y las temperaturas extremas.

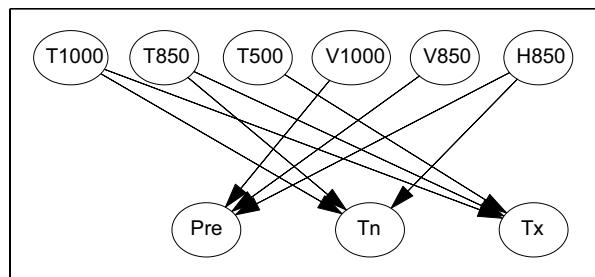


Figura 4.14: Grafo que muestra la relación de dependencia entre variables fenomenológicas (abajo) y variables atmosféricas (arriba).

En este grafo es interesante resaltar el hecho de que los modos de precipitación dependen del flujo meridiano atmosférico en niveles medios y bajos y de la humedad atmosférica en niveles medios, mientras que los modos de temperatura máxima dependen exclusivamente de la temperatura atmosférica en todos los niveles; finalmente los modos de temperatura mínima dependen de la temperatura atmosférica en niveles medios y bajos así como de la humedad atmosférica en niveles medios. Otro aspecto interesante es el hecho de que aunque a priori las variables fenomenológicas son dependientes, resultan inde-

pendientes dado el estado de la atmósfera, lo cual significa que el estado de la atmósfera contiene implícitamente la información para generar un pronóstico físicamente consistente.

A continuación se muestra el código de Matlab necesario para realizar este ejemplo (los datos están en MeteoLab/BayesNets/BNConcep):

```
%SELECCION DE VARIABLES DISCRETIZADAS
%se carga el estado de la atmósfera de la cuenca
%norte dividido en 15 campos con 10 estados cada uno
load ('clasATM10')
%se carga la fenomenología de la zona central de la cuenca
%norte para la Prec, Tmín y Tmax con 10 estados cada uno
load ('clasFEN10')

%Se unen las variables discretizadas en una única matriz de datos
dato=[PreCan TnCan TxCan ATM];

%Se etiquetan las variables para identificarlas en el grafo
labels={'Pre','Tn','Tx',...
'Z1000','Z850','Z500',...
'T1000','T850','T500',...
'U1000','U850','U500',...
'V1000','V850','V500',...
'H1000','H850','H500'};

%APRENDIZAJE DE LA ESTRUCTURA
novar=size(dato,2);
ns = max(dato);
pot_par=novar-1;
max_par=3;

%Correlación para elegir los padres potenciales
N=abs(corrcoef(pstd(dato)));
[a,b]=sort(N);
%lista ordenada de los padres potenciales
lis=b(end-pot_par+1:end,:');

[dag,score]= L_B(dato',ns,lis,pot_par,max_par);

%PRESENTACION GRAFICA
%Se eliminan las variables que no intervienen
lista=find(sum(dag)>0 | sum(dag')>0);
dag=dag(lista,lista);labels=labels(lista);
figure; draw(dag,labels);
```

Este ejemplo ilustra cómo el patrón atmosférico global que se ha venido utilizando en diferentes ejemplos puede ser redefinido de acuerdo con la estructura de dependencia del modelo, pues puede ocurrir que no todas las variables incluidas en el patrón global resulten realmente relevantes para la modelización

conjunta de los fenómenos incluidos en el modelo, aunque esto dependerá del fenómeno y del área de interés.

4.8.2. Dependencia Espacial en la Red Pluviométrica

En el Ejemplo 2.9 del Cap. 2 se mostró la dificultad existente para caracterizar las relaciones de dependencia espacial de una misma variable (por ejemplo, la precipitación) en la red de estaciones pluviométricas de la geografía nacional. Por una parte, si se utiliza la correlación para cuantificar estas relaciones, se capturan sólo las dependencias lineales “a priori” entre las variables, pero no los patrones de dependencia condicional que resultan al introducir evidencia en el modelo (por ejemplo, “ha llovido en Palencia”), o los patrones de dependencia no lineal (otros métodos lineales, como la correlación canónica, también tienen esta limitación). Como se verá en esta sección, las redes Bayesanas proporcionan un modelo apropiado para caracterizar conjuntamente las relaciones a priori y condicionales más relevantes entre las variables del modelo (las estaciones de la red pluviométrica).

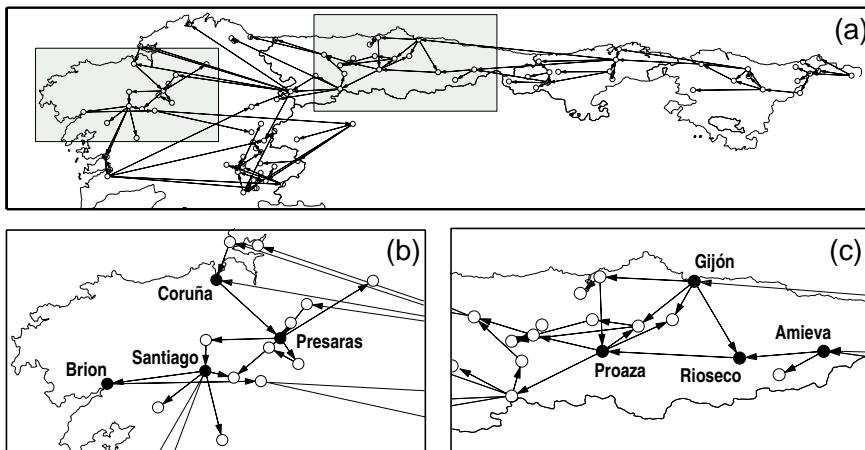


Figura 4.15: Grafo dirigido acíclico que codifica las relaciones de dependencia entre la precipitación en 100 estaciones de la cuenca Norte. Para simplificar la figura sólo se etiquetan algunas ciudades.

La Fig. 4.15 muestra un grafo dirigido acíclico correspondiente a la precipitación en la red de estaciones mostrada en la Fig. 4.12(b) obtenido con el algoritmo *LK2* (ver el fichero *MeteoLab/BayesNets/SpatialDependence.m* para los detalles de la implementación). La precipitación se considera una variable discreta $\{y^1, \dots, y^n\}$, con cuatro estados ($0 = \text{"seco"}$, $1 = \text{"lluvia débil"}$, $2 = \text{"lluvia moderada"}$ y $3 = \text{"lluvia fuerte"}$), que corresponde con los umbrales $0, 2, 10$ y 20 mm/día , respectivamente (el superíndice indica el número de estación). Las estaciones (nodos) están conectadas mediante aristas que, como se vio en la sección anterior, representan relaciones de dependencia; si existe una arista del nodo y^j al nodo y^k , se dice que y^j es un parente de y^k , o que y^k es

un hijo de y^j . Por ejemplo, en la Figura 4.15(c), los nodos *Gijon* y *Amieva* son ambos padres de *Rioseco* y *Proaza* es un hijo de *Rioseco*.

Este grafo implica una factorización de la FPC (4.1) mediante el producto de 100 funciones de probabilidad condicionada, una para cada nodo, dados sus padres (por ejemplo, la Tabla 4.3 muestra una de estas probabilidades condicionadas, que han sido estimadas a partir del registro histórico de datos).

$\Pi_{rioseco}$	Estado de Rioseco			
	0	1	2	3
{0, 0}	0.9377	0.0552	0.0047	0.0024
{0, 1}	0.5909	0.3727	0.0273	0.0091
{0, 2}	0.3333	0.4444	0.1111	0.1111
{0, 3}	0.0000	0.0000	0.6667	0.3333
{1, 0}	0.4026	0.4935	0.0909	0.0130
{1, 1}	0.1736	0.5069	0.2500	0.0694
{1, 2}	0.0968	0.6452	0.2581	0.0000
{1, 3}	0.0000	0.6000	0.4000	0.0000
{2, 0}	0.0909	0.4545	0.4545	0.0000
{2, 1}	0.1429	0.3333	0.3095	0.2143
{2, 2}	0.0000	0.1212	0.3939	0.4848
{2, 3}	0.0000	0.1111	0.6667	0.2222
{3, 0}	0.5000	0.1000	0.4000	0.0000
{3, 1}	0.0000	0.0357	0.3929	0.5714
{3, 2}	0.0000	0.1481	0.1111	0.7407
{3, 3}	0.0000	0.0000	0.0909	0.9091

Tabla 4.3: Probabilidad del nodo *Rioseco*, condicionada a su conjunto de padres $\Pi = \{Gijon, Amieva\}$.

El modelo resultante permite calcular la probabilidad de que llueva en cada estación, y actualizar estas probabilidades cuando se dispone de evidencia. Por ejemplo, antes de disponer de ninguna información acerca de la precipitación en las distintas estaciones, las probabilidades “a priori” de las variables se obtienen calculando $P(y^k = i)$, $i = 0, 1, 2, 3$; $k = 1, \dots, 100$, tal como muestra la Tabla 4.4(a) para algunas de ellas. Esta tabla muestra que los regímenes de precipitación en el área de estudio son similares, dado que todas ellas pertenecen a una misma cuenca hidrográfica. La estación más lluviosa es *Santiago* con casi la mitad de probabilidad lluvia. A continuación, a medida que se recibe cierta información o evidencia e relevante para el problema (por ejemplo, la precipitación ocurrida en las estaciones automáticas conectadas online), se pueden calcular las nuevas probabilidades $P(y^k|e)$ para actualizar el conocimiento. Por ejemplo, las Tablas 4.4(b)-(d) muestran el efecto producido por distintas evidencias. Comparando estas probabilidades con las probabilidades iniciales (Tabla 4.4(a)) se pueden analizar los efectos de la información sobre las relaciones entre las distintas variables. De esta forma se puede obtener importante información climatológica como, por ejemplo, las combinaciones de estados más probables (a partir de las realizaciones con mayor probabilidad del

conjunto de variables).

En este ejemplo, se ha mostrado que las redes Bayesianas permiten implementar de forma eficiente en un único modelo las relaciones probabilísticas de dependencia espacial existente entre un gran número de estaciones pluviométricas. Estos modelos pueden contribuir en gran medida a la realización coherente de estudios estadísticos que requieren de la dependencia espacial para su coherencia. En las dos secciones siguientes mostramos dos ejemplos que utilizan esta idea para, por un lado desarrollar generadores estocásticos de tiempo con coherencia espacial y, por otro, desarrollar modelos de predicción local coherentes espacialmente.

4.8.3. Generadores de Tiempo (*Weather Generators*)

En la sección 2.3 se han descrito en detalle los algoritmos estándar para la simulación de series de tiempo para una estación dada. El problema de estos algoritmos es que cuando son aplicados a un conjunto de estaciones, no tienen en cuenta la dependencia espacial entre estaciones para generar series espacialmente consistentes con la climatología. En los últimos años han sido desarrollados algunos modelos para la simulación conjunta de series en un conjunto de estaciones, pero las soluciones propuestas son demasiado complejas y parciales (ver, por ejemplo, Wilks, 1999c). Las redes Bayesianas pueden ser utilizadas para este propósito. En este caso, la idea básica es generar una muestra de realizaciones a partir de la FPC conjunta de las variables de la red Bayesiana. La simulación se ve facilitada por la factorización de la probabilidad, pues cada variable puede ser simulada de forma independiente (de acuerdo con los valores de los padres), siguiendo el adecuado orden ancestral (ver Castillo et al. (1997), Cap. 6, para más detalles). Uno de los métodos de simulación más simples e intuitivos consiste en ir simulando variable a variable empezando de padres a hijos; sólo se simula el valor de una variable cuando han sido simulados todos y cada uno de sus padres Henrion (1988) y se consideran sólo aquellas realizaciones compatibles con la evidencia (método de aceptación/rechazo). El algoritmo finaliza cuando se tiene el número de realizaciones requerido.

Este algoritmo actúa como un generador estocástico de tiempo con consistencia espacial. En lugar de realizar simulaciones independientes para cada variable, como en un generador de tiempo estándar, este tipo de simulación tiene en consideración todas y cada una de las dependencias espaciales codificadas en el grafo. Se ilustra la aplicación de estos algoritmos de simulación para la precipitación en el caso discreto, aunque el esquema es análogo para el caso continuo. La Fig. 4.16 ilustra una muestra de valores de precipitación (1, 2, 3, o 4) generados para 200 días consecutivos para tres observatorios, apreciándose la consistencia espacial de los valores generados.

La simulación ha de ser compatible con las evidencias, caso de que las haya. Por ejemplo, las Figs. 4.17 muestran simulaciones de precipitación generadas bajo la evidencia de eventos de ‘seco’ o ‘lluvia’ simultáneamente en los observatorios de *Madrid* y *La Coruña* (los generadores de tiempo estándar no poseen esta flexibilidad, Wilby and Wilks (1999)).

(a) Probabilidad marginal (inicial), $P(y^k)$							
Estado	Amieva	Gijon	Rioseco	Proaza	Coruña	Santiago	Brion
0	0.680	0.667	0.648	0.707	0.596	0.545	0.570
1	0.198	0.235	0.207	0.200	0.250	0.218	0.201
2	0.063	0.071	0.090	0.059	0.106	0.103	0.120
3	0.059	0.027	0.055	0.034	0.048	0.134	0.109

(b) $P(y^k Amieva = 3)$							
Estado	Amieva	Gijon	Rioseco	Proaza	Coruña	Santiago	Brion
0	0.000	0.250	<u>0.125</u>	0.211	0.334	0.368	0.429
1	0.000	0.382	<u>0.073</u>	0.311	0.351	0.243	0.229
2	0.000	0.229	<u>0.293</u>	0.222	0.202	0.152	0.169
3	1.000	0.139	<u>0.509</u>	0.256	0.113	0.237	0.173

(c) $P(y^k Santiago = 3)$							
Estado	Amieva	Gijon	Rioseco	Proaza	Coruña	Santiago	Brion
0	0.556	0.502	0.482	0.466	0.142	0.000	<u>0.052</u>
1	0.239	0.308	0.250	0.320	0.369	0.000	<u>0.130</u>
2	0.100	0.133	0.151	0.127	0.309	0.000	<u>0.313</u>
3	0.105	0.057	0.117	0.087	0.180	1.000	<u>0.505</u>

(d) $P(y^k Amieva = 3, Santiago = 0)$							
Estado	Amieva	Gijon	Rioseco	Proaza	Coruña	Santiago	Brion
0	0.000	0.365	0.210	0.361	0.657	1.000	0.894
1	0.000	0.371	0.089	0.329	0.252	0.000	0.083
2	0.000	0.175	0.316	0.167	0.070	0.000	0.018
3	1.000	0.089	0.385	0.143	0.021	0.000	0.005

Tabla 4.4: (a) Probabilidades marginales (iniciales) de los nodos; (b)-(d) probabilidades condicionadas a la evidencia: $e_1 = \{Amieva = 3\}$, $e_2 = \{Santiago = 3\}$, y $e_3 = \{Amieva = 3, Santiago = 0\}$. La evidencia se indica en negrita y la probabilidades que más varían se han subrayado.

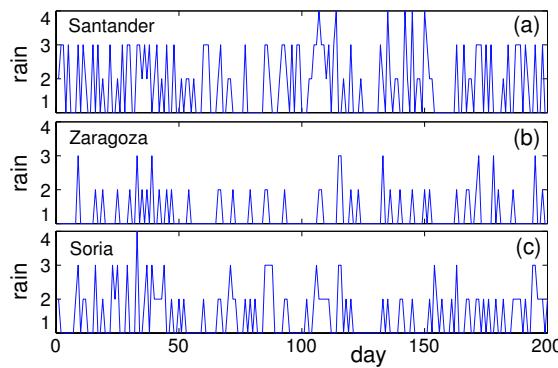


Figura 4.16: Simulación de series temporales de precipitación espacialmente distibuidas para 200 días aplicando el método de aceptación rechazo a la distribución marginal (el orden de simulación es el mismo que el establecido previamente en el algoritmo $K2$): (a) Santander, (b) Zaragoza, (c) Soria.

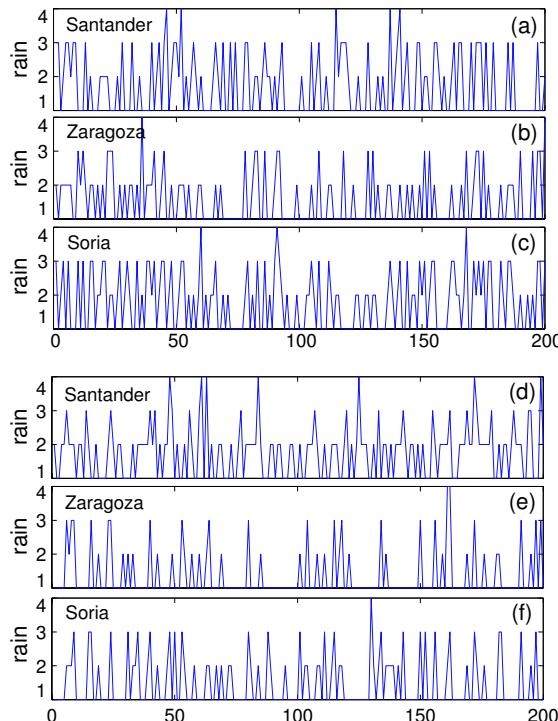


Figura 4.17: Simulación de series temporales de precipitación espacialmente distibuidas para 200 días considerando la evidencia $\text{Madrid}=4, \text{Coruña}=4$ (figuras (a)-(c)), y $\text{Madrid}=1, \text{Coruña}=1$ (d)-(f).

4.8.4. Predicción Probabilística Local (Downscaling)

Otro interesante problema donde aplicar una red Bayesiana es la *predicción meteorológica local* (p.e., precipitación diaria en la ciudad de Santander). La principal herramienta para la predicción meteorológica son las salidas de los ACMs, ya que estos simulan el futuro estado de la atmósfera sobre una rejilla 4D. Por desgracia, como ya se ha comentado, estos modelos globales tienen muy limitada su resolución y deben parametrizar todos los procesos físicos cuya escala es inferior a su resolución; que son precisamente procesos muy ligados al suelo y que tienen gran importancia en la fenomenología local. Las técnicas locales, como el *método de los análogos* (Lorenz, 1969) o vecinos más cercano (Nearest Neighbors), ofrecen una solución muy simple a este problema, ya que los modelos son entrenados localmente (en el sentido de similitud en los patrones atmosféricos), bajo la consideración de que patrones atmosféricos similares conducen a eventos meteorológicos similares. Así, una predicción local se obtiene a partir de un patrón previsto \mathbf{X}_t , buscando el conjunto de vecinos cercanos, o conjunto de análogos, en la base de datos de estados de la atmósfera según una métrica predefinida (la euclídea ha demostrado ser la más apropiada). Hecho esto, las observaciones correspondientes al conjunto de análogos son procesadas para obtener la predicción. Otra alternativa más directa se describe en Cofiño et al. (2002).

Una variante del método de análogos desarrollada recientemente (ver Gutiérrez et al., 2001, 2004, para más detalles), mejora los tiempos de cálculo y la calidad de las predicciones definiendo el conjunto de análogos mediante un algoritmo de agrupamiento sobre una discretización del estado de la atmósfera S . El predictando tendrá tantos estados diferentes como grupos ($S = c$ se refiere al subconjunto de estados de la atmósfera pertenecientes al grupo c). De esta manera, basta con asignar un grupo a cada nuevo estado de la atmósfera c_t , para obtener automáticamente la predicción probabilística local $P(Y_t = i)$ para la ocurrencia de $Y = i$ el instante t ; con otras palabras, es la probabilidad condicionada de que ocurra el fenómeno dado un estado de la atmósfera (o sea, dado el valor de la variable $S = c$):

$$P(y^k = i | \mathbf{x}_t) \approx P(y^k = i | s = s_t), \quad k = 1, \dots, 100, \quad i = 1, 2, 3, 4. \quad (4.7)$$

donde k son los puntos de observación considerados e i son los posibles valores de y (cuatro en el caso de la precipitación). Como se resaltó en Cano et al. (2002) estos modelos son redes Bayesianas ‘ingenuas’ que clasifican de manera simple sin considerar las dependencias espaciales entre las variables. Es decir, cuando se trabaja con series temporales multivariantes, la técnica de análogos asume implícitamente relaciones de independencia espacial y temporal que ignoran una parte, eventualmente muy importante, de la información potencial de la base de datos, ya que predice cada localidad independiente del resto. En su versión ingenua, una red Bayesiana es una generalización del método de análogos, ver Fig.4.18(a), ya que éste último calcula las probabilidades para un estado de la atmósfera. Por tanto la distribución conjunta de precipitación en

los puntos y_t^1, \dots, y_t^n dado $s = s_t$ se factoriza de acuerdo con

$$P(y_t^1, \dots, y_t^n | s = s_t) = \prod_{k=1}^n P(y_t^k | s = s_t).$$

Es decir, el cálculo de la probabilidad de cada estación sólo depende del estado de la atmósfera. Este problema se puede resolver considerando la red Bayesiana descrita en el Ejemplo 4.8.2 para representar la dependencia espacial entre estaciones, como se ilustra en la Fig.4.18(b):

$$P(y_t^1, \dots, y_t^n | s = s_t) = \prod_{k=1}^n P(y_t^k | \pi^k, s = s_t).$$

En este caso, las variables están acopladas de acuerdo con las dependencias espaciales relevantes de forma que el método de análogos queda como un caso particular de esta red. La Fig. 4.19 muestra la red Bayesiana resultante.

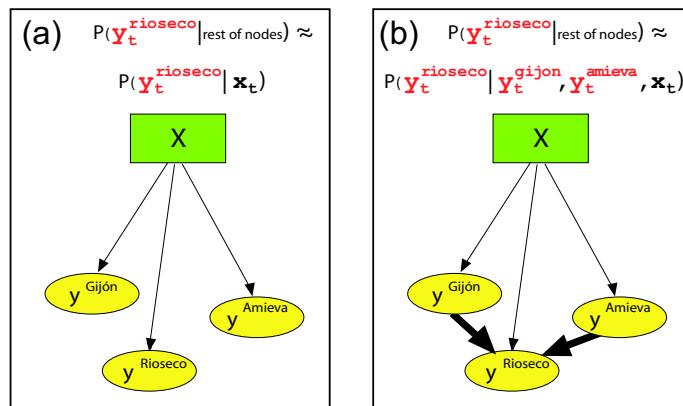


Figura 4.18: (a) Método de análogos y (b) método generalizado

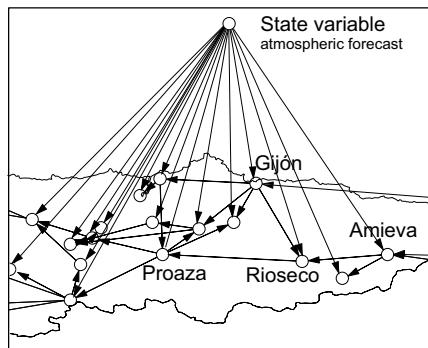


Figura 4.19: Añadiendo la variable estado de la atmósfera al grafo.

En la Fig. 4.20 se muestra una comparativa entre el método clásico, es decir el que no contempla dependencia espacial respecto a la climatología: (a) 2mm, (b) 10mm, y (c) 20mm durante el periodo DJF de 1999, obteniéndose valores positivos del BSS, con medias de 0.31, 0.22 y 0.12, respectivamente.

Por otra parte, en la Fig. 4.21 se compara el método generalizado respecto del clásico obteniéndose una apreciable mejora con valores medios de BSS de 0.06, 0.11, y 0.19 para los umbrales (a) 2mm, (b) 10mm, y (c) 20mm respectivamente. Se puede ver que la pericia del método generalizado aumenta con el umbral de precipitación lo cual supone que la dependencia espacial cobra especial relevancia cuando se trata de eventos raros debido fundamentalmente al fuerte sesgo climático del método clásico, es decir, el método clásico suaviza más los resultados.

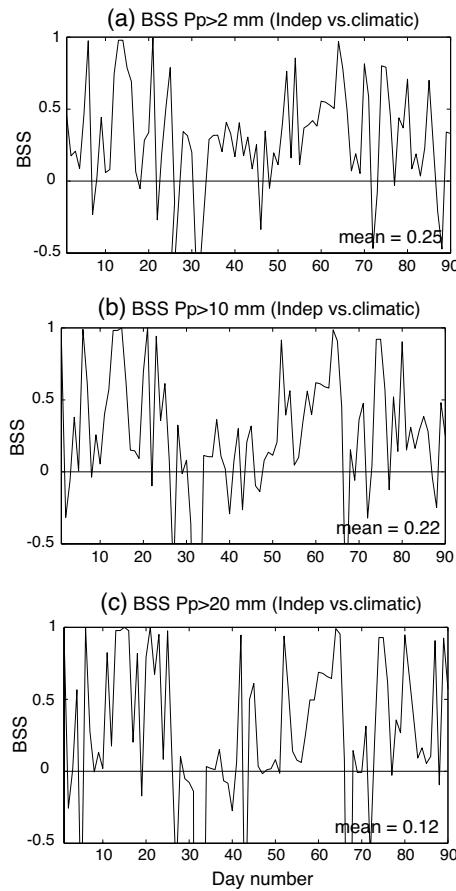


Figura 4.20: Validación utilizando el BSS (Brier Skill Score) del método clásico (BN independiente) respecto a la climatología para los umbrales de precipitación (a) 2mm, (b) 10mm, y (c) 20mm. El periodo de validación es DJF 1999 (90 días).

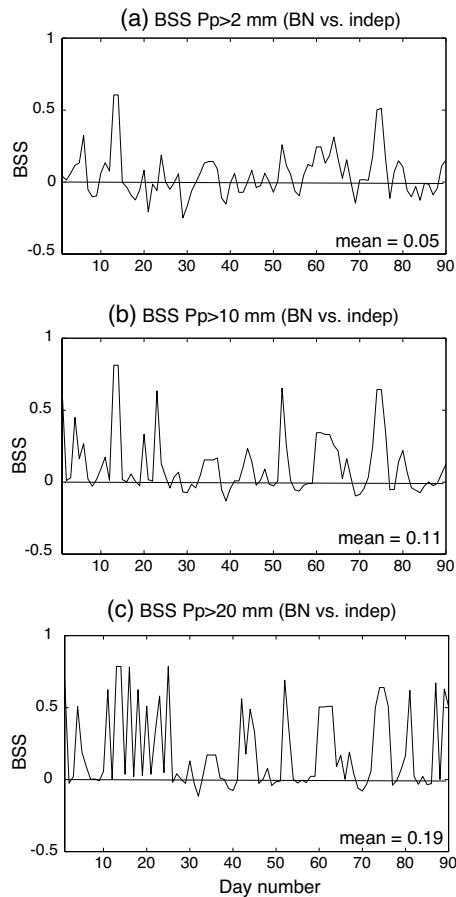


Figura 4.21: Validación utilizando el BSS (Brier Skill Score) del método generalizado (BN con dependencia espacial) respecto al clásico para los umbrales de precipitación (a) 2mm, (b) 10mm, y (c) 20mm. El periodo de validación es DJF 1999 (90 días).

4.8.5. Aplicación al Relleno de Lagunas

Un problema que se afronta en la práctica al trabajar con series temporales de observaciones es que es muy común encontrar defectos como la existencia de lagunas (falta de datos) y/o inhomogeneidades (derivadas de cambios en los parámetros de observación). Por una parte, las lagunas dificultan la aplicación de las distintas técnicas estadísticas (en la mayoría de los casos obligan a utilizar algoritmos iterativos costosos, del tipo de Esperanza-Maximización, EM). Por otra parte, las inhomogeneidades provienen de series no estacionarias que, en realidad, no representan a un único “modelo”. Por todo ello, es muy útil disponer de series de referencia completas y homogéneas que permitan llevar a cabo estudios climatológicos de forma simple y consistente. Los dos problemas anteriores, aunque distintos en esencia, comparten muchas características comunes. Por ejemplo, un método de homogeneización marcará las pautas para el desarrollo de métodos de relleno de lagunas que preserven la homogeneidad elegida. En la literatura, el interés por estos problemas se ha centrado casi exclusivamente en la escala mensual y han sido numerosos los métodos de homogeneización Alexandersson (1986) y relleno de lagunas (Stepanek, 2001) propuestos. Por ejemplo, las medias mensuales de precipitación pueden considerarse aproximadamente normales y, además, se dan correlaciones significativas entre estaciones a distancias de cientos de kilómetros.

En este ejemplo se tratan estos problemas a escala diaria, lo cual dificulta el análisis ya que no se cumple la hipótesis de normalidad de los datos, pues éstos requieren una distribución apropiada (gamma, doble exponencial, etc.) y además, las correlaciones entre estaciones cercanas son más débiles, con un alcance significativo del orden de decenas de kilómetros. Sin embargo, si se dispone de una red de observatorios suficientemente densa (como la red pluviométrica del INM, con más de 10.000 observatorios), se tienen correlaciones muy significativas entre estaciones cercanas; además, se puede seguir trabajando con distribuciones normales, considerando distribuciones truncadas. A partir de este hecho, se propone un nuevo método de relleno de lagunas basado en una red Bayesiana gaussiana de todas las estaciones que permite obtener la esperanza de la precipitación para un día dado en cualquier estación a partir de cualquier subconjunto de las observaciones disponibles. Para ello, se considera la precipitación acumulada en 24 horas entre las 07Z y 07Z. Se opera considerando como unidad las cuencas parciales, o subcuenca, que contienen entre 5 y 50 estaciones.

Desde un punto de vista estadístico, el problema de relleno de lagunas a escala diaria es muy similar al problema de la predicción, con la salvedad de que en el primer problema se dispone de más información que en el segundo. Para ilustrar estos problemas, considérese el valor diario acumulado de la precipitación, $X = \{x_1, \dots, x_n\}$, en un conjunto dado de estaciones. Ambos problemas tratan de estimar el valor de la precipitación $x_k(T)$ en una estación k y en un instante de tiempo T . Para ello, en un problema de predicción se dispondría como única información, del valor de la precipitación en tiempos anteriores $t < T$ en todas las estaciones (en este caso se supone que no hay lagunas). En cam-

bio, en el caso del relleno de lagunas, se dispondría del valor de la precipitación $x_i(T), i \in D(T)$, en el subconjunto de estaciones $D(T)$ con dato para el día T , así como los valores en fechas anteriores, $x_i(t), i \in D(t), t < T$. Este aspecto de disponibilidad o no de la información “circundante” resulta de capital importancia para realizar la estimación de la laguna, ya que este problema puede ser extremadamente difícil en el caso de parámetros y localidades con gran variación espacial y sin información cercana. Existen numerosos métodos de relleno de lagunas en series temporales; sin embargo, la mayoría de ellos utilizan de una u otra forma la hipótesis de normalidad y, por tanto, se aplican preferentemente en datos mensuales. El más popular de los métodos de relleno es la regresión multivariada. Sin embargo, cuando las lagunas están repartidas por la serie de forma inhomogénea (es decir, los conjuntos $D(t)$ presentan mucha variabilidad) sería necesario estimar toda una batería de modelos de regresión para cada estación, a fin de poder estimar el valor en el tiempo T a partir de cualquiera que fuese $D(T)$. Cuando se trabaja con una red de estaciones muy densa, este es un esfuerzo que resulta inabordable. En la siguiente sección se propone un nuevo método que utiliza un único modelo conjunto de todas las variables para llenar las lagunas (obsérvese que este hecho garantiza ya una cierta coherencia en el relleno de las lagunas).

El principal inconveniente de trabajar con datos diarios de precipitación es la carencia de normalidad e incluso de correlación significativa entre los datos. Para constatar este hecho se han calculado las correlaciones entre el observatorio de Santander y los observatorios de la red secundaria circundantes, para precipitación y temperatura. La Fig. 4.22 muestra que el alcance de la correlación para un umbral del 95 % es de 20 km para la precipitación y de 60 km para la temperatura máxima. Este resultado es válido para la banda litoral del Cantábrico oriental donde el tipo de precipitación típico es de escala frontal y las temperaturas en la franja litoral son muy uniformes; en otras zonas esto es diferente, aunque se ha constatado un comportamiento similar. Los alcances más pequeños corresponden a observatorios aislados y/o con precipitaciones típicamente convectivas para los cuales la metodología propuesta es menos adecuada.

Por tanto, dada la gran densidad de observatorios disponibles se propone, para cada subcuenca, aprender una red Gaussiana para modelizar la precipitación, donde la probabilidad asociada al estado de precipitación nula (único valor con probabilidad no nula) se supone distribuido en una cola izquierda ficticia de la densidad de probabilidad. Es decir, la densidad de probabilidad de la precipitación se aproxima por una distribución Gaussiana truncada (Glasbey and Nevison, 1997). En este ejemplo, dado que el número de estaciones en cada subcuenca no es excesivo, se considera una red Bayesianas global (es decir, un modelo Gaussiano multivariado para definir los datos). Una vez aprendidos los parámetros (medias y covarianzas) del modelo, interesa obtener las distribuciones condicionadas $N(X \setminus D(t)|D(t))$ para cada día t (donde $X \setminus D(t)$ es el conjunto diferencia en X y $D(t)$). De esta forma, se podrá estimar el valor de la precipitación en las estaciones con lagunas $X \setminus D(t)$, a partir de la

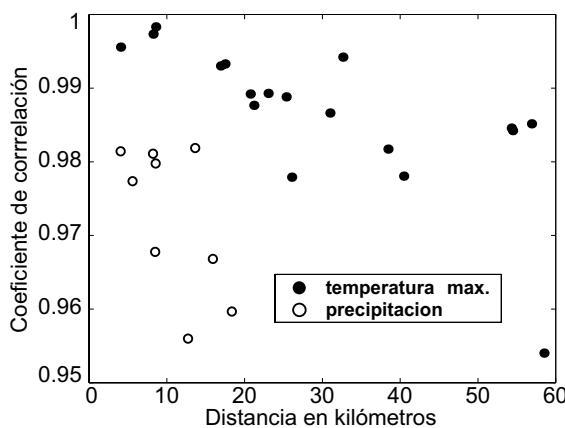


Figura 4.22: Correlación entre el observatorio de Santander y los observatorios circundantes en función de la distancia, para precipitación y temperatura.

información existente para esa misma fecha (por ejemplo, se puede considerar la media de la distribución condicionada de cada estación). Supongamos que un día dado queremos calcular la función de probabilidad de $Y = X \setminus D(t)$ condicionado al valor conocido de $Z = D(t) = z$. Entonces, dado que la distribución conjunta es normal, la probabilidad de Y condicionada a Z también es normal y su media y matriz de covarianzas vienen dadas por (para más detalles ver Castillo et al., 1997, Cap. 6):

$$\mu_{Y|Z=z} = \mu_Y + \Sigma_{YZ}\Sigma_{ZZ}^{-1}(z - \mu_Z), \quad (4.8)$$

$$\Sigma_{Y|Z=z} = \Sigma_{YY} - \Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZY}. \quad (4.9)$$

El algoritmo de relleno de lagunas está basado en estas fórmulas. Aplicado al dato diario de la red pluviométrica del INM, para la serie 1950-1999, consta de los siguientes pasos:

1. Se divide la red pluviométrica en subcuencas hidrográficas; de ésta manera solamente se relacionarán series pertenecientes a la misma subcuenca hidrográfica, independientemente de su distancia.
2. Se determina un umbral para limitar el número máximo de lagunas de cada serie; por ejemplo, seleccionando series con más de 12000 observaciones válidas en el periodo 1950-1999; es decir, si hay más de un 35% de lagunas, la serie se rechaza.
3. Se calcula la matriz de covarianzas a partir de la muestra, ignorando las lagunas.
4. Para cada fecha se extrae la submatriz de covarianza de las series sin laguna (que actuarán como predictores) y el correspondiente vector de

covarianzas cruzadas para generar la matriz de regresión que servirá para estimar simultáneamente las lagunas del día.

Con éste método, las lagunas son inferidas utilizando todas las observaciones válidas en observatorios circundantes. El método es equivalente a una regresión múltiple, aunque es más eficiente ya que contiene todas las relaciones de dependencia lineal entre todas las variables, y por tanto todas las posibles regresiones múltiples que se pueden plantear sobre un conjunto de variables. Esta ventaja cobra especial importancia en el caso del problema del relleno de lagunas donde las variables dependientes y las independientes cambian en función de la disponibilidad observacional de la base de datos, exigiendo un tipo diferente de regresión para cada caso. Se han hecho pruebas de relleno de lagunas en series de precipitación utilizando este método con éxito en áreas donde la proximidad entre estaciones es tan grande que la linealización de la precipitación es, a escala local, una buena aproximación (ver Fig. 4.23).

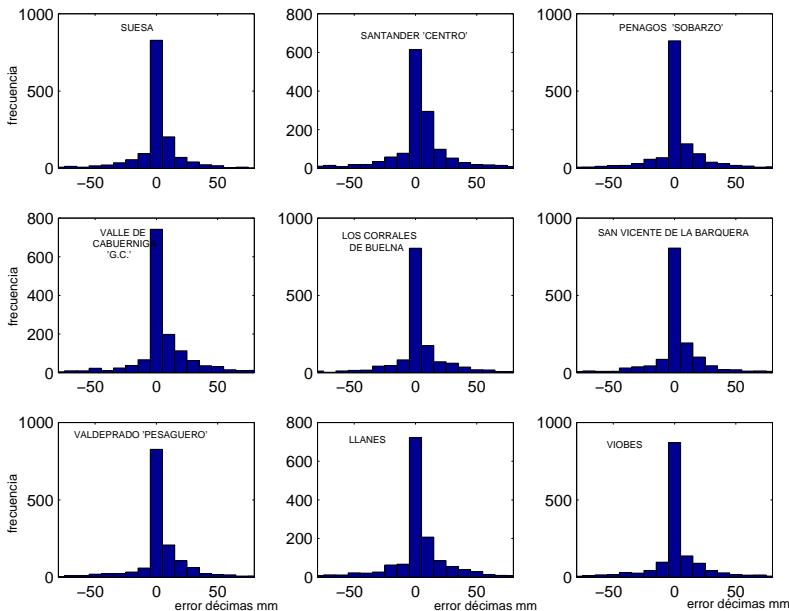


Figura 4.23: Distribución de los errores en el relleno de lagunas de la precipitación diaria para varias estaciones de la Cuenca Norte aplicando el método de relleno de lagunas descrito. El conjunto de test consta de 1500 días tomados al azar de la serie diaria 1950 – 1999.

La Fig. 4.24 muestra la serie original y la obtenida aplicando el algoritmo de relleno a la estación de Santander.

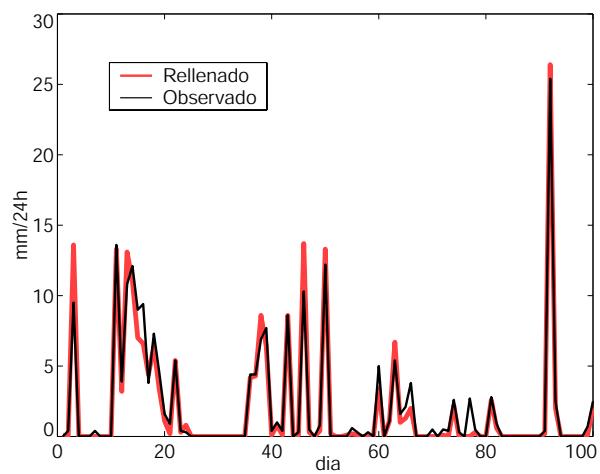


Figura 4.24: Relleno de lagunas de precipitación para Santander. Se han tomado como conjunto de test, los primeros 100 días de 1979.

CAPÍTULO 5

Redes Neuronales

5.1. *Introducción*

La computación paralela y las redes neuronales son dos nuevos paradigmas que han despertado un gran interés en los últimos años. La novedad de estos modelos es una nueva estructura computacional compuesta de un gran número de elementos de cómputo simples (neuronas) interconectados en paralelo. Estos modelos se inspiran en la Biología, en contraposición a los sistemas inteligentes tradicionales basados en reglas, probabilidad, etc. (ver Sec. 4.2), por lo que su forma de representar el conocimiento es abstracta (en base a la arquitectura de la red y las conexiones que se establecen entre las neuronas). Estos modelos se construyen por medio de algoritmos automáticos de aprendizaje, que ajustan las conexiones entre neuronas para modelizar un conjunto de datos dado. Las redes neuronales son métodos de tipo “caja negra”, pues el modelo resultante es difícil de interpretar a partir de los datos. Una descripción detallada de este campo se puede consultar en Bishop (1995); Haykin (1998).

Las redes neuronales han probado su valía para resolver problemas complejos en diversas áreas, incluyendo en la predicción meteorológica y oceánica (Schizas et al., 1994; Hsieh and Tang, 1998). En los últimos años se han desarrollado diversas extensiones de estos modelos para cubrir deficiencias en los mismos, y especializarlos en problemas concretos; por ejemplo, las redes funcionales permiten incluir conocimiento cualitativo del problema en la estructura de la red (ver Castillo et al., 1999, para más detalles); también se han aplicado técnicas modernas de optimización, como los algoritmos genéticos, para buscar una estructura más conveniente de la red para un problema dado (Cofiño et al., 2004). Actualmente las redes neuronales son un campo muy consolidado en el que continuamente se desarrollan nuevos métodos y aplicaciones.

En este capítulo se pretende mostrar una visión práctica de los modelos más populares y los problemas que permiten resolver. En la Sec. 5.2 se describen distintas estructuras de red neuronal, así como el funcionamiento de una neurona. En la Sec. 5.3 se presentan las principales metodologías de aprendizaje de redes (supervisado y no supervisado) y se analizan las particularidades de tres arquitecturas. A continuación se describen varios modelos concretos: Los perceptrones (Sec. 5.4), las redes multicapa (Sec. 5.5), las redes de base radial (Sec. 5.6), y las redes competitivas (Sec. 5.7), incluyendo las redes autoorganizativas SOM (Sec. 5.8). Por último, la Sec. 5.9 ilustra la aplicación de estas técnicas en el análisis e interpretación de la predicción por conjuntos (EPS).

5.2. Estructura de una Red Neuronal

En analogía a la estructura del cerebro de un ser vivo, los modelos computacionales de redes neuronales están compuestos por un gran número de unidades simples de proceso (neuronas) conectadas entre sí en base a una topología definida. Normalmente, las neuronas se agrupan en capas, comenzando con una capa de entrada, para adquirir los datos, y terminando con una capa de salida, para devolver los resultados. Cada topología, o arquitectura, responde a un tipo de problema concreto que se quiere resolver.

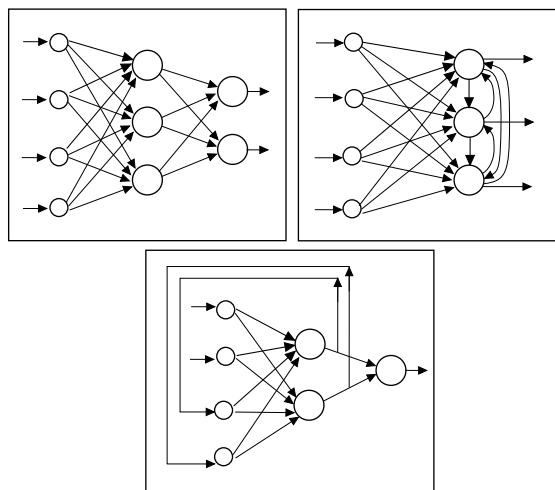


Figura 5.1: Tres topologías distintas de red neuronal: (a) red neuronal multicapa, (b) red competitiva, y (c) red recurrente.

Por ejemplo, la Fig. 5.1 muestra tres modelos de red neuronal distintos. La primera de ellas se denomina *red multicapa* y sólo tiene conexiones entre neuronas de capas consecutivas. La red de la Fig. 5.1(b) es una *red competitiva* y, además de conexiones entre neuronas de capas consecutivas, también posee conexiones entre las neuronas de la última capa. Por último, la red de la Fig. 5.1(c) es una *red recurrente* y posee conexiones entre capas no consecutivas (en

el ejemplo mostrado, algunas neuronas de entrada toman su valor retrasado de las neuronas de salida).

Una vez elegida una topología de conexión, la funcionalidad de la red viene dada por la función concreta que realiza cada neurona (actividad neuronal), y por los pesos de conexión de unas neuronas con otras. La topología de conexión y la actividad neuronal definen el tipo de red neuronal concreta, y los pesos de las conexiones son los parámetros que, ajustados al conjunto de datos que define un problema concreto, permiten a la red “aprender” y generalizar el conocimiento aprendido. En este sentido, las redes neuronales son sistemas inteligentes (ver Sec. 4.2) cuya base de conocimiento es la propia estructura de la red y los pesos aprendidos (conocimiento abstracto) y cuyo módulo de razonamiento consiste en propagar entradas por las distintas capas de la red, hasta la capa de salida.

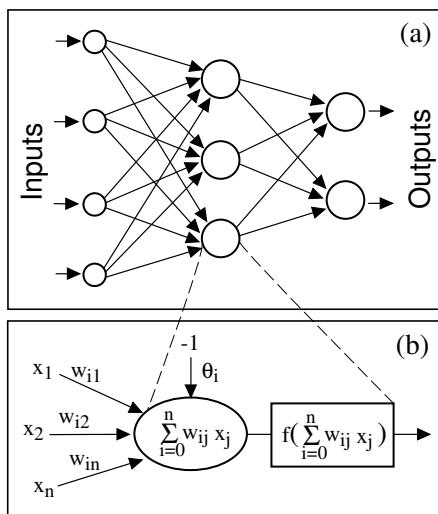


Figura 5.2: (a) Red neuronal multicapa y (b) la función procesadora de una única neurona con sus dos módulos: actividad lineal, y función de activación.

La Figura 5.2(b) muestra un esquema de la función de actividad neuronal típica, que también está inspirado en la Biología: “cuando el estímulo total que recibe una neurona supera un cierto umbral, entonces la neurona emite un impulso; en caso contrario, permanece en reposo”. Esta idea se implementa en la práctica calculando la suma ponderada de las entradas que recibe la neurona (actividad lineal), y filtrándola posteriormente con una función de activación sigmoidal de tipo umbral (disparo/no disparo) para obtener la salida:

$$y_i = f\left(\sum_{j=1}^n w_{ij} x_j - \theta_i\right) = f\left(\sum_{j=0}^n w_{ij} x_j\right), \quad (5.1)$$

donde $f(x)$ es la función activación y θ_i es el umbral de activación de la neurona. Obsérvese que el umbral de activación se puede incluir en el sumatorio

considerando una nueva neurona auxiliar $x_0 = -1$ conectada a y_i con un peso $w_{i0} = \theta_i$. Por tanto, la salida de una neurona y_i se obtiene simplemente transformando la suma ponderada de las entradas que recibe usando la función de activación (ver Fig. 5.2(b)). Las funciones de activación continuas más populares son:

- *Funciones sigmoidales:* Son funciones monótonas acotadas que dan una salida gradual no lineal para las entradas. Las funciones sigmoidales más populares son:

1. La función logística, con rango de 0 a 1 (ver Figura 5.3):

$$f_c(x) = \frac{1}{1 + e^{-cx}}.$$

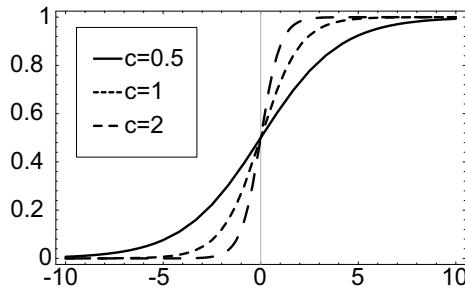


Figura 5.3: Función de activación sigmoidal logística $f_c(x) = (1 + e^{-cx})^{-1}$.

2. La función tangente hiperbólica, con rango de -1 a 1 (similar a la función logística, pero con valores positivos y negativos):

$$f_c(x) = \tanh(cx).$$

- *Funciones núcleo:* Localizadas alrededor de un punto, como la distribución Gaussiana.
- *Funciones lineales:* Dan una salida lineal con un rango infinito:

$$f(x) = x; x \in R.$$

5.3. Aprendizaje y Validación

Una de las principales propiedades de las redes neuronales es su capacidad de aprender a partir de datos. Una vez que ha sido elegida la arquitectura de red para un problema particular, los pesos de las conexiones se ajustan para codificar la información contenida en un conjunto de *datos de entrenamiento*.

Las redes multicapa y las redes recurrentes son apropiadas para problemas de *aprendizaje supervisado*, donde se dispone de un conjunto de patrones de

entrenamiento de la forma $(\mathbf{x}_p, \mathbf{y}_p) = (x_{1p}, \dots, x_{mp}; y_{1p}, \dots, y_{np})$, $p = 1, \dots, a$, donde cada patrón de entrada \mathbf{x}_p tiene asociado el correspondiente patrón de salida \mathbf{y}_p . El entrenamiento está basado en que la red sea capaz de reproducir estos patrones con el menor error posible. Este esquema responde a problemas de clasificación y aproximación o ajuste de datos, y desde un punto de vista estadístico se puede considerar como una generalización de la regresión no lineal, donde el modelo es la propia actividad de la red completa y los parámetros a ajustar son los pesos de las conexiones. Como se verá más adelante, el problema de aprender este tipo de redes se reduce a obtener los pesos apropiados utilizando algún algoritmo de aprendizaje apropiado.

Por otra parte, las redes competitivas responden a problemas de *aprendizaje no supervisado*, donde se dispone de un conjunto de patrones de entrenamiento de la forma $\mathbf{x}_p = (x_{1p}, \dots, x_{mp})$, $p = 1, \dots, a$. En este caso, no se conoce el valor de salida de cada patrón de entrada \mathbf{x}_p y la red tiene que auto-organizarse para hallar las salidas óptimas para cada entrada (en base a algún criterio de aprendizaje que corresponda a maximizar la entropía de la salida, etc.). Normalmente este tipo de redes se utilizan en problemas de agrupamiento, donde cada una de las neuronas de salida representa un grupo distinto, y la red tiene que distribuir los patrones de entrenamiento entre los grupos de forma automática.

Una vez terminado el proceso de aprendizaje y calculados los pesos de la red neuronal, es importante comprobar la calidad del modelo resultante. Por ejemplo, en el caso de aprendizaje supervisado, una medida de la calidad puede darse en términos de los errores entre los valores de salida deseados y los obtenidos por la red neuronal. Algunas medidas estándar del error son (ver Cap. 7 para más detalles):

1. La suma de los cuadrados de los errores (Sum of Square Errors, SSE), definida como

$$\sum_{p=1}^a \| y_p - \hat{y}_p \|^2 . \quad (5.2)$$

2. La raíz cuadrada del error cuadrático medio (Root Mean Square Error, RMSE) definida como

$$\sqrt{\sum_{p=1}^a \| y_p - \hat{y}_p \|^2 / n} . \quad (5.3)$$

3. El error máximo,

$$\max_{p=1, \dots, a} \| y_p - \hat{y}_p \| . \quad (5.4)$$

También es deseable realizar una *validación cruzada* para obtener una medida de la capacidad de generalización del modelo; es decir, la capacidad de reproducir nuevos patrones del problema, no utilizados en el proceso de aprendizaje. Con este propósito, los datos disponibles se pueden dividir en dos partes:

una parte destinada al entrenamiento de la red y otra parte a la validación. Cuando el error de validación es significativamente mayor que el error de entrenamiento, entonces se produce un *problema de sobreajuste* durante el proceso de entrenamiento que puede ser debido a un excesivo número de parámetros. Por tanto, es necesario mantener un compromiso entre la complejidad del modelo, y el número de datos disponible, de forma que el modelo sea más simple que el conjunto de datos que representa. En caso contrario, el modelo sólo memoriza la información no siendo capaz de generalizar lo aprendido a nuevas situaciones.

5.4. Perceptrones (Redes de una Capa)

Los perceptrones (ver Rosenblat, 1962) son las arquitecturas más simples y consisten en una capa de entrada, $\{x_1, \dots, x_m\}$, y una de salida, $\{y_1, \dots, y_n\}$, de forma que las neuronas de la capa de salida están conectadas con las de entrada y no hay conexiones entre las neuronas de una misma capa (ver Fig. 5.4). Para indicar el número de entradas y de salidas, este tipo de redes suelen denominarse de forma abreviada mediante $m : n$, siendo m y n el número de neuronas de entrada y salida respectivamente (obsérvese que las neuronas de entrada no realizan ningún cálculo y, por tanto, no contabilizan). Los perceptrones suelen denominarse también redes de retro-propagación de una única capa (one layer feedforward network).

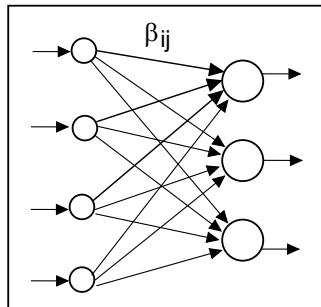


Figura 5.4: Estructura de un Percetrón.

En un perceptrón, las neuronas de salida típicas, realizan el cálculo:

$$\hat{y}_j = f\left(\sum_{i=0}^m \beta_{ji} x_i\right) = f(\boldsymbol{\beta}_j^T \mathbf{x}), \quad j = 1, \dots, n. \quad (5.5)$$

donde $f(\cdot)$ es la función de activación y $\boldsymbol{\beta}_j$ el correspondiente vector de pesos. \hat{y}_p denota el valor de salida obtenido insertando el correspondiente patrón de entrada en la red (5.5), mientras y_p denota el valor real de salida conocido.

5.4.1. Algoritmo de Aprendizaje. La Regla Delta

En los algoritmos de aprendizaje para este tipo de redes, se usan métodos de optimización matemática para obtener los pesos $\boldsymbol{\beta}_j$ que minimizan una

cierta función de error. Obsérvese que los pesos son los únicos parámetros desconocidos de la red y son los que proporcionan flexibilidad a la misma para ajustarse a distintas situaciones caracterizadas por un conjunto de patrones entrada-salida.

Los algoritmos de aprendizaje más populares se basan en minimizar la suma de los cuadrados de los errores (otras medidas distintas de error han dado lugar a algoritmos de aprendizaje diferentes que se comentan más adelante):

$$E(\boldsymbol{\beta}) = \sum_{j,p} (y_{jp} - \hat{y}_{jp})^2 = \sum_p \|\mathbf{y}_p - \hat{\mathbf{y}}_p\| \quad (5.6)$$

$$= \sum_{j,p} (y_{jp} - f(\sum_i \beta_{ji} x_{ip}))^2 = \sum_p \|\mathbf{y}_p - f(\boldsymbol{\beta}^T \mathbf{x}_p)\| \quad (5.7)$$

Dado que esta función es no lineal, no existe ningún método exacto para obtener su solución (los pesos óptimos), aunque recientemente Castillo et al. (2002) han presentado un nuevo método de aprendizaje para este tipo de redes que transforma la función de error haciéndola lineal en los parámetros del modelo (los pesos).

Uno de los algoritmos de optimización más simples para este problema es el método del descenso de gradiente (también llamado “regla delta” en este caso). Se trata de un algoritmo iterativo que en cada etapa trata de modificar incrementalmente los pesos, de forma que disminuya el error (inicialmente se toma un valor aleatorio para los pesos). En este caso concreto, el incremento de los pesos se obtiene en base a los vectores en los que la función de error decrece más rápidamente, que corresponde al opuesto del gradiente de la función de error respecto de los pesos, $-\nabla E$ (método del descenso de gradiente). Por tanto, en cada paso de iteración cada uno de los pesos β_{ji} se modifica mediante un incremento $\Delta\beta_{ji}$ proporcional al gradiente del error:

$$\begin{aligned} \Delta\beta_{ji} &= -\eta \frac{\partial E(\boldsymbol{\beta})}{\partial \beta_{ji}} = -\eta \sum_p (y_{jp} - \hat{y}_{jp}) \frac{\partial \hat{y}_{jp}}{\partial \beta_{ji}} \\ &= -\eta \sum_p (y_{jp} - \hat{y}_{jp}) f'(\sum_i \beta_{ji} x_{ip}) x_{jp}, \end{aligned} \quad (5.8)$$

$$\Delta\boldsymbol{\beta}_j = -\eta \nabla E(\boldsymbol{\beta}) = -\eta \sum_p (y_{jp} - \hat{y}_{jp}) f'(\boldsymbol{\beta}_j^T \mathbf{x}_p) x_{jp}, \quad (5.9)$$

donde el parámetro η es la *tasa de aprendizaje*, es decir, la constante que regula la intensidad de la variación incremental de los pesos (obsérvese que la aproximación de la superficie de error mediante el gradiente es sólo válida en un sentido local y, por tanto, el rango de la tasa de aprendizaje está limitado por este hecho).

Algunas funciones de activación permiten definir su derivada en función de sí mismas, simplificando la fórmula (5.8) al no involucrar derivadas formales:

$$f(s) = \frac{1}{1 + e^{-cs}} \Rightarrow f'(s) = c f(s) (1 - f(s)),$$

$$f(s) = \tanh(cs) \Rightarrow f'(s) = c(1 - f(s)^2).$$

Si la función de activación fuese lineal ($f(s) = s$), las salidas de la red dadas en (5.5) se reducirían a una combinación lineal de las entradas, resultando los siguientes incrementos de los pesos:

$$\Delta\beta_{ji} = -\eta \frac{\partial E}{\partial \beta_{ji}} = \eta \sum_p (y_{jp} - \hat{y}_{jp}) x_{ip}, \quad (5.10)$$

que proporcionan la solución global del problema lineal. Obsérvese que este problema puede resolverse en un sólo paso con técnicas de optimización lineal, no siendo necesario el uso de una técnica iterativa. De hecho, el modelo resultante es equivalente a un modelo de regresión lineal, donde los parámetros son los pesos de la red.

5.4.2. Mejoras y Modificaciones

Han sido varias las modificaciones propuestas en la literatura con el propósito de mejorar la eficiencia del método de aprendizaje anterior. A continuación se describen las más populares:

- *Término de inercia.* Este término extra se introduce en la expresión de $\Delta\beta_{ji}$ para acelerar la convergencia teniendo en cuenta no sólo el gradiente local, sino las distintas tendencias en la “superficie” de error. Con ello se evita que la red caiga en pequeños mínimos locales. La nueva regla de actualización viene dada por

$$\Delta\beta_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} + \mu \Delta'\beta_{ji},$$

donde $\Delta'\beta_{ji}$ hace referencia a los valores previos de $\Delta\beta_{ji}$ (en el paso de iteración previo) y μ es el *parámetro de inercia*.

- *Tasa de aprendizaje variable.* En el método de descenso de gradiente, la tasa de aprendizaje es constante en todo el proceso de entrenamiento, pero la eficiencia del algoritmo es muy sensible a este valor; si la tasa es muy grande, el algoritmo oscila y se vuelve inestable, y si es demasiado pequeña, entonces la convergencia es muy lenta. Tampoco es posible saber cuál es el valor más conveniente de la tasa antes de comenzar el entrenamiento, e incluso puede cambiar durante el proceso de aprendizaje. Una tasa de aprendizaje adaptativa varía durante el entrenamiento según la complejidad local de la superficie de error.
- *Métodos de regularización.* Estos métodos incluyen términos en la función de error para penalizar pesos grandes:

$$E(\boldsymbol{\beta}) = \sum_{p=1}^a (\mathbf{y}_p - \hat{\mathbf{y}}_p)^2 + \lambda \sum_{i,j} \beta_{ji}^2, \quad (5.11)$$

donde λ es un parámetro de regularización, que controla el equilibrio entre el modelo ajustado y la penalización. El efecto de esta regularización de los pesos es suavizar la función de error, ya que los pesos grandes están usualmente asociados a valores de salida altos (para una descripción más detallada desde un punto de vista estadístico ver Hoerl and Kennard, 1970). Esta técnica también está relacionada con el método de *descomposición de los pesos* que consiste en recortar las conexiones de la red que tengan poca importancia (pesos muy bajos).

Otros métodos de aprendizaje más eficientes se describen en Haykin (1998).

Ejemplo 5.1 (Clasificación con Perceptrones). *En este ejemplo se ilustra la aplicación del perceptrón en problemas de clasificación. Concretamente se analiza un problema que consiste en la clasificación de la mezcla de dos muestras obtenidas de distribuciones gaussianas con desviación 0.5 y medias 0.5 y -0.5, respectivamente (ver Fig. 5.5). Para una primera aproximación al problema (clasificación lineal), se puede considerar una función de activación lineal $f(x) = x$. El óptimo de la red neuronal se muestra en la Fig. 5.6 y, en este caso, es equivalente al óptimo que se obtendría directamente por un método de regresión lineal.*

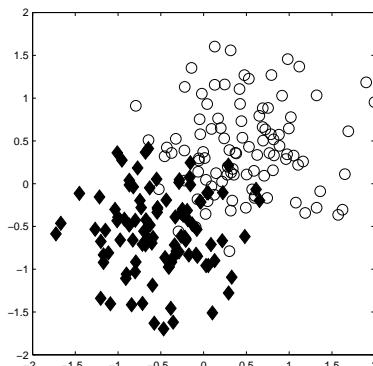


Figura 5.5: Muestra de 100 puntos obtenida de la mezcla de dos distribuciones gaussianas de desviación 0.5 y media 0.5 y -0.5

Para obtener un criterio de separación no lineal se puede considerar un perceptrón con una función de activación sigmoidal; en este ejemplo se utiliza la función logística. El óptimo obtenido se muestra en la Fig. 5.7. Obsérvese que en este último ejemplo la red es equivalente a un modelo de regresión o clasificación logística.

A continuación se muestra la forma de entrenar una red neuronal con MeteoLab, utilizando la Toolbox de redes neuronales de Matlab con el código descrito en MeteoLab/NeuralNets/ (en MeteoLab/NeuralNets/netlab también se incluyen los códigos para realizar los mismos ejemplos utilizando la Toolbox de software libre NetLab www.ncrg.aston.ac.uk/netlab/).

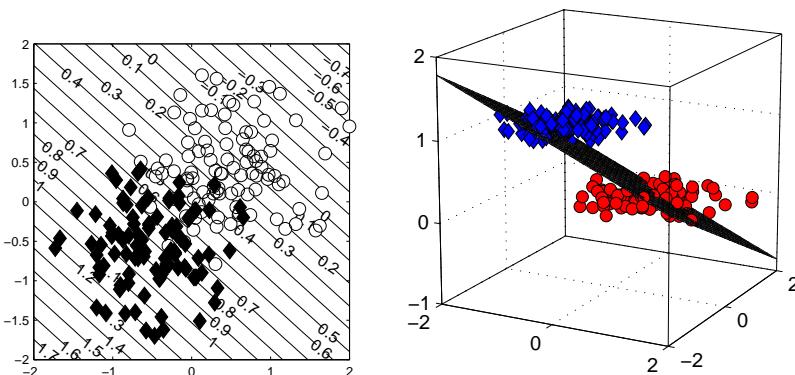


Figura 5.6: Superficie generada por el perceptrón lineal: Hiperplano de separación óptimo, en el sentido de mínimos cuadrados

```

load ejemplo5.1.mat %load prepared data

nnetL=newff([[-2 2];[-2 2]],1',{'purelin'});
nnetL=train(nnetL,X,Y);

nnetS=newff([[-2 2];[-2 2]],1',{'logsig'});
nnetS=train(nnetS,X,Y);

[A,B]=meshgrid(linspace(-2,2,25),linspace(-2,2,25));
SL=reshape(sim(nnetL,[A(:),B(:)]'),size(A));
SS=reshape(sim(nnetS,[A(:),B(:)]'),size(A));

figure
pcolor(A,B,SL) % For the lineal net
pcolor(A,B,SS) % For the sigmoidal net
colorbar

```

En el ejemplo anterior se definen dos redes utilizando el comando `newff`, que crea una nueva red multicapa (feed-forward). Por ejemplo, la instrucción `newff([[-2 2];[-2 2]],1,'purelin')` indica que se quiere crear una red con dos entradas (ambas con rango $[-2, 2]$) y una capa de salida, que tiene función de activación lineal. Las dos entradas corresponden a las coordenadas x e y de los puntos de entrenamiento, mientras que la salida es cero (para un grupo) y uno (para el otro). Una vez definida la red, el comando `nnetL=train(nnetL,X,Y)` lleva a cabo el entrenamiento de la red, un número predefinido de pasos (cien por defecto). ■

Ejemplo 5.2 (Clasificación de Patrones de Precipitación). En este ejemplo se aplican las mismas técnicas que en el ejemplo anterior para clasificar patrones atmosféricos en base a la precipitación local asociada en un cierta

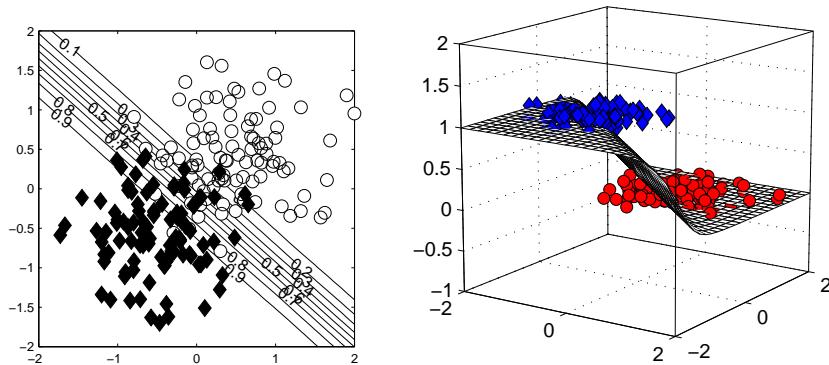


Figura 5.7: Superficie generada por el perceptrón sigmoidal no lineal.

estación (la ciudad de Santander). Para ello se ha considerado el patrón atmosférico en la zona Ibérica definido en la Sección 1.9.4 (ver Fig. 1.22(b)). En la Fig. 5.8 se muestran dos gráficas correspondientes a distintas combinaciones de componentes principales de estos patrones. Se han dibujado con símbolos distintos aquellos patrones que corresponden a eventos de precipitación ($Precip > 0.5mm$) en la localidad de Santander. En la primera figura puede observarse que el poder de discriminación de las dos primeras CPs es muy bajo, mientras que la segunda figura muestra cómo la tercera CP aporta información que permite discriminar parcialmente ambas categorías. Dado que las CPs siguen distribuciones aproximadamente normales, el parecido de este problema con el anteriormente expuesto es notable; sin embargo, en este caso las clases están muy mezcladas y la capacidad de discriminación es relativa. En secciones posteriores se analiza este problema en detalle. ■

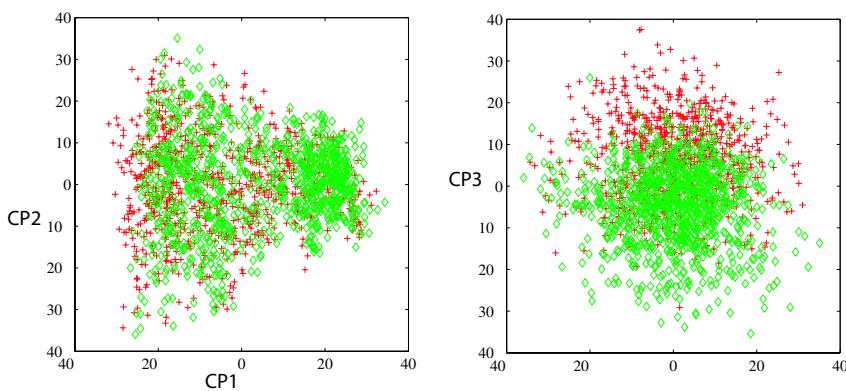


Figura 5.8: Gráficos de (a) CP1 vs CP2 y (b) CP2 vs CP3 mostrando en distintas clases los eventos lluvia ($Precip > 0.5mm$) y no lluvia en Santander.

5.5. Redes Multi-Capa

Los perceptrones son modelos bastante limitados en la práctica, y son demasiados sencillos tanto desde un punto de vista biológico, como matemático (en realidad, estas redes son equivalentes a modelos de regresión lineal y logística). Por ello, se han desarrollado modelos más generales (perceptrones multicapa, o redes multicapa) que incorporan capas internas ocultas, además de la capa de entrada y la de salida. Los nodos de la capa de entrada alimentan la red hacia delante donde cada nodo en las capas ocultas y de salida recibe de forma progresiva una entrada de los nodos de las capas previas y calcula un valor de salida para la siguiente capa. La Fig. 5.9 muestra la estructura de este tipo de redes, con un número arbitrario de capas ocultas (izquierda) y con sólo una capa oculta (derecha). Normalmente se denota la estructura de la red indicando el número de neuronas en cada capa. Así la red de la derecha se representaría como una red 4:3:2.

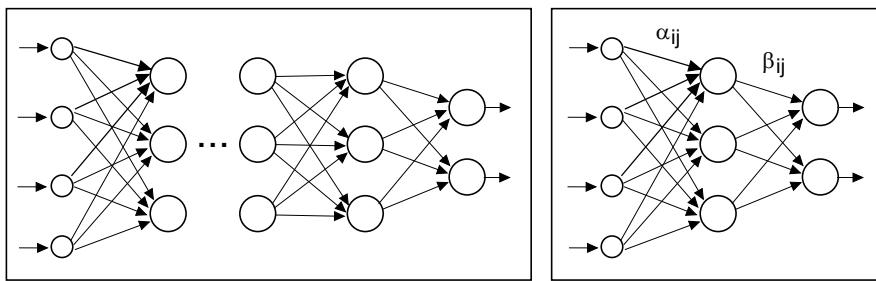


Figura 5.9: Estructura de una red multicapa, o perceptrón multicapa, (izquierda) y una popular estructura de dos capas de topología 4:3:2.

Para cada una de las salidas, y_i , el perceptrón de dos capas (una oculta y la de salida, ya que la de entrada no cuenta) mostrado en la Fig.5.9(derecha) calcula un valor $y_i = F_i(x_1, \dots, x_m)$ a partir de las entradas de la forma:

$$y_i = f\left(\sum_i \beta_{ij} f\left(\sum_k \alpha_{ik} x_{kp}\right)\right). \quad (5.12)$$

Las capas ocultas dan una mayor flexibilidad al modelo resultante, que puede considerarse un modelo no paramétrico de regresión no lineal (obsérvese que el número de parámetros, o pesos, para una red $m:k:n$ es $(m+n)k$). De esta forma, se pueden generalizar los resultados de predicción obtenidos con la regresión en el Cap. 3 (para una descripción más general, incluyendo otros tipos de redes, ver Hastie et al., 2001). Sin embargo, el problema de este tipo de redes es su diseño, pues a priori no existe ninguna regla orientativa sobre el número de capas óptimo para un problema dado, o el número de neuronas en cada capa.

Se ha demostrado que un perceptrón con dos capas ocultas puede aproximar con un grado de exactitud dado cualquier función continua; además, cuando las funciones no son continuas sólo se requiere una capa adicional: tres capas (dos ocultas) (Cybenko, 1989). Este resultado resuelve parcialmente el problema

del diseño de una estructura de red apropiada para un problema dado. Para funciones continuas, el problema se reduce a elegir un número apropiado de neuronas en la capa oculta para ajustar el modelo. Este número también tiene una cota superior si se quiere evitar que la red se sobreajuste; el número total de parámetros ha de ser mucho menor que el número de datos disponible. Sin embargo, aún así, en la práctica es necesario probar redes con distintas configuraciones para obtener un modelo apropiado.

Una vez elegido un diseño para la red neuronal, es necesario utilizar un algoritmo de aprendizaje para ajustar los parámetros (pesos) a los datos. El método de aprendizaje más popular para redes multicapa es conocido como *retro-propagación (backpropagation)* y está basado en minimizar la función de error cuadrático total usando el método del descenso de gradiente de forma apropiada, dado que los pesos de las dos capas dependen unos de otros.

5.5.1. El Algoritmo de Retro-propagación

Supóngase que se tiene un conjunto de entradas $\{x_{1p}, \dots, x_{mp}\}$ y sus correspondientes salidas $\{y_{1p}, \dots, y_{np}\}$, $p = 1, \dots, a$, como patrones de entrenamiento. Como en el caso del perceptrón, se considera la función que da el error cuadrático total:

$$E(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{j,p} (y_{jp} - f(\sum_i \beta_{ji} h_{ip}))^2 \quad (5.13)$$

$$= \sum_{j,p} (y_{jp} - f(\sum_i \beta_{ji} f(\sum_k \alpha_{ik} x_{kp})))^2 \quad (5.14)$$

$$= \sum_p \|\mathbf{y}_p - f(\boldsymbol{\beta}^T f(\boldsymbol{\alpha}^T \mathbf{x}_p))\| \quad (5.15)$$

donde $\boldsymbol{\beta}$ tiene por columnas los vectores de pesos de la neurona j -ésima de la capa de salida, y $\boldsymbol{\alpha}$ los de la k -ésima neurona de la capa oculta.

El algoritmo de retro-propagación está basado en la misma idea del descenso de gradiente usado en el método de la regla delta. Por tanto, se han de modificar incrementalmente los pesos de forma proporcional al gradiente de la función de error:

$$\Delta\beta_{ik} = -\eta \frac{\partial E}{\partial \beta_{ik}}; \quad \Delta\alpha_{kj} = -\eta \frac{\partial E}{\partial \alpha_{kj}}, \quad (5.16)$$

donde el término η es la tasa de aprendizaje, que podría ser diferente para ambos conjuntos de pesos.

Dado que (5.15) involucra los pesos de las neuronas ocultas y las de salida, el problema de actualizar iterativamente sus valores no es tan simple como en el caso de los perceptrones de una capa. Una solución a este problema fue dada con el algoritmo de retro-propagación de dos pasos (ver, por ejemplo, Rumelhart and McClelland, 1986). En primer lugar, la entrada de un patrón \mathbf{x}_p se propaga hacia delante, obteniendo el valor de las unidades ocultas, $\hat{\mathbf{h}}_p$, y la salida, $\hat{\mathbf{y}}_p$, y, por tanto, el error asociado. Los valores obtenidos se usan

para actualizar los pesos β_{ik} de la capa de salida usando (5.13). Más tarde, los pesos obtenidos se utilizarán para actualizar los pesos de la capa oculta, α_{kj} , usando (5.14) para propagar hacia atrás los errores anteriores. La forma final de algoritmo resulta como sigue (ver Castillo et al., 1999, Cap. 1, para más detalles):

1. Iniciar los pesos con valores aleatorios.
2. Elegir un patrón de entrenamiento y propagarlo hacia adelante obteniendo los valores $\hat{\mathbf{h}}_p$ e $\hat{\mathbf{y}}_p$ para las neuronas de las capas ocultas y de salida.
3. Calcular el error asociado a las unidades de salida:

$$\delta_{jp} = (y_{jp} - \hat{y}_{jp}) f'(\boldsymbol{\beta}_j^T \hat{\mathbf{h}}_p).$$

4. Calcular el error asociado a las unidades ocultas:

$$\psi_{jp} = \sum_k \delta_{jp} \beta_{jk} f'(\boldsymbol{\alpha}_k^T \mathbf{x}_p).$$

5. Calcular:

$$\Delta\beta_{jk} = \eta \hat{\mathbf{h}}_p \delta_{jp},$$

y

$$\Delta\alpha_{ki} = -\eta \sum_j x_{ip} \delta_{jp} \psi_{jp},$$

y actualizar los pesos de acuerdo con los valores obtenidos.

6. Repetir los pasos anteriores para cada patrón de entrenamiento.

El algoritmo anterior se suele utilizar en la modalidad denominada *batch*, donde se acumulan los incrementos para cada patrón de entrenamiento y sólo se modifica el valor de los pesos después de cada ciclo completo.

Ejemplo 5.3 (Series Temporales: El Sistema de Lorenz). En la Sec. 1.6 se ha analizado el problema de la predecibilidad en sistemas no lineales, analizando el sistema de Lorenz. Uno de los fenómenos más sorprendentes relacionados con estos sistemas es su comportamiento aparentemente impredecible y errático. Dado que estos sistemas son deterministas, pueden ser predichos a corto plazo utilizando una técnica de ajuste no paramétrico para inferir la estructura funcional del sistema a partir de los datos disponibles. Sin embargo, dada la sensibilidad de estos sistemas a las perturbaciones en sus condiciones iniciales, una predicción a largo plazo sólo es posible en términos probabilísticos.

En este ejemplo se analiza la eficiencia de las redes multicapa para predecir la dinámica caótica del sistema de Lorenz a partir de una serie temporal del mismo. En primer lugar se podrían aplicar las técnicas de series temporales descritas en la sec. 3.3; sin embargo, la no linealidad del sistema hace que estas técnicas no resulten apropiadas (como se mostrará más adelante). Por

tanto, las redes neuronales multicapa pueden verse como generalizaciones no lineales de los modelos autoregresivos.

Se consideran los valores de los parámetros $\sigma = 10$, $b = 8/3$, y $r = 28$, para el sistema de Lorenz (1.18), que presenta dinámica caótica para estos valores. Para generar el conjunto de datos de entrenamiento, se integra el sistema a partir de la condición inicial $\mathbf{u}_0 = (x(0), y(0), z(0)) = (-10, -5, 35)$ utilizando un método de integración de Runge-Kutta de cuarto orden y paso de integración $\tau = 10^{-2}$. Esto es equivalente a extraer una muestra del sistema $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_N$ a intervalos de tiempo equi-espaciados $t_n = n\tau$, $n = 0, 1, 2, \dots$.

Se está interesado en la aproximación del modelo funcional de la forma $\mathbf{u}_{n+p} = f(\mathbf{u}_n)$, para un horizonte de predicción p (en este ejemplo $p = 1$). Para este propósito se considera una red multicapa con funciones de activación sigmoidales para las capas ocultas y lineal para la capa de salida (de esta forma, la salida no está acotada, y no es necesario normalizar los datos). El proceso de entrenamiento se lleva a cabo considerando pares entrada-salida de la forma $(\mathbf{u}_n, \mathbf{u}_{n+p})$, donde p es el horizonte de predicción. Al tratarse de un sistema continuo se han considerado distintas redes multicapa 3:k:3 con una única capa oculta con k neuronas, además de tres neuronas de entrada (x_n, y_n, z_n) y tres neuronas de salida ($x_{n+p}, y_{n+p}, z_{n+p}$). Para cada una de estas redes (con k variando entre 1 y 20) se realizaron diez experimentos con diferentes pesos iniciales. Por ejemplo, la Figura 5.10(a) muestra los errores obtenidos para la variable de predicción x con el mejor modelo de red de 6 neuronas en la capa oculta. El error cuadrático medio (RMSE) obtenido por esta red fue de 0.133 para el proceso de aprendizaje y de 0.149 para la validación. Estos resultados muestran que no se produce sobreajuste en los datos.

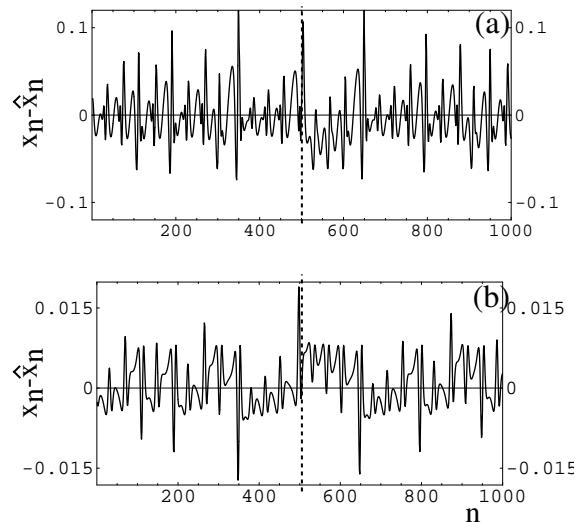


Figura 5.10: Residuos $x_n - \hat{x}_n$ para dos modelos neuronales (a) 3:6:3 y (b) 3:5:5:3. Los 500 primeros puntos se utilizan para entrenamiento y los restantes para validación.

A pesar del buen comportamiento de la red de seis neuronas ocultas para la predicción a un paso, no está claro que el modelo neuronal obtenido pueda reproducir la dinámica del sistema de Lorenz. Por ejemplo, en el caso anterior, el modelo obtenido para \hat{x}_{n+1} fue:

$$\begin{aligned} -3768.18 - \frac{0.34}{1 + e^{9.31+0.53x_n-0.68y_n-0.21z_n}} + \frac{0.92}{1 + e^{7.64-0.121x_n-0.149y_n-0.13z_n}} - \\ \frac{2.75}{1 + e^{6.19+0.15x_n+0.0451y_n-0.09z_n}} - \frac{2.04}{1 + e^{1.13+0.06x_n+0.0119y_n-0.06z_n}} + \\ \frac{7164.31}{1 + e^{-0.12+0.00021x_n-0.0002y_n+0.000021z_n}} - \frac{63.52}{1 + e^{-0.24+0.08x_n-0.016y_n+0.0049z_n}}, \end{aligned}$$

La Figura 5.11 ilustra este hecho mostrando la evolución de la red neuronal para dos condiciones iniciales distintas; en el primer caso, el sistema neuronal converge a una trayectoria periódica (Fig. 5.11(a)), mientras que en el segundo caso converge a un punto fijo (Fig. 5.11(b)); ninguno de ellos refleja el comportamiento caótico del sistema de Lorenz.

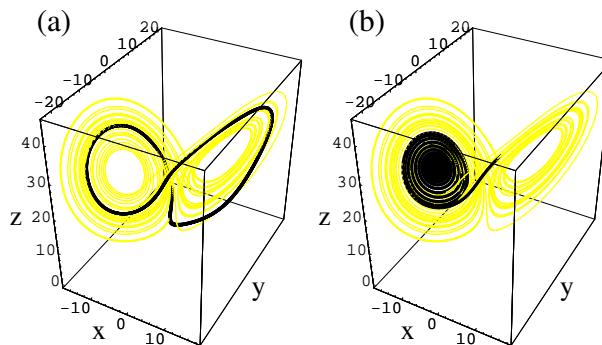


Figura 5.11: Evolución en el espacio de fases de un modelo neuronal 3:6:3 con dos condiciones iniciales diferentes. La parte sombreada en el fondo corresponde a la órbita caótica original y se muestra con propósito ilustrativo.

Dado que la función de activación de las neuronas de salida es lineal, también se consideraron redes con dos capas ocultas de la forma 3:k:k:3 para aprender los datos. Por ejemplo, para $k = 5$, el error disminuye y el comportamiento dinámico del modelo neuronal obtenido se asemeja al sistema caótico original. Por ejemplo, la Figura 5.10(b) muestra el error de entrenamiento y de test asociados a una red multicapa de tipo 3:5:5:3 (este error es un orden de magnitud menor que el asociado con el modelo de 6 neuronas ocultas mostrado en la Fig. 5.10(a)). El RMSE del entrenamiento y test fueron de 0.0221 y 0.0237, respectivamente, los cuales indican que no hubo sobre-ajuste. La Figura 5.12 muestra la evolución de los sistemas original y neuronal, comenzando en la misma condición inicial. El punto donde ambos sistemas comienzan a separarse ($t \approx 3$) es aproximadamente el umbral impuesto por el comportamiento caótico en la precisión numérica de los cálculos realizados. Por tanto, la red no sólo se ajusta a los datos, sino que es capaz de reproducir la dinámica del sistema.

Finalmente, si se sigue incrementando el número de neuronas en la capa oculta por encima de veinte, el error de entrenamiento continúa decreciendo, pero los modelos neuronales comienza a sobre-ajustarse a los datos. Como consecuencia, el comportamiento de estos modelos presenta diferencias significativas con el sistema original (la mayoría de las veces los modelos neuronales divergen asimptóticamente). ■

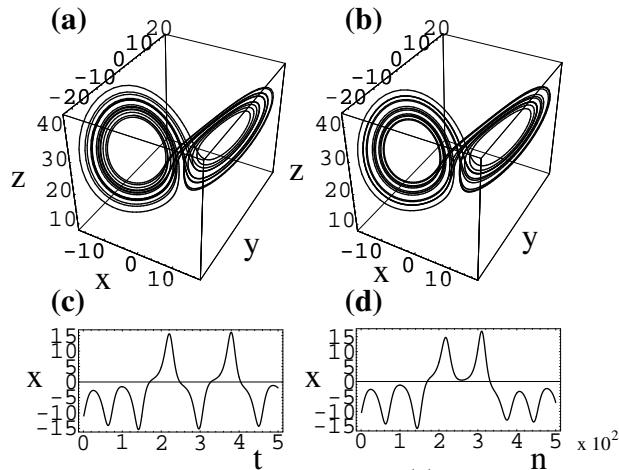


Figura 5.12: Evolución en el espacio de fases de (a) el modelo de Lorenz y (b) un modelo neuronal de tipo 3:6:6:3.

Ejemplo 5.4 (Serie Temporal de Lorenz. Embedding). Para ilustrar la técnica de inversión descrita en la Sec. 3.3 se considera de nuevo el modelo de Lorenz descrito en el ejemplo anterior, y se supone que sólo se dispone de un observable del sistema (por ejemplo, la variable x). La Fig. 5.13 muestra una proyección del espacio de embedding dado por $\tau = 10$ y $m = 4$; se observa la similitud topológica con el atractor real del sistema. Se puede aplicar una red neuronal para tratar de modelizar la dinámica del sistema en el espacio de retrasos. La Fig. 5.14 muestra el error obtenido por una red neuronal 4:3:3:1 entrenada para estimar x_n utilizando $x_{n-\tau}, x_{n-2\tau}, x_{n-3\tau}$ y $x_{n-4\tau}$. En este caso, el tiempo de predicción es τ (en este ejemplo τ corresponde al mínimo retraso tomado: $n = 10$ ó $t = 10 * 10^{-2} = 0.1$ unidades de tiempo).

Si se trata de utilizar este mismo modelo para predecir el valor del sistema a tiempos más largos (iterando el sistema con las predicciones obtenidas), existe un horizonte a partir del cual el sistema aproximado ya no sigue la misma dinámica del original. Por ejemplo, en la Fig. 5.15 se muestra la predicción obtenida por la red neuronal para umbrales de tiempo mayores. En esta figura, tanto la red neuronal como el sistema de Lorenz son inicializados en un instante de tiempo y prolongados de forma independiente. Las trayectorias de ambos sistemas siguen la misma evolución hasta un tiempo $t = 1.8$ en que comienzan a diverger, marcando el umbral de predicción en este caso. ■

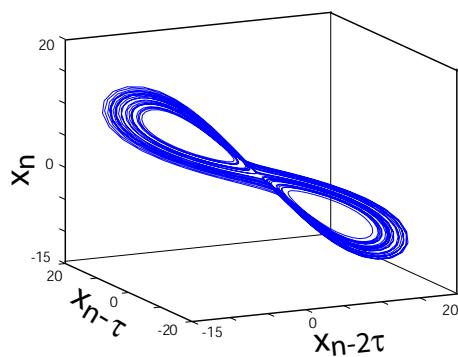


Figura 5.13: Atractor en el espacio de embedding obtenido a partir de un serie temporal de la variable x del sistema de Lorenz.

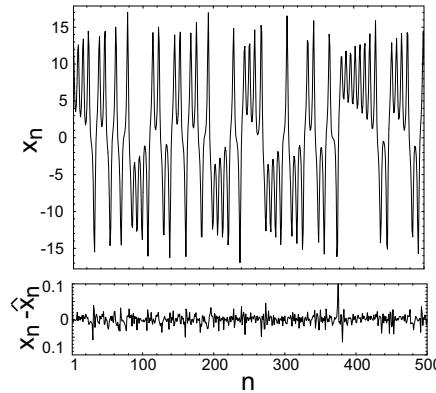


Figura 5.14: Serie temporal real x_n y error obtenido con la estimación \hat{x}_n dada por una red neuronal 4:3:3:1 (la abscisa de la figura inferior está magnificada un factor 50 para claridad de la visualización).

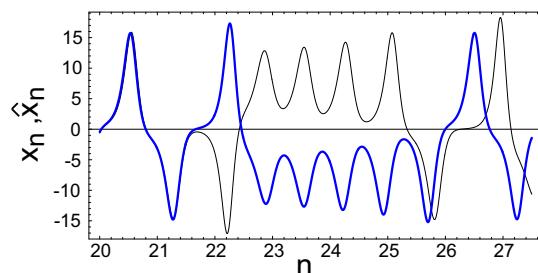


Figura 5.15: Serie original del modelo de Lorenz (trazo grueso) y serie obtenida prolongando la red neuronal a partir de la misma condición inicial.

5.5.2. MeteoLab: *El Sistema de Lorenz*

En esta sección se describe el código necesario para desarrollar un modelo aproximado de Lorenz utilizando redes neuronales. Para ello, primero se integra el sistema real a partir de una condición dada utilizando el comando `ode45`. A continuación, se entrena un modelo neuronal y un modelo lineal y se comparan los resultados (el código está disponible en el directorio MeteoLab/NeuralNets de la toolbox MeteoLab).

```
%Integrating the Lorenz model
[t,data]=ode45('LorenzEq',[0:0.1:50],[10,10,10]);
plot(data)

h=1; X=data(1:end-h,:); Y=data(h+1:end,:);

% Training a 3:4:4:3 neural net
figure nnet=newff([min(X,[],2),max(X,[],2)],...
[4 4 3],{'tansig','tansig','purelin'});
nnet.trainParam.epochs=500;
nnet=train(nnet,X,Y);
pred=sim(nnet,X);

figure; plot(pred(1,:),Y(1,:),'k');

% Training a linear model
nnetl=newff([min(X,[],2),max(X,[],2)],[3],{'purelin'});
nnetl.trainParam.epochs=100;
nnetl=train(nnetl,X,Y);
predl=sim(nnetl,X);

figure; plot(predl(1,:),Y(1,:),'b');
```

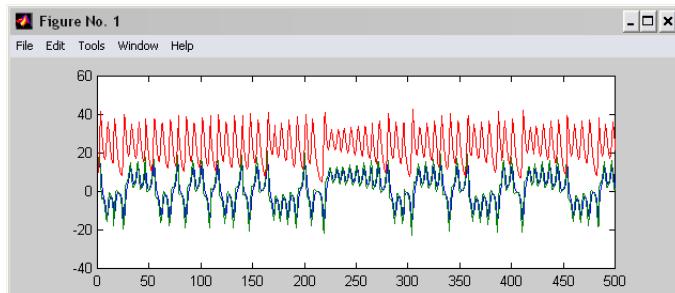


Figura 5.16: Serie temporal del sistema de Lorenz utilizada para entrenar una red neuronal.

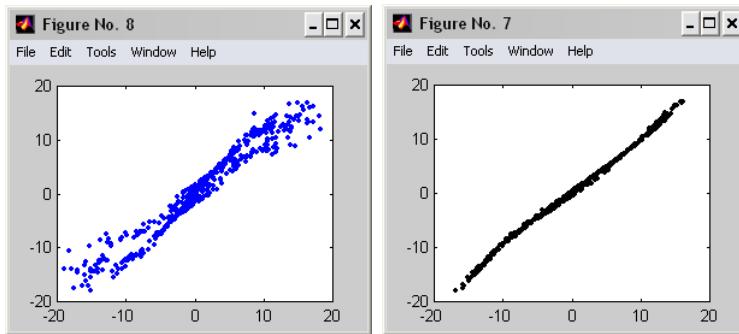


Figura 5.17: Valor real vs. predicho utilizando un modelo lineal (izquierda) y una red neuronal (derecha). La diagonal corresponde a una predicción perfecta.

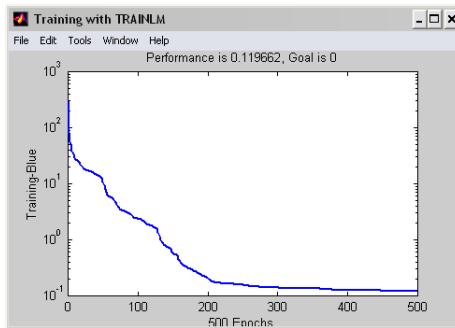


Figura 5.18: Evolución del error en 500 iteraciones del algoritmo de aprendizaje.

5.6. Redes de Función de Base Radial

Modificando la topología de la red, es posible crear nuevos modelos de red neuronal para resolver problemas concretos. Sin embargo, también se tiene flexibilidad para cambiar el modelo de neurona. En los casos anteriores, las funciones de activación de las neuronas eran sigmoidales (definidas en la recta real). En esta sección se estudia otro tipo importante de redes neuronales, denominadas redes de función de base radial. La topología de la red es idéntica a una red multicapa con una sola capa oculta, pero con distinta función de activación para las neuronas de la capa oculta. En éste caso, la función de activación es una función núcleo cuyo valor se calcula en base a la distancia entre los patrones de entrada y los centros de las funciones base consideradas. Se pueden considerar diferentes tipos de funciones de activación o funciones base, la más común es la función gaussiana centrada en c_j y de anchura σ_j ,

$$h_j(\mathbf{x}) = \exp\left(-\frac{(\mathbf{x} - \mathbf{c}_j)^2}{2\sigma_j^2}\right) \quad (5.17)$$

Otra elección, aunque menos común, son las funciones de Cauchy

$$h_j(\mathbf{x}) = \frac{\sigma_j^2}{(\mathbf{x} - \mathbf{c}_j)^2 + \sigma_j^2} \quad (5.18)$$

Por otra parte, la función de activación de la capa de salida es lineal. De esta forma, la red calcula una combinación lineal de núcleos definidos en el espacio de los datos:

$$f(\mathbf{x}) = \sum_{j=1}^m w_j h_j(\mathbf{x}). \quad (5.19)$$

Los parámetros de estos modelos son los centros y varianzas de los núcleos ($c_j, \sigma_j, j = 1, \dots, m$) y los pesos $w_j, j = 1, \dots, m$. El primer conjunto de parámetros definen la forma de la red, mientras que los pesos determinan su ajuste a los datos. Los algoritmos de aprendizaje para este tipo de redes tienen en cuenta ambos términos.

5.6.1. Aprendizaje de los Parámetros

Dado un conjunto de entrenamiento con a patrones, $\{(\mathbf{x}_i, y_i), i = 1, \dots, a\}$, fijadas las funciones base $h_j, j = 1, \dots, m$ (es decir \mathbf{c}_j y $\sigma_j, j = 1, \dots, m$), se puede encontrar el vector de pesos óptimo minimizando la suma de errores al cuadrado (problema de mínimos cuadrados),

$$E = \sum_{i=1}^a (f(x_i) - y_i)^2 \quad (5.20)$$

cuya solución viene dada por

$$\hat{\mathbf{w}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} \quad (5.21)$$

donde \mathbf{H} es la matriz de diseño ($H_{ij} = h_j(x_i)$) e $\mathbf{y} = (y_1, \dots, y_a)^T$ es el vector de salida esperado.

Como todos los problemas de mínimos cuadrados, al trabajar con matrices de grandes dimensiones, se puede tener un problema mal condicionado. Para solventar este problema, en la ecuación de errores que debe ser minimizada, se puede considerar un parámetro de penalización de la forma

$$E = \sum_{i=1}^a (f(x_i) - y_i)^2 + \lambda \sum_{j=1}^m w_j^2 \quad (5.22)$$

con un *parámetro de regularización* λ , que permite obtener una vector de pesos mucho más robusto. En este caso, suponiendo que λ es conocido, el vector de pesos óptimo viene dado por

$$\hat{\mathbf{w}} = \mathbf{A}^{-1} \mathbf{H}^T \mathbf{y}, \quad (5.23)$$

$$\mathbf{A} = (\mathbf{H}^T \mathbf{H})^{-1} + \lambda \mathbf{I}_m, \quad (5.24)$$

donde \mathbf{I}_m es la matriz identidad de orden m .

Un problema de aprendizaje más complicado se tiene cuando no se conocen los parámetros asociados a las funciones núcleo $\{h_j, j = 1, \dots, m\}$; es decir, no se conocen los centros \mathbf{c}_j asociados a cada una de las funciones y/o no se conocen sus anchuras σ_j . Por ejemplo, cuando no se conocen los centros, se pueden aplicar distintos algoritmos para estimarlos a partir de los datos. Uno de los más populares está basado en los algoritmos de agrupamiento descritos en la Sec. 2.6. De esta forma, los centros estarán asociados a un conjunto de prototipos obtenidos mediante, por ejemplo, k-medias a partir de los datos de entrada de la red.

Estas técnicas sólo permiten obtener los centros de las funciones, pero la anchura tiene que ser elegida de algún otro modo. Una forma de elegir la anchura es tomarla fija para todas las funciones como un múltiplo de la distancia promedio entre los centros. De esta forma se garantiza que las funciones se solapan, y ofrecen una representación suave de la distribución de los datos.

Hasta este punto se han tratado este tipo de redes como dos elementos separados; por un lado se optimiza el error de salida junto con el parámetro de regularización, y por otro los parámetros de las funciones de base radial; es decir, los centros y la anchura de la función. También se pueden optimizar todos los parámetros libres de la red en un mismo paso, considerando un método de entrenamiento supervisado, aunque son computacionalmente costosos (Hastie et al., 2001).

5.6.2. MeteoLab: Funciones de Base Radial

En esta sección se utiliza la toolbox *Neural Nets* de Matlab para mostrar el uso de una red neuronal de base radial. Se intenta ajustar una superficie a partir de un conjunto de puntos. También se pueden consultar los ejemplos desarrollados en *NetLab*, incluidos en el directorio correspondiente.

```
a=20;
[x,y,z]=peaks(a); %Peaks surface on a a*a lattice
figure;
surf(x,y,z)

range=6/10; %spread
goal=0.5; %training error
netrb= newrb([x(:) y(:)]',z(:)',goal,range);

[x2,y2] = meshgrid(-3:6/(a-1):3);
snetrb=reshape(sim(netrb,[x2(:),y2(:)]'),size(z));
figure; surf(x2,y2,snetrb)
```

El código anterior inicializa una nueva red de base radial utilizando el comando **newrb** que toma como entrada el valor objetivo del error, y la anchura

común de las funciones núcleo: σ_j . El algoritmo calcula el número de funciones núcleo necesarias para describir los datos con el error requerido, y también estima los centros óptimos de las mismas. La Fig. 5.19 muestra los resultados obtenidos al entrenar dos redes con funciones objetivo distintas, mientras que la Fig. 5.20 muestra los centros correspondientes.

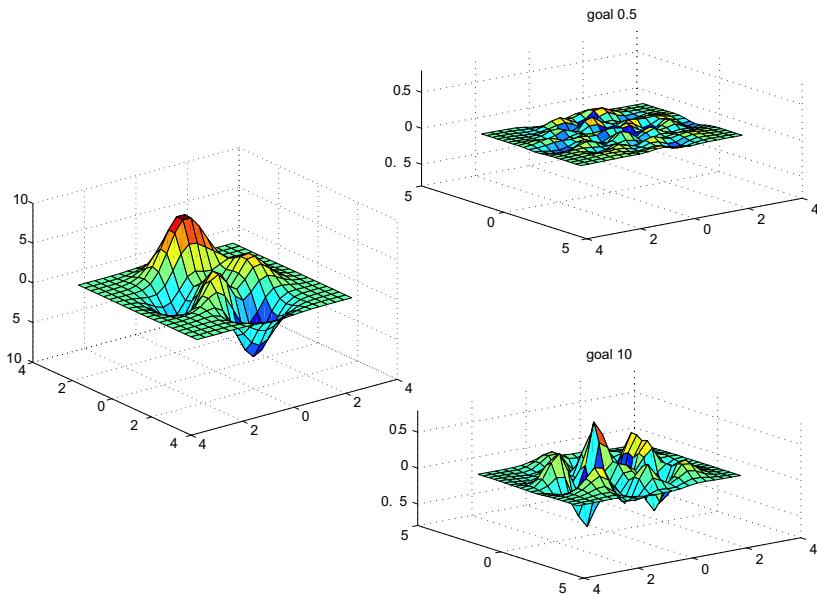


Figura 5.19: (izquierda) Superficie original y (derecha) superficies diferencia entre la superficie original y la simulada, al considerar dos cotas de error diferente en el entrenamiento.

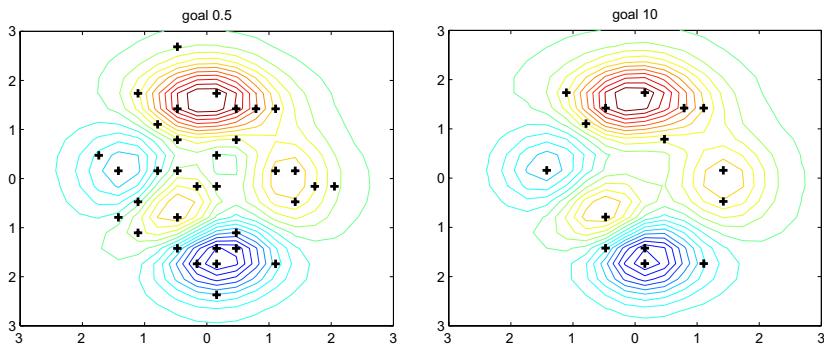


Figura 5.20: Contornos de la superficie simulada y centros considerados al construir la red considerando dos cotas de error diferentes en el entrenamiento.

5.7. Redes Competitivas

Las redes competitivas son un tipo de arquitecturas de red muy populares, utilizadas para detectar automáticamente grupos homogéneos en el conjunto de datos disponible. En estos problemas, la única información disponible son los patrones de entrada, que representan un punto en el espacio de datos donde se quieren obtener clases, pero no se tiene un vector de salida asociado a cada patrón; por tanto, son técnicas de aprendizaje no supervisado. Una red neuronal competitiva simple tiene la misma estructura que un perceptrón. La capa de entrada tiene una neurona para cada dimensión del espacio de datos, mientras que la capa de salida contiene tantas neuronas como grupos distintos se quieran obtener en el conjunto de datos. El mecanismo de aprendizaje no supervisado se crea dejando que las neuronas de salida compitan entre sí; para ello se incluyen conexiones laterales entre las neuronas de la capa de salida (ver Fig. 5.21), de forma que el valor de salida de cada neurona puede depender del resto de valores (por ejemplo, se puede hacer que sólo una neurona, la “ganadora”, tenga un valor de salida no nulo).

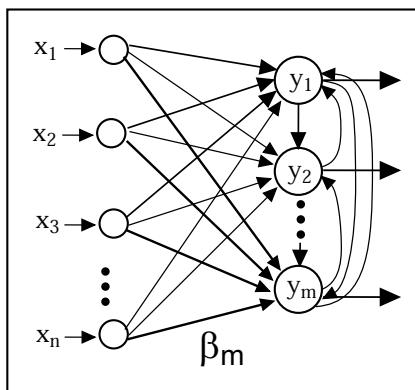


Figura 5.21: Una red neuronal competitiva para agrupar un espacio n dimensional en m grupos.

La Fig. 5.21 muestra la estructura de una red competitiva. Se han destacado las conexiones de una de las neuronas de salida, y_m , para poner de manifiesto que el vector de pesos asociado, $\beta_m = (\beta_{m1}, \dots, \beta_{mn})$ tiene la misma dimensión que el espacio de datos original. Por tanto, cada vector de pesos define un *prototipo* en el espacio de datos que será representativo del grupo asignado a la neurona de salida correspondiente. Este tipo de arquitectura se entrena normalmente con un algoritmo competitivo del tipo ‘winner takes all’, que consiste en asignar cada patrón de entrada a aquella neurona con el prototipo (vector de pesos) más cercano al patrón, según alguna métrica predefinida; la neurona resultante se denomina ‘ganadora’. Así, dado un patrón de entrada, cada neurona y_k calcula su distancia (la de su vector de pesos) al patrón y produce como salida un uno, cuando $\hat{y}_k = \min\{\hat{y}_1, \dots, \hat{y}_m\}$, y un cero en caso contrario.

Considérense los datos de entrenamiento dados por un conjunto de patrones

de entrada $(x_{1j}, \dots, x_{nj}), j = 1, \dots, a$. Inicialmente, se asignan valores aleatorios pequeños a los pesos de la red. Para cada uno de los patrones de entrada \mathbf{x}_j , se calcula la neurona ganadora, $y_{w(j)}$. Para incrementar la ventaja de esta neurona sobre el patrón de entrada, los pesos de la neurona son modificados de acuerdo con

$$\Delta\beta_{w(j)} = \eta(\mathbf{x}_j - \beta_{w(j)}), \quad (5.25)$$

donde η es la tasa de aprendizaje. El efecto de esta alteración es acercar el vector pesos hacia el patrón de entrada en el espacio de datos. El algoritmo procede iterativamente en ciclos, en los cuales se recorren todos los datos de entrada. La modalidad *batch* de este algoritmo incrementa los pesos después de cada ciclo, y no después de cada patrón. Una vez entrenada la red, la salida ganadora para un patrón dado indica el grupo correspondiente.

Nótese que en el caso de clasificación supervisada usando perceptrones multi-capa, el usuario proporciona ejemplos de las diferentes categorías. No obstante, esta nueva arquitectura proporciona un método de clasificación que detecta automáticamente grupos de patrones similares en el espacio de las configuraciones. En esencia, el algoritmo básico es muy similar a la técnica de agrupamiento de k -medias descrita en la Sec. 2.6. Sin embargo, existen numerosas modificaciones de este algoritmo con distintas características y propiedades. Por ejemplo, para evitar el problema de trabajar con un número fijo de unidades de salida (grupos) se han propuesto algunas arquitecturas competitivas más generales. Por ejemplo, la Teoría de Resonancia Adaptativa (Adaptive Resonance Theory, ART) introducida por Grossberg (1976) utiliza la sensibilidad de la clasificación efectuada por la red para determinar el número óptimo de salidas necesarias para clasificar los datos.

Ejemplo 5.5 (Redes competitivas). Considerese el conjunto de puntos generados por tres distribuciones Gaussianas bivariadas independientes con varianza $\sigma^2 = 0.08$ mostradas en la Figura 5.22(a). Las áreas sombreadas de la figura indican las desviaciones estándar de las distribuciones Gaussianas correspondientes. En este ejemplo se ilustra el uso de redes neuronales competitivas para agrupar este conjunto de datos en un número de grupos predeterminado. Se tienen dos entradas, las coordenadas x e y , y se considerarán dos redes diferentes con dos y tres unidades de salida, respectivamente (ver Figura 5.23).

Los vectores de pesos de las redes construidos con los datos anteriores se muestran en la Figura 5.23. La Figura 5.23(a) muestra también los vectores $(\beta_{1i}, \beta_{2i}), i = 1, 2$ correspondientes a dos unidades de salida. Similarmente, la Figura 5.23(b) muestra los tres vectores de pesos asociados. Nótese que, en el caso de dos unidades de salida, los puntos correspondientes a las distribuciones Gaussianas mostradas en la parte superior de la Figura 5.22(b) se asignan a la misma clase. No obstante, cuando se incrementa el número de categorías, cada uno de los datos de las tres distribuciones diferentes se asigna a un grupo distinto, como se muestra en la Fig. 5.22(c). ■

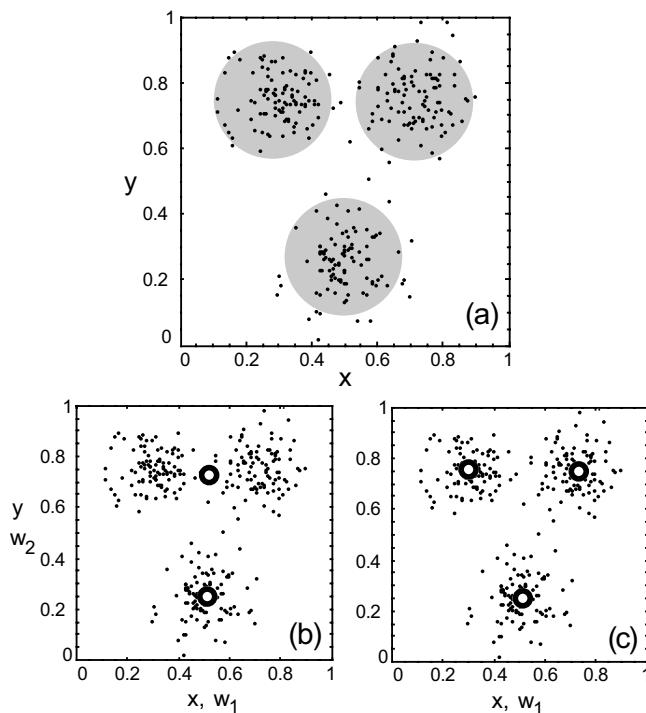


Figura 5.22: (a) Datos aleatorios consistentes en 300 puntos generados por tres distribuciones Gaussianas bivariadas independientes con varianza $\sigma^2 = 0.08$ y medias $(0.3, 0.75)$, $(0.5, 0.25)$, y $(0.75, 0.75)$, respectivamente. (b) Ejemplo de aprendizaje competitivo con dos y (c) tres unidades de salida. Los círculos muestran el vector de peso de cada unidad después del entrenamiento con estos datos.

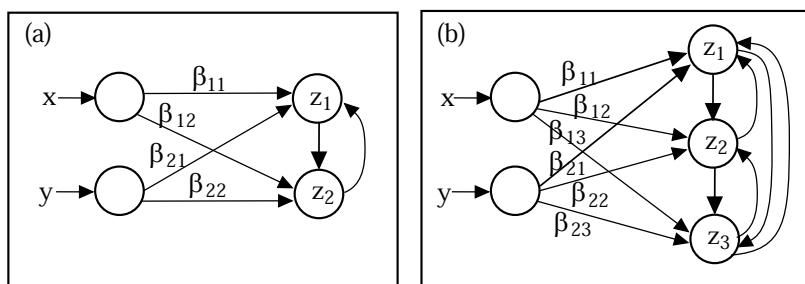


Figura 5.23: Dos redes neuronales competitivas con (a) dos y (b) tres unidades de salida.

5.8. Redes Auto-Organizativas (SOM)

Las redes auto-organizativas (Self-Organizing Maps, SOM) son redes competitivas especialmente indicadas para trabajar en espacios de alta dimensionalidad, ya que permiten organizar y visualizar los datos de forma intuitiva y eficiente proyectándolos en un espacio arbitrario (normalmente una rejilla 2-dimensional). Existen distintos métodos empíricos y/o subjetivos para visualizar datos meteorológicos (ver Macedo et al. (2000)), pero las SOM tienen una serie de ventajas que serán utilizadas en este capítulo.

A pesar de que esta técnica surgió en el contexto de la computación neuronal (Kohonen, 2000), las SOM también puede verse como una generalización de la técnica de k -medias descrita en la Sec. 2.6. En este caso, cada una de las neuronas de salida (o prototipos) y_k de la SOM está representado por una neurona en la capa de salida, y tiene asociados dos vectores: uno en el espacio de los datos (el vector de pesos β_k) y otro en el espacio base bidimensional de proyección (la posición relativa en la rejilla, $\mathbf{r}_k = (i_k, j_k)$). La característica de esta técnica es la inclusión de un núcleo espacial de vecindad cuyo efecto es mantener unidos en el espacio de los datos aquellos centroides vecinos en la rejilla 2D. La amplitud del núcleo decrece durante el entrenamiento haciendo local la noción de vecindad cuando finaliza el entrenamiento. De esta forma, una vez finalizado el entrenamiento, los centroides vecinos están también cercanos en la rejilla 2D. Por tanto, el proceso de aprendizaje es similar al descrito para redes competitivas, pero con la inclusión de un término de vecindad que modifica no sólo los pesos de las neuronas ganadoras, sino también los de sus vecinas en la rejilla. Así, el proceso de aprendizaje proyecta la estructura topológica del espacio de datos en la rejilla prefijada. En los últimos años, han sido numerosas las aplicaciones que han utilizado las ventajas de esta técnica (ver Oja and Kaski, 1999, y las referencias incluidas). En meteorología la aplicación ha sido más reciente (ver Hewitson and Crane, 2002).

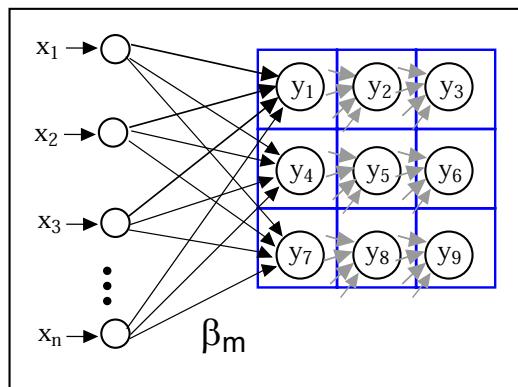


Figura 5.24: Esquema de una red auto-organizativa (SOM) con $3 \times 3 = 9$ neuronas organizadas en una red 2D.

Como se muestra en la Figura 5.24, una SOM está formada por un número arbitrario de grupos representados por las neuronas de salida de la red y_1, \dots, y_m , localizados sobre una retícula regular en un espacio de baja dimensión, usualmente una red 2D para propósitos de visualización (en este caso $m = s \times s$). El vector $r_k = (i, j)$ representa la posición de la neurona en la retícula, donde $1 \leq i, j \leq s$. Además, como se vio en la sección anterior, cada neurona y_k tiene asociada un vector de pesos con la capa de entrada $\beta_k = (\beta_{k1}, \dots, \beta_{kn})$, el cual puede asociarse con la posición del centro del grupo sobre espacio n -dimensional de los datos (miles de dimensiones o, cientos de componentes principales para patrones atmosféricos).

Siguiendo un procedimiento similar al algoritmo ‘winner takes all’ (o al algoritmo de k -medias descrito en la Sec. 2.6), los vectores prototipo (pesos) de la SOM son inicializados a valores aleatorios. El objetivo del algoritmo de entrenamiento es adaptar iterativamente los vectores, de forma que cada prototipo final represente a un grupo de datos (aquellos que están más cerca a ese prototipo que a ningún otro). Lo que hace a la SOM diferente de otros algoritmos de agrupamiento es que el proceso de entrenamiento incluye un mecanismo auto-organizativo de vecindad tal que los grupos vecinos en la retícula 2D son también similares en el espacio real.

Una implementación del algoritmo de entrenamiento se realiza en ciclos sucesivos; en cada ciclo se analiza cada uno de los vectores \mathbf{x}_i , $i = 1, \dots, a$, calculando el prototipo más cercano (o “ganador”) $\mathbf{y}_{w(i)}$, como aquel que minimiza la distancia al vector de datos:

$$\|\mathbf{x}_i - \mathbf{y}_{w(i)}\| = \min_k \|\mathbf{x}_i - \mathbf{y}_k\|, \quad k = 1, \dots, m. \quad (5.26)$$

Después de cada ciclo, los prototipos se recalculan en base al centroide del grupo correspondiente y de los grupos vecinos:

$$\mathbf{y}_j = \frac{\sum_{i=1}^a h(\|\mathbf{r}_j - \mathbf{r}_{w(i)}\|) \mathbf{x}_i}{\sum_{i=1}^a h(\|\mathbf{r}_j - \mathbf{r}_{w(i)}\|)}, \quad j = 1, \dots, m. \quad (5.27)$$

donde la función $h(r)$ es un núcleo de vecindad que mide las distancias de los grupos en la retícula 2D y determina la tasa de cambio de un prototipo en base a los grupos vecinos (normalmente se usa una función Gaussiana):

$$h(r) = e^{-r/s(t)}.$$

El radio de vecindad $s(t)$ decrece monótonamente en el tiempo, suavizando las restricciones topológicas (se suele elegir un decaimiento lineal a cero para estas funciones). Para una descripción detallada de diferentes implementaciones del método, el lector puede consultar Oja and Kaski (1999).

Hay que tener en cuenta que la ventaja de tener los grupos organizados en la red tiene un coste, ya que comparada con la técnica estándar de k -medias, la SOM pierde parte de la variabilidad de los grupos, en favor de la restricción topológica impuesta. Esto se debe a que en los métodos clásicos los centroides se mueven libremente en el espacio y sólo tienen que minimizar la varianza

intra-grupos, mientras que en una SOM la relación de vecindad entre centroídes supone un recorte de libertad de movimiento. Es decir, si no es necesario disponer de una topología para los prototipos es más conveniente y sencillo utilizar un algoritmo estándar, pero si se quieren estudiar transiciones o posiciones relativas entre diferentes prototipos, entonces la SOM resulta de gran utilidad.

Ejemplo 5.6 (Clasificación de Patrones Atmosféricos). *Los problemas de regionalización y clasificación ya han sido abordados con técnicas de agrupamiento en la Sec. 2.6. La ventaja de utilizar una SOM es que, además de obtener los grupos, se obtendrá también una organización de vecindad de los mismos. Esto proporciona una útil visualización de las posibles transiciones e interrelaciones entre clases. En la Fig. 5.25 se muestra el resultado de aplicar este algoritmo al modelo atmosférico correspondiente a la zona NAO del Ejemplo 2.14, considerando redes de $100 = 10 \times 10$ prototipos. En esta figura se pueden observar los prototipos finales, así como la red proyectada en el espacio de los datos (sólo se muestran las dos primeras CPs).*

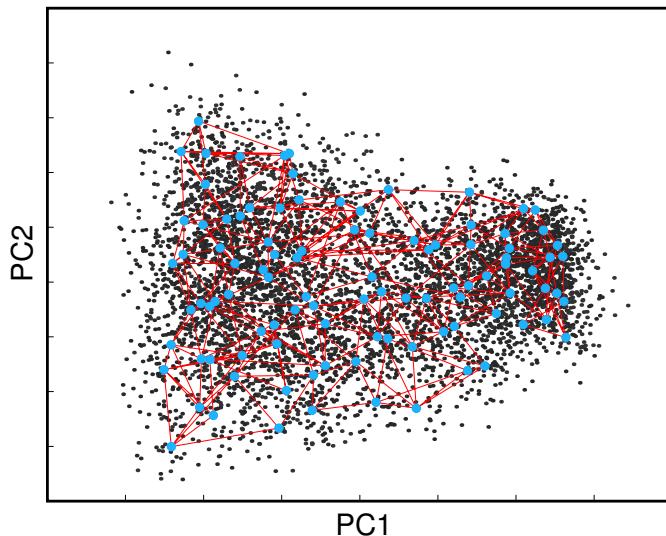


Figura 5.25: Proyección sobre las dos primeras CPs de los patrones atmosféricos de ERA-40 definidos con una SOM 10×10 para los patrones atmosféricos de la zona NAO.

La Figura 5.26 muestra los patrones atmosféricos de los prototipos resultantes (temperatura en 500mb) para el primer cuadrante de la cuadrícula. Esta figura muestra que los patrones cercanos en la cuadrícula son parecidos entre sí, mientras que los lejanos corresponden a situaciones diferentes. Cada prototipo representa un grupo de días cuyo patrón atmosférico (según el modelo utilizado) es cercano al prototipo. ■

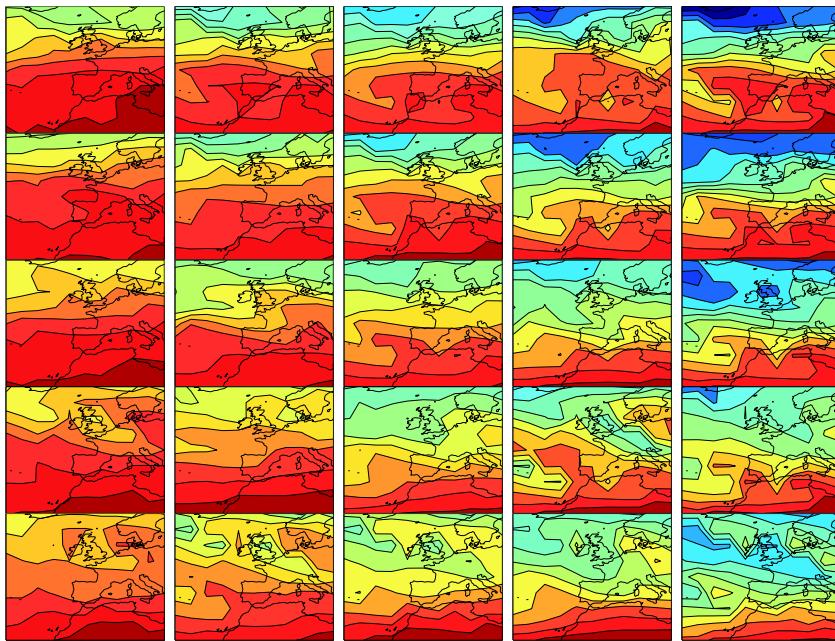


Figura 5.26: Campos de temperatura en 500mb correspondientes a los prototipos de una subejilla 5×5 de la SOM mostrada en la Figura 5.25.

Ejemplo 5.7 (Clasificación de Fenómenos Adversos). En este ejemplo se analiza el problema de la clasificación de situaciones atmosféricas relacionadas con fenómenos adversos de precipitación diaria. La clasificación se ha restringido a aquellas fechas en las que se ha observado un fenómeno adverso, $\text{Precip} > 40 \text{ mm}$, en alguno de los observatorios de la red de estaciones completas del INM en la Península y Baleares. Se ha realizado una clasificación de los patrones atmosféricos en 16 clases con una SOM para el período ERA40, considerando una configuración de los patrones atmosféricos similar a la dada por el Modelo 1 (zona NAO) de la Sección 1.9.4. La Figura 5.27 muestra los campos de geopotencial en 1000 mb para los prototipos de los grupos obtenidos; por otra parte, la Fig. 5.28 muestra los patrones de precipitación asociados a cada grupo; estos patrones se han obtenido interpolando los valores de las estaciones de la red de estaciones completas del INM. En esta figura pueden observarse modos claros de precipitación en Galicia, Levante, etc., mostrando una clara conexión entre los patrones atmosféricos y los fenomenológicos.

Este ejemplo muestra la posibilidad de obtener modelos de predicción local que relacionen la configuración de los campos atmosféricos con eventos observados en superficie de especial interés. Por ejemplo Cavazos (1997) utiliza un esquema similar para predecir la precipitación observada en superficie en México. ■

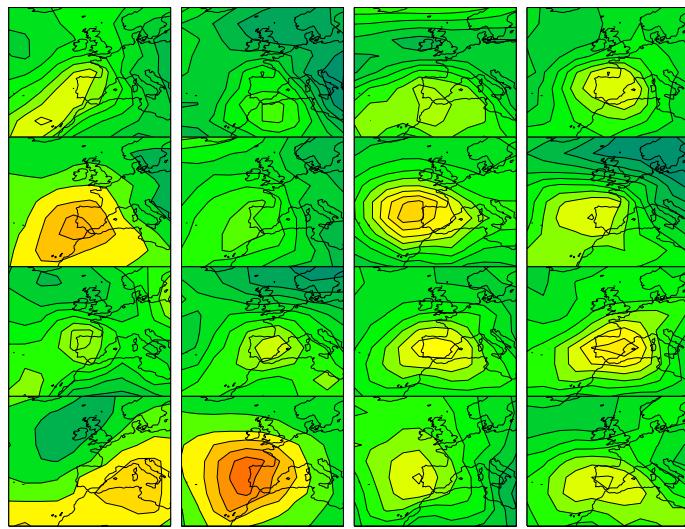


Figura 5.27: Campos de geopotencial en 1000mb correspondientes a los prototipos de una SOM 4×4 entrenada con los días del período ERA40 asociados a precipitaciones fuertes en superficie.

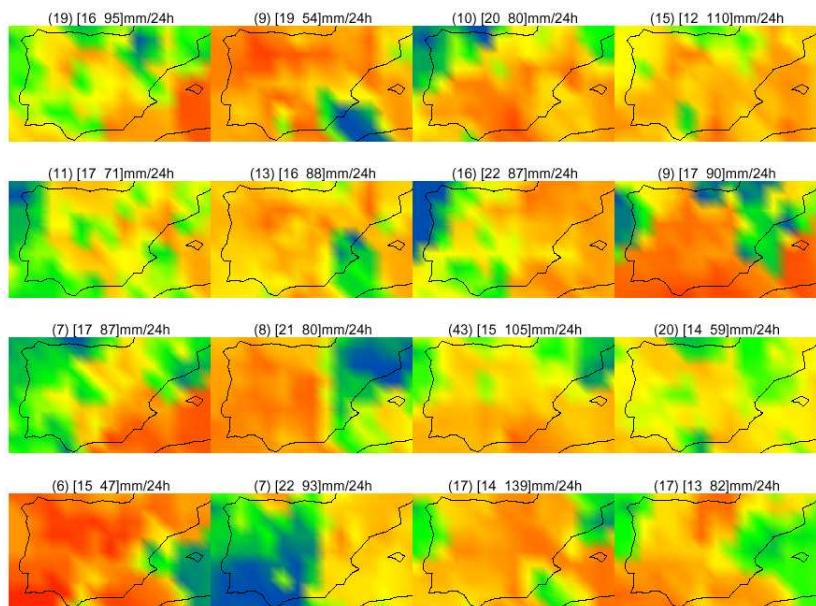


Figura 5.28: Patrones fenomenológicos de precipitaciones fuertes asociados a los grupos de la SOM. Entre paréntesis se muestra el número de elementos de cada grupo y entre corchetes el rango de precipitación observada.

5.8.1. MeteoLab: Predicción con Redes Neuronales

En esta sección se ilustra la aplicación de las redes neuronales en el mismo ejemplo de predicción local analizado con técnicas de regresión en la Sec. 3.4.4. En este caso se considera la precipitación para analizar el potencial de los modelos no lineales en comparación con la regresión lineal. Se trata de obtener predicciones locales diarias de precipitación en distintas estaciones españolas del GSN a partir de patrones atmosféricos definidos sobre la península Ibérica.

Lo primero que se hace una vez cargados los datos es estandarizarlos, para evitar un mal condicionamiento del problema. A continuación, se selecciona un conjunto de entrenamiento y un conjunto de test para poder validar la red, y comprobar que no se produce sobreajuste (un 60 % de los datos se utilizan para entrenamiento y el resto para test). Se considera una red neuronal con 10 neuronas de entrada (las 10 primeras CPs de los patrones) y una neurona en la capa de salida, con función de activación lineal para dar una predicción determinista de la precipitación. El código completo del ejemplo está contenido en el directorio **MeteoLab/NeuralNets** de MeteoLab.

```

dmn=readDomain('Iberia');
[EOF,CP]=getEOF(dmn,'ncp',10);

Example.Network={'GSN'}; Example.Stations={'Spain.stn'};
Example.Variable={'Precip'};
dates=[dmn.startDate,dmn.endDate];
[data,Example]=loadStations(Example,'dates',dates,'ascfile',1);
Example.Info.Name(4,:)
data=data(:,4); %we take data for Navacerrada station

j=find(isnan(data)==0); % Selecting days with no missing data
X=CP(j,:); Y=data(j,:);
[X,mX,dX,Y,mY,dY]=prestd(X,Y); % Data standarization

e=fix(size(X,2)*0.60); % Selecting 75 percent of data for training
Xt=X(:,1:e); Yt=Y(:,1:e);

nnet=newff([min(Xt,[],2),max(Xt,[],2)],...
[8 4 1],{'logsig','logsig','purelin'});
nnet.trainParam.epochs=200;
nnet=train(nnet,Xt,Yt);

P=sim(nnet,X);
Pred=((P.*repmat(dY,[1 size(P,2)]))+repmat(mY,[1 size(P,2)]))';
Obsr=data(j,:);

figure; plot(Obsr(e+1:end,:),'b');
hold on; plot(Pred(e+1:end,:),'r');
figure, plot(Obsr,Pred,'.')

```

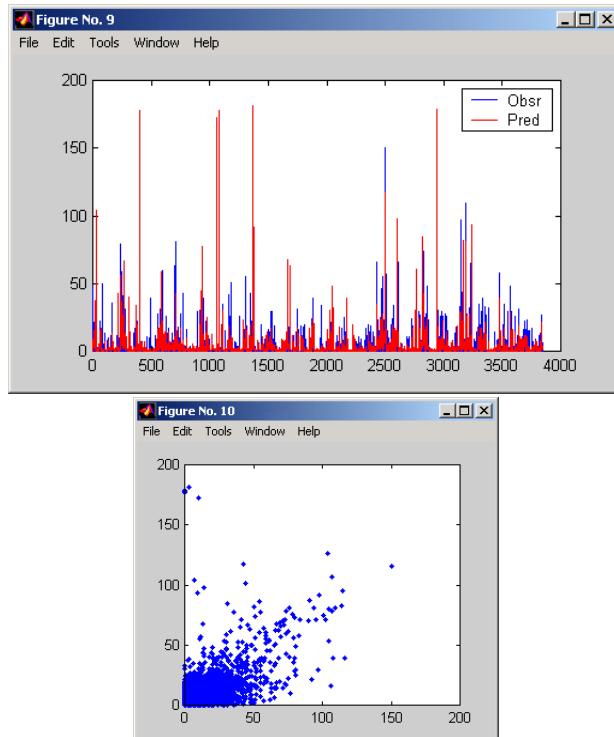


Figura 5.29: (superior) Predicción determinista de la precipitación en Navacerrada (Madrid) usando una red neuronal 10:8:4:1; en rojo se muestra el valor observado y en azul el previsto; (inferior) valor previsto frente al valor observado.

También se pueden utilizar las redes neuronales para dar una predicción probabilística de eventos discretos. Para ello, basta considerar una función de activación de tipo logístico (con valores entre 0 y 1) en la capa de salida. A continuación se muestra el código para la predicción probabilística del evento *Precip > 0.1 mm* en la estación de Navacerrada (Madrid) usando una red neuronal de tipo 10:8:4:1 con salida de tipo logístico. También se hace uso de las herramientas de validación descritas en el Cap. 7 para verificar las predicciones resultantes durante un período de test de aproximadamente 10 años.

```

Y=data(j,:); Y=Y>0.1;
nnet=newff([min(Xt,[],2),max(Xt,[],2)],...
[8 4 1],{'tansig','tansig','logsig'});
nnet.trainParam.epochs=200;
nnet=train(nnet,Xt,Yt);
P=sim(nnet,X);
Pred=P'; Obsr=Y';

```

```
makevalidation(Obsr,Pred)
```

La Fig.5.30 muestra las curvas ROC y de valor económico, y la Fig. 5.31 muestra la resolución y fiabilidad, de las predicciones. Se puede observar que las predicciones obtenidas tienen gran pericia para predecir el evento, además de una buena fiabilidad y resolución.

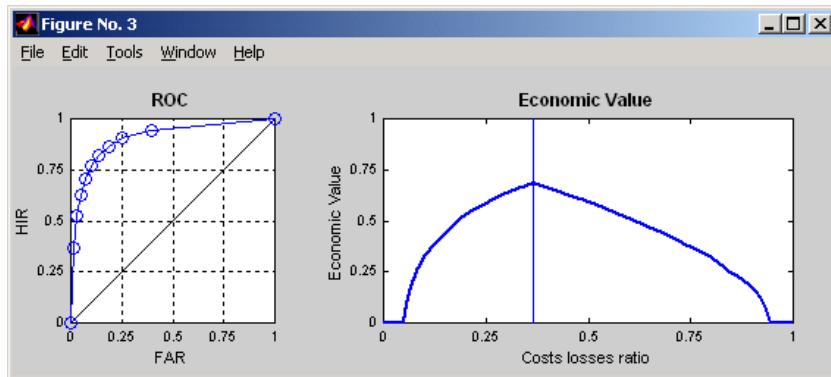


Figura 5.30: Validación (curvas ROC y de valor económico) para predicción probabilística del evento $Precip > 0.1mm$ en Navacerrada (Madrid) usando una red neuronal 10:8:4:1.

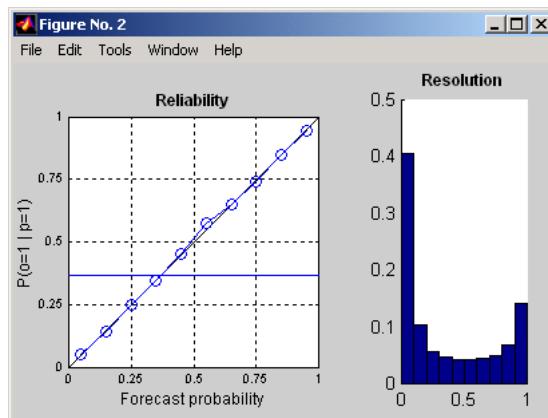


Figura 5.31: Validación (fiabilidad y resolución) para predicción probabilística del evento $Precip > 0.1mm$ en Navacerrada (Madrid).

5.9. Predicción a Plazo Medio con SOM

En la Sec. 3.7 se describió el método estándar de predicción basado en análogos. Una limitación importante de esta técnica es que siempre considera la misma cantidad de análogos, sin tener en cuenta la estructura del espacio donde se buscan los vecinos cercanos del patrón previsto. Este hecho no es consistente con la distribución de patrones atmosféricos, ya que éstos no siguen una distribución uniforme, sino que algunas regiones del espacio están más pobladas que otras (ver Toth, 1991, para un análisis detallado de la distribución de patrones atmosféricos). Por tanto, se requieren conjuntos de análogos de tamaño variable que se adapten a las características de la región del espacio donde se halle el patrón previsto. Por otra parte, los algoritmos de análogos son computacionalmente costosos, ya que la búsqueda de vecinos involucra el cálculo de las distancias del patrón previsto con todos los patrones de la base de datos de re-análisis.

Los algoritmos de agrupamiento proporcionan una solución simple para estos problemas. Estas técnicas permiten dividir la base de datos en grupos C_i , caracterizados por un patrón prototipo \mathbf{v}_i . Cada grupo resultante puede ser utilizado como conjunto de análogos para aquellos patrones que sean posteriormente clasificados en el grupo (patrones más cercanos a \mathbf{v}_i , que a cualquier otro prototipo). Por tanto, el cálculo de distancias sólo involucra al patrón previsto y a los prototipos, lo que reduce significativamente el tiempo de computación. La técnica de agrupamiento también resuelve los problemas causados por las inhomogeneidades de la distribución de re-análisis, ya que el número de elementos en cada uno de los grupos se puede adaptar automáticamente a la distribución de los patrones atmosféricos utilizando un algoritmo apropiado. En la literatura se han aplicado distintas técnicas de agrupamiento en este problema. Por ejemplo, Hughes et al. (1993) describe una algoritmo de downscaling que utiliza una técnica de agrupamiento llamada CART (ver también Zorita et al., 1995), Gutiérrez et al. (2004) describen un método para predicción a corto plazo utilizando k -medias. Por otra parte, Cavazos (1997) y Cofiño (2004) aplican redes auto-organizativas (SOM) para agrupar patrones atmosféricos similares en una base de datos de reanálisis. Esta técnica permite generalizar el método de análogos mostrado en la Fig. 3.22 y aplicarlo de forma óptima para analizar predicciones por conjuntos (EPS). También permite estimar modelos de regresión locales para cada uno de los grupos resultantes (Cavazos, 2000) (por una parte, este método podría ser mejorado aplicando algún método de estimación no lineal a los grupos resultantes (ver, por ejemplo, Fontela-Romero et al., 2002)).

La Fig. 5.32 muestra el esquema de downscaling basado en SOM, y su aplicación a las predicción por conjuntos. En la parte derecha, cada miembro de la predicción (por ejemplo, cada uno de los 51 miembros del sistema de predicción a plazo medio del ECMWF; ver Sec. 1.7.3) corresponde a un grupo de la retícula de la SOM. Por tanto, el conjunto de predicciones define una distribución de probabilidad sobre la retícula de la SOM, indicando la probabilidad prevista para los distintos grupos. A continuación, para predecir una variable (precipita-

ción, etc.) en una localidad de interés, se asocia a cada grupo un valor previsto (la media, etc.) en base a las observaciones de los días de reanálisis correspondientes. Esto permite traducir la probabilidad sobre la retícula dada por el EPS en una probabilidad sobre la variable a predecir. Además, la dispersión del EPS se puede cuantificar en términos de la dispersión de la probabilidad resultante sobre la retícula.

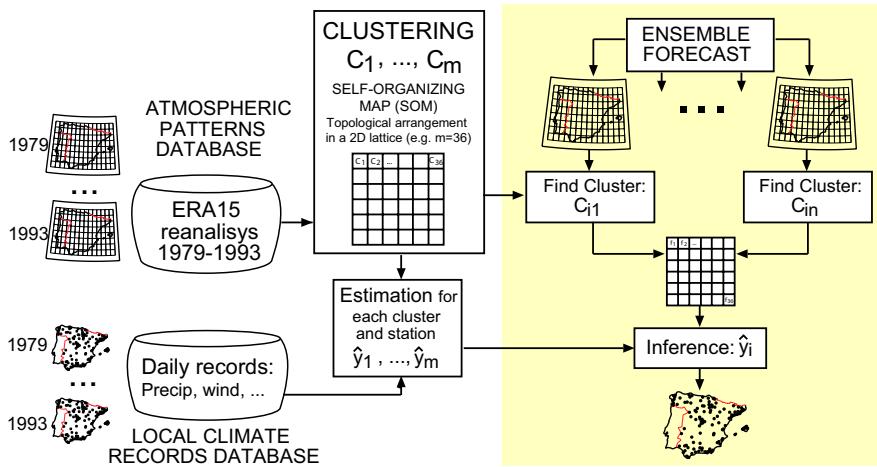


Figura 5.32: Esquema del algoritmo de downscaling por agrupamiento utilizando redes auto-organizativas en una retícula 2D.

Por tanto, el uso de la SOM como técnica de agrupamiento permite obtener:

- *Una predicción probabilística y numérica*, a partir de algún estadístico central, o algún percentil, de la función de probabilidad resultante de combinar la probabilidad de la SOM y la probabilidad fenomenológica de cada grupo.
- *Una medida de predecibilidad* (o confianza) basándose en algún estadístico de dispersión de la probabilidad de la SOM. Obsérvese, que la dispersión estadística de la función de probabilidad, está relacionada con la dispersión de los patrones de los miembros de la predicción por conjuntos, ya que la SOM conserva las distancias del espacio real en la retícula 2D, que sirve de soporte a la función de probabilidad. Por tanto, la dispersión de los miembros de la predicción en el espacio real redundará en una mayor dispersión de la probabilidad sobre la retícula de la SOM (Eckert and Ambühl, 1996).

A continuación se analiza la aplicación de esta técnica a la predicción por conjuntos de plazo medio. Para ello se consideran las salidas operativas del modelo de predicción por conjuntos del ECMWF, para el período DEF 1998 y 2000 (180 días). Para cada uno de los días, este modelo realiza una predicción hasta $D + 9$ a partir de una condición inicial, a esta predicción se la denomina

control. Después se generan otras 50 predicciones, a partir de perturbaciones de la condición inicial, aplicando el método de vectores singulares. Por tanto, para cada alcance de predicción entre $D + 1$ y $D + 9$ se tienen 51 patrones atmosféricos que se consideran igualmente probables, para la descripción del patrón atmosférico previsto.

Tal como se comentó anteriormente, cada conjunto de 51 patrones define una función de probabilidad que tiene como soporte la retícula de la SOM. Por ejemplo, la Fig. 5.33 muestra las funciones de probabilidad dadas por los 51 patrones para cada uno de los alcances obtenidos a partir del día 24 de febrero de 1998. Como puede observarse en esta figura, la predicción para los primeros dos días se centra en un sólo grupo, mientras que la dispersión comienza a crecer a partir del tercer día, a medida que aumenta la incertidumbre sobre la predicción. Los dos números que aparecen sobre la distribución, indican la entropía (izquierda) y la desviación típica (derecha).

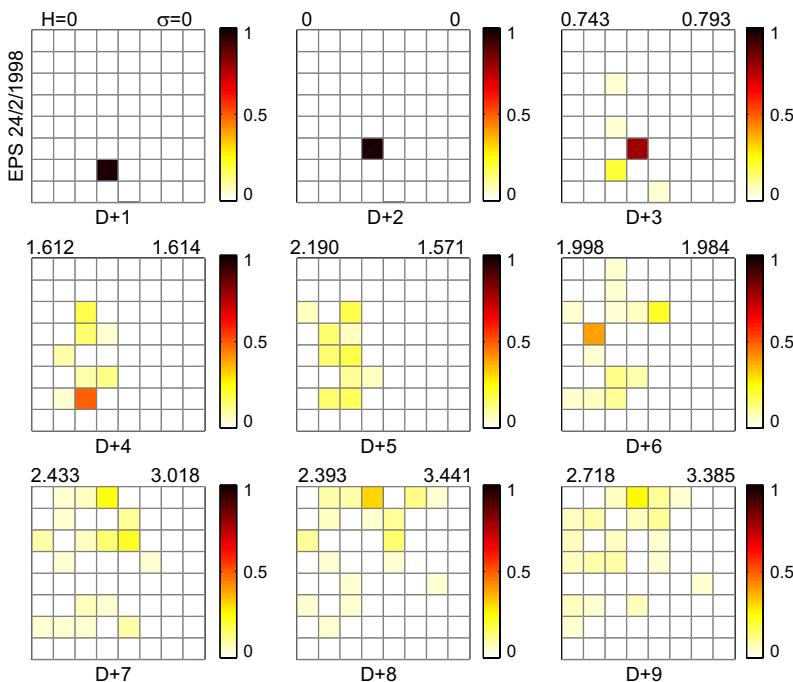


Figura 5.33: Probabilidades definidas sobre la SOM por los miembros de la predicción por conjuntos del día 24/2/1998 para los alcances entre $D + 1$ y $D + 9$.

Sin embargo, no todas las situaciones son similares a la anterior, ya que se puede encontrar un período estable donde los miembros del conjunto de predicciones tienen escasa variabilidad (ver Fig. 5.34), y también una situación más impredecible, en la que la dispersión es considerable desde el primer día de alcance (ver Fig. 5.35).

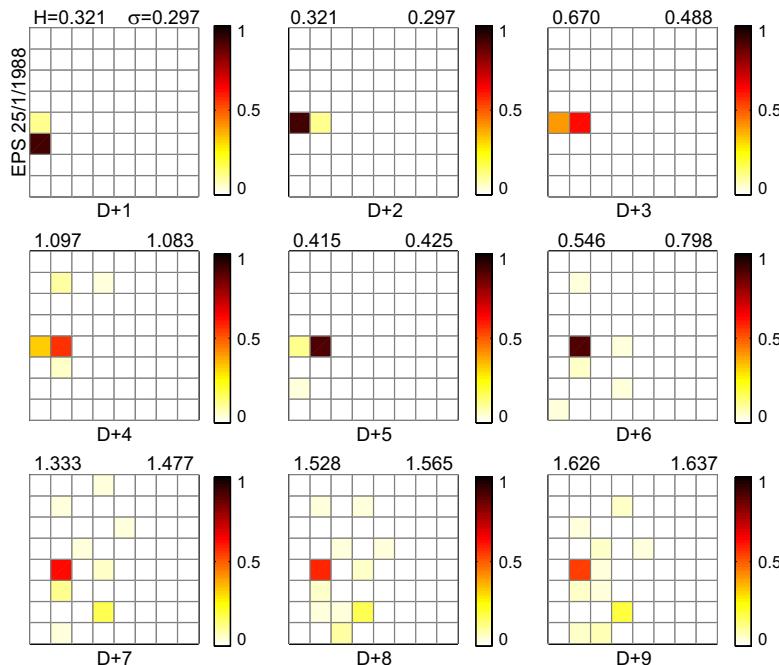


Figura 5.34: Probabilidades definidas sobre la SOM por los miembros de la predicción por conjuntos del día 25/1/1998 para los alcances entre $D + 1$ y $D + 9$.

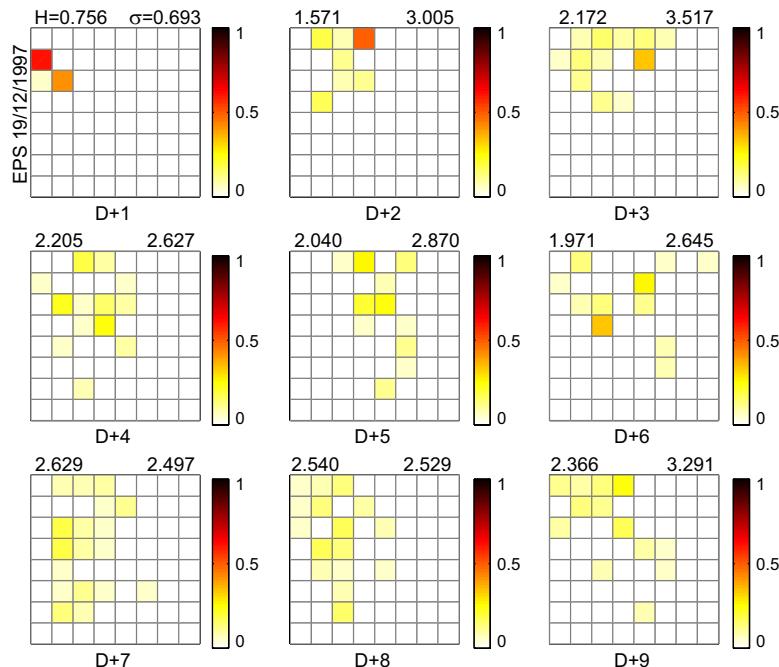


Figura 5.35: Probabilidades definidas sobre la SOM por los miembros de la predicción por conjuntos del día 19/12/1997 para los alcances entre $D + 1$ y $D + 9$.

A pesar de la variabilidad de unas situaciones a otras, se puede esperar que, en promedio, la dispersión de las funciones de probabilidad crezca a medida que aumenta el alcance de la predicción (y, por tanto, su incertidumbre). Para constatar este hecho se han calculado los respectivos cuartiles de las dispersiones para los distintos alcances, considerando las predicciones realizadas en los 180 días. Las Figs. 5.36 (a) y (c) muestran la entropía y la desviación típica de la probabilidad para el alcance D . En estas figuras puede observarse que la desviación tiene algunos valores fuera de rango durante los primeros días de la predicción y comienza a saturar a partir del quinto día, y por otra parte, el comportamiento de la entropía es más monótono. Cada una de estas dos medidas da una información distinta sobre la variabilidad de cada uno de los patrones. Por ello, se considera también un parámetro de dispersión definido como la suma de los parámetros anteriores. En la Fig. 5.36(d) se muestra que este parámetro presenta la característica de saturación de la desviación, mientras que reduce el problema de los valores fuera de rango. Por último, la Fig. 5.36(b) muestra la entropía relativa de un día respecto al anterior. Como se puede observar en la figura, este valor decrece de forma continua a medida que se incrementa el alcance de la predicción; se observa un gran crecimiento incremental de la dispersión los primeros cinco días, mientras que la dispersión aumenta más lentamente a partir del quinto día de alcance. Este hecho coincide con la saturación indicada por la desviación típica.

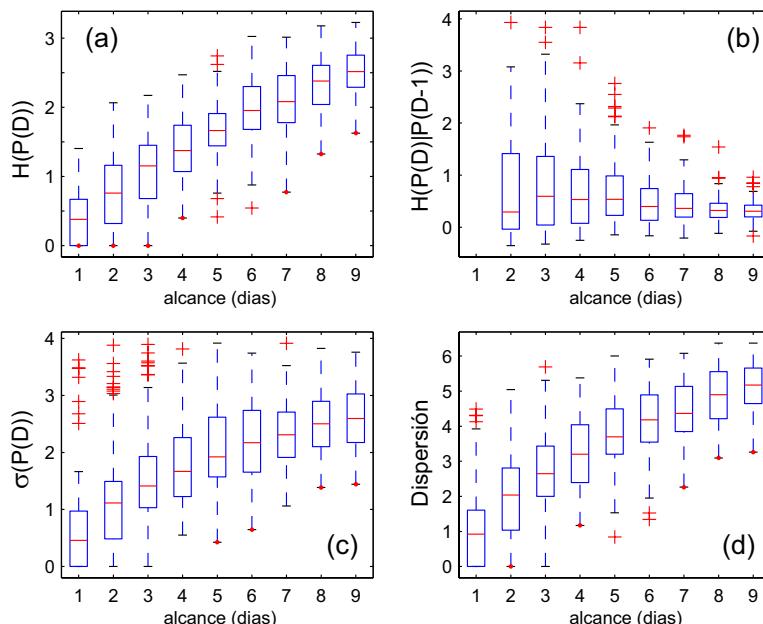


Figura 5.36: (a) Entropía de la probabilidad para el alcance D , $H(P(D))$; (b) Entropía relativa de la probabilidad de un día, respecto a la probabilidad del día anterior, $H(P(D)|P(D-1))$; (c) desviación típica de la probabilidad y (d) parámetros de dispersión, definido como la suma de la desviación típica y la entropía.

La validación del sistema de predicción probabilística por conjuntos ha mostrado ser superior a un sistema determinista clásico basado en una única integración (ver, por ejemplo, Richardson, 2000) en el medio plazo. A continuación se comprueba que este resultado también es cierto cuando se utilizan los patrones atmosféricos para realizar predicciones locales. Para ello, se ha entrenado una SOM considerando el patrón de la Cuenca Norte de la Península Ibérica mostrado en la Fig. 1.22(c) y se ha aplicado el método de downscaling descrito en la Fig. 5.32 para predecir el evento $P(Precip > 0.1mm)$ en 100 estaciones de la Cuenca Norte. La Fig. 5.37 muestra los Brier Scores (BS) diarios promediados para todas las estaciones obtenidos con los 50 miembros, perturbados, de la predicción por conjuntos (etiqueta EPS) y una única predicción dada por el control (el patrón correspondiente a la condición inicial sin perturbar) (etiqueta control). En esta figura puede verse que el downscaling realizado con la predicción por conjuntos es claramente superior al downscaling tradicional, basado en una única predicción entre los días quinto y octavo, siendo ligeramente mejor el resto de los días. En concreto el noveno día el comportamiento del sistema de predicción por conjuntos comienza a perder pericia respecto al control.

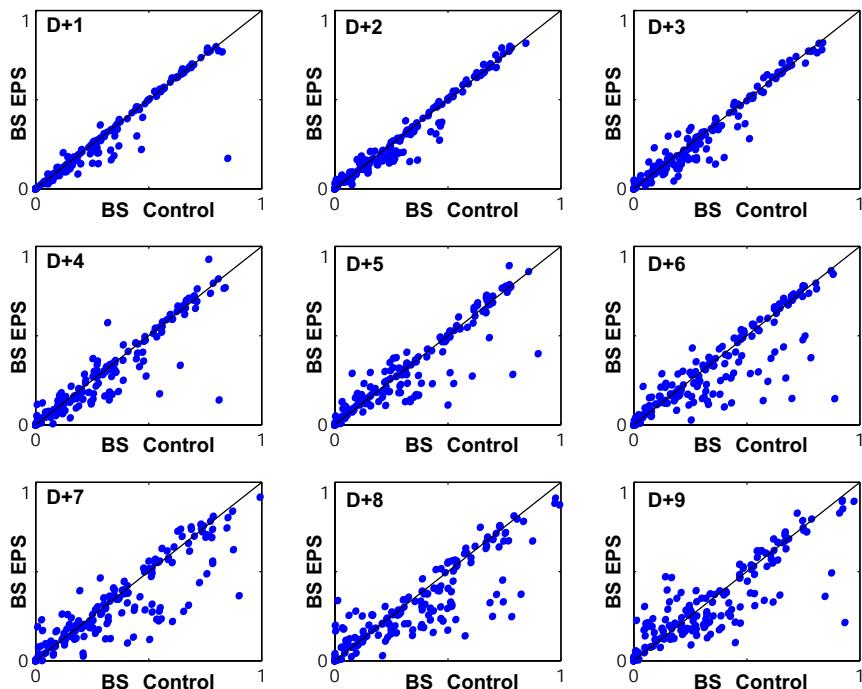


Figura 5.37: Brier Score (BS) de la predicción local realizada con todos los miembros del conjunto (BS EPS) y sólo con el control (BS Control) para alcances entre uno y nueve días. Los valores representados son las medias para todas las estaciones principales de la Cuenca Norte.

Por último, se muestra que la medida de dispersión de la probabilidad de la

SOM, definida como la suma de la desviación típica y de la entropía, está correlacionada con la predecibilidad de la situación correspondiente. Para ello, en la Fig. 5.38 se muestra la dispersión frente al error BS obtenido en la respectiva predicción realizada, para los distintos alcances del modelo. Se puede observar que a partir del tercer día de predicción, aparece una relación entre, la medida de dispersión definida a partir de la SOM, y el error que se produce en la predicción (confianza en la predicción, o predecibilidad). Esta relación indica que el rango de errores posibles crece linealmente al crecer la dispersión.

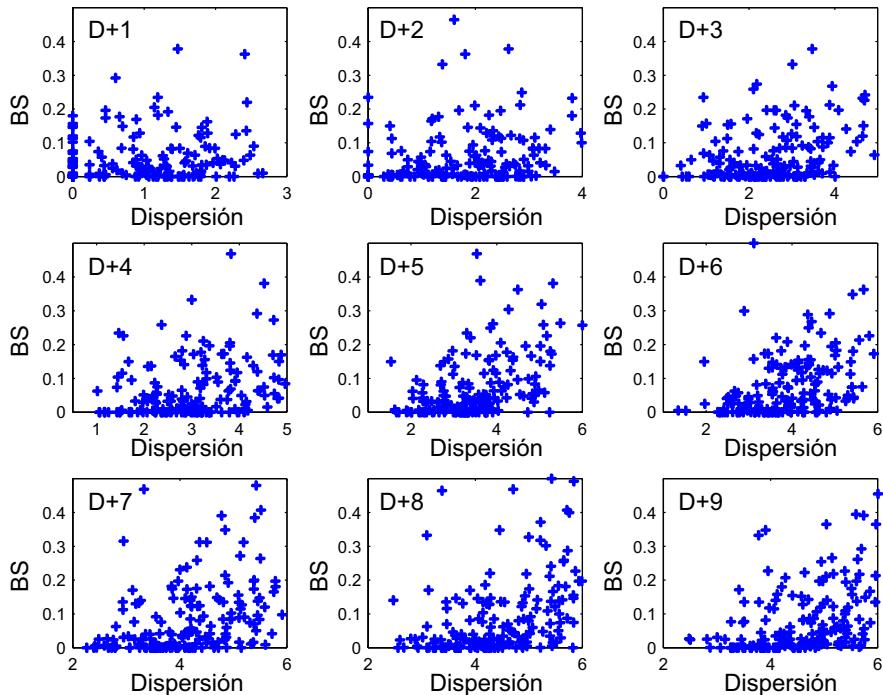


Figura 5.38: Brier Score (BS) de la predicción local realizada con el conjunto de predicciones frente a la medida de dispersión (desviación típica más entropía) para alcances entre uno y nueve días. Los BS representados son las medias para todas la estaciones principales de la Cuenca Norte.

5.10. Predicción Estacional

En secciones anteriores el objetivo de la predicción a corto y medio plazo era estimar numéricamente o probabilísticamente el valor de una cierta variable atmosférica con una anticipación de entre uno y nueve días. Este es el límite que se suele fijar como umbral de predecibilidad de la atmósfera, aunque este límite no es un valor fijo que haya sido obtenido teóricamente (obsérvese que además este valor oscilará temporal y espacialmente). A tiempos superiores, la no-linealidad de los modelos de circulación atmosférica, junto con los errores

asociados a las observaciones y las aproximaciones de los modelos utilizados, impiden una predicción numérica acertada. Por tanto, en principio, el término “predicción estacional” puede parecer incorrecto. Sin embargo, aunque no es posible precisar el valor concreto de una cierta variable con una antelación de un mes (o un trimestre), en ocasiones sí resulta posible dar algún tipo de información útil asociada a la misma; por ejemplo, se podría tratar de predecir si el valor medio del siguiente mes, o trimestre, será significativamente inferior o superior a la media climatológica correspondiente (es decir, hablando en términos de precipitación, si se espera un mes o estación más húmeda o seca de lo normal). Así pues, en este tipo de predicciones, se trabaja con anomalías (estimaciones de desviaciones mensuales, o trimestrales, respecto de la climatología) o con su carácter (positivo, normal, o negativo) y no con valores concretos de las variables (ver Carson, 1998, para una descripción del estado de la predicción estacional).

Los avances producidos en las técnicas de predicción por conjuntos han dado un enorme impulso a los modelos acoplados océano-atmósfera, de predicción estacional, que actualmente permiten predecir con cierta fiabilidad anomalías mensuales o estacionales durante ciertas épocas y en ciertas regiones del globo (asociadas principalmente a anomalías en la ENSO). Estos resultados han motivado la puesta en marcha operativa, de distintos productos de predicción estacional en distintos centros internacionales. Por ejemplo, el modelo System-II del ECMWF produce predicciones estacionales operativas. Asimismo también se dispone de un re-análisis (DEMETER), de un conjunto de seis modelos Europeos globales de predicción por conjuntos con distintos esquemas de perturbación (www.ecmwf.int/research/demeter/) (Palmer et al., 2004).

No sólo se ha puesto de manifiesto la posibilidad de predecir en cierta medida los patrones estacionales de circulación sinóptica, sino que también, se ha mostrado su asociación con anomalías regionales y locales de precipitación, y de temperatura, tanto de la NAO como de la ENSO (ver, por ejemplo, Hurrell, 1995; Rodríguez-Fonseca and Serrano, 2002). Por ejemplo, si se considera por una parte la temperatura media mensual del agua del mar (SST) frente a la costa de Chicama en el Norte de Perú (representante de la oscilación sinóptica del ENSO), la precipitación en la ciudad de Piura (medias mensuales) y el caudal del río Piura, se puede comprobar fácilmente que existe una relación entre estas variables, que permite trasladar anomalías sinópticas a anomalías locales (ver Ejemplo 2.8). Sin embargo, los efectos en otras regiones cercanas son distintos e incluso opuestos (épocas de sequía asociadas con los episodios fuertes de El Niño en la sierra de Perú y en el altiplano). Con este ejemplo se pone de manifiesto la utilidad de disponer de técnicas de downscaling, que permitan trasladar las anomalías predichas por los modelos estacionales a anomalías regionales y locales de variables de interés para la actividad humana.

Una dificultad para realizar estudios de predicción local a partir de predicciones mensuales y estacionales es que, hasta la fecha, no se disponía de una base de datos con predicciones durante un período de tiempo suficientemente representativo. Durante los últimos años, el proyecto Europeo DEMETER ha creado una base de datos con las predicciones estacionales de seis modelos

acoplados atmósfera-océano con nueve miembros en cada caso (obtenidos aplicando distintos esquemas perturbativos). Para ello se ha utilizado información del re-análisis ERA-40 (1957-2002). Los modelos estacionales se inicializan con los datos de ERA40 cuatro veces al año (Febrero, Mayo, Agosto y Noviembre) y son integrados para un alcance de seis meses, almacenando las variables que proporcionan una detallada descripción de la atmósfera predicha día a día. En esta sección se realiza un estudio sistemático de la predecibilidad estacional en latitudes tropicales, teniendo en cuenta el fenómeno de El Niño.

Para ello, se considera la región Norte de Perú que se muestra en la Fig. 5.39, donde se dispone de datos diarios de precipitación para dos estaciones: Sausal de Culucán y Morropón.

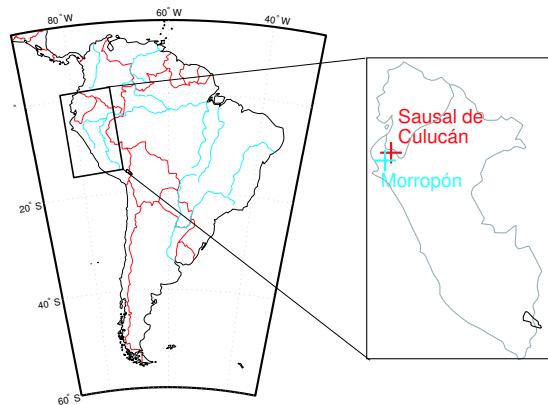


Figura 5.39: Localización geográfica de Perú y de dos estaciones de la zona norte.

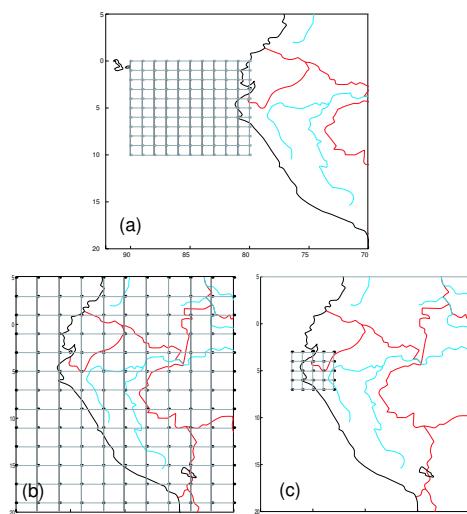


Figura 5.40: Rejillas consideradas para la definición del patrón atmosférico en la zona Norte de Perú.

Para aplicar la técnica basada en redes auto-organizativas descrita en la sección anterior se consideran distintas rejillas sobre la zona de interés de diferentes resoluciones y coberturas (ver Fig. 5.40). Sobre cada una de ellas se entrenaron distintas redes auto-organizativas con los patrones de re-análisis del período 1957-2002. Una vez realizado todo el estudio se comprobó que, de nuevo, los mejores resultados se obtuvieron con el patrón local (en este caso, con la rejilla dinámica de la Fig. 5.40(c)). Por ejemplo, la Fig. 5.41 muestra los histogramas fenomenológicos definidos por la precipitación observada en cada una de las estaciones sobre una SOM entrenada con la rejilla local. Estas dos “huellas digitales” muestran las distintas formas de llover que se producen en dos estaciones tan cercanas.

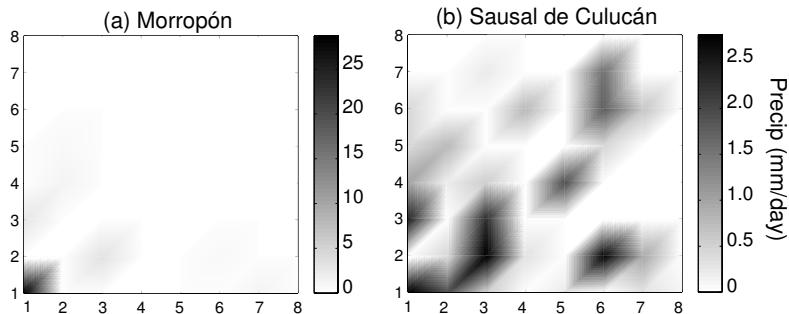


Figura 5.41: Histogramas fenomenológicos de la precipitación en (a) Morropón y (b) Sausal de Culucán obtenidos sobre una SOM (el valor de cada grupo se obtiene como el promedio de los días pertenecientes al grupo).

En este trabajo se utilizan sólo los modelos del ECMWF, Météo France, UKMet Office y Max Plank Institute de DEMETER, con nueve miembros en cada uno. Dado que el fenómeno de El Niño tiene especial incidencia en el Norte de Perú a principios de año, se consideran las predicciones de los modelos realizadas en Noviembre para el trimestre Diciembre-Febrero; por tanto, se tratan de predecir las anomalías locales de precipitación a partir de predicciones obtenidas entre uno y tres meses de anticipación. En este caso, se tienen varias opciones para combinar esta información. Se puede considerar cada modelo por separado y añadir un nuevo modelo, que consiste en la mezcla de todos (multi-modelo), y luego comparar la eficiencia de todos los modelos frente al multi-modelo. La Fig. 5.42 muestra el proceso seguido en este caso.

En la Fig. 5.43 se muestran los resultados obtenidos. Para cada trimestre Diciembre-Febrero de cada año del período 1983-1998 se consideraron las predicciones obtenidas para los 90 días por cada miembro de cada modelo. Por tanto, cada trimestre se obtuvieron 9 predicciones para cada modelo de las cuales se obtuvo el valor promedio, que es considerado como la predicción del modelo; por otra parte, se consideró la predicción conjunta dada por todos los modelos (la media de las predicciones de cada uno de los modelos). La observación real (cuando está disponible) se muestra en la línea central con un cuadrado, mientras que la predicción del multi-modelo se muestra con una

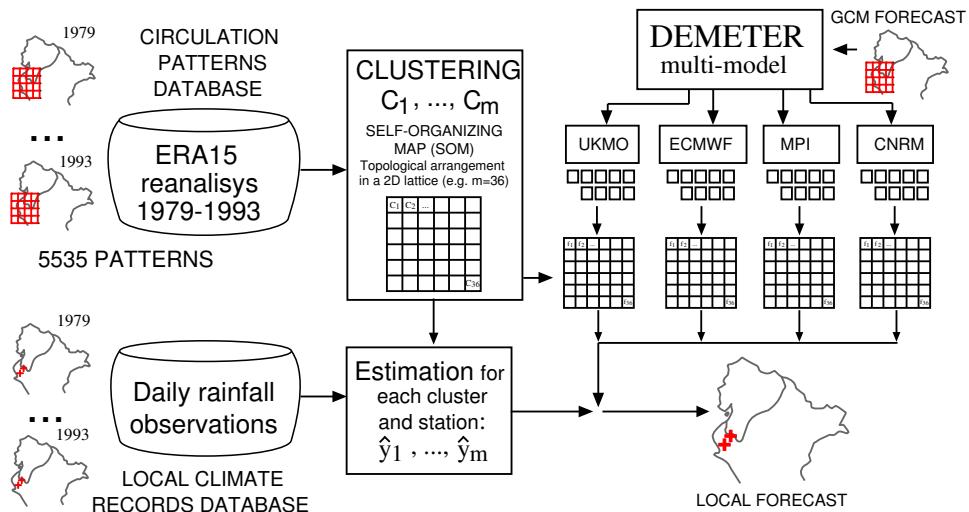


Figura 5.42: Esquema de downscaling para DEMETER. Cada modelo produce 9 patrones atmosféricos diarios durante un período de seis meses.

cruz. Las cajas muestran los respectivos cuartiles de precipitación en la estación considerada, para todo el período de estudio. Por tanto, cuando la predicción está por encima de cuartil 75, o por debajo del 25, se considera que se está prediciendo una anomalía positiva, o negativa, de precipitación, respectivamente. A partir de esta figura se puede observar que el sistema predice con acierto la anomalía positiva de precipitación durante los dos episodios de El Niño fuertes de los años 1982/83 y 1997/98. Además también se ha probado que durante estos dos episodios de El Niño, se podía predecir la anomalía positiva con mayor anticipación. Por ejemplo, el modelo Francés CNRM y el multi-modelo son capaces de predecir la anomalía de precipitación para el año 1982/83 en la estación de Moropón con cuatro meses de antelación, mientras que todos los modelos proporcionan el valor correcto de la anomalía para el año 1997/98 con cuatro meses de antelación (ver Cano et al., 2003, para más información).

De forma análoga al estudio realizado en el caso de predicción a medio plazo para caracterizar la predecibilidad de las predicciones, en este se consideran las probabilidades definidas en la SOM por cada uno de los modelos. La Fig. 5.44 muestra los histogramas definidos sobre la SOM por los patrones previstos para el período Diciembre-Febrero por el multi-modelo DEMETER (primera fila) y cada uno de los cuatro modelos individuales con 1-3 meses de anticipación. La columna de la izquierda muestra la climatología (período 1983-1998); la columna central muestra el año no Niño 1989/90 (similar a la climatología); la columna de la derecha muestra el trimestre del año Niño 1997/98. La Fig. 5.45 muestra la entropía relativa del histograma de cada modelo respecto al histograma de la climatología de DEMETER. En esta figura se puede comprobar que este índice es un buen indicador de la predecibilidad de cada situación.

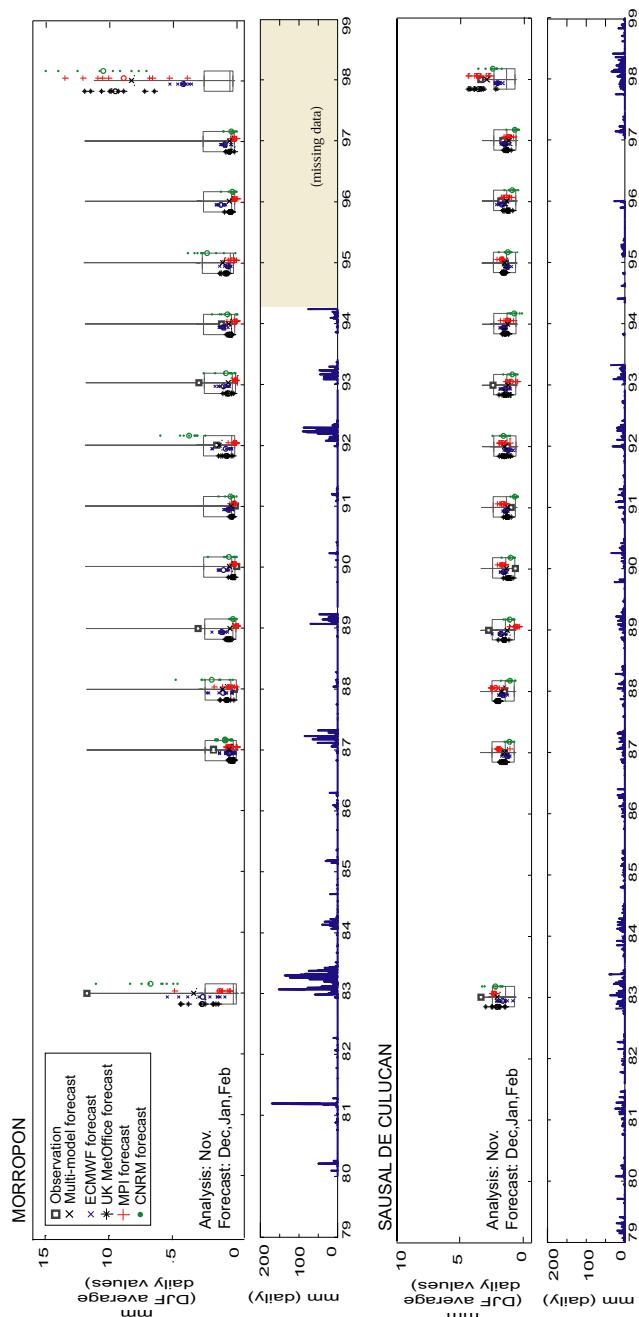


Figura 5.43: Predicción local de precipitación en Morropón y Sausal de Culucán para los cuatro modelos y para el multi-modo en el período 1983-1998.

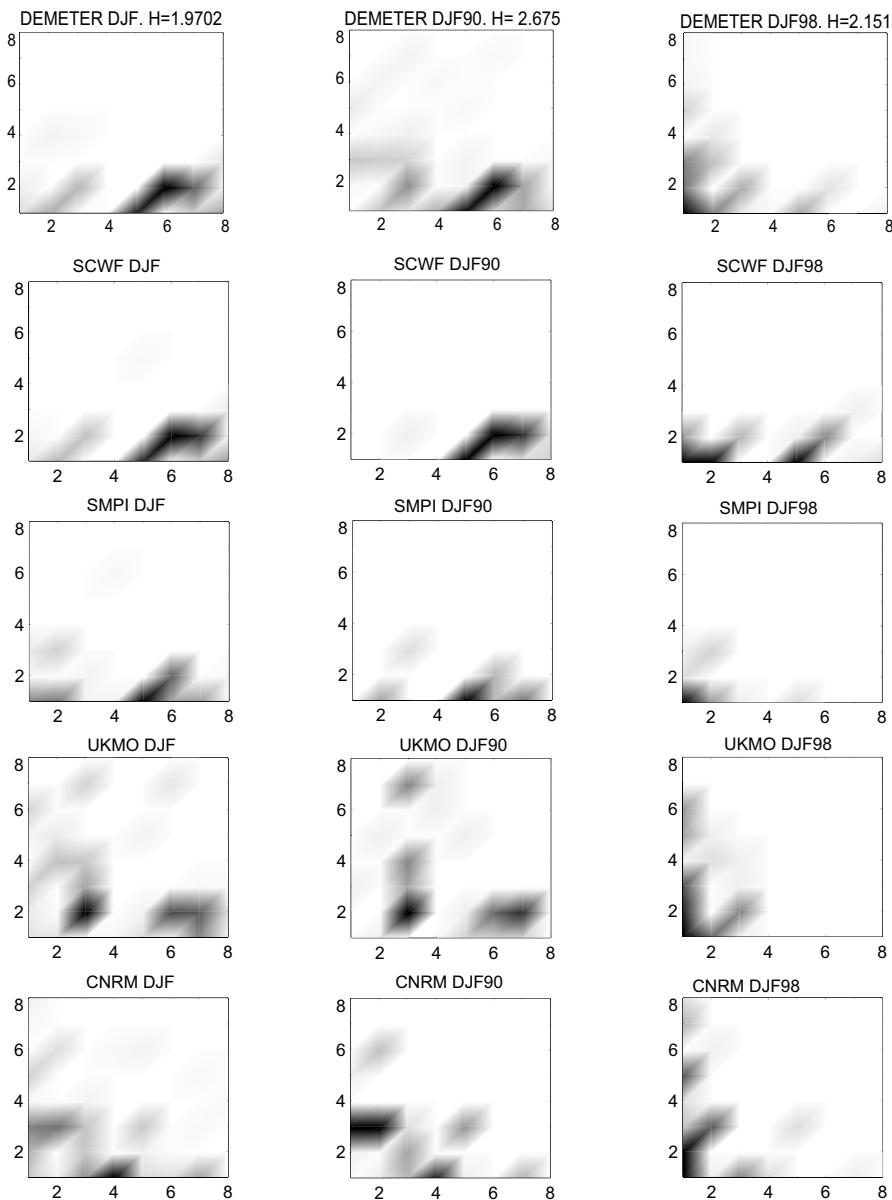


Figura 5.44: Histogramas definidos sobre la SOM por los patrones previstos para el periodo Diciembre-Febrero por el multi-modelo DEMETER (primera fila) y cada uno de los cuatro modelos individuales. La columna de la izquierda muestra la climatología (periodo 1983-1998); la columna central muestra un año no Niño (similar a la climatología); la columna de la derecha muestra el trimestre de un año Niño. En la primera fila se muestran las entropías de cada probabilidad.

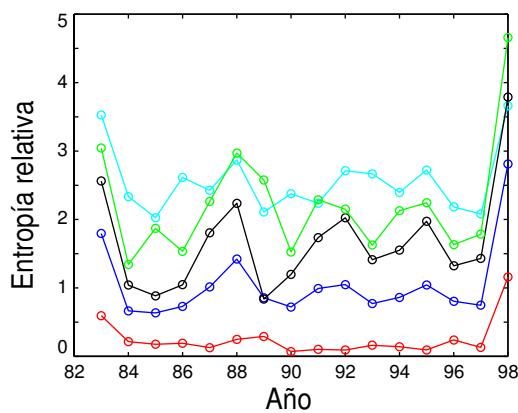


Figura 5.45: Entropía relativa de cada modelo cada año respecto a la climatología de DEMETER. Los colores de los modelos son los mismos que en figuras anteriores.

Parte IV

Apéndices

CAPÍTULO 6

Implementación Operativa. El Sistema PROMETEO

6.1. *Introducción*

En capítulos anteriores se han descrito los aspectos teóricos de los métodos de aprendizaje automático y su aplicación a la predicción local probabilística en el corto plazo. También se ha mostrado la utilidad práctica de estos nuevos métodos en la predicción operativa. En este capítulo se describe el sistema *Prometeo*, que tiene como núcleo la Toolbox *MeteoLab* descrita en los distintos capítulos del libro y permite configurar de forma sencilla distintos experimentos “a medida” de predicción local (con técnicas lineales, redes neuronales, etc.). Este sistema también incorpora una aplicación Web con la que consultar las predicciones en tiempo real, y validar las predicciones obtenidas para un problema dado, de forma que se pueda comprobar la utilidad del sistema. Además, se ha hecho uso de las tecnologías actuales de Internet que, por una parte, permiten acceder a bases de datos (que almacenan las predicciones) y, por otra, permiten ejecutar la aplicación con unos parámetros concretos de interés (región geográfica, patrón atmosférico, etc...) obteniendo las predicciones como resultado. En la Sec. 6.2.3 se describe este desarrollo.

Como ejemplo del sistema, que ayudará a entender su estructura, en la Sección 6.2 se describe la adaptación hecha para el Instituto Nacional de Meteorología (INM) para la predicción local diaria en una red de 2500 estaciones en la Península, Baleares y Canarias. El sistema final está basado en el método descrito en Gutiérrez et al. (2004). En la actualidad, esta aplicación forma parte de la cadena operativa del INM (como herramienta de apoyo) y genera diariamente predicciones operativas con algunos de los modelos que más se

utilizan en el INM (HIRLAM, con sus pasadas de 00UTC y 12UTC para alcances de D+0 y D+1; y el modelo operativo del ECMWF para la pasada de las 12UTC y los alcances D+2 y D+3). A partir de las predicciones numéricas de estos modelos, se infieren predicciones locales para las doce cuencas hidrográficas (diez en la Península, una para Baleares y otra en Canarias), cada una de las cuales tiene asociado su correspondiente dominio atmosférico. En total se realizan unas 162500 predicciones diarias (5 salidas numéricas, 2500 observatorios, y 13 predicciones para cada estación: 6 para precipitación y 7 meteoros). El manejo interactivo, rápido y eficiente en modo operativo de tal volumen de información exige una meticulosa organización de los datos.

6.2. Sistema Operativo en el INM

En esta sección se describe la adaptación del sistema *PROMETEO* realizada para el Instituto Nacional de Meteorología. Estas predicciones se pueden consultar en meteo.macc.unican.es/prometeo y también en la intranet del INM nbdc-sa.inm.es:9180/prometeo/mysqlDB/index.jsp. Para describir este sistema, se comienza describiendo la fase de preproceso y las tareas involucradas, y se finaliza con el sistema de consulta Web para predicciones operativas. La Fig. 6.1 muestra la estructura global de la aplicación, destacando sus cuatro módulos (A), (B), (C) y (D). En las siguientes secciones se describen en detalle cada uno de estos módulos.

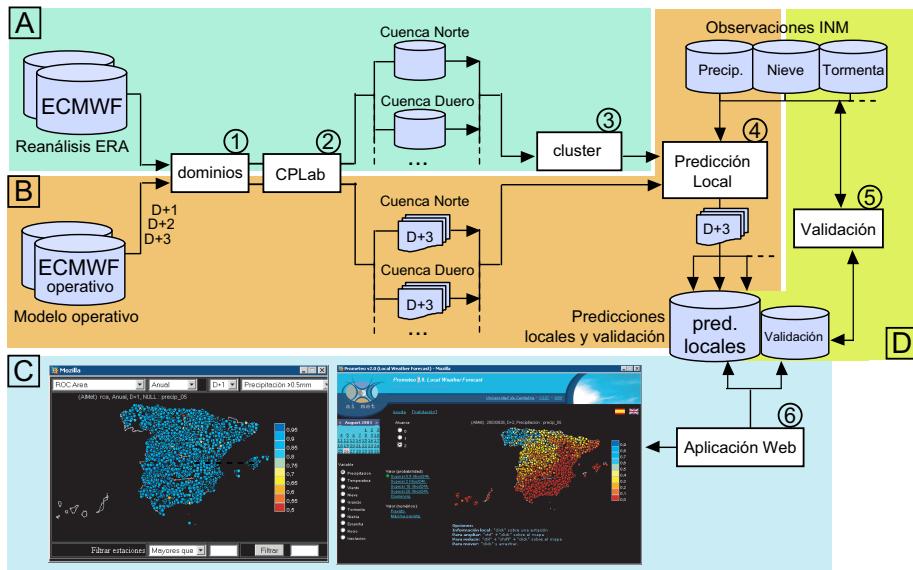


Figura 6.1: Estructura de la aplicación *PROMETEO* con las etapas: (A) configuración, (B) producción operativa, (C) consulta a través de la Web, y (D) validación.

Para evitar problemas de confusión en la descripción del sistema, únicamente se citarán las salidas del modelo operativo y el re-análisis del ECMWF. Además, la base de datos de observaciones estará limitada a aproximadamente 2500 estaciones operativas de la red secundaria del INM.

6.2.1. Configuración e Inicialización del Sistema

En esta sección se describe la fase inicial de configuración y adaptación del sistema a un entorno concreto (zona geográfica, variables, bases de datos, etc.). Concretamente se describen los elementos que intervienen en la configuración de la aplicación para la predicción local sobre la red secundaria del INM a partir de las predicciones numéricas del ECMWF. En este módulo se utiliza la base de datos de re-análisis ERA-15 del ECMWF, descrita en la Sec. 1.9, que comprende el período 1979-1994 (5569 días en total). Primero, se especifica el dominio sobre el que se definirá el estado de la atmósfera (Tarea 1 en Fig. 6.2); a continuación, en la Tarea 2, se aplica el algoritmo de componentes principales a la base de datos de reanálisis ERA40 (Sec. 2.5) para reducir la dimensión del estado de la atmósfera y, finalmente, se aplica una técnica de agrupamiento sobre los patrones obtenidos (Tarea 3). Este módulo es el que contiene el núcleo del sistema y será necesario ejecutarlo una sola vez para cada dominio como configuración previa a la fase operativa de la aplicación que es la obtención diaria de predicciones.

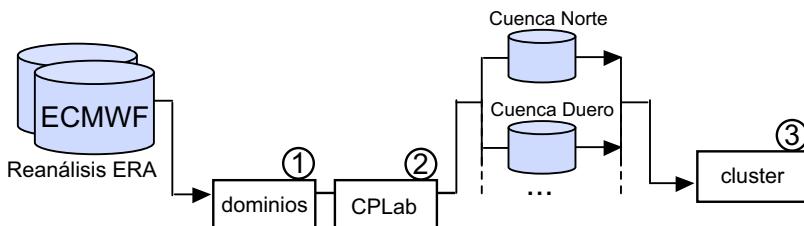


Figura 6.2: Módulo de configuración e inicialización, con la definición del *patrón atmosférico* abarcando un dominio dado (Tarea 1), la compresión de información aplicando componentes principales (2), y la ejecución del agrupamiento (3).

La figura 6.2 muestra esquemáticamente este módulo, donde se puede ver la base de datos de re-análisis y los 3 módulos siguientes:

1. *Definición de Dominios*. Es el encargado de, a partir de los dominios definidos por el usuario, extraer los datos necesarios del reanálisis para definir el patrón atmosférico y agrupar la información en forma vectorial. En la Sección 1.9.4 se había visto la forma de definir un patrón atmosférico concreto ilustrando su aplicación sobre la península Ibérica. Los mejores resultados de predicción local se obtienen con un patrón 4D de área limitada. Por ello, se han considerado 12 dominios distintos (uno para cada cuenca).

Para ello se han definido los dominios necesarios de *MeteoLab*, tal como se muestra en la Fig. 1.22. Por ejemplo, a continuación se muestra un ejemplo de definición de configuración atmosférica para la Cuenca Norte:

```
lon=-10,-9,-8,-7,-6,-5,-4,-3,-2,-1,0,1,2,3,4
lat=44,43,42,41,40,39,38,37,36
lvl=1000,850,500
tim=06,12,18,24,30
par=129,130,131,132,157

nod=2,3,4,5,6,7,8,9,17,18,19,20,21,22,23,24,32,33,34
src=../SourceDB/era10
startDate: '01-Sep-1957'
endDate: '30-Aug-2002'
path: '../PatternsData/Iberia/CuencaNorte/'
```

Figura 6.3: Definición de la configuración atmosférica de un patrón que abarca la cuenca norte en la península Ibérica.

Este fichero define un patrón atmosférico de 1° de resolución en longitud y latitud; tres niveles en altura 1000mb, 850mb, y 500mb; cinco horas de análisis 06, 12, 18, 24, y 30 UTC; y cinco variables: Temperatura (T, parámetro 129), geopotencial (Z, parámetro 130), coordenadas del viento (U y V, parámetros 131 y 132, respectivamente) y humedad relativa (H, parámetro 157). El patrón 4D resultante es:

$$\mathbf{x} = (\mathbf{x}_{06}, \mathbf{x}_{12}, \mathbf{x}_{18}, \mathbf{x}_{24}, \mathbf{x}_{30}), \quad (6.1)$$

donde $\mathbf{x}_t = (T_t^{1000}, T_t^{850}, T_t^{500}, \dots, H_t^{1000}, H_t^{850}, H_t^{500})$.

2. *Compresión de Datos.* Una vez se ha “filtrado” la base de datos de reanálisis extrayendo los patrones definidos en (6.1) con el módulo *dominio*, se aplica un análisis de componentes principales con el objetivo de reducir la dimensión de los patrones, pues estos normalmente contienen mucha información redundante, debido a la dependencia espacio-temporal. Es recomendable estandarizar los datos para evitar que las variables con mayor magnitud dominen al resto. Esta tarea es llevada a cabo por *MeteoLab*.
3. *Agrupamiento (estado de la atmósfera).* Este módulo es el encargado de realizar y almacenar el resultado de agrupar los datos de reanálisis después de haber sido filtrados y comprimidos en los módulos anteriores. Esta tarea se lleva cabo con el algoritmo k-medias de *MeteoLab*. Dados los problemas de inicialización en los algoritmos iterativos de agrupamiento, el sistema realiza varios entrenamientos que son automáticamente evaluados con el módulo de validación para contrastar la calidad de cada uno de ellos.

En la aplicación para el INM, los procesos anteriores se han aplicado a cada una de las 12 cuencas hidrográficas, almacenando las CPs del reanálisis, la matriz de transformación, y los agrupamientos realizados. Esta fase de configuración e inicialización es la que exige una mayor capacidad de cómputo y de almacenamiento, y la que condiciona la interactividad del sistema, ya que es utilizada en el proceso de explotación operativa.

6.2.2. Explotación Operativa

Este módulo es el encargado de obtener las predicciones locales. En la figura 6.4 se puede observar su estructura, incluyendo su tarea principal, denominada “predicción local”. Este módulo requiere la configuración e inicialización previa del sistema, que conlleva la definición del dominio (o dominios) y de las correspondientes componentes principales y del agrupamiento a utilizar (tareas marcadas con línea discontinua en la citada figura).

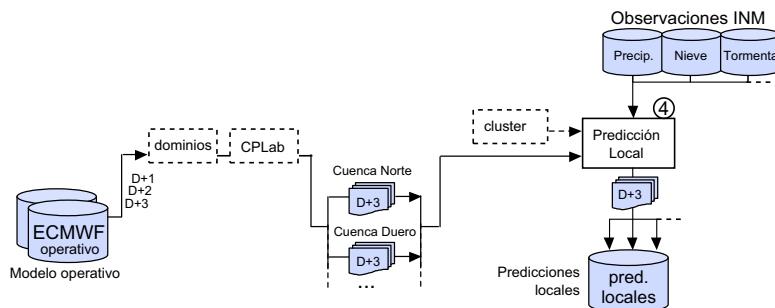


Figura 6.4: Módulo de explotación operativa de la aplicación *Prometeo*.

Este módulo de explotación operativa se alimenta con las salidas diarias (predicciones numéricas) de un modelo operativo. En esta aplicación se utiliza el modelo operativo del ECMWF en la pasada de las 12 UTC y se puede extender en su versión determinista hasta un alcance de 10 días, aunque a partir del cuarto día tiene mayor valor predictivo el EPS (con salida cada 6 horas y un dominio espacial global). De estas predicciones se deben extraer las mismas variables, niveles y área geográfica previamente especificadas en el módulo *dominios*. Aunque sólo es necesario disponer de la predicción para el día de interés y, por tanto, no es necesaria una base de datos sino un único fichero de entrada. El sistema actual contiene predicciones desde el año 2002 para desarrollar una herramienta interactiva que permita explorar y validar un periodo amplio de predicciones. Esta base de datos de predicciones se actualiza diariamente agregando las predicciones realizadas.

Este módulo de explotación operativa involucra las siguientes tareas:

- (1,2) *dominio, CPLab*. Al igual que en la fase de configuración, se filtran los datos operativos sobre la región geográfica indicada. Ahora hay que tener

en cuenta que para cada día hay distintos alcances (en la figura se indican los alcances D+1, D+2 y D+3, correspondientes a las predicciones hechas un día para cada uno de los tres días siguientes). Las componentes principales de los patrones atmosféricos resultantes son obtenidas aplicando las transformaciones calculadas en la fase de configuración. Este módulo es muy simple y rápido en la fase operativa y por ello no es necesario almacenar la salidas, que se calculan en tiempo real para obtener la predicción.

- (4) *Predicción local.* Esta es la tarea central de la aplicación ya que en ella se combinan: los resultados del agrupamiento (que condensan los datos de re-análisis), las salidas de los modelos numéricos (en forma de vectores de CPs) y las observaciones (almacenadas en bases de datos). Todo ello es necesario para obtener una predicción local a partir del método descrito en Gutiérrez et al. (2004). En este módulo también se realizan las distintas tareas de inferencia probabilística y numérica. Por ejemplo, se obtienen predicciones probabilísticas de la precipitación para distintos umbrales y se obtiene una predicción numérica de la cantidad estimada de precipitación (usando tanto la media como un percentil superior para compensar la pérdida de resolución del grupo).

Un fichero de configuración define los observatorios de la base de datos de observaciones donde se desea obtener una predicción (2500 en la aplicación del INM), y asocia cada uno de ellos a un dominio (en este caso, la cuenca hidrográfica a la que pertenece). Estas predicciones se almacenan posteriormente en la base de datos de *predicciones locales* para un posterior acceso y tratamiento.

Una vez obtenidas las predicciones, sería deseable poder acceder a ellas de forma interactiva. En la siguiente subsección se explica el módulo encargado de establecer un sistema de acceso interactivo a esta base de datos de predicciones locales y a los resultados de las validaciones de estas predicciones.

6.2.3. Acceso Web a las Predicciones

Un forma versátil de acceder a las predicciones locales es a través de Internet, utilizando un navegador. Para ello, se ha diseñado una aplicación Web para que usuarios anónimos y autorizados puedan acceder a esta información usando cualquier navegador de internet (ver Fig. 6.5). Para ello se ha hecho uso de tecnología de acceso dinámico a bases de datos a través de internet. En particular se ha usado tecnología JAVA www.sun.com/java, tanto servlets como JSP (Java Server Pages), debido a su difusión, funcionalidad, y carácter abierto de su código. El software de conexión utilizado ha sido el desarrollado por el grupo de trabajo Apache (www.apache.org) y se denomina Jakarta. Esta tecnología ha permitido desarrollar una aplicación WEB (Módulo 5) con el objetivo de que el usuario pueda interaccionar con las predicciones local de forma sencilla, seleccionando día, alcance, variable, etc. En la figura 6.5 se muestra un ejemplo de la página web, en la que se puede seleccionar la información

deseada para una fecha concreta y mostrarla en un mapa dinámico donde se representan las predicciones numéricas y probabilísticas con un código de colores (en el ejemplo mostrado se muestra la probabilidad de que la precipitación supere 0.1mm). Además se puede acceder a la información puntual para una estación concreta sin más que pulsar sobre ella con el ratón.

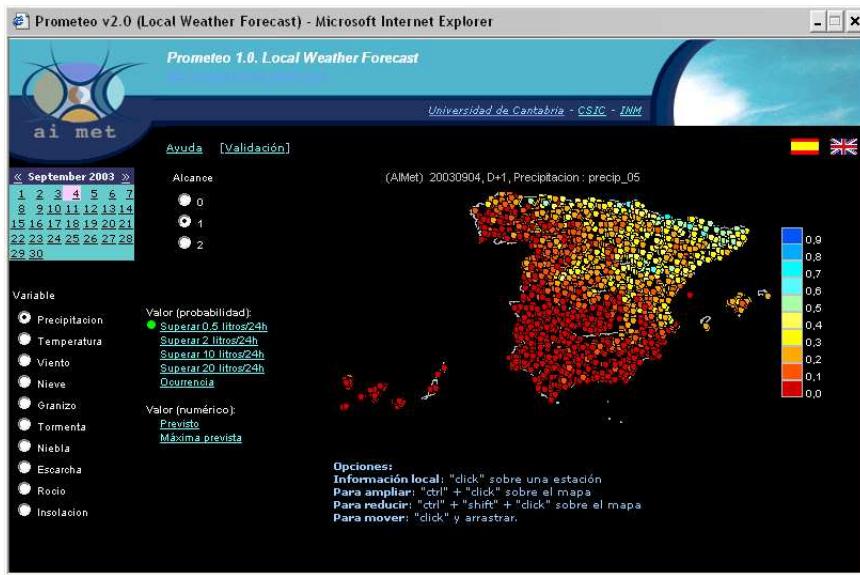


Figura 6.5: Aplicación Web de acceso interactivo a predicciones locales del sistema *Prometeo* adaptado a la predicción local en la red secundaria del INM meteo.macc.unican.es/prometeo/

6.3. Casos de Estudio

A continuación se muestran distintos casos de estudio que ilustran el funcionamiento del sistema.

En la figura 6.6 se muestran las predicciones de precipitación y tormenta realizadas la última semana de Enero de 2003. En concreto el día 30 de Enero se registraron grandes nevadas en los sistemas central, pirenaico y en los picos de Europa, así como tormentas en la mitad norte de la Península Ibérica. También se registraron precipitaciones que superaron los 20mm en algunos puntos de Cantabria y el País Vasco. Las predicciones del sistema reflejan esta situación claramente con un día de antelación (son predicciones a D+1). La figura 6.6 muestra los eventos *Precip* $> 0.5\text{mm}$ (lluvia débil), *Precip* $> 20\text{mm}$ (lluvia fuerte) y *Tormena*, mientras que la Fig. 6.7 muestra las predicciones de nieve, granizo, y la predicción numérica de la temperatura máxima, donde se observan valores muy bajos en muchas localidades.

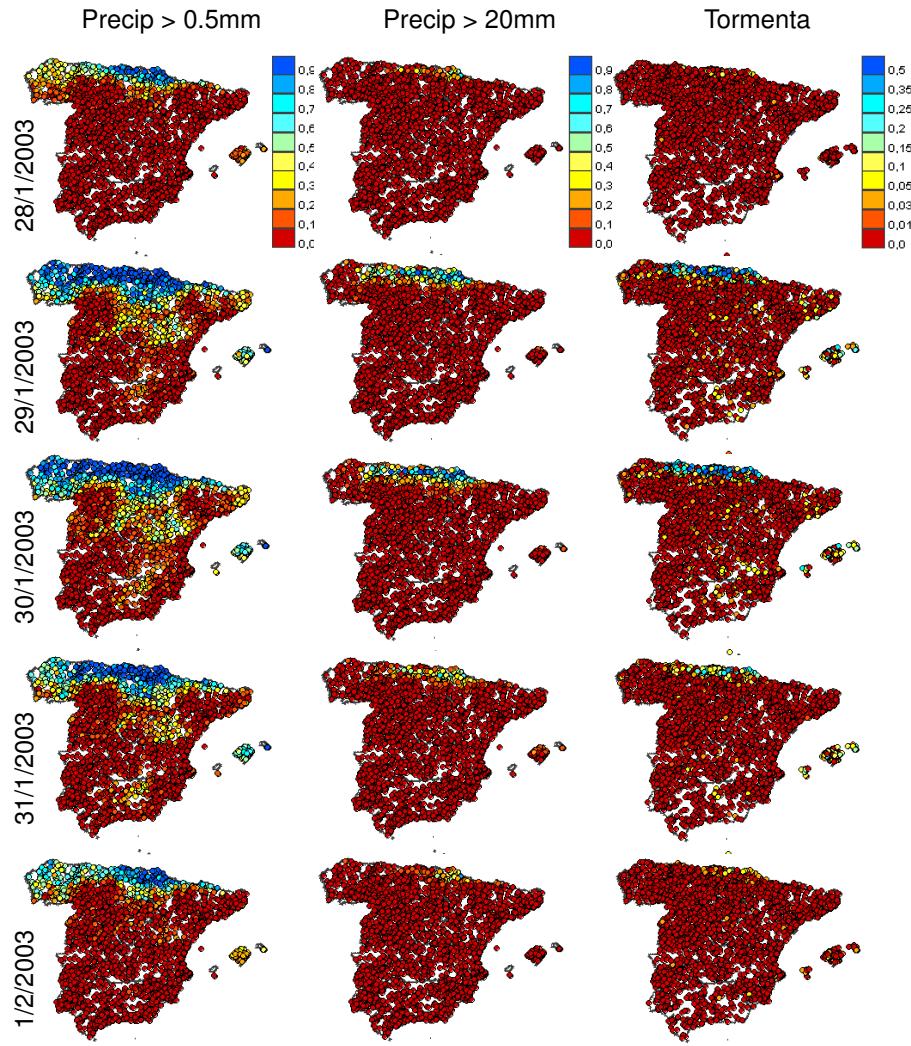


Figura 6.6: Predicciones probabilísticas del 28/1/2003 al 1/2/2003 a D+1. En columnas se muestran los eventos $P(\text{Precip} > 0.5\text{mm})$, $P(\text{Precip} > 20\text{mm})$ y $P(\text{Tormenta})$.

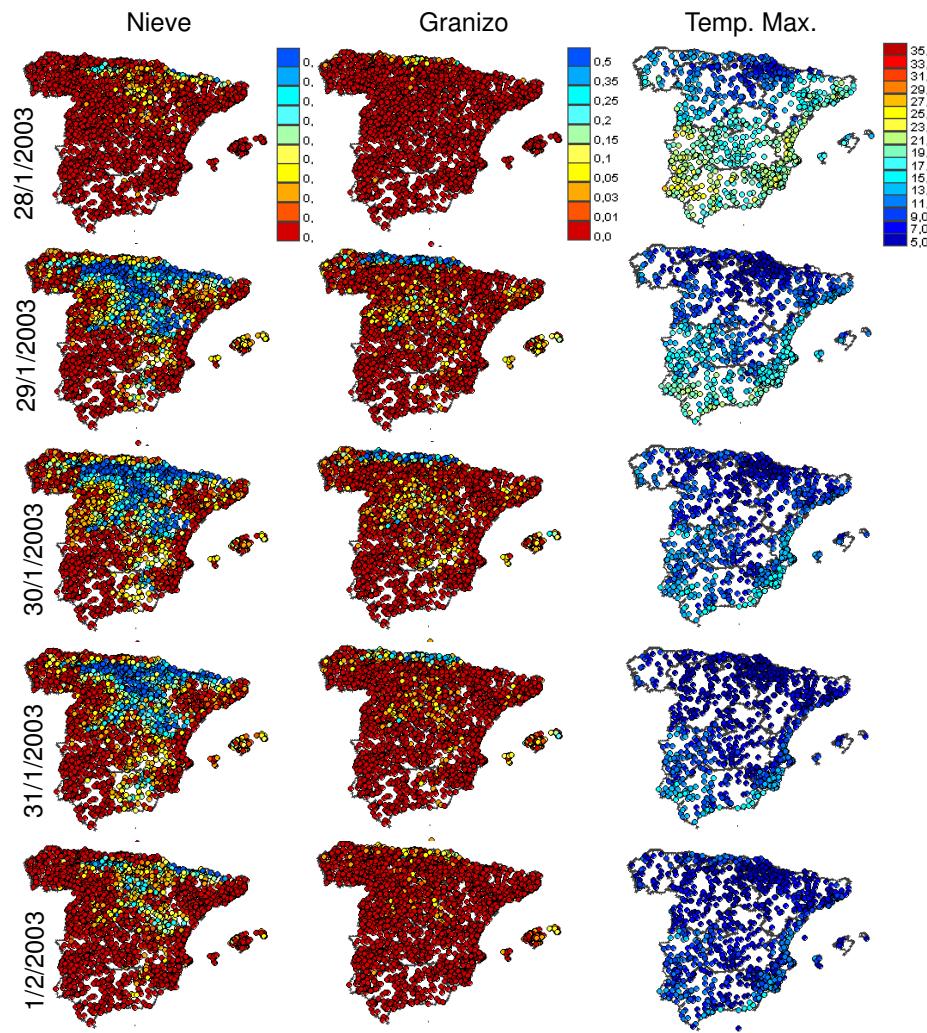


Figura 6.7: Predicciones del 28/1/2003 al 1/2/2003 a D+1. En columnas se muestran los eventos $P(Nieve)$, $P(Granizo)$ y valor de temperatura máxima previsto.

Cabe también destacar la predicción realizada para las islas Baleares en las que se observaron nevadas en localidades próximas al mar.

En el ejemplo anterior, las predicciones se han obtenido con alcance de un día. Sin embargo, tiene interés analizar cómo varían éstas cuando se aumenta el horizonte de predicción. En la primera columna de la Fig. 6.8 se muestra la predicción de precipitación débil para el día 19/3/2003, obtenida con 1, 2, y 3 días de antelación. En esta figura puede comprobarse que este evento era altamente predecible y las predicciones no varían sustancialmente en un lapso de tres días. Sin embargo, existen otras situaciones menos predecibles en las que los distintos alcances de predicción muestran distintas situaciones. Por ejemplo, en la segunda columna de la Fig. 6.8 se muestran las predicciones para 7/3/2003; en esta figura puede comprobarse cómo la predicción se va afinando con el paso del tiempo. La tercera columna de la Fig. 6.8 muestra un caso más extremo, en el que las predicciones no coinciden.

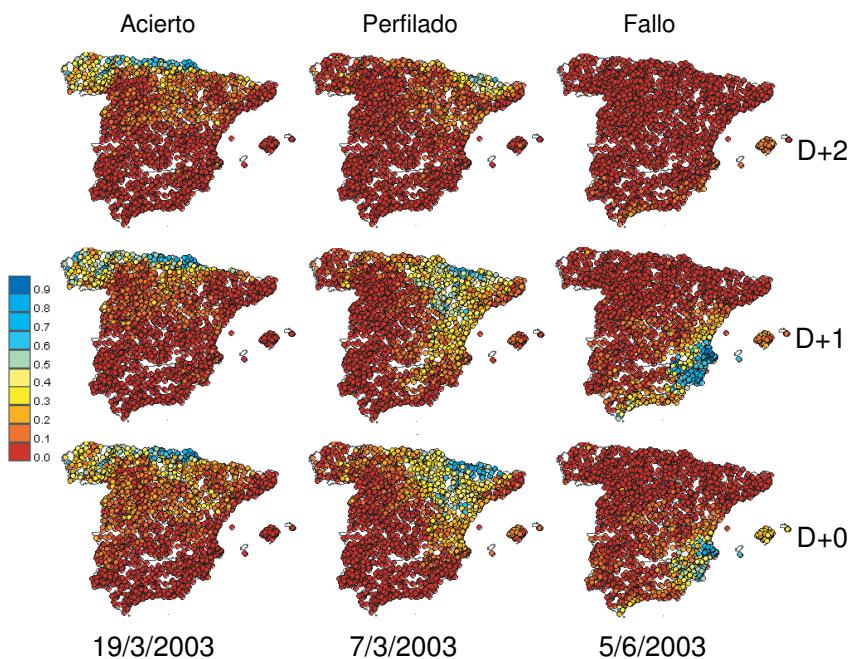


Figura 6.8: Predicciones probabilísticas de $P(\text{Precip} > 0.5\text{mm})$. En filas se muestran los distintos alcances de la predicción ($D+1$, $D+2$ y $D+3$); la primera columna corresponde a la fecha 19/3/2003, mientras que la segunda corresponde a la fecha 7/3/2003 y la tercera 5/6/2003

Otro aspecto interesante del sistema es su reflejo de la dinámica y evolución de distintas situaciones sinópticas (como las entradas de frentes en la península). En la figura 6.9 se puede ver la evolución de la probabilidad de lluvia débil, desde el 24 de Marzo de 2003 al 1 de Abril de 2003. En ese período puede

observarse (a través de sus efectos en la probabilidad de lluvia) la entrada de un frente por el noroeste de la península y su posterior evolución hasta que finalmente desaparece.

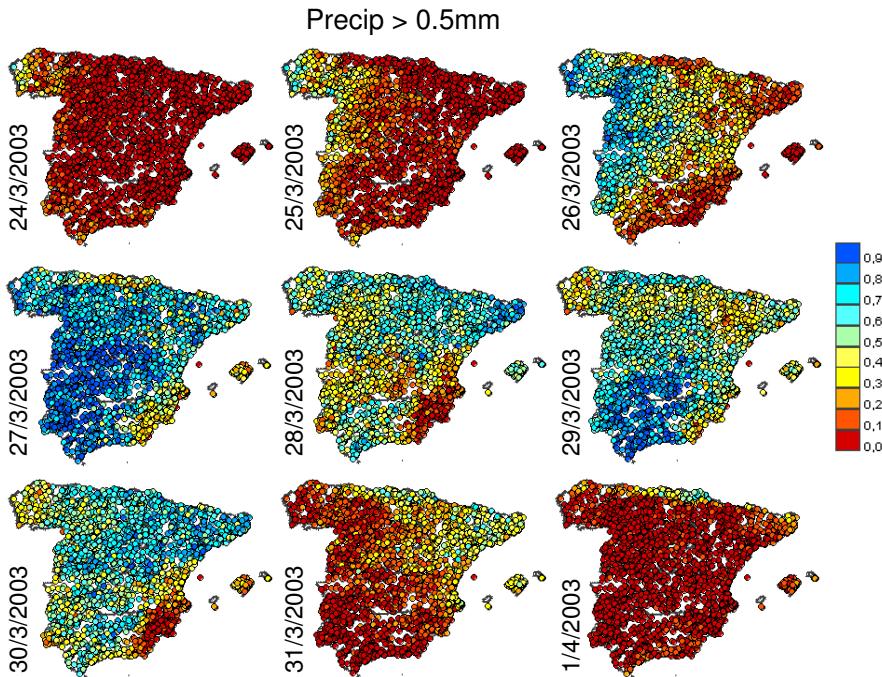


Figura 6.9: Predicciones probabilísticas de $P(Precip > 0.5mm)$ entre el 24/3/2003 y el 1/4/2003 a D+1.

Todos los mapas mostrados en esta Sección se han obtenido directamente de la página Web del sistema (meteo.macc.unican.es/prometeo) y han sido predicciones realizadas en modo operativo, que pueden ser consultadas en tiempo pasado para comprobar diferentes aspectos de las predicciones realizadas por el sistema. La predicción operativa utilizando esta herramienta se ofrece en la Intranet del INM (y próximamente también en Internet). Fruto de esta comprobación también surge la necesidad de realizar y almacenar distintos elementos de validación, los cuales ofrecen valiosa información que permite evaluar las distintas configuraciones dependientes de los parámetros del sistema, y que es utilizada para mejorar las predicciones.

6.4. Validación. Retroalimentación del Sistema

Una parte importante de la aplicación es el módulo de validación. Este módulo tiene un valor tanto informativo (mostrando la calidad de las distintas predicciones), como técnico (proporcionando una medida de calidad para tomar

decisiones en el momento de construir o modificar el sistema; por ejemplo, a la hora de seleccionar el número de vecinos óptimo, o un agrupamiento eficiente para el algoritmo). Para que las validaciones sean representativas, es necesario considerar un periodo suficientemente largo de predicciones operativas. Para ello, el módulo de validación puede ejecutar el módulo de producción operativa para un período dado y almacenar los resultados en la base de datos; o bien puede trabajar con un conjunto de fechas ya almacenadas en la base de datos. Como se comentó anteriormente, en el sistema se han cargado las predicciones de los años 2002-2003 y en base a ellas puedan realizarse las validaciones oportunas. En la Fig. 6.10 se muestra la estructura de este módulo, que accede a las predicciones locales y las valida frente a las observaciones almacenadas en la base de datos. Después de aplicar distintos scores de validación (Brier Score, Brier Skill Score, fiabilidad, resolución, ROC Area, error cuadrático) almacena los resultados en la base de datos *validación* para cada uno de los alcances del modelo numérico. Los resultados de validación son posteriormente promediados por estación (Invierno, Primavera, Verano u Otoño) y también se almacenan las medias anuales y la climatología del evento, para que el usuario pueda comprobar cual es la frecuencia del evento.

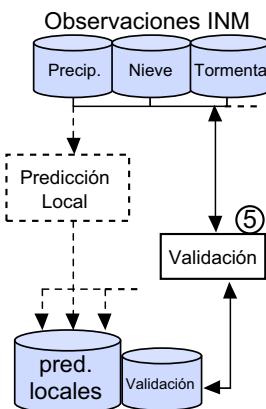


Figura 6.10: Módulo de validación de la aplicación *Prometeo*.

Estas validaciones sirven para conocer qué estaciones superan una cierta capacidad predictiva, qué configuración atmosférica de patrones ofrece mejores resultados, etc. Todo esto se calcula de forma simple utilizando este módulo.

6.4.1. Validación Operativa de Prometeo

En esta sección se muestra una detallada validación del sistema Prometeo realizada durante el período 1987-1988. Para ello se utilizan las medidas estándar como la pericia, valor económico, etc., descritas en el Capítulo 7. El resto de información puede consultarse directamente en la página de Prometeo (meteo.macc.unican.es/prometeo).

Las predicciones se obtienen aplicando la técnica de análogos (o de los vecinos más cercanos) utilizando patrones con distinta información tomada de los campos atmosféricos integrados en ERA40. Los Brier Skill Score (BSS), diagramas de fiabilidad y resolución, y curvas ROC y de valor económico se han obtenido para el conjunto de los dos años (*anual*), y separadamente para cada estación del año (DEF: *Invierno*, MAM: *Primavera*, JJA: *Verano*, SON: *Otoño*). Estas medidas de validación se han calculado para cuatro umbrales distintos de la precipitación 0.5, 2, 10 y 20 mm.

Con el fin de comprobar la eficiencia del método en distintas zonas de la península, se calcularon los BSS anuales para $Prec > 0.5\text{mm}$ considerando como modelo de referencia la climatología (un valor positivo indicará una superioridad del método frente a la climatología, mientras que un valor negativo indicará la situación contraria). La Figura 6.11 muestra la distribución espacial de valores BSS en la península.

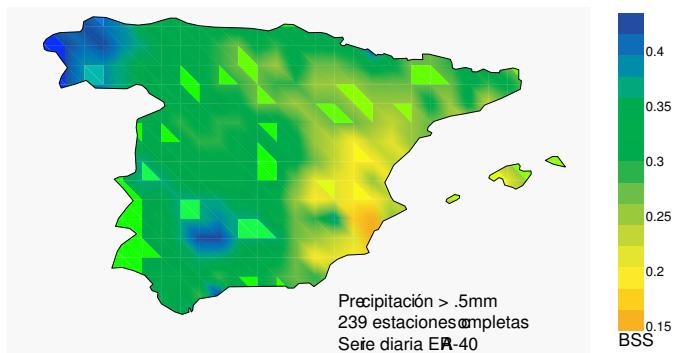


Figura 6.11: BSS para evento de precipitación $> 0.5\text{mm}$ en el periodo 1987-1988 de PROMETEO.

Se constata que PROMETEO tiene una pericia buena en toda la península (con la excepción de algunas estaciones aisladas). La zona donde tiene menor pericia es la zona del Mediterráneo (concretamente en Levante), siendo el Noroeste peninsular la zona con pericias más elevadas; este es el lógico decrecimiento de la pericia con la frecuencia del evento. Para el resto de umbrales, los resultados son similares.

En las figuras 6.12 y 6.13 se muestran medidas alternativas al BSS para validar los pronósticos probabilísticos: las curvas de fiabilidad y resolución y las curvas ROC, respectivamente. Finalmente, las dos últimas figuras muestran las curvas de valor económico para los veranos e inviernos de los dos años de validación.

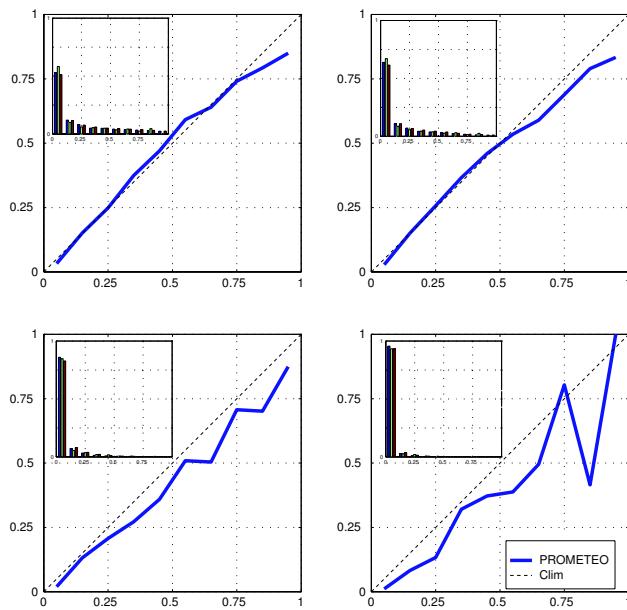


Figura 6.12: Diagramas de Fiabilidad y de resolución para los años 1987-1988.

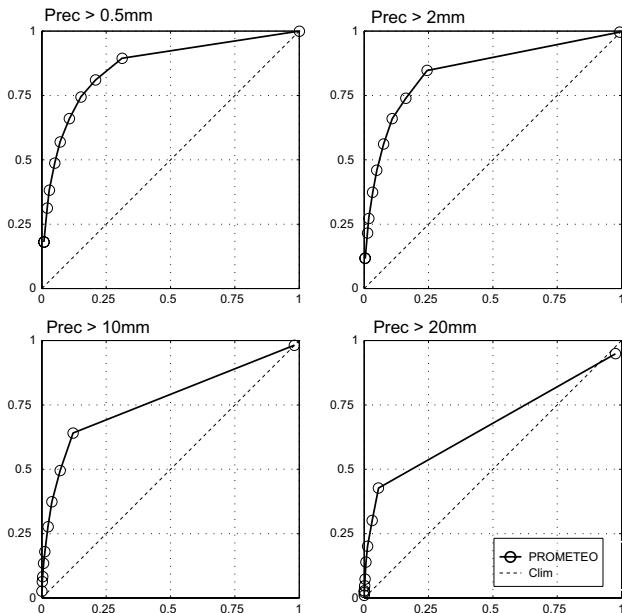


Figura 6.13: Curvas ROC (1987-1988).

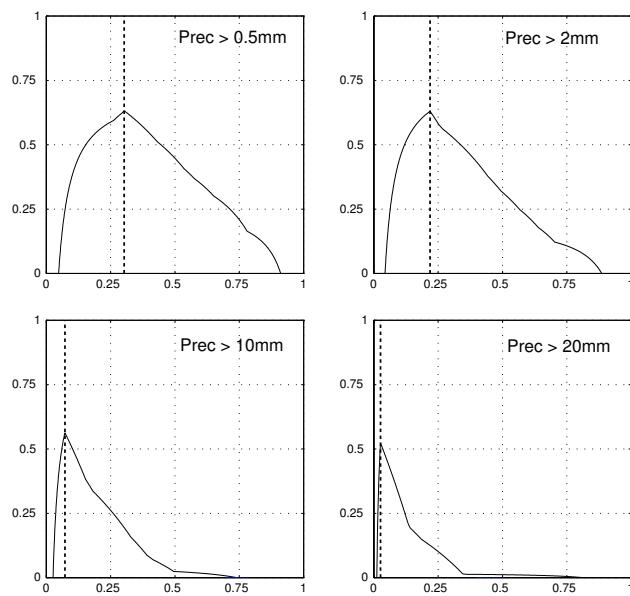


Figura 6.14: Valor económico para Invierno de los años 1987-1988.

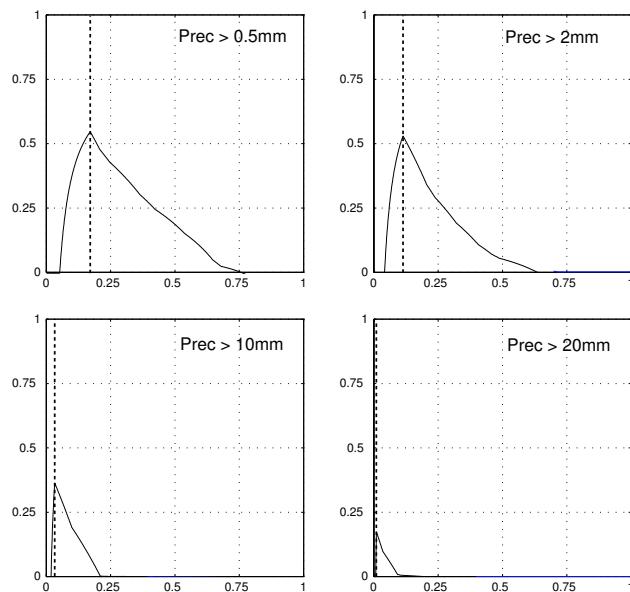


Figura 6.15: Valor económico para Verano de los años 1987-1988.

CAPÍTULO 7

Verificación de Sistemas de Predicción Probabilística

7.1. *Introducción*

En la actualidad se producen una gran variedad de predicciones meteorológicas de distintos fenómenos, utilizando una amplia gama de modelos y métodos. Tradicionalmente estas predicciones se obtenían de forma determinista, es decir, sin tener en cuenta la incertidumbre asociada. Sin embargo, en las últimas décadas se ha puesto de manifiesto la necesidad de formular la predicción meteorológica en términos probabilísticos, para tener en cuenta la influencia de las distintas fuentes de incertidumbre que afectan a la predicción (ver Cap. 1 para más detalles). Este cambio ha hecho necesario el desarrollo de nuevas medidas de validación para cuantificar la calidad de este tipo de predicciones, teniendo en cuenta su carácter probabilístico, así como la propia naturaleza de la predicción (a corto o medio plazo, estacional, multi-modelo, etc.). La siguiente tabla ilustra la complejidad de la predicción numérica actual, con las distintas opciones que se utilizan de forma operativa.

Alcance	Fenómeno	Tipo	Modelo	Método
Corto	Precipitación	Determinista	Área limitada	Estadístico
Medio	Racha máxima	Probabilística	Global	Dinámico
Mensual	Temp. max.	Discreta	Acoplado	Ensemble
Estacional	Temp. min.	Continua	...	Multimodelo
Climático	Meteoros

En este capítulo se analizan los métodos más comunes para evaluar la *bondad* de las predicciones, sin tener en cuenta el sistema empleado para su obtención, sino sólo su naturaleza (discreta/continua, determinista/probabilística). Murphy (1993) define la bondad de una predicción en base a tres aspectos:

- *Consistencia*: no se generan predicciones contradictorias,
- *Calidad*: concordancia entre la predicción y la observación y
- *Utilidad*: valor real (económico, etc.) de la predicción para un usuario.

Para tener en cuenta estos aspectos a la hora de validar un sistema, es necesario disponer de criterios de verificación (o validación) apropiados, que proporcionen medidas objetivas de los distintos aspectos de la bondad. Además, cuando las predicciones están distribuidas espacialmente (por ejemplo, un conjunto de predicciones locales en una red de estaciones), la verificación también ha de considerar esta característica (ver Ebert and McBride, 2000). Por otra parte, también es importante tener en cuenta que la calidad de un sistema puede variar según la adversidad del evento a predecir, la geografía de la zona, etc. Por tanto, en general es necesario cuantificar tanto la calidad global espacio-temporal, como la calidad condicionada a determinados eventos y zonas. Finalmente, también hay que distinguir los diferentes usuarios a quienes va dirigida la validación (medidas fácilmente interpretables para el público en general, o medidas más informativas para una audiencia técnica). Las medidas de verificación resultantes se utilizan para propósitos tan distintos como proporcionar información del error de los modelos disponibles a los usuarios de productos meteorológicos, comparar distintos modelos operativos en una época o situación dada, analizar la eficiencia de las nuevas versiones de los modelos operativos, estudiar la viabilidad económica de un cierto modelo, etc.

En este capítulo se describen algunas de las medidas e índices de verificación más utilizados en las ciencias atmosféricas, que permiten tener en cuenta las características mencionadas anteriormente. En la Sec. 7.2 se analizan los distintos tipos de predicciones en base a su carácter discreto/continuo y determinista/probabilístico; en la Sec. 7.3 se describen distintos aspectos sobre la calidad de las predicciones. A continuación, las secciones 7.4 y 7.5 analizan la calidad de las predicciones en el caso determinista y probabilístico, respectivamente. Finalmente, las secciones 7.6.1 y 7.6.2 describen dos métodos específicos para valorar la calidad de una predicción probabilística teniendo en cuenta sus aciertos y fallos para un determinado evento y el valor económico de los mismos para un usuario específico. Finalmente, la Sec. 7.7 describe el módulo de validación implementado en la herramienta *MeteoLab*, que incluye los distintos índices descritos en este capítulo.

Para una descripción más completa de este tema, el lector puede consultar Jolliffe and Stephenson (2003), o Wilks (1995) (Cap. 7).

7.2. *Tipos de Predicciones*

En el ámbito de la predicción meteorológica se trabaja con variables de naturaleza y características muy diferentes: variables *binarias* (nieve, no nie-

ve), *categóricas* (seco, lluvia débil, lluvia moderada y lluvia fuerte), variables *continuas* (temperatura), e incluso *mixtas* (como la precipitación, ver Ej. 2.2). Además, una misma variable puede considerarse como binaria, categórica o continua según la situación particular de estudio. Por ejemplo, es muy común transformar una variable en binaria considerando un umbral de interés para la misma: temperaturas inferiores a 0°C , $viento > 90\text{km/h}$, etc. Para cada uno de estos grupos de variables se efectúan distintos tipos de predicción.

7.2.1. Predicciones Deterministas

Una predicción es determinista cuando proporciona el estado de una variable discreta, o el valor de una continua, en una situación dada; por ejemplo, $viento > 90\text{km/h}$, o $precipitacion = 18.2\text{mm}$.

Algunos sistemas de predicción determinista elementales, utilizados como sistemas de referencia para validar otros métodos, son:

- La *persistencia*, que consiste en pronosticar para un instante t lo que ocurrió en el instante $t - 1$ (por ejemplo, la precipitación predicha para hoy será la que se haya observado ayer):

$$\hat{o}(t) = o(t - 1). \quad (7.1)$$

También se puede considerar el caso más general de la $\text{persistencia}(\tau)$, donde el valor de $\hat{o}(t)$ se obtiene a partir de $\{o(t - 1), \dots, o(t - \tau)\}$.

- La *climatología*, que consiste en pronosticar utilizando algún parámetro derivado de la distribución climatológica obtenida empíricamente para un evento dado. Por ejemplo, la precipitación predicha para un día de Enero será la media de los valores diarios de precipitación en Enero en los últimos diez años:

$$\hat{o}(t) = \langle o(t_i) \rangle_{i=1\dots,N}, \quad (7.2)$$

donde $\langle \rangle$ denota el valor medio para un período dado, y $\hat{o}(t)$ denota un valor estimado (predicho) para $o(t)$.

7.2.2. Predicciones Probabilísticas

Una predicción es probabilística cuando describe en términos cuantitativos la incertidumbre asociada con la predicción. El pronóstico realizado es una distribución de probabilidad sobre el rango total de valores o categorías de la variable. La forma específica de esta distribución vendrá dada por las características de la variable analizada. Por ejemplo, si la variable es continua y la predicción se supone Gaussiana, entonces una predicción probabilística vendría dada por su media y varianza; por otra parte, si la variable es discreta, entonces la predicción vendría dada por su función de probabilidad. Por ejemplo, la Fig. 7.1 muestra una predicción probabilística particular asociada a la variable

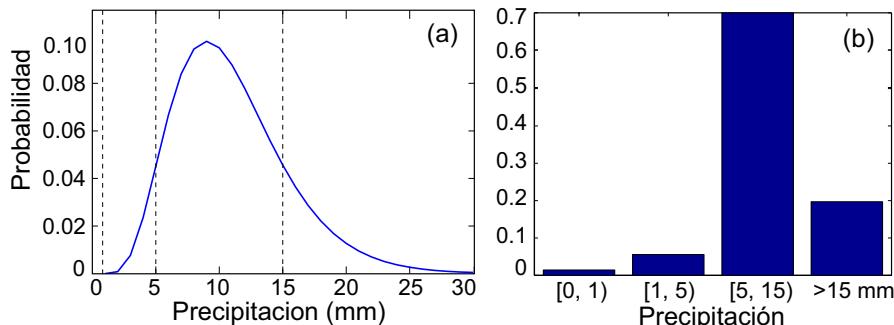


Figura 7.1: (a) Predicción probabilística para la variable continua precipitación. (b) Predicción probabilística correspondiente a la variable categórica: seco, lluvia débil, lluvia moderada, y lluvia fuerte.

precipitación, considerada de dos formas distintas: (a) como variable continua y (b) como variable categórica con cuatro estados.

Los métodos de predicción anteriores, basados en persistencia y climatología, también pueden utilizarse como métodos de predicción probabilística, utilizando la distribución de probabilidad definida por los días anteriores (persistencia) o por el período climatológico (climatología) para la predicción.

Desde un punto de vista operativo, cualquier predicción determinista puede considerarse como un caso particular de una probabilística sin más que asignar probabilidad uno al estado o valor previsto y cero al resto. Recíprocamente, cualquier predicción probabilística puede convertirse en determinista de diversas maneras: bien estimando un estado para una variable discreta (el más probable, el que supere un umbral de probabilidad, etc.), o un valor para una continua (la media, el percentil 75 %, etc.). Estos recursos se utilizan con mucha frecuencia para simplificar la predicción y la verificación.

7.3. Aspectos de la Calidad de una Predicción

Son numerosos los aspectos que se pueden tener en cuenta para analizar la calidad de un sistema de predicción, especialmente en los sistemas de predicción probabilística. Por ello, no hay ninguna medida de verificación que proporcione una información completa de la calidad del sistema, sino que cada índice de verificación describe algún atributo particular de la relación entre observaciones y predicciones. Algunos de estos atributos caracterizan a un único sistema, mientras que otros son comparativos e indican la diferencia entre sistemas de predicción distintos (o entre un sistema concreto y otro sistema de referencia como la climatología o la persistencia).

Murphy and Winkler (1987) establecen que toda la información de una verificación está contenida en la distribución conjunta $P(o, \hat{o})$ de observaciones y predicciones. En el caso de predicciones categóricas, para una variable discre-

ta con categorías $\{C_1, \dots, C_m\}$, esta información está recogida en la tabla de contingencia:

		Observado			
		c_1	c_2	\dots	c_m
Previsto	\hat{c}_1	p_{11}	p_{12}	\dots	p_{1m}
	\hat{c}_2	p_{21}	p_{22}	\dots	p_{2m}
	\dots	\dots			
	\hat{c}_m	p_{m1}	p_{m2}	\dots	p_{mm}

En base a esta información conjunta se puede obtener cualquier probabilidad marginal o condicional asociada a predicciones y observaciones: $P(o|\hat{o})$, $P(\hat{o}|o)$, $P(\hat{o})$, $P(o)$. Los distintos índices de verificación se basan en alguna de estas probabilidades para poner de manifiesto algún aspecto de interés de la predicción. A pesar de las relaciones obvias que existen entre estas probabilidades:

$$p(\hat{o}, o) = p(o|\hat{o})p(\hat{o}) = p(\hat{o}|o)p(o), \quad (7.3)$$

su interpretación y utilidad meteorológica es muy distinta y, por ello, los índices de verificación son complementarios y aportan información interesante desde distintos puntos de vista.

En función del tipo de probabilidad que utilizan (conjunta, marginal o condicionada), los índices de verificación se pueden dividir en tres grandes grupos Wilson (2001):

- Los que utilizan la función de probabilidad conjunta para analizar globalmente la correspondencia entre pares observación-predicción. El *bias*, la *correlación*, el *RMSE* y el *Skill*, descritos a continuación, pertenecen a este grupo.
- Los que utilizan probabilidades condicionadas a valores concretos de la predicción. La *fiabilidad*, la *resolución* y la *definición* pertenecen a este grupo.
- Los que utilizan probabilidades condicionadas a valores concretos de la observación. La *discriminación* y la *incertidumbre*, son de este tipo.

A continuación se describen algunas de las medidas de calidad más importantes correspondientes a estos tres grupos (ver Murphy, 1993):

- *Sesgo* (bias), o desviación sistemática: Referido a la concordancia entre la predicción media y la observación media; las medidas de sesgo, como la diferencia promedio entre la predicción y la observación, se calculan a partir de $p(o)$ y $p(\hat{o})$, y se definen de modo que un sesgo positivo indica una sobreestimación del valor a predecir, mientras un sesgo negativo indica una subestimación (por ejemplo, se predice menos cantidad de lluvia de la que realmente ocurre).

- *Asociación*: Indica el grado de relación lineal entre observación y predicción. La covarianza y la correlación son medidas de asociación.
- *Precisión* (accuracy): Relativa a la concordancia entre el valor previsto y el observado realmente, promediada sobre una muestra de parejas individuales de predicciones y observaciones; se calculan a partir de $P(\hat{o}, o)$. Medidas de precisión son, por ejemplo, el error absoluto medio, el error cuadrático medio, el Brier Score, etc.
- *Habilidad* (skill), o precisión relativa: Es la precisión de un sistema referida a otro que se toma como referencia (por ejemplo, climatología o persistencia). Se definen de modo que un valor positivo (negativo) indica que el sistema es más (menos) hábil que el de referencia. El Brier Skill Score es el índice más conocido de esta categoría.

En el caso particular de predicciones asociadas a eventos binarios (por ejemplo *si* y *no*), un único valor numérico de la probabilidad (por ejemplo $p(\hat{o} = si) = \hat{p}$) define la predicción, ya que $p(\hat{o} = si) + p(\hat{o} = no) = 1$. En ese caso, existen dos medidas especiales que caracterizan los dos aspectos más importantes de la predicción, que determinan si es o no perfecta:

- *Fiabilidad* (reliability), o bias condicional, o calibración: Para que la predicción de un cierto evento binario sea fiable, la probabilidad prevista y la observada deben coincidir lo máximo posible en todo el rango de valores de probabilidad; cuando no es así, se habla de bias condicionado. La fiabilidad se puede representar gráficamente considerando una discretización del rango de probabilidades previstas \hat{p} (por ejemplo en 10 intervalos: $[0, 0.1], [0.1, 0.2], \dots$) y dibujando $\hat{p} = p(\hat{o} = si)$ frente a $p(o = si|\hat{p})$, es decir, frente a la frecuencia relativa de las ocasiones en que ocurrió el evento cuando la probabilidad predicha para el mismo era \hat{p} . Por ejemplo, la Fig. 7.2 muestra la fiabilidad para distintas predicciones obtenidas con el sistema Prometeo (ver Cap. 6). Cuanto más cerca esté la curva de la diagonal, más fiable será el sistema; si la curva está por debajo de la diagonal, entonces la predicción sobreestima el evento, mientras que si está por encima entonces la predicción subestima (ocurrió más veces de lo que se prevé). En el caso de la Fig. 7.2, se puede decir que el sistema subestima la observación. Por otra parte, son menos fiables las predicciones de eventos raros (poco frecuentes), como $precip > 20\ mm$. Es muy importante estudiar la fiabilidad de los sistemas para diferentes eventos, ya que los eventos raros casi nunca se predicen con probabilidades altas, manifestando una notable pérdida de fiabilidad que debe ser corregida.

Cuando se tiene una predicción extendida espacialmente (como en la Fig. 7.2, que muestra el resultado para 184 estaciones distintas), entonces existen dos formas de obtener una única curva de fiabilidad. La más sencilla consiste en promediar las curvas individuales; sin embargo, en este caso se pueden compensar efectos de sobreestimación y subestimación dado lugar a una curva promedio falsa. La otra alternativa es considerar las 184

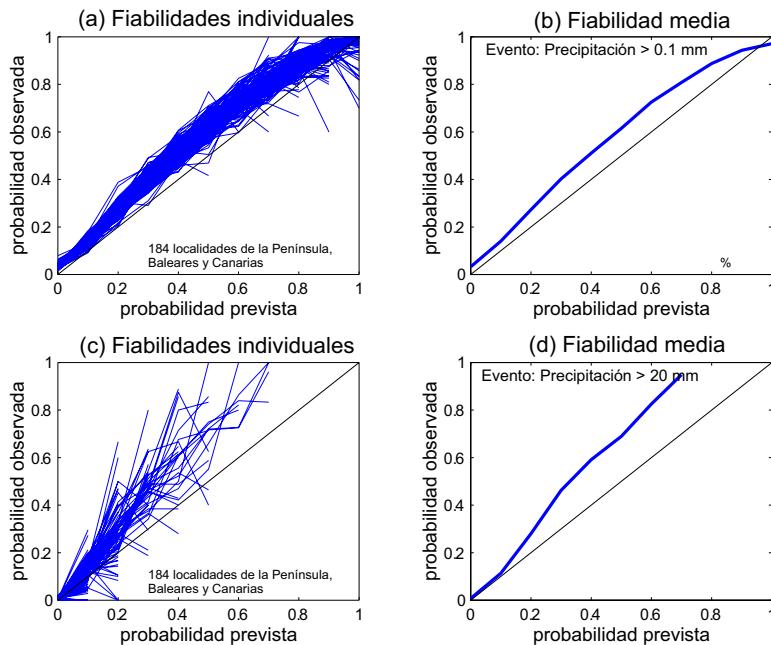


Figura 7.2: Curvas de fiabilidad para la predicción sobre 184 estaciones de los eventos (a)-(b) $\text{precip} > 0.1 \text{ mm}$ y (c)-(d) $\text{precip} > 20 \text{ mm}$.

estaciones como una única estación y obtener la fiabilidad conjunta. Sin embargo, de esta forma, los casos individuales con mala fiabilidad también quedan enmascarados en la curva global resultante (ver Fig. 7.2(d)).

- *Resolución:* Cuantifica la desviación de la predicción respecto de la predicción constante dada por la probabilidad climatológica del evento: $\hat{p} = p_c$. Una predicción climatológica constante tiene varianza nula y, por tanto, la desviación puede medirse por medio de la varianza de la predicción \hat{p} en el período de validación. La resolución se suele representar gráficamente mediante la función de probabilidad de la variable *probabilidad predicha*, $p(\hat{p}) = p(\hat{o} = \text{si})$; para ello, se suele considerar una partición del rango posible de probabilidades posibles y representar un histograma de las frecuencias relativas correspondientes. Por ejemplo, la Fig. 7.3 muestra la resolución de una de las predicciones anteriores del sistema Prometeo. Es deseable que el histograma tenga forma de ‘U’, para que la varianza sea la máxima posible y coincida con la resolución ideal (aquella que predice con probabilidad cero los casos en los que no ocurrió el evento y con probabilidad uno los casos en que ocurrió). De esta forma, se penaliza la predicción de valores ambiguos como $\hat{p} = 0.5$.

Existen diversas técnicas denominadas de *inflado* (inflating) para corregir la resolución y adecuarla a la resolución ideal. La forma más simple con-

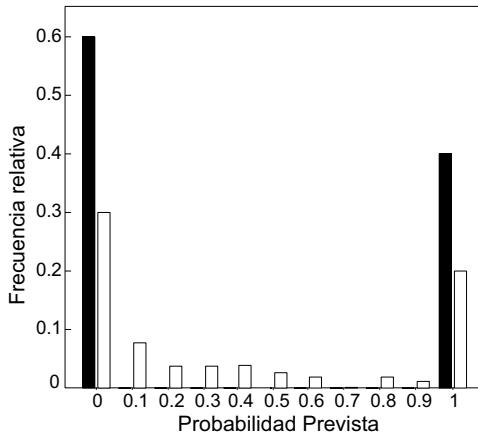


Figura 7.3: En negro la resolución ideal y en blanco un ejemplo de resolución real.

siste en hacer un cambio de escala para aumentar o reducir la varianza de la probabilidad prevista; sin embargo, en este caso se empeora el sesgo de la predicción. Existen otras técnicas más sofisticadas, como la denominada *expanded downscaling* (Burguer, 1996), que utiliza la correlación canónica para corregir de una forma global la varianza de las predicciones. Sin embargo, dada la relación constante que existe entre la varianza de una predicción y su sesgo, no es sencillo poder corregir ambas simultáneamente en la dirección óptima, variando el método de predicción (ver von Storch, 1999, para más detalles).

7.4. Verificación de Predicciones Deterministas

Cuando el resultado de la predicción es una estimación del valor de una variable continua, la calidad se mide con relativa facilidad mediante medidas de precisión (funciones de error) y de habilidad. Estas funciones son medidas globales de precisión, en el sentido de que sólo consideran la distribución conjunta, calculando la distancia entre los datos previstos y los observados. Por tanto, resultan mejor representados los eventos más frecuentes. Dado un conjunto de observaciones $\{o_1, o_2, \dots, o_n\}$ sobre las que se realizan las predicciones $\{\hat{o}_1, \hat{o}_2, \dots, \hat{o}_n\}$, los índices de verificación más habituales son:

- El error medio (bias):

$$\frac{1}{n} \sum_{i=1}^n (o_i - \hat{o}_i) = <(o_i - \hat{o}_i)>. \quad (7.4)$$

- El error absoluto medio (MAE): $<|o_i - \hat{o}_i|\>$.
- El error cuadrático medio (MSE, Mean Square Error): $<(o_i - \hat{o}_i)^2>$.

- La raíz cuadrada del error cuadrático medio (RMSE): $\sqrt{<(o_i - \hat{o}_i)^2>}$.

- La correlación:

$$\frac{<(o_i - <o_i>)(\hat{o}_i - <\hat{o}_i>)>}{\sigma(o)\sigma(\hat{o})}. \quad (7.5)$$

Cuando este índice se aplica a anomalías, y no a los valores reales, entonces se denomina Coeficiente de Correlación de Anomalías (ACC) y, como su nombre indica, mide la correlación entre el campo de anomalías previsto y el observado. En ese caso, se tiene $o_i = o_i - \mu_c$ y $\hat{o}_i = \hat{o}_i - \mu_c$, donde μ_c es la media climatológica. Esta medida es muy utilizada para la verificación de campos espaciales (Miyakoda et al., 1972).

- El LEPS (Linear Error in Probability Space): Es el error absoluto medio medido en unidades de probabilidad acumulada:

$$<|CDF - \widehat{CDF}|>, \quad (7.6)$$

donde CDF y \widehat{CDF} son los valores de la función de distribución de la variable observada correspondientes a o y \hat{o} , respectivamente. La idea básica de LEPS es que la magnitud absoluta del error no tiene la misma importancia en las colas que en la zona media de la distribución. Por ejemplo, un error de $2^{\circ}C$ en la predicción de temperatura es más grave cuando se trata de una anomalía pequeña que cuando la anomalía es grande (ver Fig. 7.4).

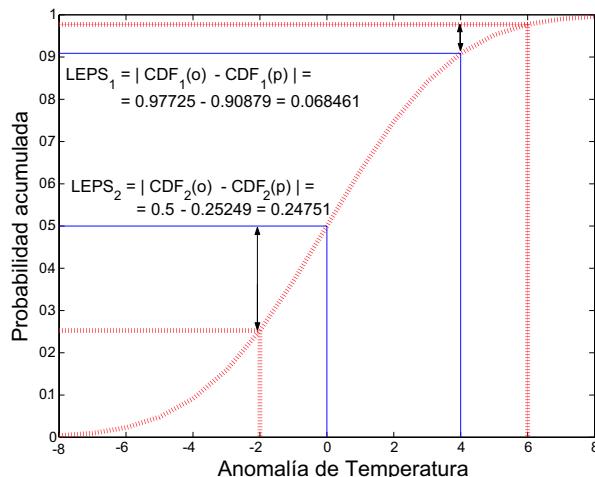


Figura 7.4: El mismo error puede ser muy diferente medido en el espacio de la variable y en el espacio de probabilidad.

En realidad el LEPS mide la probabilidad de obtener un valor entre el dato real y el previsto (Potts et al., 1996; Ward and Folland, 1991). Este índice se puede aplicar para variables tanto continuas como categóricas y también se utiliza en verificación de campos espaciales.

Estas medidas de precisión se calculan a partir de una muestra concreta y, por tanto, dependen de factores relativos a la misma como el lugar geográfico, época del año, etc. Para poder comparar distintos sistemas de predicción se hace necesario relativizar estos índices de acuerdo a un sistema de predicción de referencia. Para ello se pueden utilizar medidas de precisión relativa o *skill scores* como:

- El RMSSSE (Root Mean Square Skill Score Error): Es una medida estándar de verificación, utilizada especialmente en pronósticos a largo plazo (más de 30 días). Su fórmula viene dada por:

$$1 - \frac{RMSE}{RMSE_{ref}}, \quad (7.7)$$

donde RMSE y $RMSE_{ref}$ son los índices correspondientes al sistema de predicción que se está verificando y al sistema de referencia, respectivamente.

7.4.1. Predicciones Binarias

En las secciones anteriores se vio que un sistema de predicción determinista para una variable discreta se puede verificar a partir de su tabla de contingencia. En el caso de variables binarias, las predicciones se refieren a la ocurrencia o no de un determinado evento y la tabla de contingencia resultante es:

		Observado	
		Si	No
Previsto	Si	α	β
	No	γ	δ

En este caso, se pueden cometer dos tipos de errores en la predicción, que corresponden a los valores de la tabla de contingencia de la diagonal secundaria. Por un lado, una *falsa predicción* (o error de Tipo I) se comete al predecir un evento que no ocurre (β en la tabla de contingencia), mientras que una *omisión* (o error de tipo II) se comete cuando no se predice un evento que sí ocurre (γ en la tabla de contingencia). Las distintas medidas para validar predicciones binarias tratan de resumir toda la información de la tabla y, por tanto, los distintos tipos de errores, en un sólo parámetro. Algunos de los índices más populares son:

- *HIR* y *FAR*. Estos índices están relacionados con los aciertos y fallos del sistema determinista. Dado que han de permitir comparar sistemas de predicción distintos, los fallos y aciertos han de medirse condicionados a los eventos observados (y no a los predichos), pues así todos los sistemas se validarán en igualdad de condiciones (condicionados a la misma muestra):

- *Hit Rate (HIR)*. Es una medida de precisión que mide la probabilidad de detección. A veces se denomina *POD* (Probability Of Detection).

$$HIR = P(\hat{o} = 1|o = 1) = \frac{\alpha}{\alpha + \gamma} \quad (7.8)$$

- *False Alarm Rate (FAR)*. También es una medida de precisión que mide la proporción de fallos cuando se predice positivamente un evento.

$$FAR(u) = P(\hat{o} = 0|o = 1) = \frac{\beta}{\beta + \delta} \quad (7.9)$$

Obsérvese que un índice de fallos más natural sería $P(o = 0|\hat{o} = 1)$, es decir, la probabilidad de que no ocurra el evento cuando se ha predicho. Sin embargo, en este caso, el índice sería específico de cada sistema.

- *Precisión*: Mide la proporción de aciertos y puede ser maximizada prediciendo siempre la categoría más común. En regiones donde el evento es muy raro, se hace prácticamente la unidad, debido al gran número de aciertos negativos. No aporta información para eventos raros. Está definida de la siguiente forma y su rango es $[0, 1]$:

$$ACC = \frac{\alpha + \delta}{\alpha + \beta + \gamma + \delta} \quad (7.10)$$

La precisión de un sistema que nunca predice el evento es $1 - p_c$.

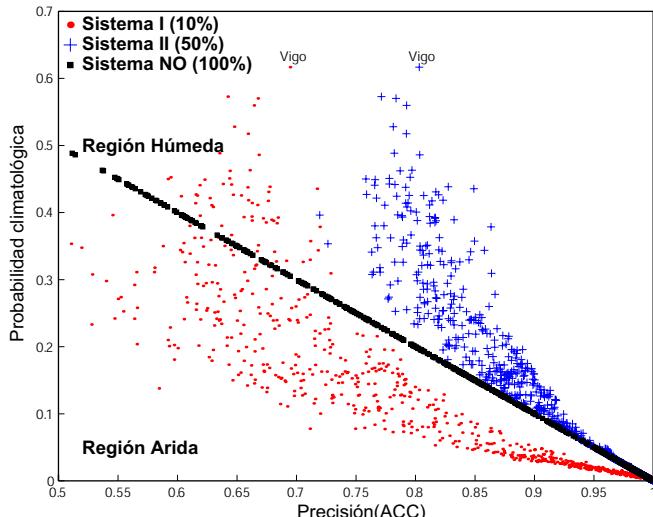


Figura 7.5: Diagrama de precisión-probabilidad para verificar la precisión de tres sistemas de predicción distintos en una red de 184 estaciones.

En la Fig. 7.5 se muestra un ejemplo de uso de este índice, aplicado a tres sistemas de predicción distintos. Estos sistemas consideran diferentes

umbrales de predicción, con el objeto de poner en evidencia problemas de inconsistencia entre diferentes medidas de bondad: el sistema I para el umbral de predicción del 10 %, el sistema II para el umbral de predicción del 50 % y el sistemas NO para el umbral de predicción 100 %. Aunque el método de predicción es el mismo para todos los umbrales, a efectos ilustrativos de validación, se pueden considerar como sistemas diferentes (en la Sec. 7.6 se verá la forma de validar el sistema en su conjunto utilizando, por ejemplo, curvas ROC).

Atendiendo a la precisión, el sistema con el umbral de predicción del 50 % es mejor; además, se ve que un promedio global de ACC para todas las localidades estará fuertemente sesgado en favor de las regiones áridas, por lo que el empleo de éste índice no es aconsejable para ello. ACC penaliza notablemente a los sistemas menos conservadores por lo que el sistema II resulta ser el más preciso en todo el rango. La recta negra representa a la precisión del sistema que nunca predice el evento.

- *Critical Success Index ó Threat Score*: Es análoga a la precisión, pero se eliminan los aciertos negativos, por lo que queda definida como:

$$CSI = \frac{\alpha}{\alpha + \beta + \gamma} \quad (7.11)$$

Aunque es más equilibrado que el ACC está definido en [0, 1], y sus mayores valores se dan en aquellas localidades donde el evento es más común. El CSI es nulo para aquellos sistemas que nunca predicen el evento, mientras que el CSI de un sistema que siempre predice el evento es $\alpha/(\alpha + \beta) = p_c$. Al igual que la precisión, es muy inestable para eventos raros.

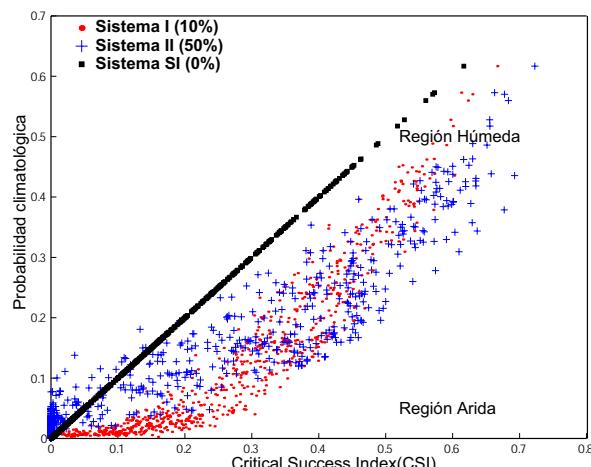


Figura 7.6: Diagrama de precisión-probabilidad para verificar el CSI de tres sistemas de predicción distintos en una red de 184 estaciones.

En la figura 7.6 no queda claro qué sistema es mejor según el criterio del CSI ya que el sistema II con el umbral de predicción del 50 % es mejor en zonas húmedas mientras que el sistema I con el 10 % resulta mejor para zonas áridas, o equivalentemente para eventos raros; como en el caso del *ACC*, no es recomendable un promedio global de *CSI* para todas las localidades ya que es muy inestable en la zona de eventos raros. Nuevamente se constata una notable dependencia respecto a la rareza del evento para diferentes umbrales de predicción en diferentes localidades. La línea negra corresponde al sistema que siempre predice el evento.

Una versión modificada de este índice muy utilizada en Estados Unidos (Wilson, 2001) se denomina Equitable Threat Score, con valores en $[-1/3, 1]$. Se trata de un *CSI* corregido para tener en cuenta el número de aciertos debidos al azar.

7.5. Verificación de Predicciones Probabilísticas

Así como las predicciones deterministas se aplican principalmente a variables continuas, las predicciones probabilísticas se suelen realizar sobre variables categóricas o categorizadas. Las medidas de verificación más simples son las generalizaciones de las correspondientes medidas deterministas.

7.5.1. Brier Score

Es el error medio calculado en unidades de probabilidad para la ocurrencia de las distintas categorías de la variable (Brier, 1950):

$$BS = \langle (p_i - o_i)^2 \rangle, \quad (7.12)$$

donde $p_i = P(\hat{o}_i = si)$, y o_i es 1 si ocurre el evento y 0 en caso contrario. El valor de *BS* es nulo para una predicción perfecta. Si se quieren comparar las precisiones de dos sistemas, es conveniente normalizar el *BS*, por ejemplo con el valor que toma para un sistema climatológico (Talagrand, 1997).

Es muy usual utilizar la descomposición del Brier score como suma de tres componentes, relacionadas con la fiabilidad, la resolución y la incertidumbre. Para ello, se descompone (7.12) de la siguiente forma:

$$(p_i - o_i)^2 = f_i(p_i - 1)^2 + (1 - f_i)p_i^2$$

donde $f_i = p(o_i = 1|p_i)$, es decir, el número de casos observados de entre los previstos con probabilidad p_i (véase Murphy (1973)). Así:

$$BS = \langle (p_i - f_i)^2 \rangle + \langle (f_i - p_c)^2 \rangle + p_c(1 - p_c) = BS_f - BS_r + I,$$

donde p_c es la probabilidad climática del evento, BS_f es la componente de fiabilidad, BS_r la componente de resolución, e I la componente de incertidumbre.

El *BS* da una idea del error promedio cometido en una predicción; por tanto, este índice es poco representativo cuando está descompensada la frecuencia de

ocurrencia y de no ocurrencia del evento. En ese caso, el BS es sólo representativo del evento más frecuente y, por tanto, es confuso utilizar este índice para validar conjuntamente la predicción del evento, ya que habría que ponderar los casos raros para que tuviesen peso en la validación.

Por otra parte, cuando las predicciones están espacial y temporalmente distribuidas, son difíciles de interpretar ya que no es fácil ofrecer parámetros que resuman la información de manera realista; la información ofrecida por los promedios espaciales suele estar fuertemente sesgada en favor de los eventos más numerosos; por tanto, la información en forma de matrices y tablas es completa pero difícil de interpretar y los mapas tienen el mismo problema. Una forma sencilla de visualizar el comportamiento extendido de dichas medidas es dibujar el valor de la medida frente a su probabilidad climatológica para cada localidad (un mismo evento puede ser muy común en unas localidades y muy raro en otras); de esta manera, los eventos raros quedan en la parte inferior del gráfico, independientemente de su localización espacial. Estos *diagramas de precisión-probabilidad* son muy útiles para poner de manifiesto muchas de las virtudes y debilidades de las diferentes medidas de verificación.

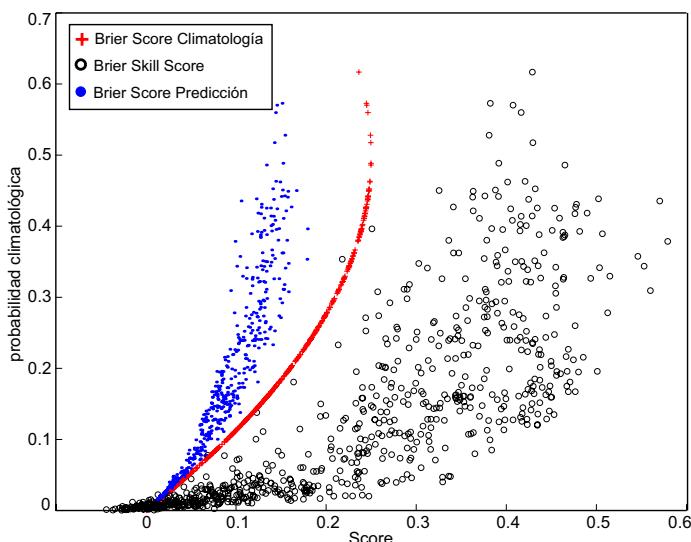


Figura 7.7: Gráfico compuesto con los brier scores y el correspondiente brier skill score para la predicción en 184 estaciones de cuatro eventos asociados a la precipitación.

La Fig. 7.7 muestra un diagrama de precisión-probabilidad para validar el BS correspondiente a la predicción de cuatro eventos distintos ($precip > 0.1\text{ mm}$, $> 2\text{ mm}$, $> 10\text{ mm}$ y $> 20\text{ mm}$) en la red de 184 estaciones completas del INM. Las predicciones se han obtenido utilizando un método de análogos similar al descrito en el capítulo anterior. Este gráfico da una idea muy particular del comportamiento del sistema de predicción.

7.5.2. Brier Skill Score

En general el *Skill Score* (SS) de un sistema con predicciones p , referido a un sistema estándar, por ejemplo la climatología, con probabilidades previstas p_c , se calcula de la forma:

$$SS = \frac{P(o = 1|p) - P(o = 1|p_c)}{1 - P(o = 1|p_c)}$$

donde $P(o = 1|p)$ es la probabilidad de que se produzca el evento dado que el sistema lo predice con probabilidad p . En términos de Brier Scores referidos a la climatología, se tiene:

$$BSS = 1 - \frac{BS}{I} \quad (7.13)$$

La interpretación es muy sencilla: si $BSS > 0$, el sistema de predicción probabilística es mejor que el climatológico ya que entonces $BS < I$; en cambio si $BSS < 0$, entonces el sistema de predicción probabilística no mejora la climatología. El caso $BSS = 1$ corresponde un sistema de predicción probabilística perfecto.

Del mismo modo se pueden definir Skill Scores para dos de las tres componentes del BS (el de incertidumbre no tiene sentido), resultando:

- Brier skill score de fiabilidad: $BSS_f = 1 - BS_f/I$.
- Brier skill score de resolución: $BSS_r = BS_r/I$.

La descomposición en términos de ‘skill’ quedará ahora:

$$BSS = 1 - BSS_f + BSS_r. \quad (7.14)$$

El BSS también tiene los mismos problemas citados anteriormente para el BS, al ser una medida de de precisión global. Por ello, tampoco es adecuada para la validación de eventos raros. Este problema se ilustra en el siguiente ejemplo:

Ejemplo 7.1 (BSS y Eventos Raros). *Supóngase que se quieren comparar dos sistemas A y B para predecir un evento raro cuya $p_c = 0.003$ (en un período de 1000 días, el evento ocurrió en tres ocasiones). El sistema A consiste en predecir siempre la no ocurrencia del evento, $p_i = 0$, por lo que su BS será igual a la probabilidad climatológica:*

$$BS_A = \frac{1}{m} \sum_{i=1}^m o_i^2 = p_c; \quad (7.15)$$

por tanto, el BSS será muy bajo $BSS_A = -p_c/(1 - p_c) \approx -p_c \approx 0$. Por otra parte, el sistema B, acierta dos de las tres ocurrencias, cometiendo otros dos fallos al predecirlo cuando en realidad no ocurrió, lo que supone un total de tres errores. Por tanto, $BSS_A = BSS_B$, con lo que ambos sistemas tienen la misma calidad. Sin embargo, para la mayoría de los usuarios el sistema B es más útil que el A. ■

7.6. Predicciones Binarias Probabilísticas

En la introducción se comentó que en meteorología es muy común transformar variables en binarias, en base a un cierto umbral crítico de su rango (por ejemplo $precip > 10 \text{ mm}$, $viento < 50 \text{ km/h}$, etc.). De forma similar, dada una predicción probabilística para una variable binaria, $p_i = p(\hat{o}_i)$, se puede obtener una predicción determinista considerando un umbral u_p para la probabilidad. Cuando la probabilidad prevista supere ese umbral, se predecirá la ocurrencia del evento ($\hat{o} = si$), y en caso contrario se predecirá la no ocurrencia. De esta forma, variando el umbral de probabilidad se puede hacer que el sistema sea más conservador o más arriesgado para predecir la ocurrencia del evento (hasta que no se tiene una probabilidad superior a u_p , no se predice el evento). Así, se pueden validar las predicciones binarias probabilísticas utilizando los recursos de los deterministas, teniendo además en cuenta la versatilidad que dan los distintos umbrales u_p al sistema de predicción. En estos casos, se tiene una tabla de contingencia diferente dependiendo del valor del umbral u_p . Por tanto, los valores asociados a los aciertos de ocurrencia (α_u), las falsas predicciones (β_u), omisiones (γ_u), y aciertos de no ocurrencia (δ_u), dependerán del umbral u_p .

Las curvas ROC y el valor económico de una predicción binaria probabilística describen de forma global las posibles tablas de contingencia en función de los distintos umbrales u_p . Para ello, utilizan dos índices HIR y FAR (ver Sec. 7.4.1) para cuantificar los “fallos” y “aciertos” del sistema.

7.6.1. Curvas ROC (Relative Operating Characteristics)

Las Curvas ROC se obtienen representando la tasa de falsas alarmas (FAR) frente a la tasa de aciertos (HIR) para los diferentes umbrales de probabilidad u_p entre cero y uno. Obsérvese que para $u_p = 0$ siempre se predice el evento, luego tanto las falsas alarmas como los aciertos son máximos, y corresponden al punto (1, 1) de la curva ROC. En el caso opuesto, $u_p = 1$, nunca se predice el evento, luego ambos índices son mínimos y definen el punto (0, 0) de la curva. Por tanto, todas las curvas ROC comienzan en (1, 1) y terminan en (0, 0). Además, un sistema aleatorio de predicción climatológica que genere 0 con probabilidad $1 - p_c$ y 1 con probabilidad p_c tendrá asociada la diagonal como curva ROC. Este último resultado es fácil de mostrar, pues la tabla de contingencia sería:

		Observado	
		Si	No
Previsto	Si	p_c^2	$p_c(1 - p_c)$
	No	$p_c(1 - p_c)$	$(1 - p_c)^2$

de donde $HIR(p_c) = FAR(p_c) = p_c$. Por otra parte, el punto correspondiente a $u_p = p_c$ en la curva ROC tiene la misma pendiente que la diagonal. Por tanto, en este punto se produce un cambio de la tasa de crecimiento. Por ejemplo,

operando con un umbral inferior a la probabilidad climatológica, el sistema de predicción probabilística comete más falsas alarmas que aciertos, ya que la pendiente de la curva ROC se hace menor que la diagonal.

Por ejemplo, en la Fig. 7.8 se muestran las curvas ROC obtenidas para la predicción de cuatro eventos relacionados con la precipitación ($precip > 0.1, 2, 10, y 20 mm$) en la red de estaciones completas del INM (184 estaciones) utilizando una técnica de análogos similar al sistema Prometeo descrito en el capítulo anterior. Las curvas se han obtenido considerando las predicciones en las 184 estaciones globalmente. Puede verse que, a medida que el evento es más extremo, las falsas alarmas disminuyen para todos los umbrales de probabilidad, dado que la frecuencia climatológica del evento disminuye y, por tanto, la probabilidad de predecir el evento es también baja (si la fiabilidad del sistema es buena). De esta forma, aunque para eventos extremos pueden alcanzarse tasas de acierto relativamente altas con un número pequeño de falsas alarmas, hay que tener en cuenta la definición de FAR para analizar los resultados.

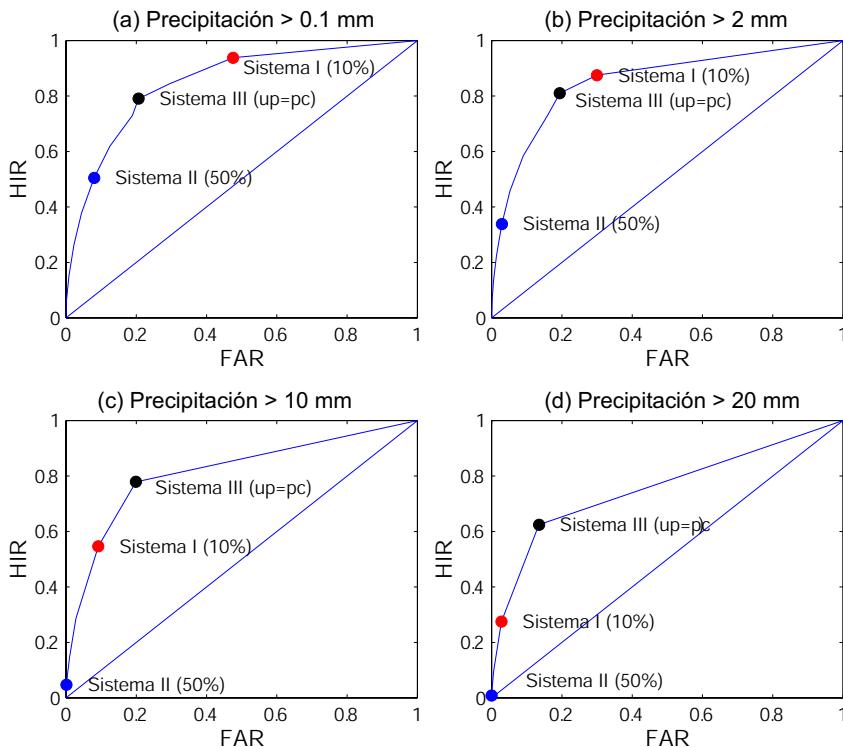


Figura 7.8: Curvas ROC para cuatro eventos de precipitación en una red de 183 estaciones completas del INM.

En la figura se destacan tres sistemas de predicción determinista concretos (asociados a distintos umbrales u_p), que han sido utilizados en secciones anteriores para describir los índices de verificación de sistemas deterministas. Sobre estas curvas se pueden comparar dos sistemas diferentes de predicción. Para

cada umbral de probabilidad, el mejor sistema será aquel cuya tasa HIR sea más alta para un mismo nivel de FAR.

Roc Skill Area (RSA)

En el ejemplo anterior, las curvas ROC proporcionan una descripción cualitativa global del sistema. Sin embargo, en aplicaciones prácticas interesa obtener un índice numérico para poder comparar unos sistemas con otros. El *área encerrada bajo la curva* es un buen índice global de la pericia del modelo. Un sistema de predicción climatológico definiría un área de 0.5, asociada a la diagonal. Por ello, se suele utilizar como índice de pericia asociado a la curva ROC el valor $2 \text{RocArea} - 1$ (Roc Skill Area, RSA). Así el valor máximo del índice es 1 (predicción perfecta) y el valor cero se alcanza para una predicción climatológica aleatoria.

La Fig. 7.9 muestra el área ROC correspondiente a las 184 estaciones por separado, indicando una fuerte dependencia del área ROC con la probabilidad climática del evento. De esto se desprende la conclusión de que un promedio espacial de áreas ROC favorece a aquellos sistemas que predigan mejor en las zonas donde el evento es más frecuente.

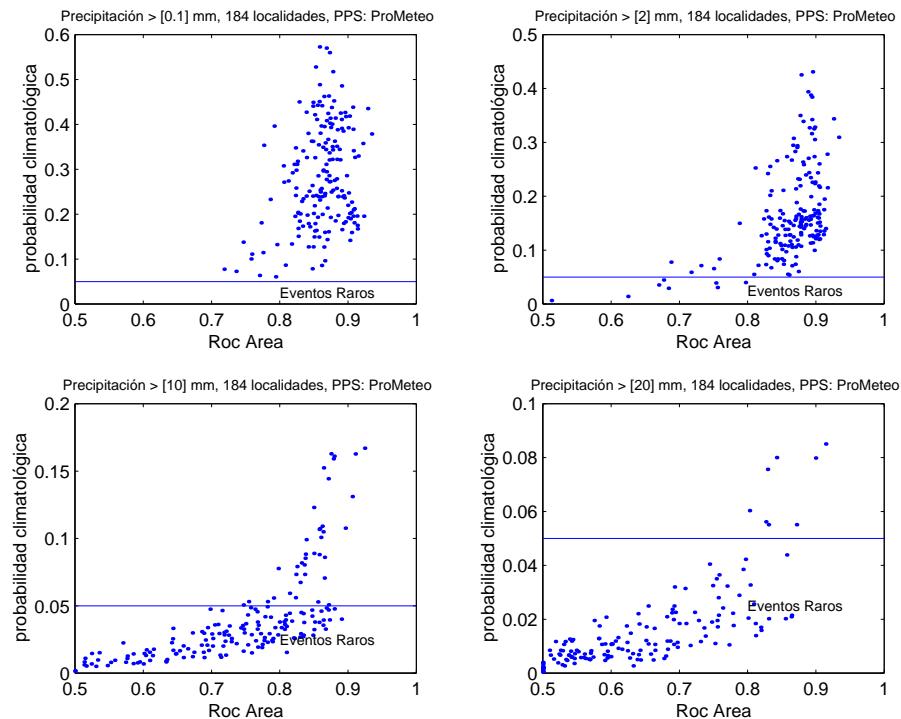


Figura 7.9: Ejemplo de áreas ROC obtenidas para cuatro eventos de precipitación en 183 estaciones completas del INM.

También se suelen utilizar otros índices de verificación basados en las tasas HIR y FAR.

Best Hanssen-Kuipers score $[-1, 1]$

Este índice tiene la ventaja de ser independiente de la distribución de los eventos en la muestra; es decir no depende de si hay muchos o pocos eventos. Para ello, considera un umbral que sea función directa de la probabilidad climatológica $u_p = p_c$ (habitualmente los casos raros requieren un umbral de predicción bajo mientras que los restantes casos requieren un umbral progresivamente mayor):

$$BHK = HIR(p_c) - FAR(p_c) \quad (7.16)$$

Este índice es más equilibrado que los anteriores (precisión y CSI) para un promedio global, especialmente para eventos raros (Doswell et al., 1990).

7.6.2. Valor Económico

La curva ROC determina la pericia de un sistema de predicción probabilístico para los distintos umbrales de probabilidad que dan lugar a distintas predicciones deterministas (más o menos conservadoras) de un cierto fenómeno. La pericia del modelo viene dada por el número de falsas alarmas y de aciertos de las predicciones para cada uno de los umbrales. Así, cada usuario específico de las predicciones puede valorar la calidad de las mismas en base a la configuración que más le interese en su problema; es decir, en función del umbral óptimo para su aplicación. Sin embargo, este sistema de verificación no tiene en cuenta el coste que le suponen a un usuario concreto las distintas fuentes de error. Por ejemplo, proteger una cosecha ante una eventual tormenta de granizo puede suponer a un agricultor un coste menor que las pérdidas que ocasionaría su destrucción, pero para ello, las predicciones tienen que ser lo bastante fiables para que el número de falsas alarmas no encarezca demasiado el precio de protegerse, frente al coste total de las pérdidas derivadas de no protegerse. Las *curvas de valor económico* tienen en cuenta estos factores y permiten calcular el valor económico de un sistema de predicción para un usuario concreto (Katz and Murphy, 1997). Esta situación es especialmente importante en el ámbito de predicción estacional, donde un usuario puede tomar acciones preventivas con suficiente antelación ante una predicción favorable o desfavorable de un fenómeno. Por ejemplo, en el Norte de Perú los agricultores algodoneros retrasan la siembra del algodón cuando esperan un fenómeno fuerte de El Niño. De esa forma, obtienen menores cosechas, pero a cambio evitan que la cosecha se pierda a consecuencia de las lluvias torrenciales de principios de año. Por tanto, el agricultor necesita un criterio sencillo que le permita adoptar aquella estrategia que maximice el beneficio, o minimice los gastos, en función de la pericia del sistema de predicción y de su relación costes/pérdidas.

El modelo de decisión que da lugar a las curvas de valor económico se basa en los siguientes supuestos. Se considera que adoptar una acción preventiva supone un coste C , independientemente de que el evento suceda o no; por otro

lado, si ocurre el evento y no se toma ninguna acción ,se produce una pérdida P . El gasto total se puede obtener a través de la tabla de contingencia del sistema:

		Ocurrencia	
		Si	No
Acción	Si	αC	βC
	No	γP	$\delta 0$

Tabla 7.1: Gastos asociados a las distintas combinaciones de los eventos: ocurrencia del fenómeno y toma o no de acciones preventivas.

De acuerdo con la tabla, el gasto total derivado de utilizar el sistema de predicción probabilística sería:

$$G = \alpha C + \beta C + \gamma P \quad (7.17)$$

Utilizando (7.8) y (7.9), se tiene:

$$G = HIR P_c C + FAR (1 - P_c) C + (1 - HIR) P_c P \quad (7.18)$$

Si el sistema de predicción probabilística fuese perfecto, entonces $HIR = 1$ y $FAR = 0$, por lo que:

$$G_{perfecto} = P_c C \quad (7.19)$$

y utilizando como sistema de predicción probabilística de referencia el climatológico se tiene:

$$G_{climatico} = \min\{C, P_c P\} \quad (7.20)$$

ya que, el gasto generado por las acciones preventivas nunca deberá superar el límite $P_c P$.

En base a todo lo anterior, el *valor económico de una predicción* para un período dado se define como la razón entre los incrementos (respecto de la climatología) de los gastos del sistema en estudio y de un sistema perfecto, y su rango es $[-\infty, 1]$:

$$V = \frac{G - G_{climatico}}{G_{perfecto} - G_{climatico}} \quad (7.21)$$

sustituyendo y poniendo V en función de $R = C/P$, es decir, de la relación *Coste/Pérdidas*, se tiene:

$$V = \frac{HIR P_c R + FAR (1 - P_c) R + (1 - HIR) P_c - \min\{R, P_c\}}{P_c R - \min\{R, P_c\}} \quad (7.22)$$

Ya que tanto HIR como FAR dependen del umbral de probabilidad u_p , V es función de u_p y de la razón coste/pérdidas R ; es decir, existe una curva $V = f(R)$ para cada valor de u_p . Por ejemplo, la Fig. 7.10 muestra las curvas de valor económico asociado a un mismo modelos de predicción probabilística,

considerando dos umbrales distintos para la probabilidad (10 % para el Sistema I y 50 % para el Sistema II). En esta figura se puede observar que mientras uno de los sistemas conserva mayor valor económico para $R < P_c$, el otro sistema se comporta del modo contrario. Por ejemplo, en el Sistema II (asociado al umbral 50 % de probabilidad), a medida que R disminuye, el sistema de predicción probabilística pierde rápidamente su valor económico ya que el coste de la acción preventiva se hace tan pequeño que llega a ser rentable una acción preventiva permanente haciendo que el sistema de predicción probabilística sea obsoleto.

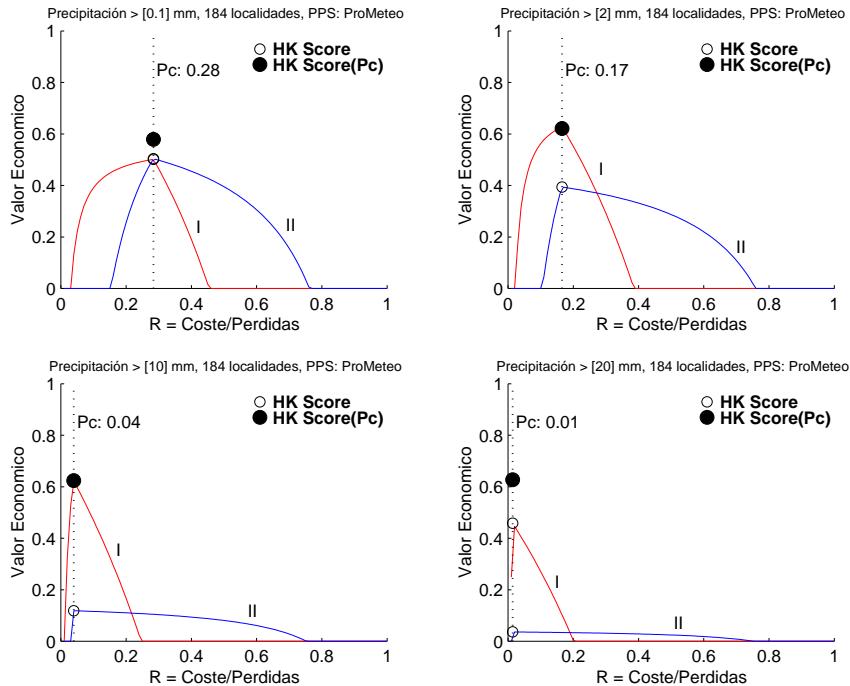


Figura 7.10: Curvas de valor económico para diferentes eventos. Para costes bajos es mejor el sistema I, pero para costes altos es mejor el sistema II.

Como se observa en las curvas anteriores, para cada valor de R existe un único valor de u_p que maximiza el valor económico del sistema de predicción probabilística, que se alcanza cuando $R = P_c$ para cualquier u_p . Además este valor coincide con el índice de Hanssen-Kuipers. Sustituyendo R por P_c en la ecuación 7.22, se tiene:

$$V(u_p, P_c) = \frac{HIR P_c^2 + FAR(1 - P_c), P_c + (1 - HIR) P_c - P_c}{P_c^2 - P_c}, \quad (7.23)$$

luego:

$$V(u_p, P_c) = HIR - FAR = BHK \quad (7.24)$$

Envolvente de la curva de valor económico

Dado que las curvas de valor económico presentan comportamientos distintos para diferentes umbrales de predicción dependiendo de la relación coste pérdidas, en la práctica se suele utilizar una medida del comportamiento global del sistema para todos los posibles umbrales de probabilidad. Para ello, se suele definir la curva de valor económico de un sistema de predicción probabilística como la envolvente de las curvas de valor económico dadas por los distintos umbrales de la probabilidad. Es decir, se toman los umbrales que maximizan el valor económico para cada relación coste pérdidas.

7.7. MeteoLab: Validación

Una paquete importante de la toolbox *MeteoLab* es el de validación, con el cual se pueden analizar aquellos aspectos más relevantes de la bondad de los diferentes experimentos realizados. Su manejo es muy sencillo, como se vio en la Sec. 3.7.2, donde se realizaba una predicción probabilística local de precipitación sobre las estaciones del GSN definidas en la península Ibérica, utilizando una técnica de análogos con 25 vecinos. A continuación se muestran los resultados obtenidos al validar la predicción probabilística obtenida utilizando la técnica de análogos con 10 y 200 vecinos para la estación de Iguelo (San Sebastián).

```
dmn=readDomain('Iberia'); Stations.Network={'GSN'};
Stations.Stations={'Spain.stn'};
Stations.Variable={'Precip'};

%Training data
dates={'1-Jan-1960','31-Dec-1998'};
[EOF,CP]=getEOF(dmn,'ncp',50,'dates',dates);
[dataE,Stations]=loadStations(Stations,'dates',dates,'ascfile',1);
%Test data
dates={'1-Jan-1999','31-Dec-1999'};
[EOF,CPT]=getEOF(dmn,'ncp',50,'dates',dates);
[dataT,Stations]=loadStations(Stations,'dates',dates,'ascfile',1);

umbral=5; i=find(~isnan(dataE)); dataE(i)=dataE(i)>umbral;
i=find(~isnan(dataT)); dataT(i)=dataT(i)>umbral;
O=[];P=[];
for j=1:1:180
    %[AnalogPat,Neig,NeigDist]=getAnalogous(CPT(j,:),CP,10,'knn',[]);
    [AnalogPat,Neig,NeigDist]=getAnalogous(CPT(j,:),CP,200,'knn',[]);
    O=[O;dataT(j,:)];
    P=[P;nanmean(dataE(Neig,:))];
end
Validation=makeValidation(O(:,1),P(:,1));
```

La función `makeValidation` utilizada en el código anterior proporciona una estructura que contiene toda la información sobre la validación probabilística, índices BS de la climatología y de la predicción, índice BSS, área de la curva ROC, HIR, FAR, etc. Por ejemplo, utilizando el método de análogos con 200 vecinos, la validación obtenida se puede obtener como:

```
Validation =
PC: 0.2889
BSP: 0.1284
BSC: 0.2054
BSS: 0.3748
HIR: [11x1 double]
FAR: [11x1 double]
RSA: 0.7136
...
```

La Fig. 7.11 muestran las curvas ROC y de valor económico obtenidas en el ejemplo anterior para la predicción probabilística del evento $Precip > 5mm$ en Igueldo (San Sebastián) para los seis primeros meses del año 1999, usando la técnica de análogos con 10 y 200 vecinos respectivamente.

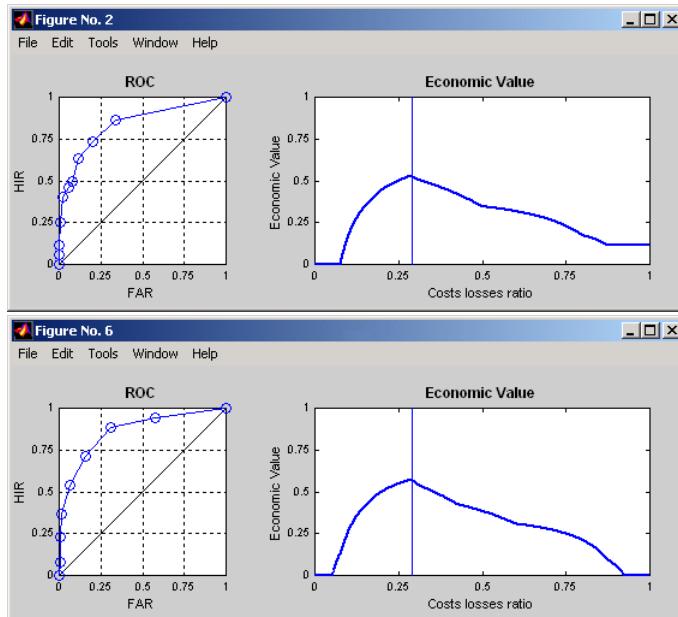


Figura 7.11: Validación (curvas ROC y de valor económico) para predicción probabilística del evento $Precip > 5mm$ en Igueldo (San Sebastián) para los seis primeros meses del año 1999, usando la técnica de análogos con 10 (superior) y 200 (inferior) vecinos.

Por otra parte, la Fig.7.12 muestra las respectivas curvas de fiabilidad y resolución. Puede observarse que la resolución del sistema con 200 análogos es claramente peor que la del sistema con 10 análogos, que muestra una estructura bimodal en forma de 'U'.

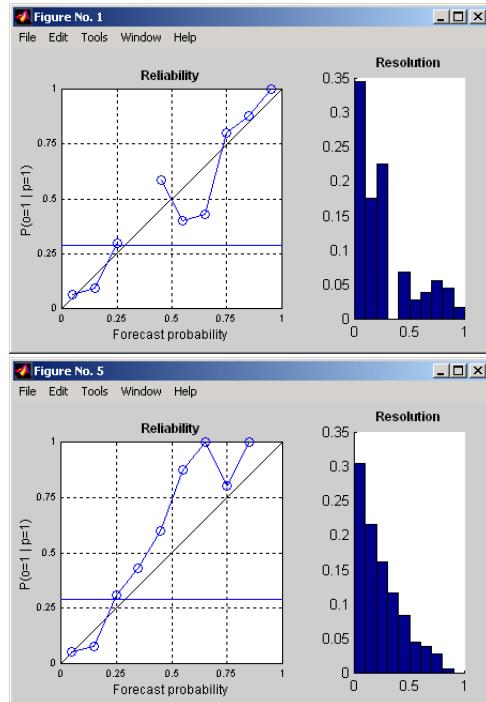


Figura 7.12: Validación (curvas de fiabilidad y resolución) para predicción probabilística del evento $\text{Precip} > 5\text{mm}$ en Igueldo (San Sebastián) para los seis primeros meses del año 1999, usando la técnica de análogos con 10 (superior) y 200 (inferior) vecinos.

Bibliografía

- ABARBANEL, H. D. I. (1995). *Analysis of Observed Chaotic Data*. Springer.
- AGRAWAL, R., IMIELINSKI, T., and SWAMI, A. (1993). Mining association rules between sets of items in large databases. In P. Buneman and S. Jalodia, eds., *Proceedings of the International Conference on Management of Data*, vol. 22(2), pp. 207–216. ACM Press.
- ALEXANDERSSON, H. (1986). Homogeneity test applied to precipitation data. *Journal of Climatology*, 6:661–675.
- ANDERBERG, M. R. (1973). *Cluster Analysis for Applications*. Academic Press, New York.
- ANDERSON, J. L. (1996). Selection of initial conditions for ensemble forecasts in a simple perfect model framework. *Journal of Atmospheric Sciences*, 53:22–35.
- AYUSO, J. J. (1994). *Predicción Estadística Operativa en el INM*, vol. B-34 of *Monografías del Instituto Nacional de Meteorología*. Ministerio de Medio Ambiente, Madrid.
- BARRY, R. and CHORLEY, R. (1998). *Atmosphere, Weather and Climate*. Routledge, London, 7th ed.
- BENISTON, M. (1998). *From Turbulence to Climate: Numerical Investigations of the Atmosphere with a Hierarchy of Models*. Springer-Verlag.
- BERGMAN, M. J. and DELLEUR, J. W. (1985). Kalman filter estimation and prediction of daily streamflow: 1- review, algorithm, and simulation experiments; 2-application to the Potomac river. *Water Resources Bulletin*, 21(5):815–832.
- BILLET, J., DELISI, M., SMITH, B. G., and GATES, C. (1997). Use of regression techniques to predict hail size and the probability of large hail. *Weather and Forecasting*, 12:154–164.

- BISHOP, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.
- BOX, G. E. P. and JENKINS, F. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, Oakland, CA, 2nd ed.
- BRIER, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3.
- BUHAMRA, S., SMAOUI, N., and GABR, M. (2003). The Box-Jenkins analysis and neural networks: Prediction and time series modelling. *Applied Mathematical Modelling*, 27:805–815.
- BUIZZA, R. and PALMER, T. N. (1995). Singular vector structure of the atmospheric global circulation. *Journal of Atmospheric Sciences*, 52:1434.
- BUIZZA, R., RICHARDSON, D. S., and PALMER, T. N. (2001). The new 80km high-resolution ECMWF EPS. ECMWF Newsletter 90, ECMWF.
- BUNTINE, W. (1991). Theory refinement on bayesian networks. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pp. 52–60. Morgan Kaufmann Publishers.
- BURGUER, G. (1996). Expanded downscaling for generating local weather scenarios. *Climate Research*, 7:111–128.
- CANO, R., COFIÑO, A. S., PRIMO, C., SORDO, C., and GUTIÉRREZ, J. M. (2003). Downscaling demeter seasonal ensemble precipitation forecasts in the tropics during el niño episodes. *Geophysical Research Abstracts*, 5:14586.
- CANO, R., COFIÑO, A. S., GUTIÉRREZ, J. M., and SORDO, C. (2002). Probabilistic networks for statistical downscaling and spatialisation of meteorological data. In *Geophysical Research Abstracts*, vol. 4, p. 194.
- CANO, R., SORDO, C., and GUTIÉRREZ, J. M. (2004). Bayesian networks in meteorology. In J. A. Gámez, S. Moral, and A. Salmerón, eds., *Advances in Bayesian Networks*, pp. 309–327. Springer Verlag.
- CARSON, D. J. (1998). Seasonal forecasting. *Quarterly Journal of the Royal Meteorological Society*, 124:1–26.
- CASTILLO, E., COBO, A., GUTIÉRREZ, J. M., and PRUNEDA, E. (1999). *An Introduction to Functional Networks with Applications*. Kluwer Academic Publishers, Boston.
- CASTILLO, E., FONTELA-ROMERO, O., GUIJARRO-BERDIÑAS, B., and ALONSO-BETANZOS, A. (2002). A global optimum approach for one-layer neural networks. *Neural Computation*, 14(6):1429–1449.
- CASTILLO, E., GUTIÉRREZ, J. M., and HADI, A. S. (1998). Modelling probabilistic networks of discrete and continuous variables. *Journal of Multivariate Analysis*, 64:48–65.

- CASTILLO, E., GUTIÉRREZ, J. M., and HADI, A. S. (1997). *Expert Systems and Probabilistic Network Models*. Springer-Verlag. Free Spanish version <http://personales.unican.es/gutierjm>.
- CAVAZOS, T. (1997). Downscaling large-scale circulation to local winter rainfall in north-eastern mexico. *International Journal of Climatology*, 17:1069–1082.
- CAVAZOS, T. (2000). Using Self-Organizing Maps to investigate extreme climate event: An application to wintertime precipitation in the Balkans. *Journal of Climate*, 13:1718–1732.
- CHATFIELD, C. (2003). *The Analysis of Time Series: An Introduction*. CRC Press, 6th ed.
- CHATTERJEE, S., HADI, A. S., and PRICE, B. (2000). *Regression Analysis by Example*. John Wiley and Sons, 3rd ed.
- CHEN, M., XIE, P., JANOWIAK, J. E., and ARKIN, P. A. (2002). Precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology*, 3:249–266.
- COFIÑO, A. S. (2004). *Técnicas Estadísticas y Neuronales de Agrupamiento Adaptativo para la Predicción Probabilística de Fenómenos Meteorológicos Locales. Aplicación en el Corto Plazo y en la Predicción Estacional*. Ph.D. thesis, Universidad de Cantabria.
- COFIÑO, A. S., GUTIÉRREZ, J. M., and IVANISSEVICH, M. L. (2004). Evolving modular networks with genetic algorithms. Application to nonlinear time series. *Expert Systems*. In press.
- COFIÑO, A. S., CANO, R., SORDO, C., and GUTIÉRREZ, J. M. (2002). Bayesian networks for probabilistic weather prediction. In *Proceedings of the 15th European Conference on Artificial Intelligence*, pp. 695–700.
- COFIÑO, A. S., GUTIÉRREZ, J. M., JAKUBIAK, B., and MELONEK, M. (2003). Implementation of data mining techniques for meteorological applications. In W. Zwiefelhofer and N. Kreitz, eds., *Proceedings of the Tenth ECMWF Workshop on the Use of High Performance Computing in Meteorology: Realizing TeraComputing*, pp. 256–271. World Scientific Publishing.
- COOPER, G. F. and HERSKOVITZ, E. (1992). A bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347.
- CORAZZA, M., KALNAY, E., PATIL, D. J., YANG, S.-C., MORSS, R., CAI, M., SZUNYOGH, I., HUNT, B. R., and YORKE, J. A. (2003). Use of the breeding technique to estimate the structure of the analysis ‘errors of the day’. *Nonlinear Processes in Geophysics*, 10:233–243.

- CORTE-REAL, J., QUIAN, B., and XU, H. (1999). Circulation patterns, daily precipitation in Portugal and implications for climate change simulated by the second Hadley Centre GCM. *Climate Dynamics*, 15:921–935.
- CYBENKO, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2:303–314.
- DE CAMPOS, L. M., FERNÁNDEZ-LUNA, J. M., and PUERTA, J. M. (2003). An iterated local search algorithm for learning Bayesian networks with restarts based on conditional independence tests. *International Journal Intelligent Systems*, 18(2):221–235.
- DEGROOT, M. H. (1989). *Probability and Statistics*. Addison Wesley, Reading, MA.
- DEVROYE, L. (1986). *Non-Uniform Random Variate Generations*. Springer Verlag, New York.
- DOSWELL, C. A., DAVIES-JONES, R., and KELLER, D. L. (1990). On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting*, 5:576–585.
- DUDA, R. O., HART, P. E., and STORK, D. G. (2001). *Pattern Classification*. Wiley-Interscience, 2nd ed.
- EBERT, E. E. and MCBRIDE, J. L. (2000). Verification of precipitation in weather systems: Determination of systematic errors. *Journal of Hydrology*, 239:179–202.
- ECKERT, C. D., P. and AMBÜHL, J. (1996). Classification of ensemble forecasts by means of an artificial neural network. *Meteor. Appl.*, 3:169–178.
- ENKE, W. and SPEKAT, A. (1997). Downscaling climate model outputs into local and regional weather elements by classification and regression. *Climate Research*, 8:195–207.
- ESTEBAN-PARRA, M. J., POZO-VÁZQUEZ, D., CASTRO-DÍEZ, Y., and TRIGO, R. M. (2003). NAO influence on maximum and minimum temperature of the Iberian Peninsula. In *14th Symposium on Global Change and Climate Variations. 83rd Annual Conference of the AMS..*
- FAUFOULA-GEORGIOU, E. and LETTENMAIER, D. P. (1987). A Markov renewal model for rainfall occurrences. *Water Resources Research*, 23:875–884.
- FEDDERSEN, H. and ANDERSEN, U. (2004). A method for statistical downscaling of seasonal ensemble predictions. To appear in Tellus A.
- FONTELÀ-ROMERO, O., ALONSO-BETANZOS, A., CASTILLO, E., PRINCIPE, J. C., and GUIJARRO-BERDIÑAS, B. (2002). Local modeling using self-organizing maps and single layer neural networks. *Lectures Notes in Computer Science*, 2415:945–950.

- FOSTER, I. and KESSELMAN, C. (2003). *The Grid 2: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Publishers.
- FOVELL, R. G. and FOVELL, M. Y. C. (1993). Climate zones of the conterminous United States defined using cluster analysis. *Journal of Climate*, 6:2103–2135.
- GABRIEL, K. R. and NEUMANN, J. (1962). A Markov chain model for daily rainfall occurrences at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society*, 88:90–95.
- GARDNER, M. W. and DORLING, S. R. (1998). Artificial neural networks (the multilayer perceptron). A review of applications in the atmospheric sciences. *Journal of Applied Meteorology*, 39:147–159.
- GEIGER, D. and HECKERMAN, D. (1994). Learning Gaussian networks. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pp. 235–243. Morgan Kaufmann Publishers.
- GEIGER, D. and HECKERMAN, D. (1995). A characterization of the Dirichlet distribution with application to learning Bayesian networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 196–207. Morgan Kaufmann Publishers.
- GLASBEY, C. A. and NEIVISON, I. M. (1997). Rainfall modelling using a latent Gaussian variable. *Lecture Notes in Statistics*, 122:233–242.
- GOLLUB, J. P. and CROSS, M. C. (2000). Chaos in space and time. *Nature*, 404:710–711.
- GRASSBERGER, P. and PROCACCIA, I. (1983). Characterization of strange attractors. *Physical Review Letters*, 50:346–349.
- GROSSBERG, S. (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23:187–202.
- GUTIÉRREZ, J. M., CANO, R., COFIÑO, A. S., and RODRÍGUEZ, M. A. (2004). Clustering methods for statistical downscaling in short-range weather forecast. *Monthly Weather Review (en prensa)*.
- GUTIÉRREZ, J. M., CANO, R., COFIÑO, A. S., and RODRÍGUEZ, M. A. (1999). Redes neuronales y patrones de analogías aplicados al downscaling climático. In *La Climatología Española en los Albores del Siglo XXI*, pp. 113–121. Asociación Española de Climatología.
- GUTIÉRREZ, J. M., COFIÑO, A. S., CANO, R., and SORDO, C. (2002a). A generalization of analogue downscaling methods by bayesian networks. In *International Conference on Quantitative Precipitation Forecasting*, pp. 87–87. The World Weather Research Programme's WWRP.

- GUTIÉRREZ, J. M., COFIÑO, A. S., CANO, R., and SORDO, C. (2002b). Probabilistic networks for statistical downscaling and spatialisation of meteorological data. *Geophysical Research Abstracts*, 4:192.
- GUTIÉRREZ, J. M. and RODRÍGUEZ, M. A. (2000). A new exact method for obtaining the multifractal spectrum of multinomial measures and IFSP measures. *Chaos, Solitons and Fractals*, 11(5):675–683.
- GUTIÉRREZ, J. M., COFIÑO, A. S., CANO, R., and RODRÍGUEZ, M. A. (2001). Aplicación de métodos de clasificación al downscaling estadístico. In *V Simposio Nacional de Predicción.*, pp. 235–240. Instituto Nacional de Meteorología.
- HASTIE, T., TIBSHIRANI, R., and FRIEDMAN, J. (2001). *The Elements of Statistical Learning*. Springer, New York.
- HAYKIN, S. (1998). *Neural Networks: A Comprehensive Foundation*. Pearson Education, 2nd ed.
- HENDON, H. H. (2001). Indonesian rainfall variability: Impacts of ENSO and local air-sea interaction. *Journal of Climate*, 16:1775–1790.
- HENRION, M. (1988). Propagation of uncertainty by logic sampling in Bayes' networks. In J. F. Lemmer and L. N.. Kanal, eds., *Proceedings of the 2nd Annual Conference on Uncertainty in Artificial Intelligence*, pp. 149–164.
- HEWITSON, B. C. and CRANE, R. G. (2002). Self-organizing maps: Applications to synoptic climatology. *Climate Research*, 22(1):13–26.
- HOERL, A. E. and KENNARD, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–82.
- HOLTON, J. (1992). *An Introduction to Dynamic Meteorology*. Academic Press.
- HOTELLING, H. (1936). Relations between two sets of variables. *Biometrika*, 28:321–377.
- HSIEH, W. W. (2000). Nonlinear canonical correlation analysis by neural networks. *Neural Networks*, 13:1095–1105.
- HSIEH, W. W. (2001). Nonlinear canonical correlation analysis of the tropical Pacific climate variability using a neural network approach. *Journal of Climate*, 14:2528–2539.
- HSIEH, W. W. and TANG, B. (1998). Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society*, 79:1855–1870.
- HUGHES, J., LETTEMAIER, D. P., and GUTTORP, P. (1993). A stochastic approach for assessing the effect of changes in regional circulation patterns on local precipitation. *Water Resources Research*, 29:3303–3315.

- HURRELL, J. W. (1995). Decadal trends in the north atlantic oscillation and relationships to regional temperature and precipitation. *Science*, 269:676–679.
- JANSÁ, A. (1990). *Notas sobre análisis meteorológico mesoescalar en niveles bajos..* Instituto Nacional de Meteorología, Madrid.
- JOLLIFFE, I. T. and STEPHENSON, D. B., eds. (2003). *Forecast Verification: A Practitioner's Guide in Atmospheric Science.* John Wiley and Sons.
- JORDAN, M. I. (1998). *Learning in Graphical Models.* MIT Press.
- JUDD, K. (2003). Nonlinear state estimation, indistinguishable states, and the extended Kalman filter. *Physica D*, 183:273–281.
- KALKSTEIN, L. S., TAN, G., and SKINDLOV, J. A. (1987). An evaluation of three clustering procedures for use in synoptic climatological classification. *Journal of Climate and Applied Meteorology*, 26:717–730.
- KALNAY, E. (2003). *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, Cambridge.
- KALNAY, E., KANAMITSU, M., KISTLE, R., COLLINS, W., DEAVEN, D., GANDIN, L., IREDELL, M., SAHA, S., WHITE, G., WOOLLEN, J., ZHU, Y., CHELLIAH, M., EBISUZAKI, W., HIGGINS, W., JANOWIAK, J., MO, K. C., ROPELEWSKI, C., WANG, J., LEETMAA, A., REYNOLDS, R., JENNE, R., and JOSEPH, D. (1996). The NMC/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77:437–471.
- KATZ, R. and MURPHY, A., eds. (1997). *Economic Value of Weather and Climate Forecasts.* Cambridge University Press.
- KLEIN, W. H. and GLAHN, H. R. (1974). Forecasting local weather by means of model output statistics. *Bulletin of the American Meteorological Society*, 55:1217–1227.
- KOHONEN, T. (2000). *Self-Organizing Maps.* Springer-Verlag, Berlin, 3rd ed.
- KÖPPEN, W. (1918). Klassifikation der klimate nach temperatur, niederschlag und jahreslauf. *Petermanns Mitt*, 64:193–203.
- KRAMER, M. A. (1991). Nonlinear principal component analysis using auto-associative neural networks. *Neural Computation*, 9(7):1493–1516.
- LINDLEY, D. V. (1987). The probability approach to the treatment of uncertainty in artificial intelligence. *Statistical Science*, 2:17–24.
- LORENZ, E. N. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Science*, 20:130–141.
- LORENZ, E. N. (1969). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, 26:636–646.

- LORENZ, E. N. (1986). On the existence of a slow manifold. *Journal of Atmospheric Science*, 43:1547–1557.
- LORENZ, E. N. (1991). Dimension of weather and climate attractors. *Nature*, 353:241–244.
- LORENZ, E. N. (1996). *The Essence of Chaos*. University of Washington Press.
- MACEZO, M., COOK, D., and BROWN, T. J. (2000). Visual data mining in atmospheric science data. *Data Mining and Knowledge Discovery*, 4:69–80.
- MANDELBROT, B. (2004). *Fractals and Chaos*. Springer Verlag.
- MARZBAN, C. (2003). A neural network for post-processing model output: ARPS. *Monthly Weather Review*, 131(6):1103–1111.
- MCGINNIS, D. L. (1994). Predicting snowfall from synoptic circulation: A comparison of linear regression and neural networks. In B. Hewitson and R. G. Crane, eds., *Neural Nets: Applications in Geography*, pp. 79–99. Kluwer Academic Publishers.
- MIDDLETON, G. V. (2000). *Data Analysis in the Earth Sciences Using MATLAB*. Prentice Hall.
- MIYAKODA, K., HEMBREE, G. D., STRICKLER, R. F., and SHULMAN, I. (1972). Cumulative results of extended forecast experiments: I. model performance for winter cases. *Monthly Weather Review*, 100(12):836–855.
- MO, K. C. and LIVEZEY, R. E. (1986). Tropical-extratropical geopotential height teleconnections during the Northern Hemisphere winter. *Monthly Weather Review*, 114:2488–2515.
- MURPHY, A. H. (1973). A new vector partition of probability score. *Journal of Applied Meteorology*, 12:595–600.
- MURPHY, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, 8:281–293.
- MURPHY, A. H. and WINKLER, R. L. (1987). A general framework for forecast verification. *Monthly Weather Review*, 115:1330–1338.
- NEAPOLITAN, R. E. (2003). *Learning Bayesian Networks*. Prentice Hall.
- NOGUER, M. (1994). Using statistical techniques to deduce local climate distributions. an application for model validation. *Meteorological Applications*, 1:277–287.
- OJA, E. and KASKI, S. (1999). *Kohonen Maps*. Elsevier, Amsterdam.
- OLIVER, J. (1991). The history, status and future of climatic classification. *Physical Geography*, 12:231–251.

- ORLANSKI, I. (1975). A rational subdivision of scales for atmospheric processes. *Bulletin of the American Meteorological Society*, 56(5):527–530.
- PALMER, T. N., ALESSANDRI, A., ANDERSEN, U., CANTELAUBE, P., DAVEY, M., DELECLUSE, P., DEQUE, M., DIEZ, E., DOBLAS-REYES, F. J., FEDDERSEN, H., GRAHAM, R., GUALDI, S., GUEREMY, J. F., HAGEDORN, R., HOSHEN, M., KEENLYSIDE, N., LATIF, M., LAZAR, A., MAISONNAVE, E., MARLETTA, V., MORSE, A. P., ORFILA, B., ROGEL, P., TERRES, J. M., and THOMSON, M. C. (2004). Development of a european ensemble system for seasonal to inter-annual prediction (DEMENTER). *Bulletin of the American Meteorological Society*, 85(6):853–872. URL <http://www.ecmwf.int/research/demeter/>.
- PEÑA, J. M., LOZANO, J. A., and LARRAÑAGA, P. (1999). An empirical comparison of four initialization methods for the k-means algorithm. *Pattern Recognition Letters*, 20:1027–1040.
- PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- PÉREZ-MUÑUZURI, V. and GELPI, I. R. (2000). Application of nonlinear forecasting techniques for meteorological modeling. *Annales Geophysicae*, 18:1349–1359.
- PHILANDER, S. G. (1990). *El Niño, La Niña, and the Southern Oscillation*. Academic Press, San Diego.
- POTTS, J. M., FOLLAND, C. K., JOLLIFFE, I. T., and SEXTON, D. (1996). Revised “LEPS” scores for assessing climate model simulations and long-range forecasts. *Journal of Climate*, 9:34–53.
- POZO-VÁZQUEZ, D., ESTEBAN-PARRA, M. J., RODRIGO, F. S., and CASTRO-DÍEZ, Y. (2000). A study of NAO variability and its possible non-linear influences on European surface temperature. *Climate Dynamics*, 17:701–715.
- PREISENDORFER, R. W. and MOBLEY, C. D. (1988). *Principal component analysis in meteorology and oceanography*. Elsevier, Amsterdam.
- PRESS, W. H., TEULOSKY, S. A., VETTERLING, W. T., and FLANNERY, B. P. (1992). *Numerical Recipies*. Cambridge University Press, Cambridge, 2nd ed.
- PU, Z.-X., KALNAY, E., PARRISH, D., WU, W., and TOTH, Z. (1997). The use of the bred vectors in the NCEP operational 3-dimensional variational system. *Weather and Forecasting*, 12:689–695.
- RIBALAYGUA, J. and BORÉN, R. (1995). Clasificación de patrones espaciales de precipitación diaria sobre la españa peninsular y baleárica. Informe nº3 del servicio de análisis e investigación del clima, INM, Madrid.

- RICHARDSON, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17:182–190.
- RICHARDSON, D. S. (2000). Skill and economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 126:649–668.
- RODRÍGUEZ-FONSECA, B. and SERRANO, E. (2002). Winter 10-day coupled patterns between geopotential height and Iberian Peninsula rainfall using the ECMWF precipitation reanalysis. *Journal of Climate*, 15:1309–1321.
- RODRIGUEZ-ITURBE, I., COX, D. R., and ISHAM, V. (1987). Some models for rainfall based on stochastic point processes. *Proc. of the Royal Society of London A*, 410:269–288.
- ROSENBLAT, F. (1962). *Principles of Neurodynamics*. Spartan, New York.
- RUBINSTEIN, R. Y. (1981). *Simulation and the Monte Carlo Method*. John Wiley and Sons, New York.
- RUMELHART, D. E. and MCCLELLAND, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. The MIT Press, Cambridge.
- SAUER, T. (1994). Time series prediction by using delay coordinate embedding. In A. S. Weigend and N. A. Gershenfeld, eds., *Time Series Prediction: Forecasting the Future and Understanding the Past*, pp. 175–193. Addison-Wesley.
- SCHIZAS, C. N., PATTICHHIS, C. S., and MICHAELIDES, S. C. (1994). Artificial neural networks in weather forecasting. *Neural Networks*, pp. 219–230.
- SCHÖLKOPF, B., SMOLA, A., and MULLER, K. R. (1998). Kernel principal component analysis. In B. Schölkopf, C. Burges, and A. Smola, eds., *Advances in Kernel Methods: Support Vector Machines*, pp. 327–352. MIT Press.
- SCHOOF, J. T. and PRYOR, S. C. (2001). Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. *International Journal of Climatology*, 21(7):773–790.
- STEPANEK, P. (2001). AnClim: Software for time series analysis. Dept. Geography, Fac. of Natural Sciences, MU, Brno.
- STERN, R. D. (1982). Computing a probability distribution for the start of the rains from a Markov chain model for precipitation. *Journal of Applied Meteorology*, 21(3):420–422.
- STROGATZ, S. (2001). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering*. Perseus Book Group, 1st ed.
- TALAGRAND, O. (1997). Evaluation of probabilistic prediction systems. In *Workshop on predictability*, pp. 1–25. ECMWF.

- TOTH, Z. (1991). Circulation patterns in phase space. A multinormal distribution? *Monthly Weather Review*, 119(7):1501–1511.
- TSONIS, A. A. and ELSNER, J. B. (1988). The wheater attractor over very short time scales. *Nature*, 333:545–547.
- VAN DEN DOOL, H. M. (1989). A new look at weather forecasting though analogs. *Monthly Weather Review*, 117:2230–2247.
- VAN OLDENBORGH, G. J., BURGERS, G., and TANK, A. K. (2000). On the El-Niño teleconnection to spring precipitation in Europe. *International Journal of Climatology*, 20:565–574.
- VERMA, U., YADAV, M., and HASIJA, R. C. (2002). A seasonal ARIMA model for monthly rainfall sequence. In *16th Australian Statistical Conference*. Canberra, Australia.
- VON STORCH, H. (1999). On the use of “inflation” in statistical downscaling. *Journal of Climate*, 12:3505–3506.
- VON STORCH, H. and ZWIERS, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge.
- WANNER, H., BRONNIMANN, S., CASTY, C., FYALISTRAS, D., LUTERBACHER, J., SCHMUTZ, C., STEPHENSON, D. B., and XOPLAKI, E. (2001). North Atlantic Oscillation. Concepts and studies. *Surveys in Geophysics*, 22:321–382.
- WARD, M. N. and FOLLAND, C. K. (1991). Prediction of seasonal rainfall in the Nordeste of Brazil using eigenvectors of sea-surface temperature. *International Journal of Climatology*, 11:711–743.
- WILBY, R. L. and WIGLEY, T. M. L. (1997). Downscaling general circulation model output. A review of methods and limitations. *Progress in Physical Geography*, 21:530–548.
- WILBY, R. L. and WILKS, D. S. (1999). The weather generation game. A review of stochastic weather models. *Progress in Physical Geography*, 23:329–357.
- WILKS, D. S. (1995). *Statistical Methods in the Atmospheric Sciences. An Introduction*. Academic Press, San Diego.
- WILKS, D. S. (1999a). Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, 93:159–163.
- WILKS, D. S. (1999b). Multisite downscaling of daily precipitation with a stochastic weather generator. *Climate Research*, 11:125–136.
- WILKS, D. S. (1999c). Multisite downscaling of daily precipitation with a stochastic weather generator. *Climate Research*, 11:125–136.

- WILSON, L. J. (2001). Verification of precipitation forecasts: A survey of methodology. Tech. rep., Environment Canada. URL www.chmi.cz/meteo/ov/wmo/src/wilson_paper_prague.pdf.
- WMO (1994). Guide to WMO binary code form GRIB 1. Technical Report 17, World Meteorological Organization (WMO).
- WUNSCH, C. (1999). The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillation. *Bulletin of the American Meteorological Society*, 80:245–255.
- YUVAL and HSIEH, W. W. (2003). An adaptive nonlinear MOS scheme for precipitation forecasts using neural networks. *Weather and Forecasting*, 18(2):303–310.
- ZORITA, E., HUGHES, J. P., P, L. D., and VON STORCH, H. (1995). Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. *Journal of Climate*, 8:1023–1042.
- ZORITA, E. and VON STORCH, H. (1999). The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. *Journal of Climate*, 12:2474–2489.
- ZWIERS, F. W. and VON STORCH, H. (1990). Regime dependent autoregressive time series modelling of the Southern Oscillation. *Journal of Climate*, 3:1347–1363.

Índice alfabético

- Adaptive resonance theory, 199
Agrupamiento, 82
 k-medias, 85
 técnicas jerárquicas, 83
Análogos, 120
 longitud reanálisis, 121
Aprendizaje
 backpropagation, 187
 no supervisado, 179, 198, 201
 redes Bayesianas, 146
 regla delta, 181
 supervisado, 178
 tasa de, 181
ARPS, 18
Artic Oscillation, 74
Autoregresivo, 95
 neuronal, 188
Backpropagation, 187
Bias, 245
 condicionado, 246
BNT, 140
BSS, 255
Caos, 18, 97
CCA, 109
 escala temporal, 115
Climatología, 243
Clustering, 82
COAMPS, 18
Componentes Principales, 68
 escala temporal, 81
 número de componentes, 79
Correlación, 64
 canónica, 109
 espacial, 114
CP, 68
D-separación, 137
Data mining, 129
DEMETER, 28
Dependencia, 61, 135
 condicional, 62, 134, 136
 espacial, 118
Descenso de gradiente, 181
Downscaling, 92
 análogos, 118
 dinámico, 24, 93
 estadístico, 94
 físico, 94
 MOS, 117
 perfect prog, 117
 redes Bayesianas, 165
El Niño, 12, 65, 73, 96, 102
 predicción de, 29, 216
ENSO, 13
EOF, 68
EPS, 21, 26
 condición inicial, 22
 multi-modelo, 23
ERA, 39
Error máximo, 179
FAR, 251
Fiabilidad, 246

- Fractal, 19
 Función de base radial, 194
 Función neuronal
 lineal, 178
 sigmoidal, 178
 HIR, 251
 HIRLAM, 24, 226
 Información mutua, 64
 k-medias, 85, 158, 228
 Lorenz, 19
 método análogos, 120
 predicción, 188
 Markov
 cadenas de, 56
 redes de, 137
 Matlab
 BNT, 140
 MeteoLab, 31
 NetLab, 183
 Matriz de covarianzas, 104
 MeteoLab, 31
 agrupamiento, 89
 análogos, 122
 correlación canónica, 111
 CP y EOF, 74
 Lorenz, 193
 modelos autoregresivos, 96
 observaciones, 34
 patrones atmosféricos, 41
 redes Bayesianas, 140, 148, 153
 regresión, 107
 validación, 262
 Minería de datos, 129
 MM5, 18, 26, 93
 red Ibérica, 26
 Modelo
 AR, 95
 ARIMA, 96
 ARMA, 95
 autoregresivo, 95
 autoregresivo neuronal, 189
 de Lorenz, 19, 188
 de Markov, 100
 de renovación, 100
 difusivo acoplado, 20
 no lineal, 18, 97
 point process, 100
 regresión, 101
 SARIMA, 96
 Modelos del Tiempo
 ARPS, 18
 COAMPS, 18
 de circulación general, 15
 HIRLAM, 24
 mesoscalares, 17
 MM5, 18
 regionales, 16
 Monzón, 14
 MOS, 117
 NAO, 12, 72, 111, 115, 203
 NetLab, 183
 Oscilación
 del Artico, 74
 del Sur, 12
 El Niño, 73
 ENSO, 13
 NAO, 12, 72
 Perfect Prog, 117
 Persistencia, 243
 Precipitación
 cantidad de, 60
 carácter dual, 50
 correlación espacial, 66
 El Niño, 13, 65, 102
 generadores de tiempo, 57
 Predecibilidad, 21
 Predicción
 climática, 30
 con análogos, 120
 corto plazo, 15, 24, 39, 116, 124
 determinista, 243
 estacional, 28, 215
 medio plazo, 26
 por conjuntos, 21
 probabilística, 30, 243
 redes neuronales, 206

- Probabilidad, 48
condicionada, 54
marginal, 54
regla de la cadena, 134
Proceso estacionario, 95

Reanálisis, 39
Red funcional, 175
Red Neuronal, 175
Red neuronal
 ART, 199
 clasificación, 183
 competitiva, 179, 198
 de base radial, 194
 multicapa, 186
 perceptrón, 180
 SOM, 201
Redes Bayesianas, 138
 aplicaciones, 155
 aprendizaje, 146
 Gaussianas, 139
 inferencia, 139
 mixtas, 139
 multinomiales, 138
 predicción local, 165
 red pluviométrica, 160
 relleno lagunas, 169
 weather generators, 162
Redes probabilísticas, 129
 Bayesianas, 138
 de Markov, 137
Regla de la cadena, 134
Regla delta, 181
Regresión, 101
 lineal, 101
 lineal multivariada, 104
 no lineal, 106
Regularización, 195
Resolución, 247
RMSE, 179
ROC, 256
 área curva, 258
RSA, 258

Series temporales, 94
 no lineales, 97
redes neuronales, 188
técnicas de inmersión, 98, 191
Sesgo, 245
Simulación, 59
Sistema experto, 129
Sistema inteligente, 129
Sobreajuste, 180
SOM, 201
Southern Oscillation, 12
SSE, 179

Teleconexión, 12, 74

V-estructura, 137
Validación
 Asociación, 246
 Best Hannan-Kuijpers, 259
 Bias, 248
 Brier Score, 253
 Brier Skill Score, 255
 BS, 253
 Correlación, 249
 Critical Success Index, 252
 Curvas Roc, 256
 FAR, 251
 Fiabilidad, 246
 Habilidad, 246
 HIR, 251
 LEPS, 249
 MAE, 248
 MSE, 248
 Precisión, 246, 251
 Resolución, 247
 RMSE, 249
 RMSSSE, 250
 RSA, 258
 sesgo, 245
 valor económico, 259
Valor económico, 259
 envolvente, 262

Weather generator, 56, 162
Winer takes all, 198