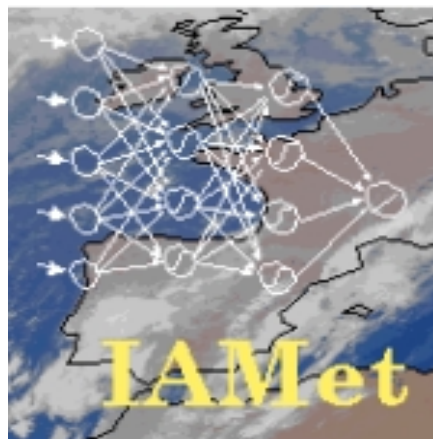




Data Mining en Meteorología

Redes neuronales y Bayesianas



José M. Gutiérrez

Antonio S. Cofiño

Carmen Sordo

Dpto. de Matemática Aplicada y C.C.
Universidad de Cantabria



Rafael Cano

Francisco J. López

Instituto Nacional de Meteorología

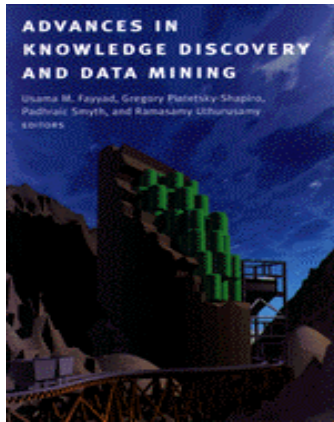


Miguel A. Rodríguez

Instituto de Física de Cantabria,
CSIC/Universidad de Cantabria.



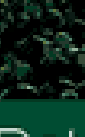
<http://grupos.unican.es/ai/meteo>



V
T
o

[illegible]

The AAAI Press



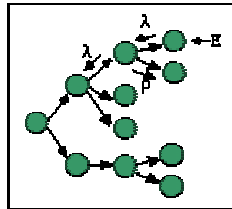
Data Mining Techniques
For Marketing, Sales,
and Customer Support

Michael J. Berry and Michael S. Broderick

[illegible]

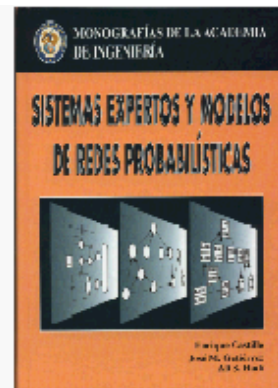
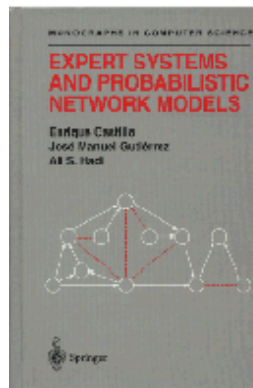
Redes Neuronales y Bayesianas. Libros en Internet

Artificial Intelligence Research Group



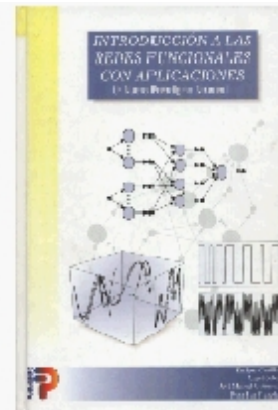
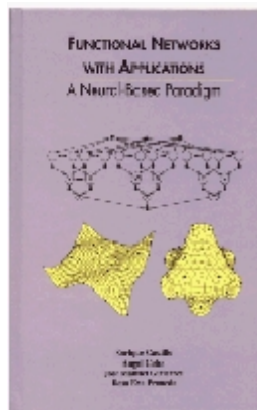
Bayesian and Neural-Based Networks

CORNELL
UNIVERSITY
UC Cantabria
University



Expert Systems and Probabilistic Network Models.
E. Castillo, J.M. Gutiérrez, y A.S. Hadi
Springer-Verlag, New York.

Monografías de la Academia Española de Ingeniería



An Introduction to Functional Networks
E. Castillo, A. Cobo, J.M. Gutiérrez and E. Pruneda
Kluwer Academic Publishers (1999).

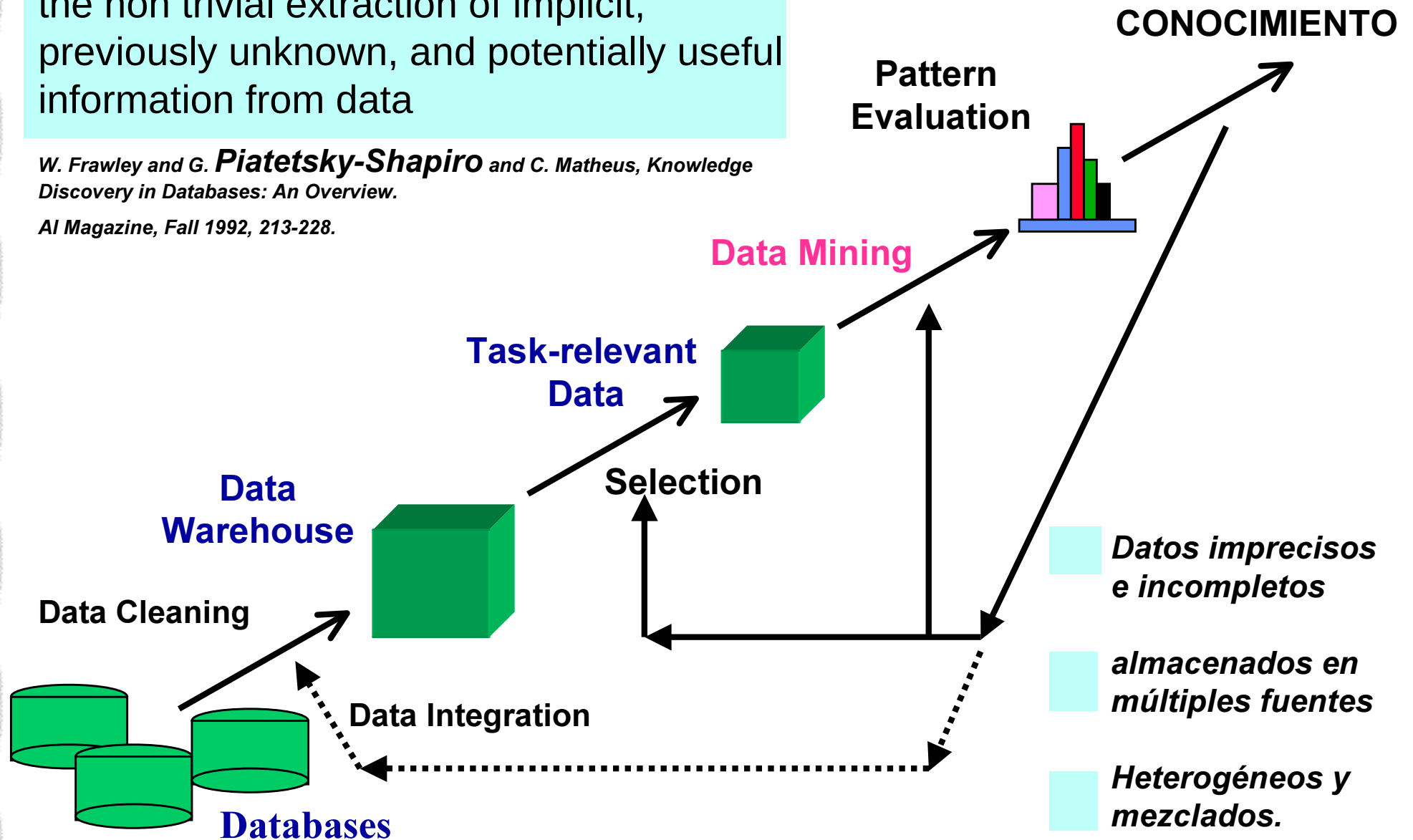
Paraninfo/International Thomson Publishing

¿Qué es Data Mining (minería de datos)?

the non trivial extraction of implicit, previously unknown, and potentially useful information from data

W. Frawley and G. **Piatetsky-Shapiro** and C. Matheus, Knowledge Discovery in Databases: An Overview.

AI Magazine, Fall 1992, 213-228.



MINISTERIO
DE CIENCIA
Y TECNOLOGIA



WP4: TESTBED



CrossGrid site - Microsoft Internet Explorer

Archivo Edición Ver Favoritos » Dirección <http://www.crossgrid.org/applications.php>

Atrás Adelante Detener Actualizar Inicio Vínculos Hotmail gratuito Personalizar vínculos Windows














Project Contribute Forums Topics Intranet Links CG online Users Account

CG nodes

Project overview
Online news
Grid Topics
Contribute
Recomend
Intranet
Calendar
Download
Grid Links
FAQ

Modules

- AcronymOTM
- Calendar
- Downloads
- Encyclopedia
- Forum
- Intranet
- Weather

Partners
Objectives
Applications
New Services
Test beds

Applications

A number of large-scale grid-enabled applications will be considered in the context of CrossGrid project.

As the first application, a **Grid-based prototype system for pre-treatment planning in vascular interventional and surgical procedures** will be developed through real-time interactive simulation of vascular structure and flow. The system consists of a distributed real-time simulation environment, in which a user interacts in Virtual Reality (VR). A 3D model of the arteries, derived using medical imaging techniques, will serve as input to a real-time simulation environment for blood flow calculations. The user will be allowed to change the structure of the arteries, thus mimicking a surgical procedure and the effects of this adaptation will be analysed in real time while the results are presented to the user in a virtual environment.

The second application will be a **Grid-based Support System for flood prevention and protection**. The kernel of this system is numerical flood modelling that requires an appropriate physical model and robust numerical schemes for a good representation of reality, as well as grid distributed supercomputing aspects for realistic simulations when dealing with large problem sizes.

As the third application we will develop **final user applications for physics analysis** running in a distributed mode in a Grid-aware environment using large distributed databases for High-Energy Physics (HEP). The challenging points are: seamless access to large distributed databases in the Grid environment, development of distributed data-mining techniques suited to the HEP field, and integration in a user-friendly interactive way, including specific portal tools. As indicated in the CERN LHC Computing Review Report, interactive data analysis tools is one of the areas where joint efforts amongst one or more experiments, resulting in common projects and products, might lead to cost savings, or decreased risk, or both. The proposed applications explore and develop the use of advanced simulation and interactive data mining techniques.

The fourth application is **weather forecast and air pollution modelling**. An important component will be a data mining system for the analysis of the archive of operational data from a mesoscale model and meteorological reanalysis DBs which include homogeneous meteorological information from a single numerical weather prediction model integrated over decades. The system will also include an atmospheric pollution chemistry module.

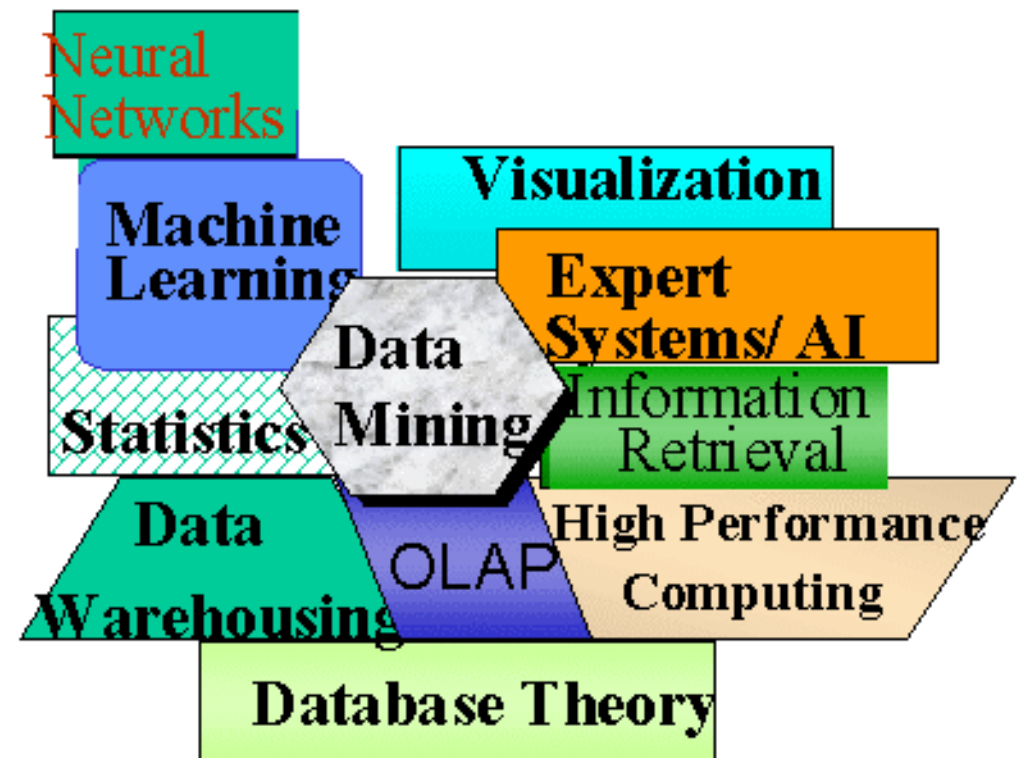
Diferencias entre DBMS y Data Mining

En los sistemas estándar de gestión de bases de datos las consultas se resuelven accediendo a distintos conjuntos de datos almacenados.

*Los sistemas de data mining infieren conocimiento de la base de datos en forma de "patrones". Este conocimiento supone un nuevo conjunto de información (**meta-datos o patrones**) en base a la cual se responden las consultas.*

¿Cómo se generan los Meta-datos?

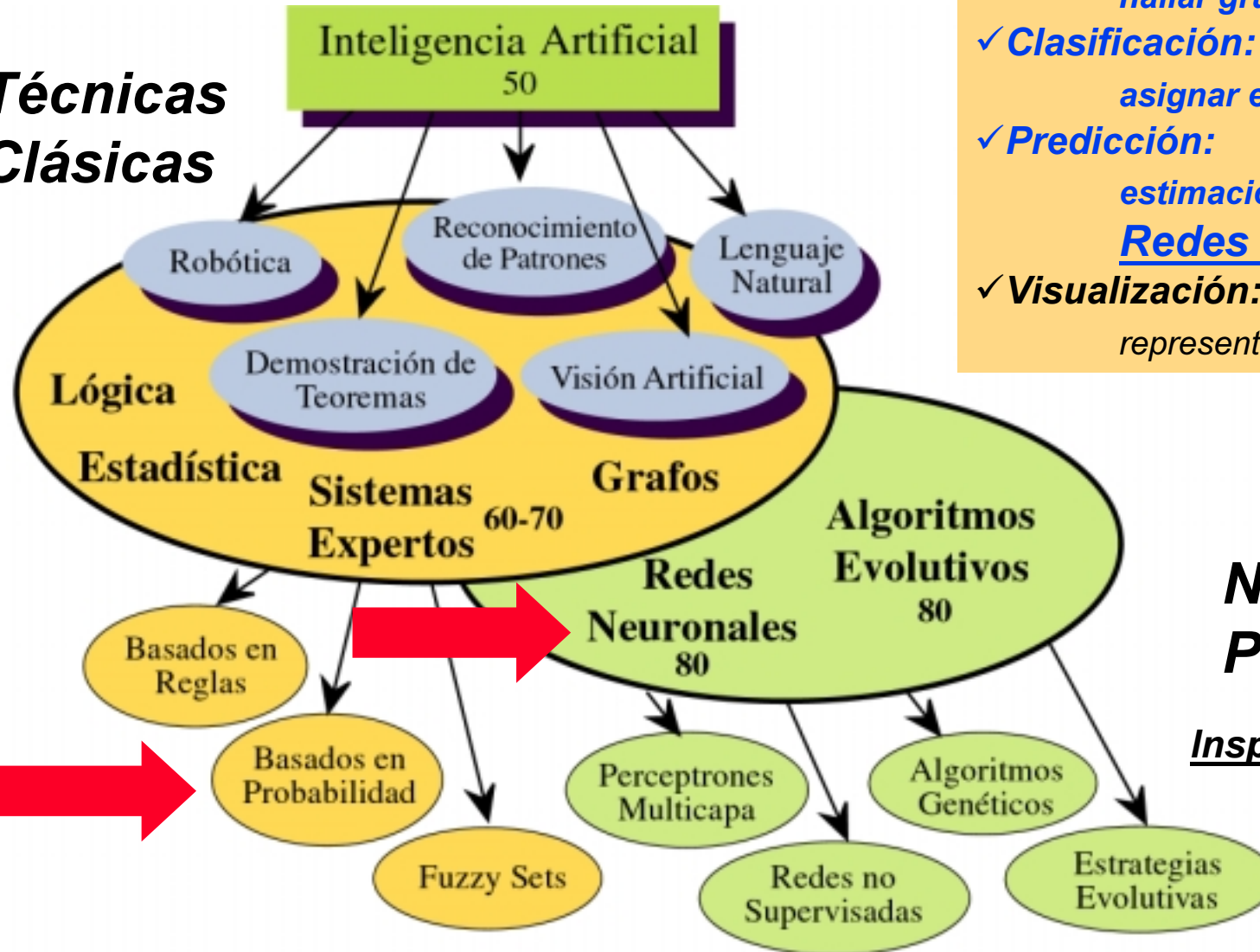
¿Cómo se accede a la Nueva información?



"Machine Learning"

- ✓ **Modelado de Dependencias:**
asociaciones entre variables
Redes Bayesianas
- ✓ **Agrupación:**
hallar grupos de elementos.
- ✓ **Clasificación:**
asignar elementos a clases.
- ✓ **Predicción:**
estimación de valores
Redes Neuronales
- ✓ **Visualización:**
representación gráfica.

Técnicas Clásicas



Nuevos Paradigmas

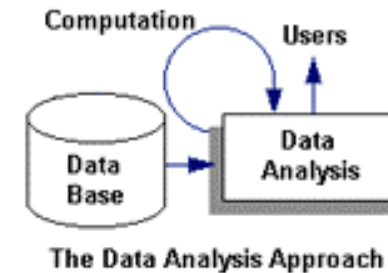
Inspiración Biológica

Tecnologías de Bases de Datos

Paradigma de Acceso a Datos:

El usuario solicita datos y procesa los datos recibidos en busca de "conocimiento".

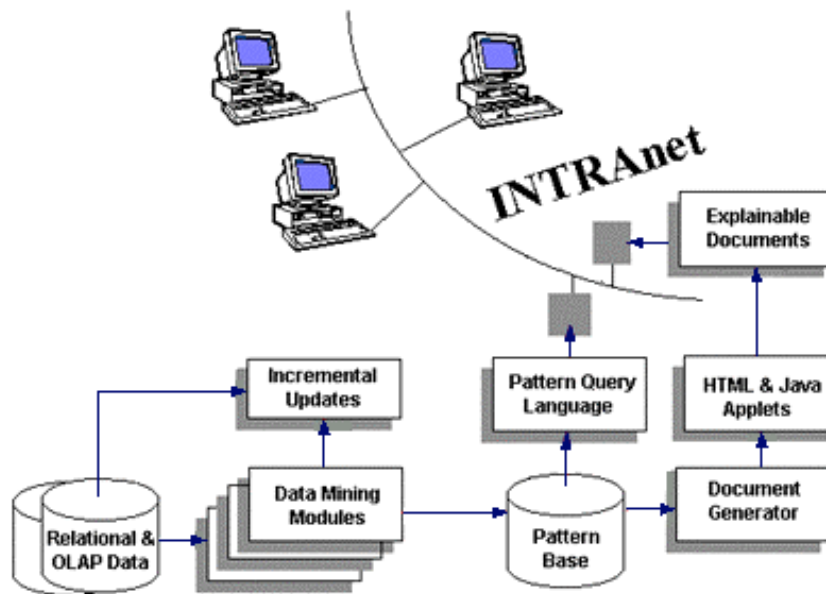
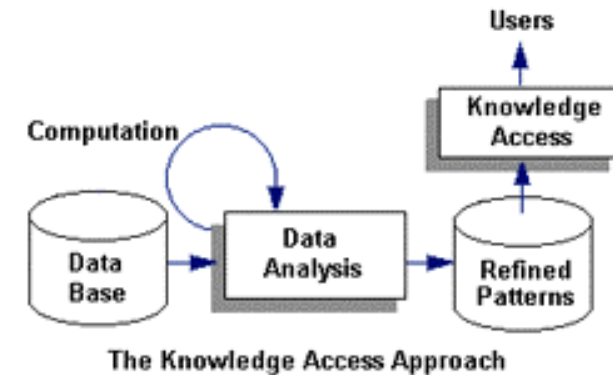
SQL + algoritmos de data mining.



Paradigma de Acceso a Conocimiento:

El sistema genera automáticamente patrones de conocimiento refinados y el usuario accede directamente a los mismos.

PQL = Pattern Query Lenguaje



PQL was designed to access patterns just as **SQL was designed to access data**. PQL resembles SQL, works atop existing SQL engines.

Information Discovery uses a *Pattern Warehouse™* of refined information and PQL works on patterns just as SQL works on a datawarehouse. While SQL relies on the relational algebra, **PQL uses the "pattern algebra"**. PQL allows pattern-based queries just as SQL allows data-based queries. And, PQL uses SQL as part of its operation, i.e. **PQL queries are decomposed into a set of related SQL queries**, the Pattern Warehouse is accessed with these queries and the results are re-combined for display to the user. The user accesses these patterns using a web browser.

IBM DB2 Intelligent Miner



[Home](#) | [Products & services](#) | [Support & downloads](#) | [My account](#)

[→ Select a country](#)

[Products & services](#) > [Software](#) > [Database and Data Management](#) > [Business Intelligence Solutions](#)

DB2 Intelligent Miner for Data

Use the DB2 Intelligent Miner for Data to gain new business insights and to harvest valuable business intelligence from your enterprise data. You can even mine high-volume transaction data generated by point-of-sale, ATM, credit card, call center, or e-commerce activities. With the Intelligent Miner for Data, you are better equipped to make insightful decisions, whether the problem is how to develop more precisely targeted marketing campaigns, reduce customer attrition, or increase revenue generated by Internet shopping.

[Buy now](#)



Features at a glance

Scalable and with support for multiple platforms, the Intelligent Miner for Data provides a single framework for database mining using proven, parallel mining techniques. Business applications for this technology vary widely, and a variety of mining algorithms is provided. Here are some typical examples of how mining algorithms available in the Intelligent Miner for Data are applied:

- You can use clustering for market segmentation, store profiling, and to reveal buying behavior.
- Associations enable you to discover product associations in a market basket analysis, site visit patterns for an e-commerce site, and combinations of financial offerings purchased in different geographical areas.
- Sequential patterns reveal buying patterns in a series of purchases made or multiple Web site visits over time.
- Classification algorithms enable you to profile customers based on a desired outcome, such as propensity, to buy high-end electronics.
- You can use predictive algorithms to score customers by factors such as likelihood of fraud, credit risk, or propensity to buy.

Related links:

- [DB2 Intelligent Miner Scoring](#)
- [DB2 Product Family](#)
- [DB2 Developer Domain](#)
- [Business Intelligence Solutions](#)
- [The data mining group](#)
- [IBM Data Mining Research](#)

Buy other versions

→ [DB2 Intelligent Miner for Data](#)

Operating systems

DB2 Intelligent Miner for Data runs on **AIX, OS/390, OS/400, Solaris, Windows 2000, Windows NT and z/OS.**

More resources

→ [IT Sharpens Data Mining's Focus](#) HCN

→ [DB2 Magazine Online: Operation Data Mining](#) HCN

→ [View IBM Webcast - Enhancing](#)

Reglas de Asociación.

Si Granizo = 4,
entonces Escarcha > 0

Si Tormenta = 1 y Precipitación > 0,
entonces Granizo = 1

860101500000000010
860102100000000010
860103500100000010
860104500000000010
860105101100000010
860106101100000010
860107300100000010
860108500000000010
860109500000001000
860110000001001100
860111001000000000



OBJETIVO. Predicción de Meteoros

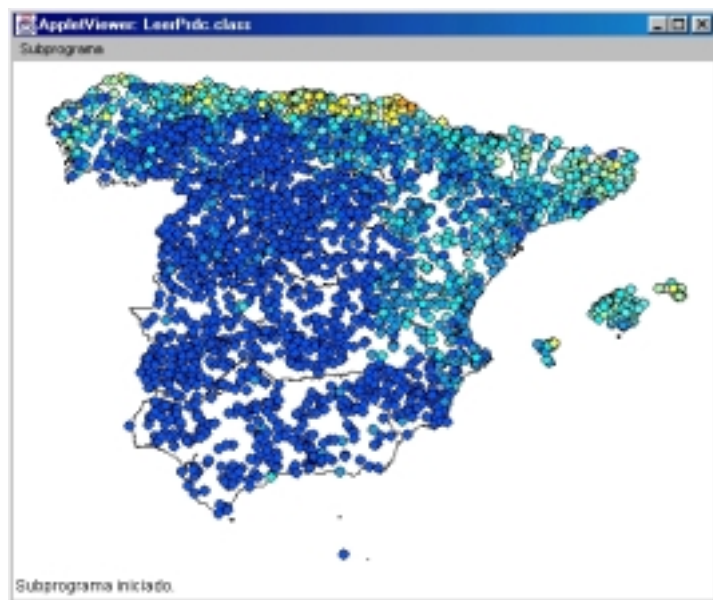
~ 2600
estaciones

FECHA:2001/07/16 VALIDO PARA D+1

C.M.T.: GALICIA

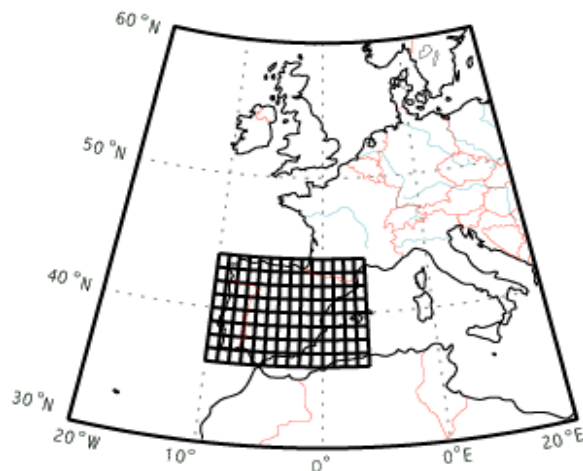
PROVINCIA	INDIC	ESTACION	PROB >= 50 km/h	PROB >= 80 km/h
LA CORUÑA	1354I	EL FERROL 'ARSENAL'	34	0
LA CORUÑA	1387E	LA CORUÑA-AEROPUERTO	13	0
LA CORUÑA	1387	LA CORUÑA-ESTAC. COMPLETA	47	0
LA CORUÑA	1360O	MONTEFARO	-	-
LA CORUÑA	1354	MONTEVENTOSO	83	20
LA CORUÑA	1428	SANTIAGO COMPOSTELA 'LABACOLLA'	44	0

*Probabilidad de Precipitación
de Tormentas
de Racha Máxima
de Nieve
de Niebla
de Granizo
de Escarcha
de Rocío*



Datos Meteorológicos Disponibles

DATOS DE PREDICCIONES (SALIDAS DE MODELOS NUMÉRICOS)



Reanálisis de modelos numéricos:

ERA-15 (1979-1993) Centro Europeo Pred. (ECMWF)

$P = (T(1000 \text{ mb}), \dots, T(500 \text{ mb}); Z(1000 \text{ mb}), \dots, Z(500 \text{ mb}); \dots)$

Para cada día el patrón es **90000 dimensional !!!!**

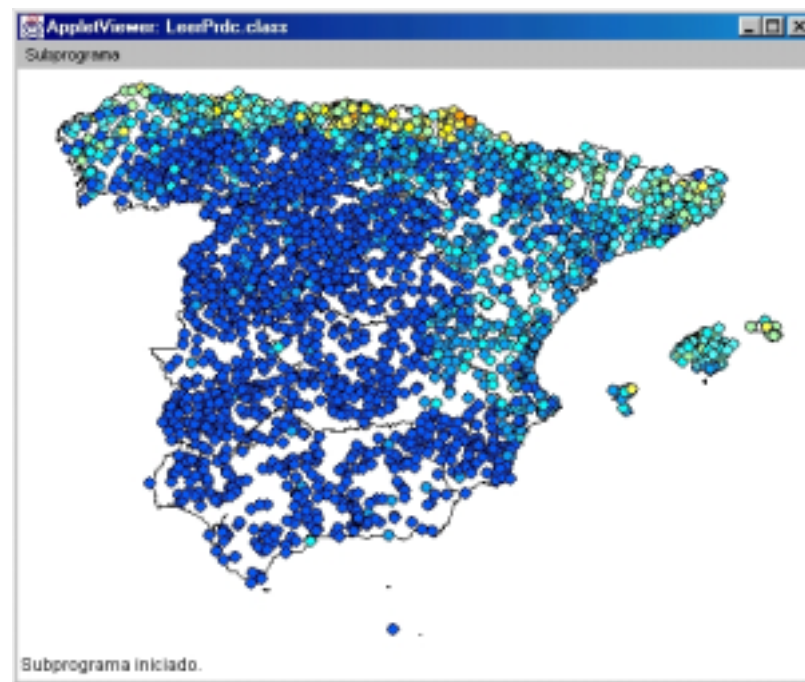
4 GB

DATOS DE OBSERVACIONES

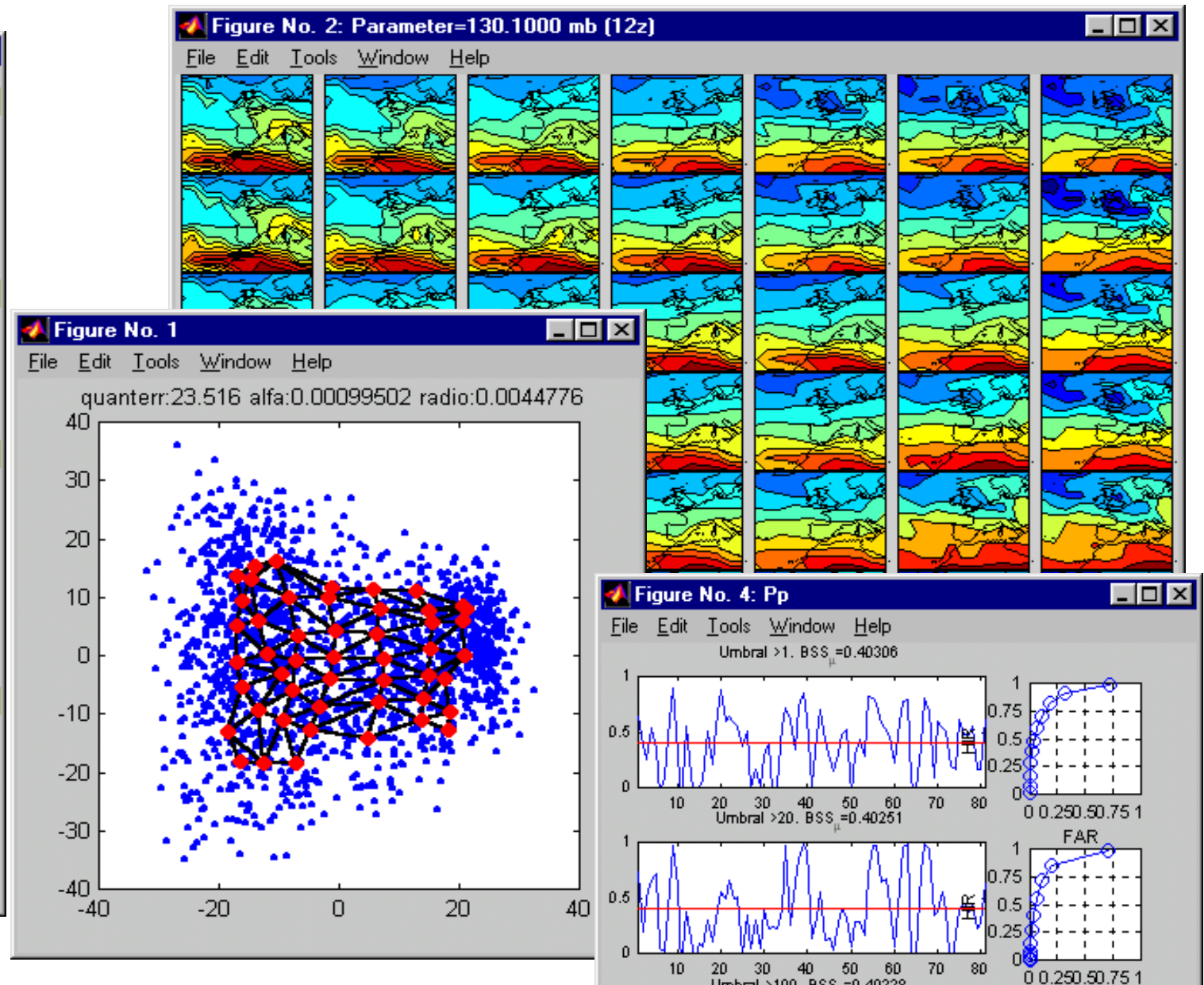
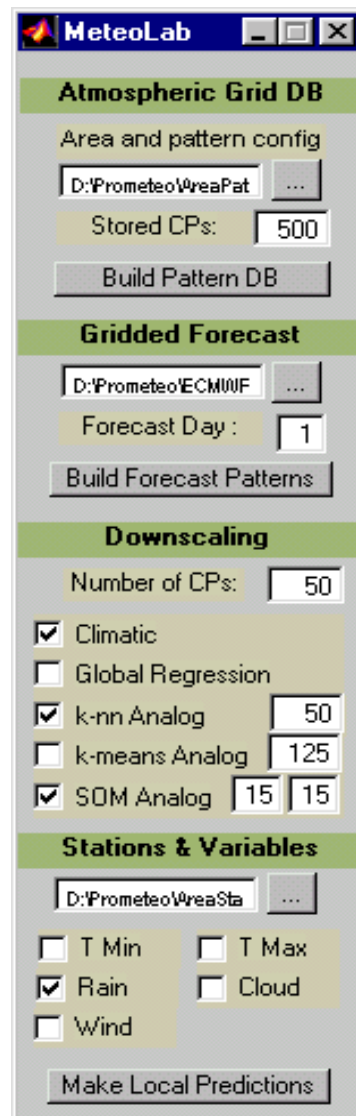
2500 estaciones de la red principal y secundaria del INM

Precipitación
Velocidad de viento
Horas de sol, ...

1 GB



The Toolbox “MeteoLab” Data Mining in Meteorology

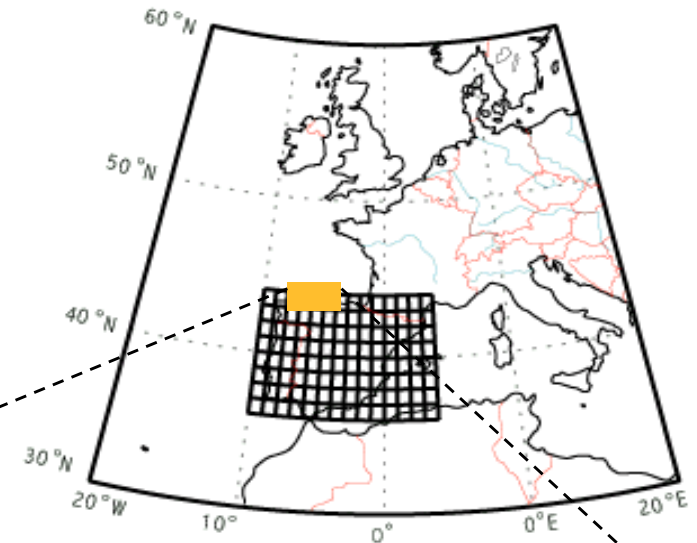


Modelos Numéricos de Predicción. **Limitaciones**

Los modelos de circulación atmosférica son las herramientas básicas para la predicción operativa.



$$\begin{cases} \frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} - \omega \frac{\partial u}{\partial p} - \frac{\partial \Phi}{\partial x} + f v + F_x \\ \frac{\partial v}{\partial t} = -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - \omega \frac{\partial v}{\partial p} - \frac{\partial \Phi}{\partial y} - f u + F_y \\ \frac{\partial \Phi}{\partial p} = -\frac{RT}{p} \\ \frac{\partial T}{\partial t} = -u \frac{\partial T}{\partial x} - v \frac{\partial T}{\partial y} + \omega \left(\frac{\kappa T}{p} - \frac{\partial T}{\partial p} \right) + \frac{\dot{H}}{c_p} \\ \frac{\partial \omega}{\partial p} = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \end{cases}$$

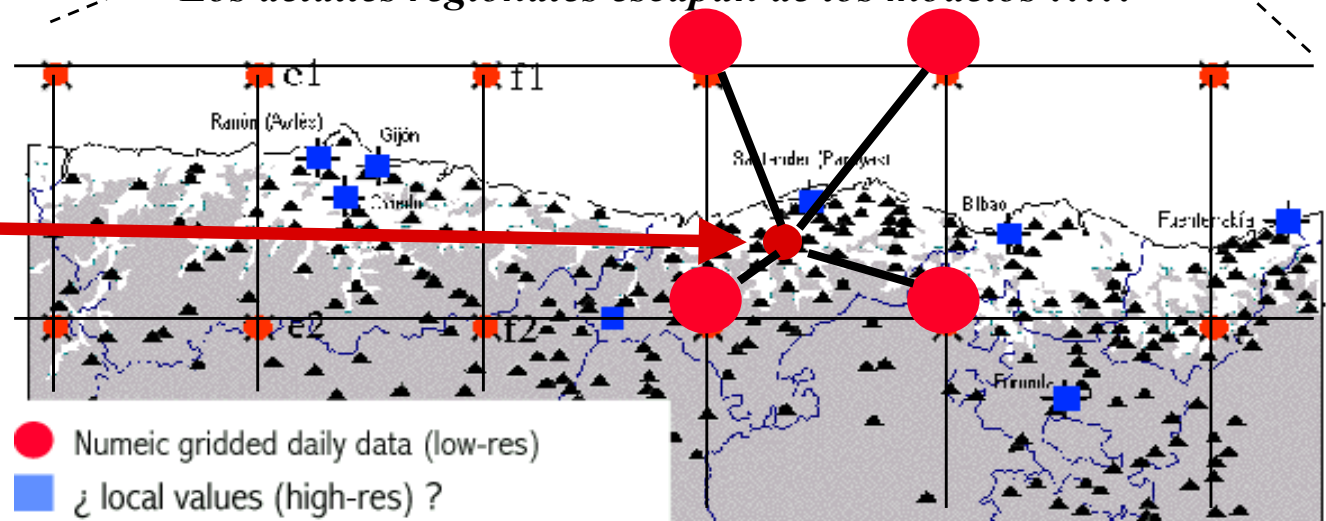


Los detalles regionales escapan de los modelos !!!!!

Cómo se puede obtener una predicción en un punto de interés (una estación)?

✓ Interpolando las predicciones de la rejilla.

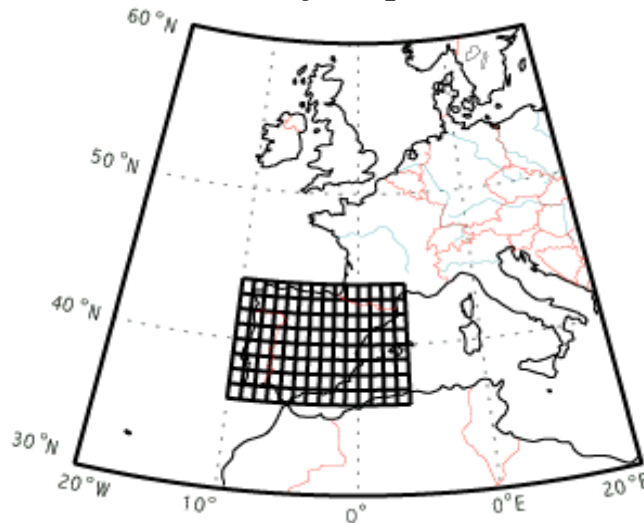
✓ Incorporando la Estadística de registros históricos: $x_1, x_2, \dots$



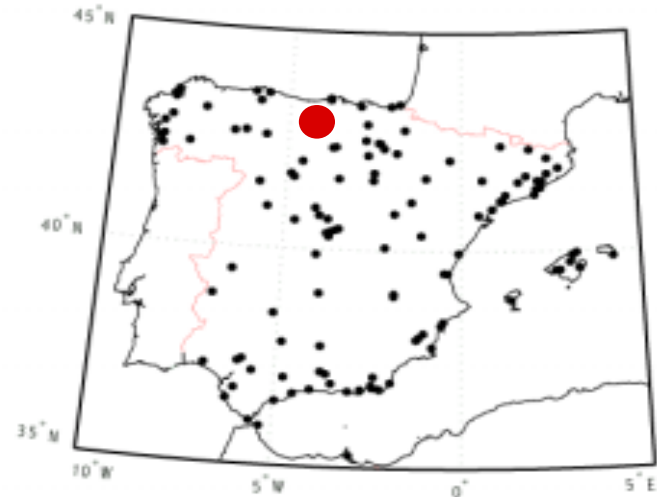
DOWNSCALING ESTADÍSTICO

Técnicas Clásicas. Regresión Lineal

Patrones de circulación atmosférica
en una rejilla para el día n

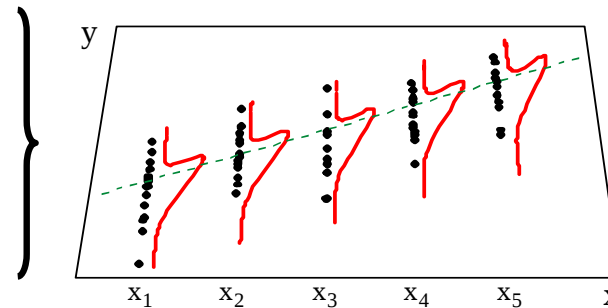


Observaciones en 122 estaciones
para el día n



$(T(1000 \text{ mb}), \dots, T(500 \text{ mb});$
 $Z(1000 \text{ mb}), \dots, Z(500 \text{ mb});$
 $\dots;)$

$H(1000 \text{ mb}), \dots, H(500 \text{ mb})) = \mathbf{X}_n$



$\mathbf{Y}_n$

Precipitación
Racha Máxima
Horas de Sol ,
 $T_{\max}$ y $T_{\min}$...

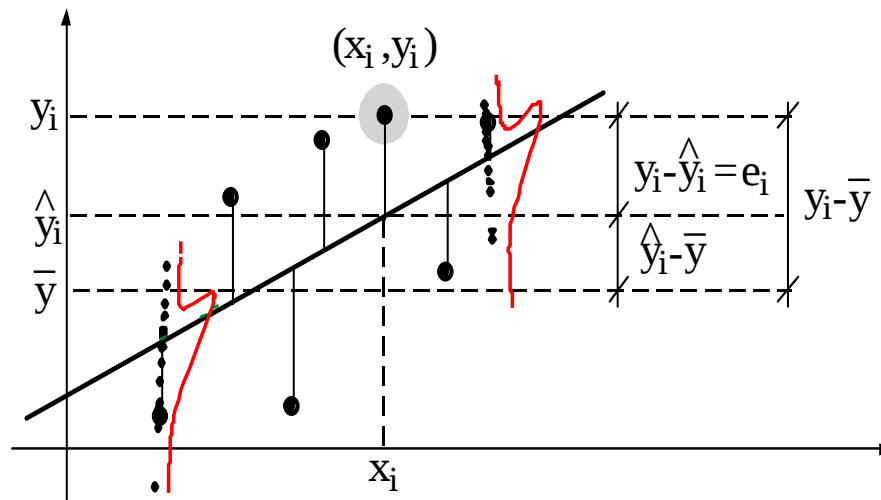
Un modelo lineal $\mathbf{Y}_n = \mathbf{W}^T \mathbf{X}_n$ se obtiene
estimando los coeficientes $\mathbf{W} = (w_1, \dots, w_k)$ a partir
de los registros históricos en días $i=1, \dots, N$
donde $\mathbf{X}_n$ e $\mathbf{Y}_i$ están disponibles.

Dado un patrón previsto $\mathbf{X}_{n+1}$
se obtiene una estimación:

$$\hat{\mathbf{Y}}_{n+1} = \mathbf{W}^T \mathbf{X}_{n+1}$$

Técnicas Clásicas de Predicción

- **Regresión Lineal:** $Y = a + b X$
 - Los parámetros **a** y **b** definen la línea recta que más se ajusta al conjunto de datos disponible; estos parámetros se estiman **directamente** de los datos utilizando el criterio de mínimos cuadrados con los valores conocidos de $Y_1, Y_2, \dots, X_1, X_2, \dots$



- **Regresión Múltiple:** una extensión de la regresión lineal para tratar vectores:

$$Y = b_0 + b_1 X_1 + b_2 X_2.$$

Numerosas funciones no lineales pueden transformarse en este tipo de ecuación.

Nonlinearities in the relationship $Y_n = f(X_n)$

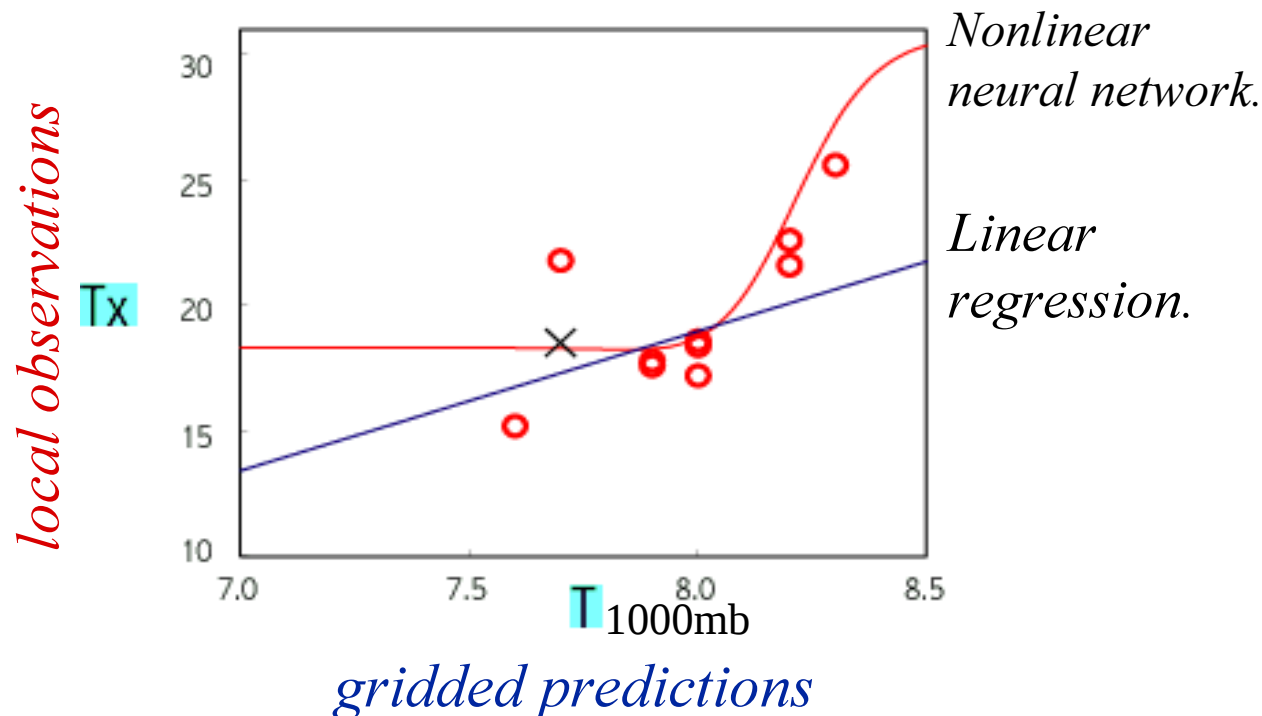
Let us consider a simple case:

$$(T(1000 \text{ mb}), \dots, T(500 \text{ mb}); Z(1000 \text{ mb}), \dots, Z(500 \text{ mb}); \dots; H(1000 \text{ mb}), \dots, H(500 \text{ mb})) = X_n$$

$TMax_n = w T1000_n$ is the simplest case for the relationship

$$TMax_n = f(T1000mb_n),$$

if any, between *local observations* and *gridded predictions*.



This example shows how a simple linear model is not appropriate in some cases:

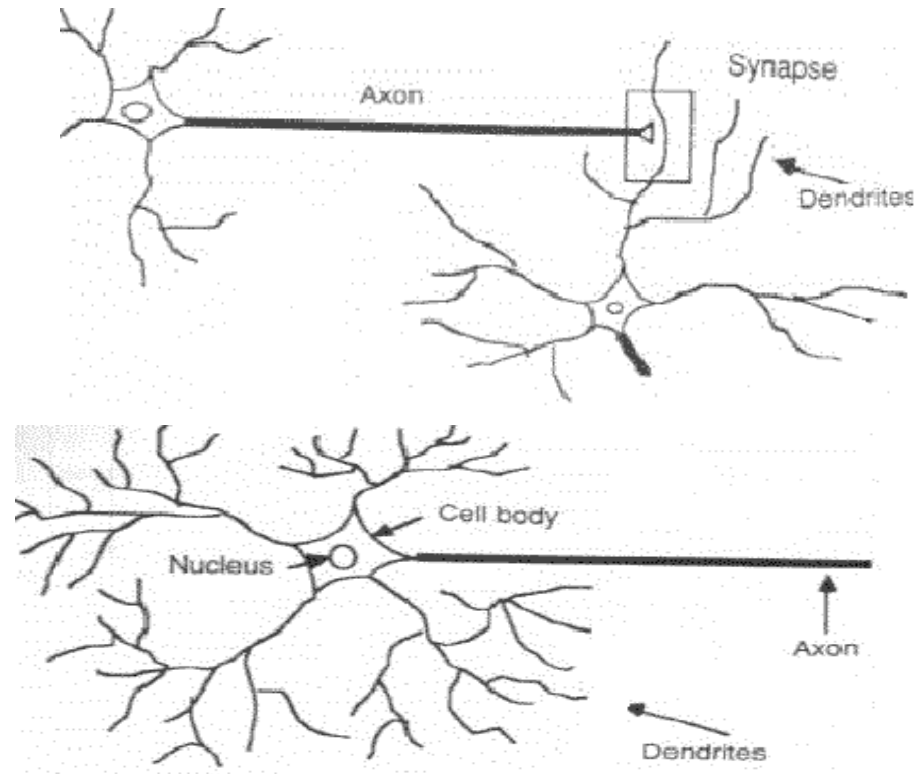
~~$$TMax_n = a T_n$$~~

$$TMax_n = f(T_n)$$

Redes Neuronales. Inspiración en la Neurofisiología

El cerebro humano está formado por un gran número de neuronas (más de 100000 millones) conectadas entre sí de forma masivamente paralela

La actividad de cada neurona se basa en descargas electroquímicas, a partir de los estímulos recibidos por neuronas vecinas a las que esté conectada.



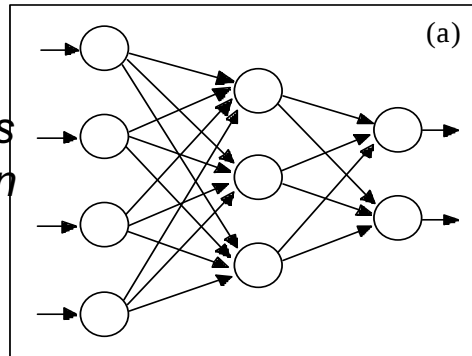
Neural Network Study (1988, AFCEA International Press, p. 60):
... a neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes.

Redes Neuronales Artificiales

Siguiendo esta metodología se desarrollaron diversas topologías de red, para organizar la conexión de los procesadores.

Redes Multicapa.

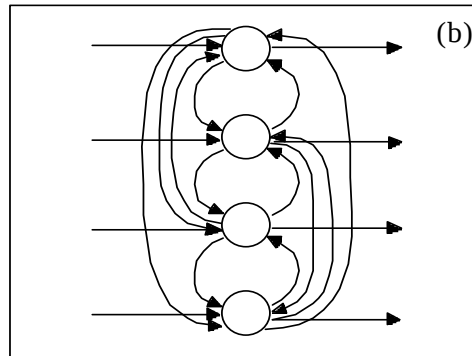
Con varias capas conectadas en serie con procesadores en paralelo.



Reconocimiento de patrones
OCR
Proc. lenguaje natural
Interpolación y extrapolación
Ajuste de funciones

Redes de Hopfield.

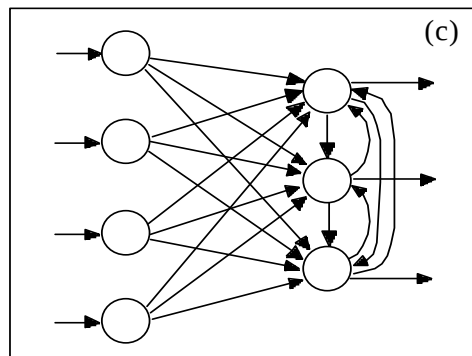
Una sólo capa totalmente interconectada y paralela.



Reconocimiento de patrones
OCR
Almacenamiento con recuperación robusta al ruido

Redes competitivas.

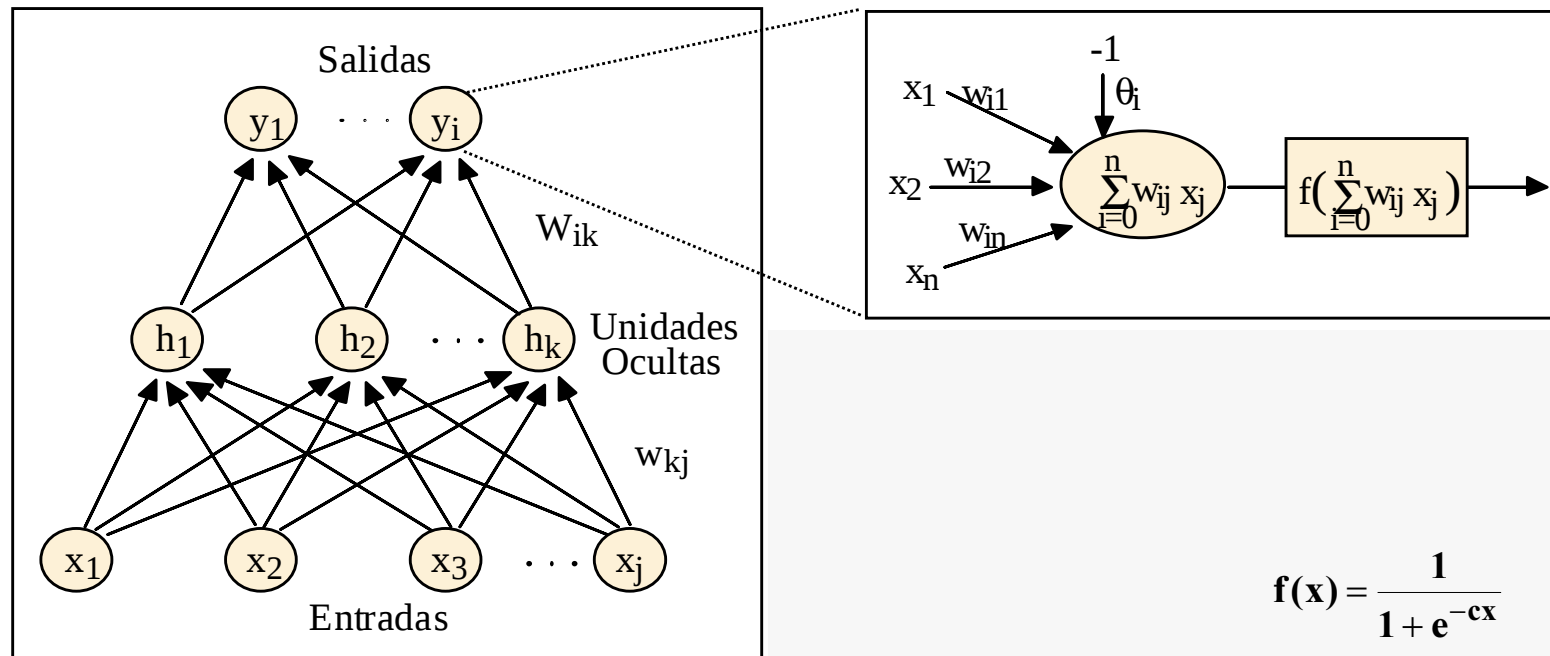
Redes multicapa con conexiones laterales en la última capa.



Clasificación autoorganizada
Reconstrucción topológica
Extracción de variables características

Redes Multicapa (Perceptrones)

Las **redes neuronales** permiten obtener una aproximación funcional de un modelo dado en base a un conjunto de datos y a operadores sigmoidales.



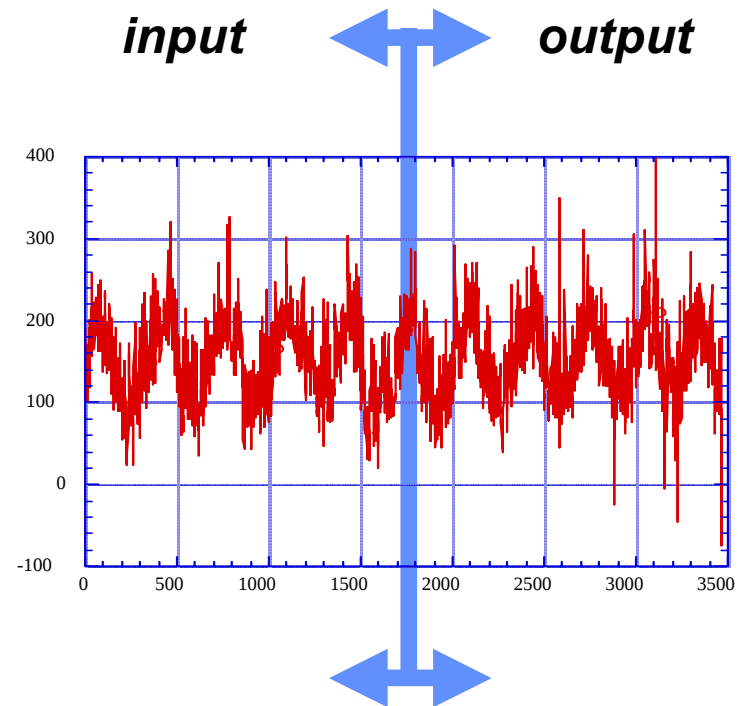
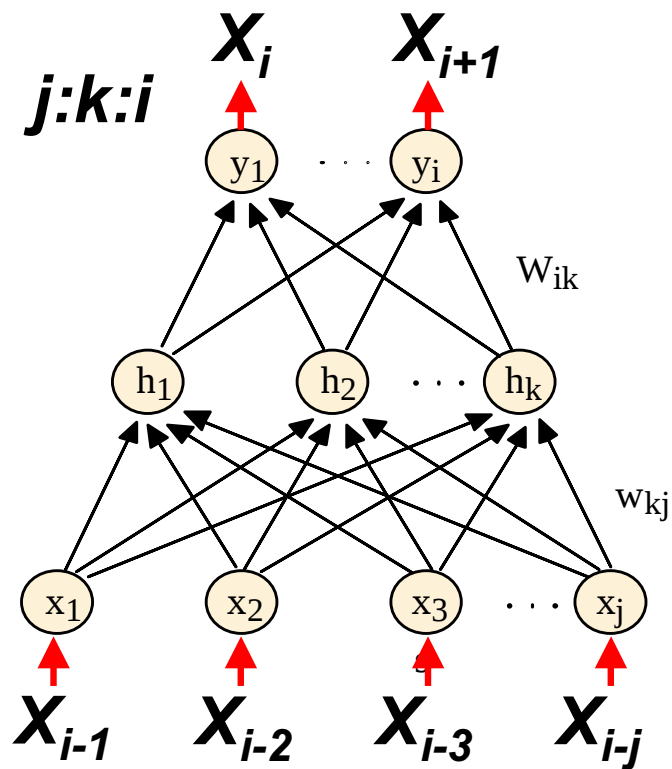
Cada procesador realiza una actividad muy simple: **combinación lineal** de las actividades recibidas por la neurona **activación * peso**.

A continuación, se calcula su actividad aplicando una **función de activación** no lineal. La salida final es un **ajuste no lineal** de las entradas.

Estimar de los datos

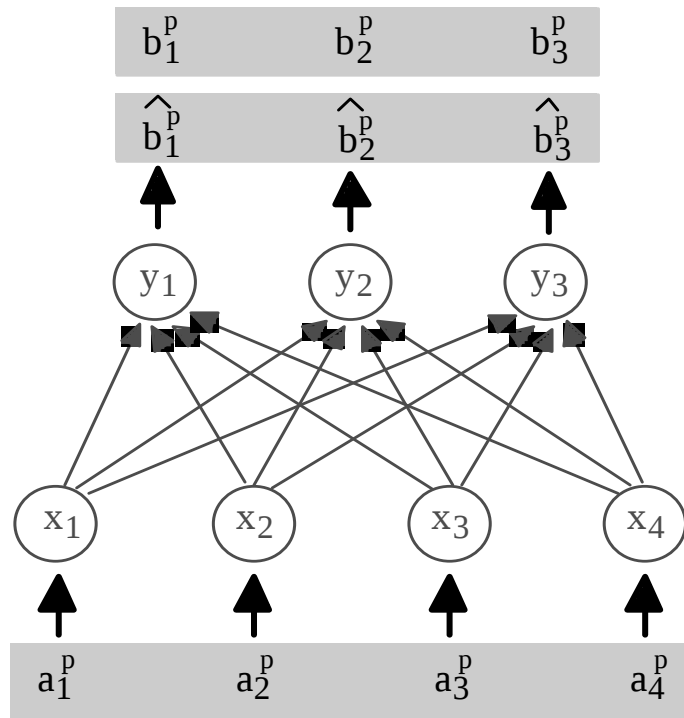
$$y_i = f \left(\sum_k W_{ik} f \left(\sum_j w_{kj} x_j \right) \right)$$

Modelización de Series Temporales



El Aprendizaje

¿Cómo se pueden hallar los pesos para que una red obtenga las salidas correctas a partir de las entradas de un conjunto de entradas-salidas dado?



Inicialmente se eligen valores aleatorios para los pesos.

Aprendizaje Hebbiano: Se modifican los pesos acorde a la correlación entre las unidades.

Se eligen los patrones (a^p, b^p) de uno en uno y se modifican los pesos w_{ij} de los procesadores con salidas incorrectas:

$$\Delta \omega_{ij} = \eta (b_i^p - \hat{b}_i^p) a_j^p$$

Descenso de gradiente: Se modifican los pesos acorde a la dirección del gradiente de una función de error.

$$\Delta \omega_{ij} = -\eta \frac{\partial E(\omega_{ij})}{\partial \omega_{ij}} = \eta \sum_p (b_i^p - \hat{b}_i^p) a_j^p$$

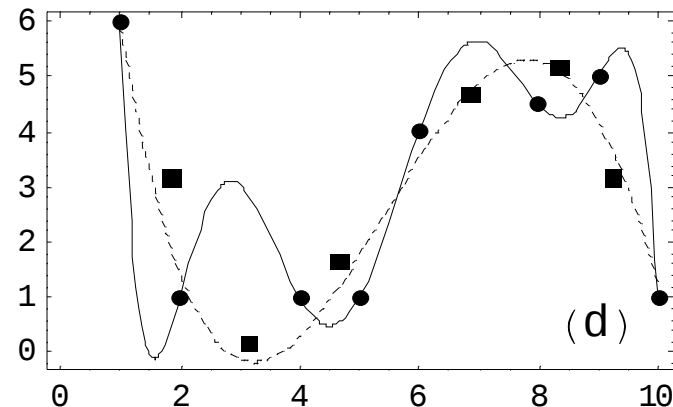
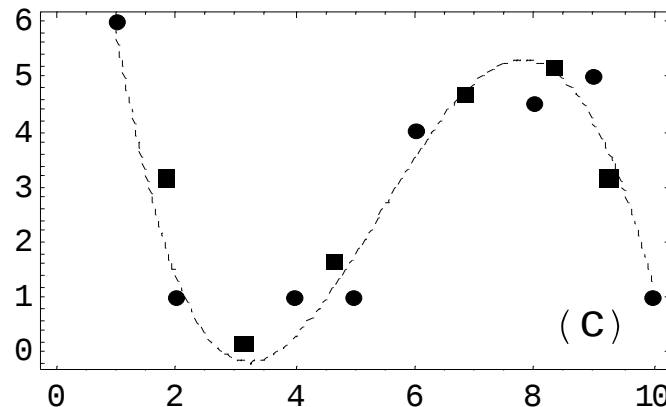
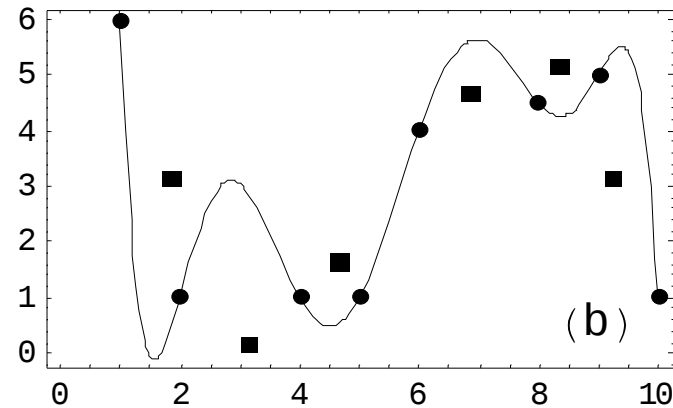
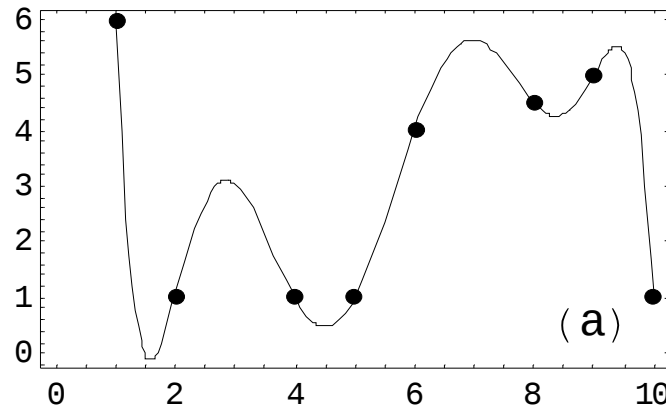
$$E(\omega_{ij}) = \sum_{i,p} (b_i^p - \hat{b}_i^p)^2$$

Regularización

$$E(\omega_{ij}) = \sum_{i,p} (b_i^p - \hat{b}_i^p)^2 + \lambda \sum_{i,j} w_{ij}^2$$

Sobreajuste (balance incorrecto de parámetros y datos)

En estadística es bien conocido que cuando se utiliza un modelo con **muchos parámetros** para ajustar un conjunto de datos procedente de proceso con **pocos grados de libertad**, el modelo obtenido puede no descubrir las tendencias reales del proceso original, aunque pueda presentar un error pequeño.

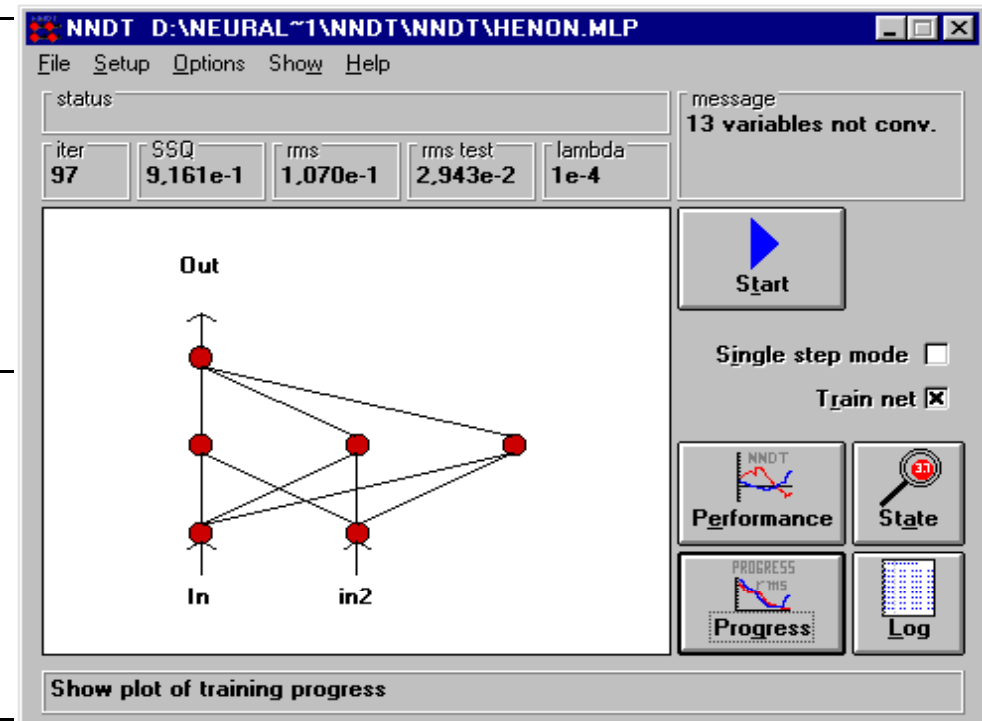
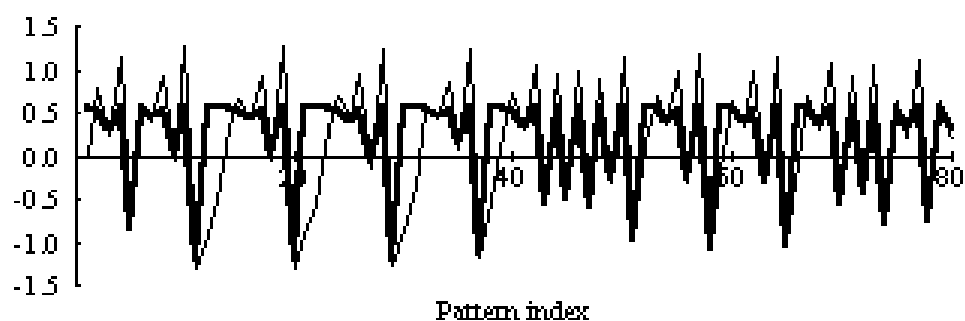
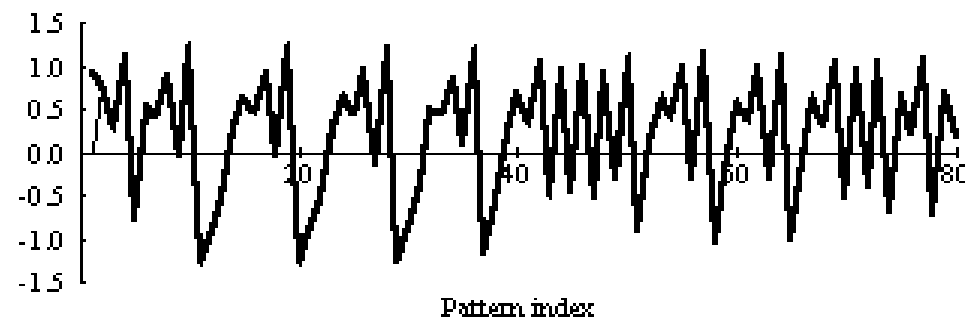
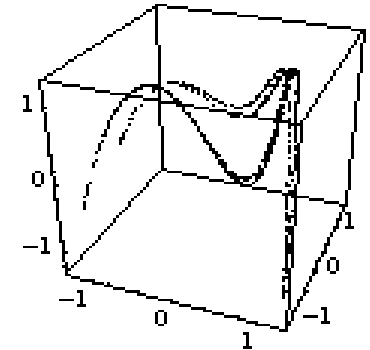
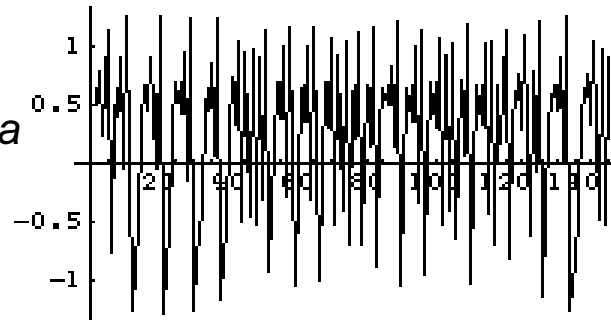


Ejemplo: Predicción No Lineal

En ocasiones es necesario tratar con series temporales caóticas, que tienen apariencia estocástica y difícilmente precicable.

Un ejemplo es la serie de Henon:

$$x_{n+1} = 1 - 1.4 x_n^2 + 0.3 x_{n-1}$$



Nonlinearities in the relationship $Y_n = f(X_n)$

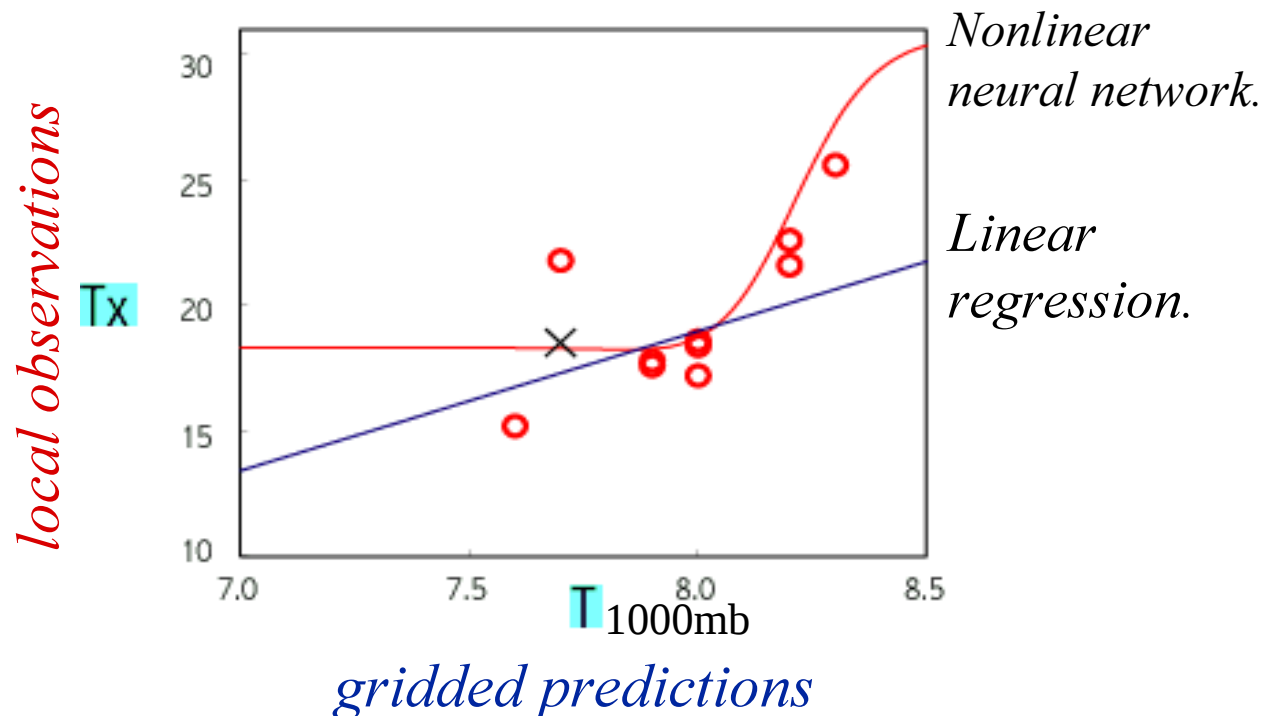
Let us consider a simple case:

$$(T(1000 \text{ mb}), \dots, T(500 \text{ mb}); Z(1000 \text{ mb}), \dots, Z(500 \text{ mb}); \dots; H(1000 \text{ mb}), \dots, H(500 \text{ mb})) = X_n$$

$TMax_n = w T1000_n$ is the simplest case for the relationship

$$TMax_n = f(T1000mb_n),$$

if any, between *local observations* and *gridded predictions*.

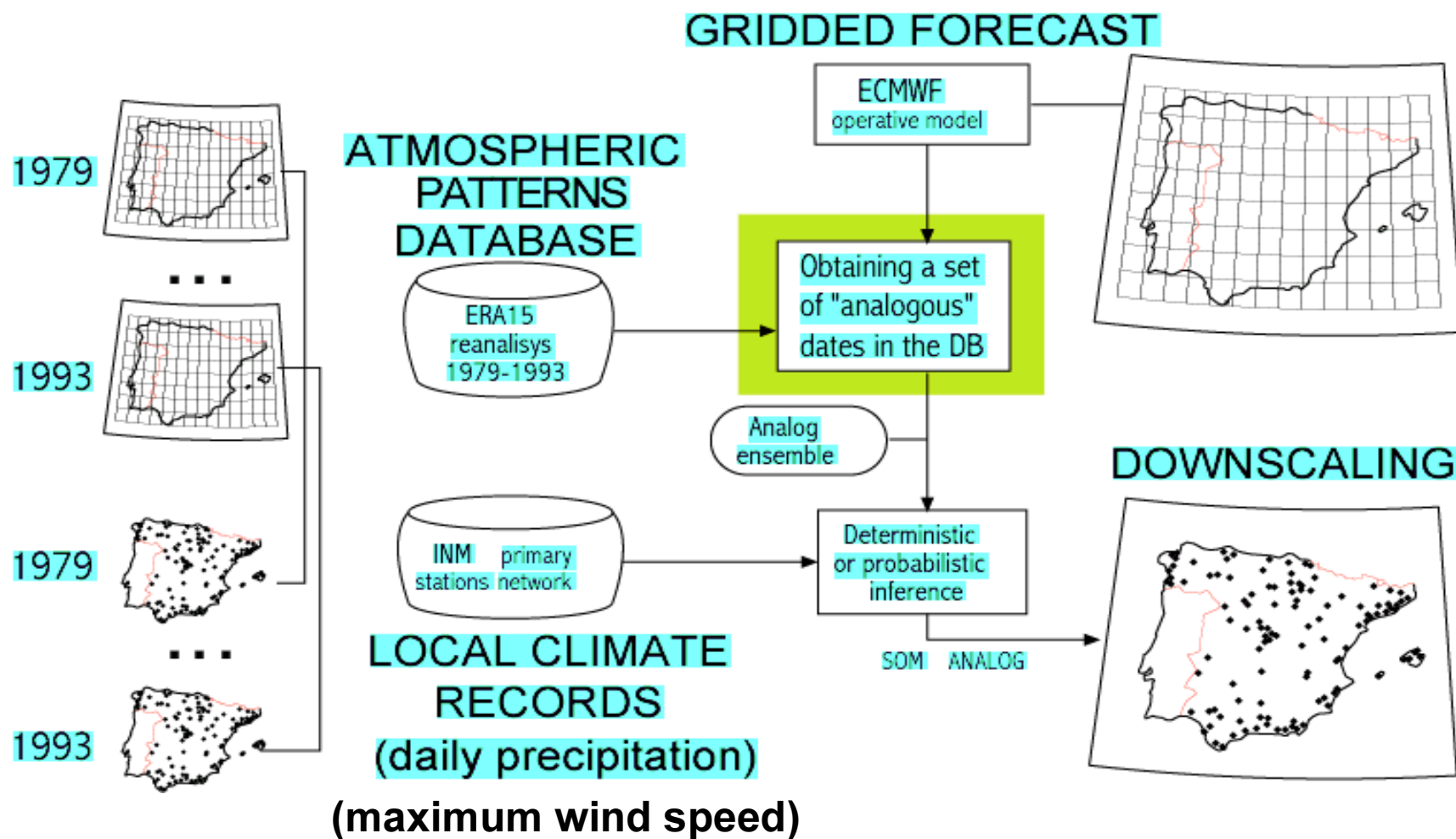


This example shows how a simple linear model is not appropriate in some cases:

~~$$TMax_n = a T_n$$~~

$$TMax_n = f(T_n)$$

Downscaling Estadístico (*con Análogos*)



Cuántos
análogos han de
tomarse?

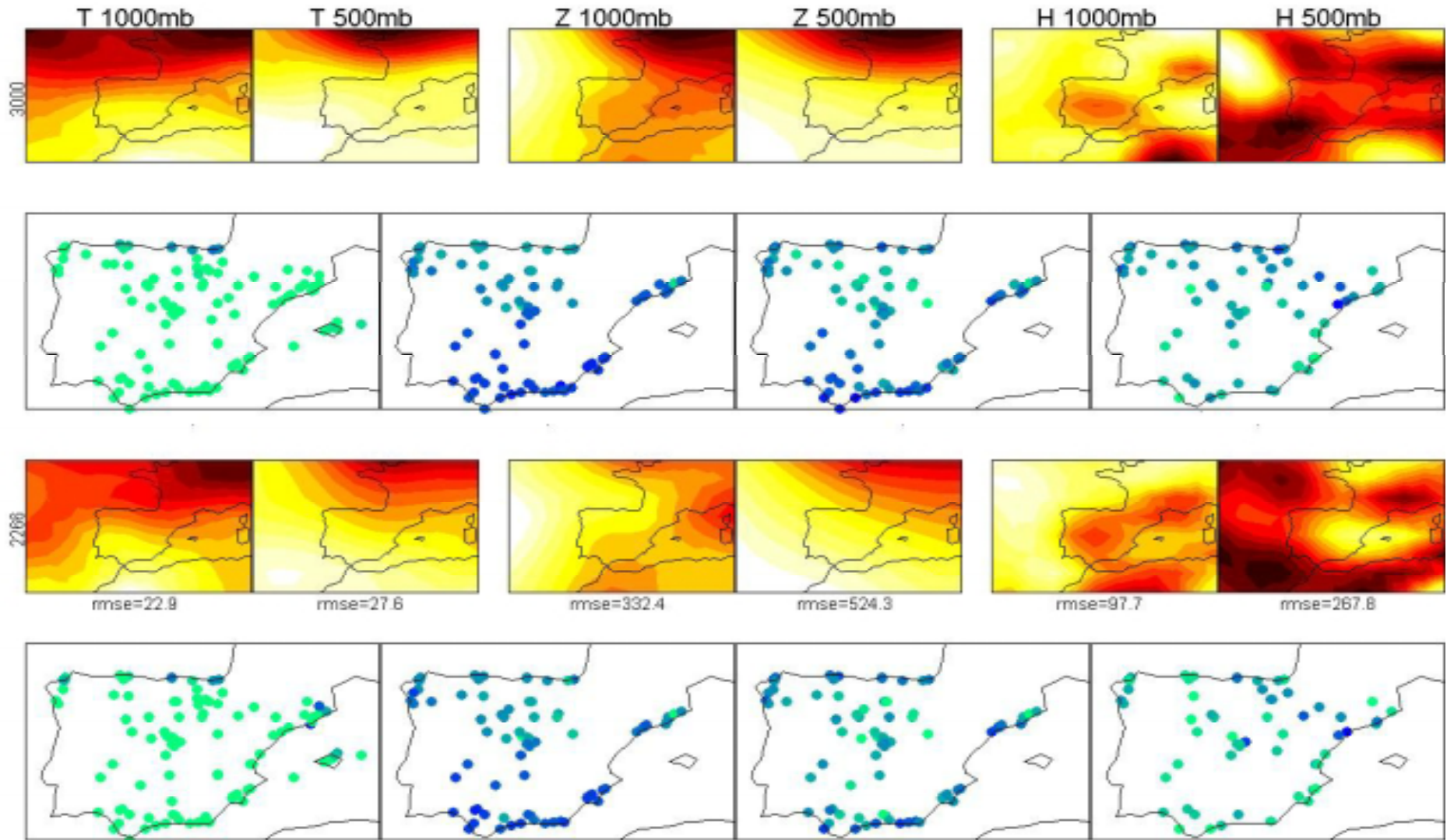


Al incrementar el número
Con un número reducido



Climatología
Sensible a error

El Modelo de Análogos. Ejemplo



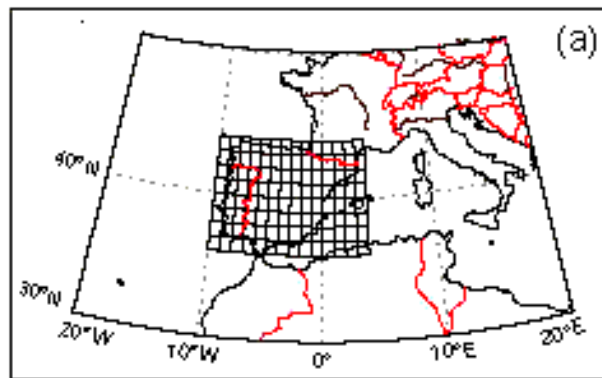
Dia Problema	Numero CP's	Analogos
3000 + -	20 + -	01-2266 ▲ ▼ 02-2267

Patrones de Circulación Atmosférica. El Reanálisis ERA-15

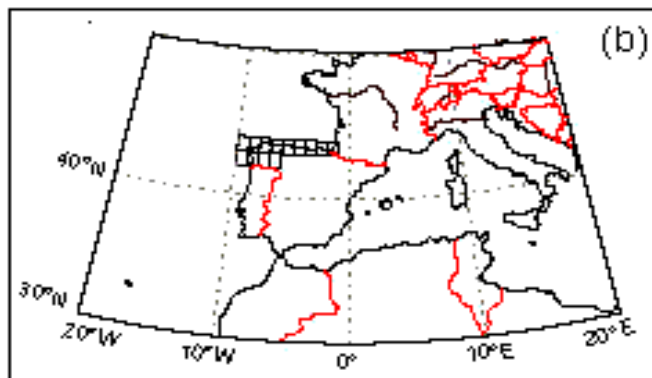
Se utilizan los patrones atmosféricos previstos a las 1200 UTC en el reanálisis ERA-15 (1979-1993) .

$$P = (T(1000 \text{ mb}), \dots, T(500 \text{ mb}); Z(1000 \text{ mb}), \dots, Z(500 \text{ mb}); \dots; H(1000 \text{ mb}), \dots, H(500 \text{ mb}))$$

Modelo 1

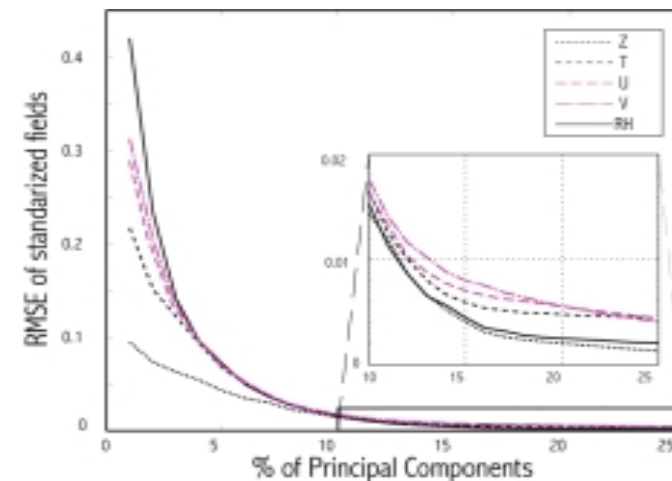


Modelo 2



P es **6000 dimensional !!!!**

Utilizando PCs **500 – 600 dim.**



$$P = (P(06), P(12), P(18), P(24), P(30))$$

Técnicas de Agrupamiento/Clasificación

Agrupamiento Local

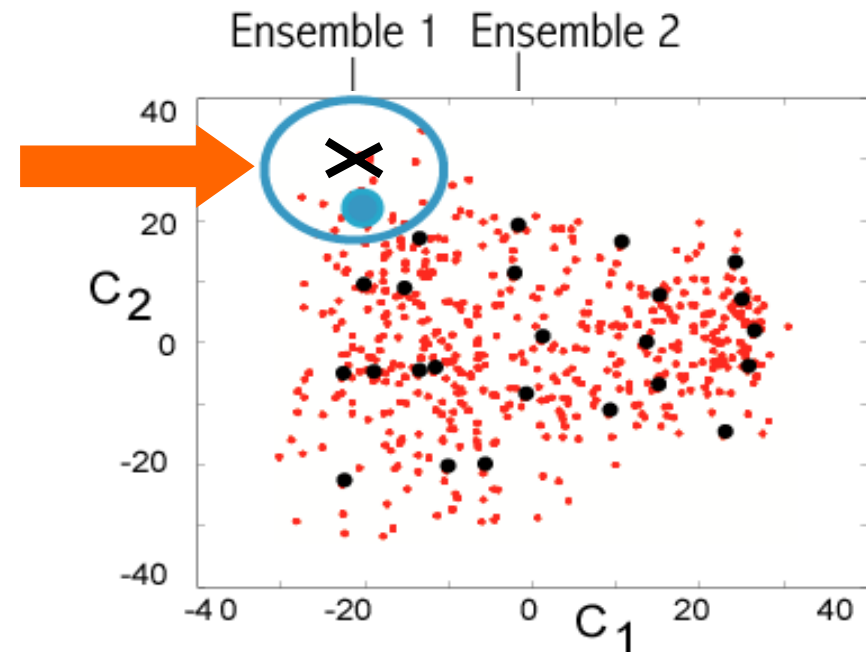
Técnica de los k -vecinos

A partir del patrón predicho se busca los k días más cercanos en el reanálisis como conjunto de análogos.

Agrupamiento Global

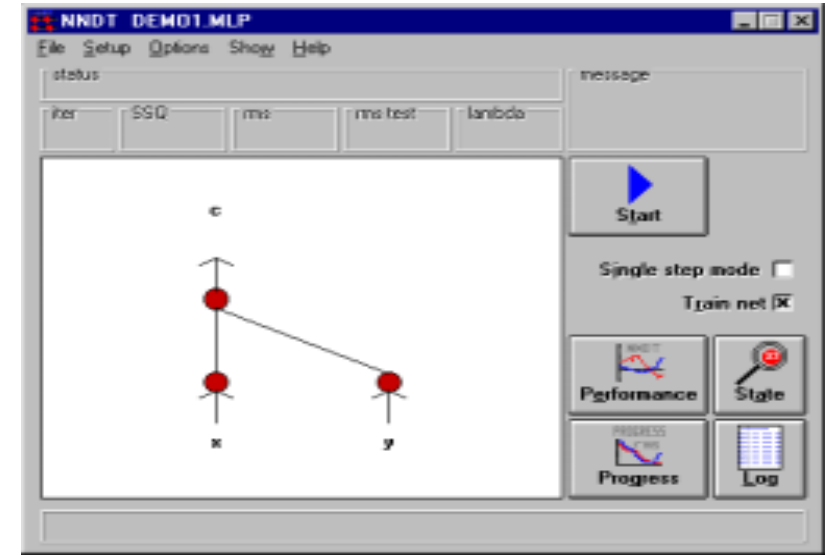
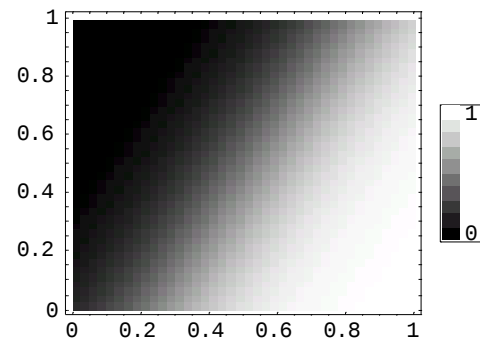
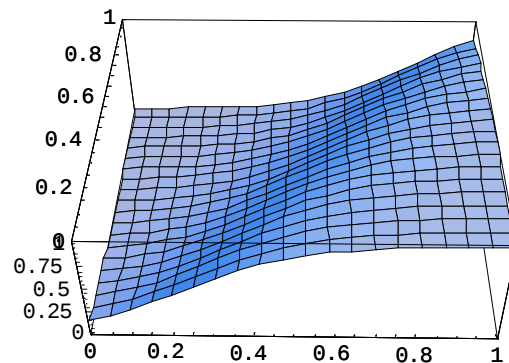
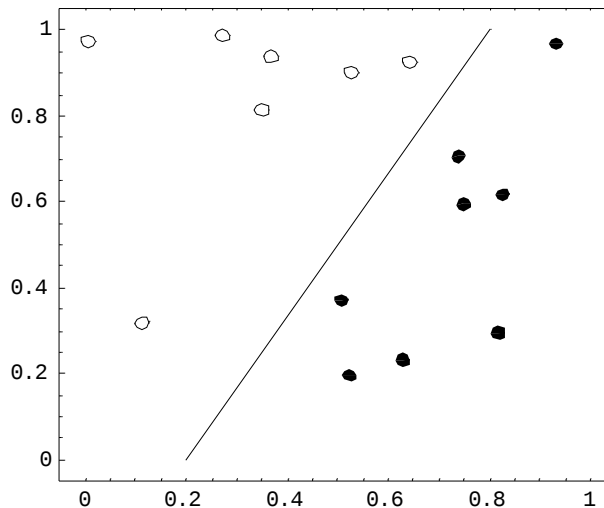
k -medias y redes auto-organizativas

- Proyección de ERA-15 usando dos CPs
- Centros de grupos (prototipos)
- ✕ Patrón atmosférico predicho
Condición inicial para el downscaling

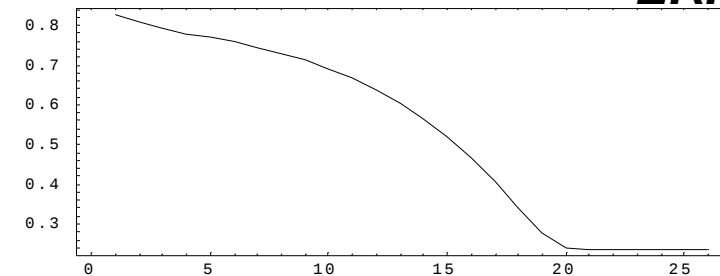


Clasificación Supervisada con Redes Neuronales

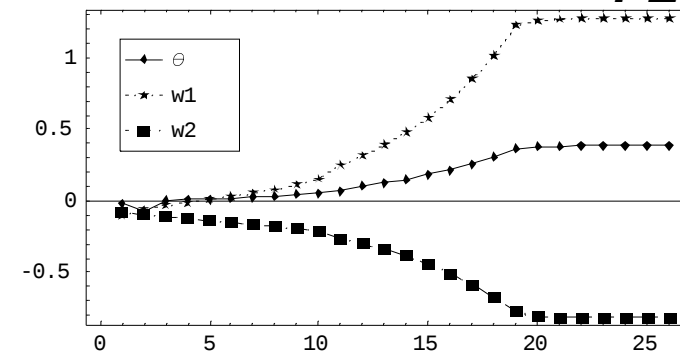
Dada una nube de puntos en el plano correspondientes a dos clases distintas, se quiere obtener un criterio de clasificación automático, que extrapole la información de estos puntos.



ERROR

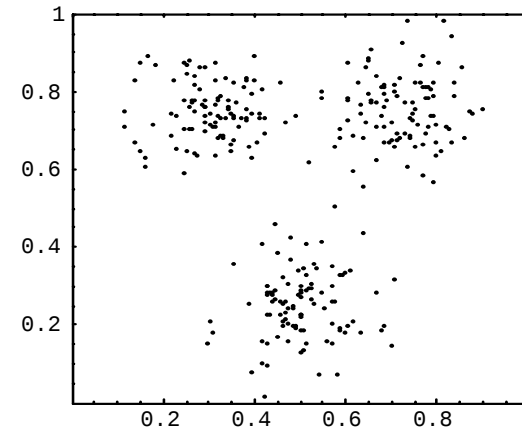


PESOS

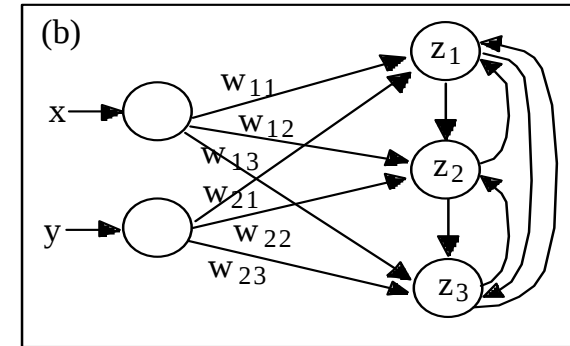
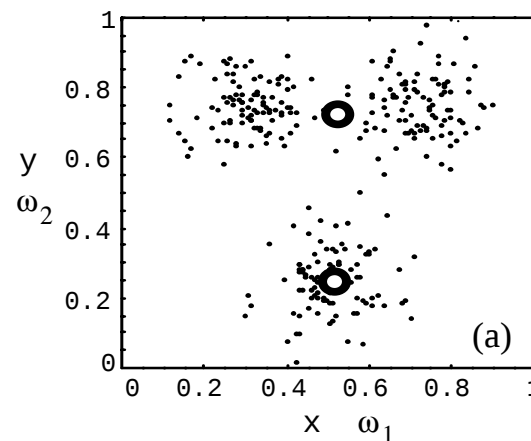
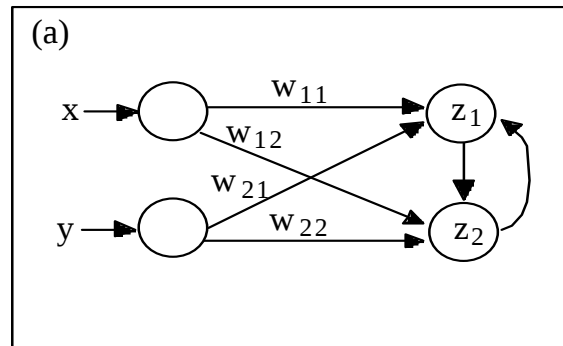


Clasificación **no Supervisada** con Redes Neuronales

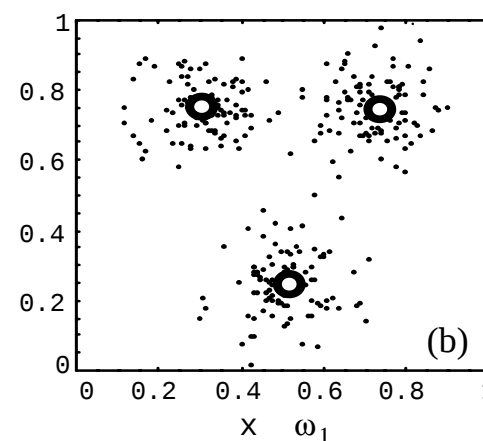
Las redes competitivas permiten obtener criterios de clasificación no supervisados (el usuario no especifica el número de clases).



Red con **dos** procesadores de salida (dos clases)



Red con **tres** procesadores de salida (tres clases)

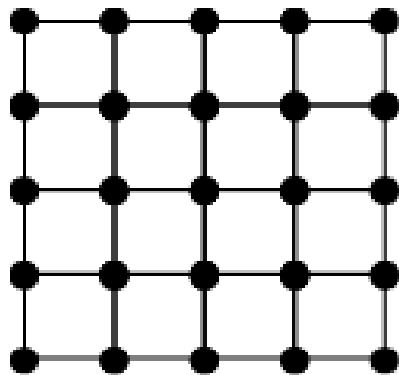


Self-Organizing Maps. Preserving Data Topology

For medium and seasonal weather forecast, we need a measure of the dispersion of the forecast ensemble members.

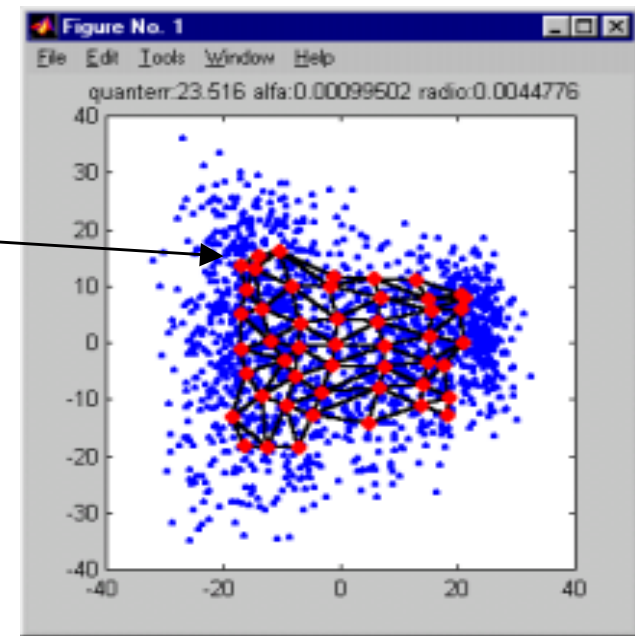
Therefore, we also need a measure of Cluster Dispersion.

Cluster units are located on a 2D lattice, each one associate with a pattern prototype (dimension 500).



Adaptive
Competitive
Learning

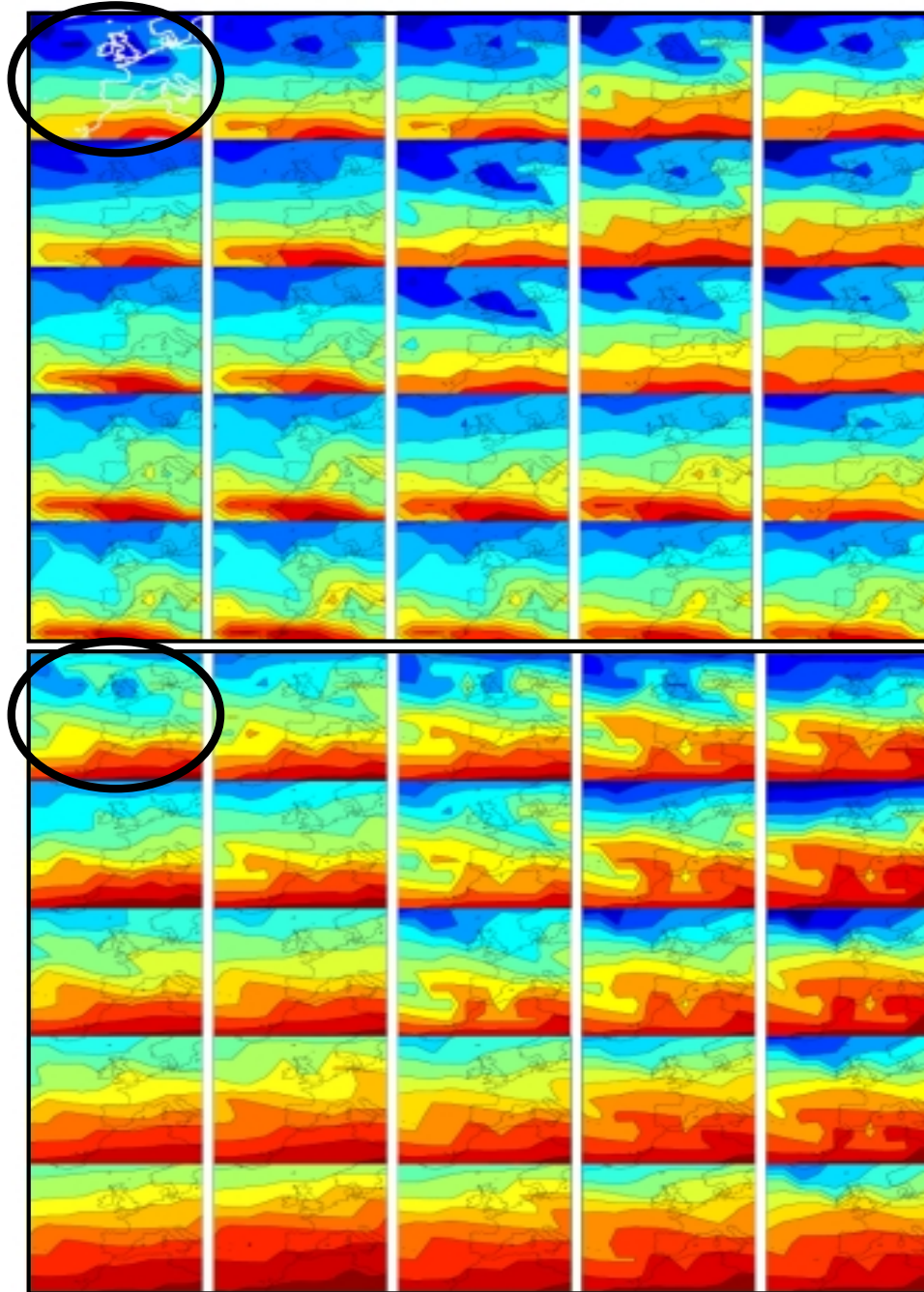
Dimension 500



Topology preserving transformation.

Close clusters in the lattice correspond to close prototypes in the high dimensional data space.

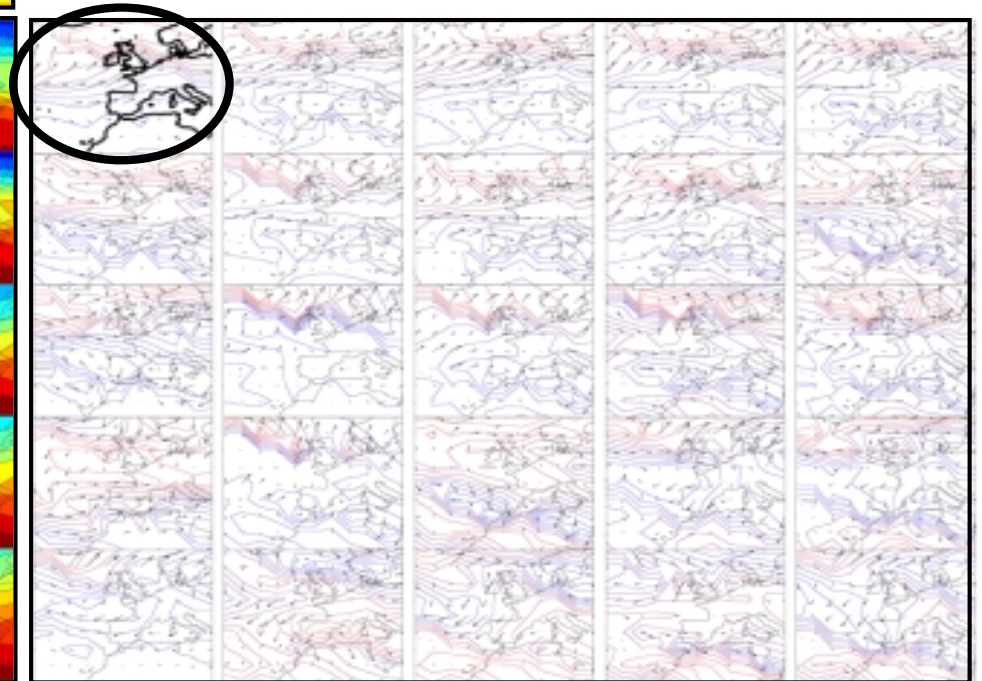
T 1000mb, T 500 mb Z,U,V 500mb patterns for a 5X5 SOM



Prototypes for a trained SOM.

Close units in the lattice are associated with similar atmospheric patterns.

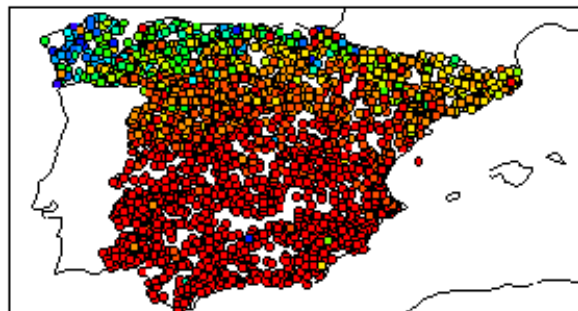
← ● **T 1000mb**
● **T 500mb**
● **Z, U, V 500mb**



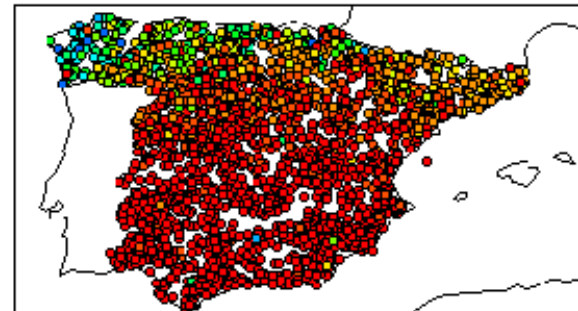
Meteoros de la Red Secundaria: Precipitación

Utilizando las salidas del HIRLAM (INM).

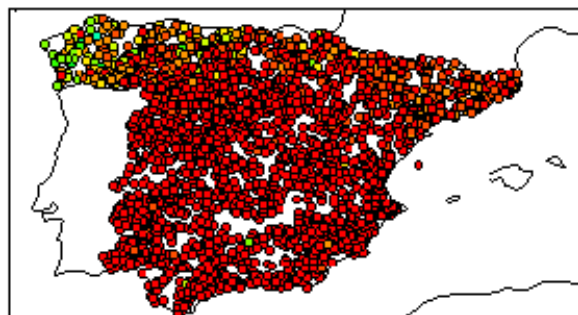
(predicción a D+1 para el 17/7/2001)



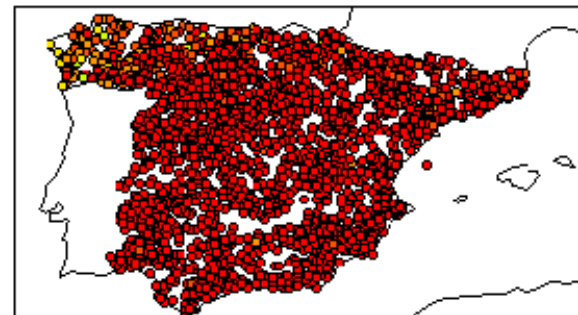
$P(pp > 0.5mm)$



$P(pp > 2mm)$



$P(pp > 10mm)$



$P(pp > 20mm)$

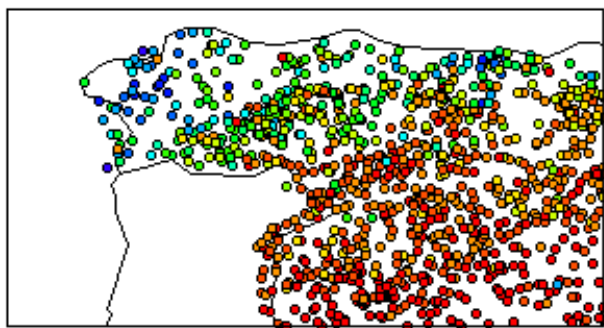


INM

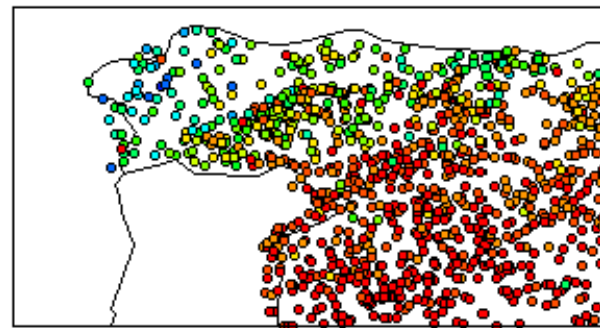
UC



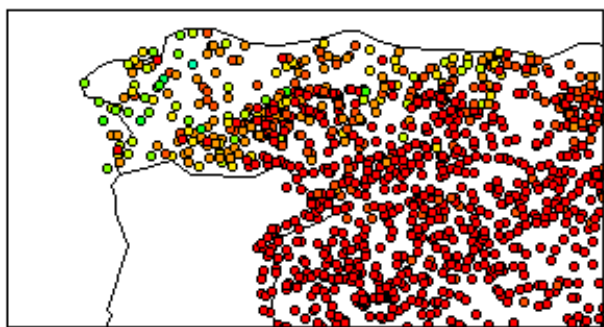
Probabilidad de Precipitación (detalle del NO)



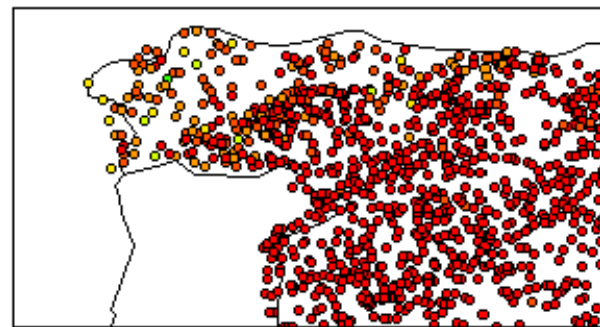
$P(pp > 0.5mm)$



$P(pp > 2mm)$



$P(pp > 10mm)$



$P(pp > 20mm)$



Cantabria y Asturias

Cataluña

País Vasco

Galicia

IM

UC

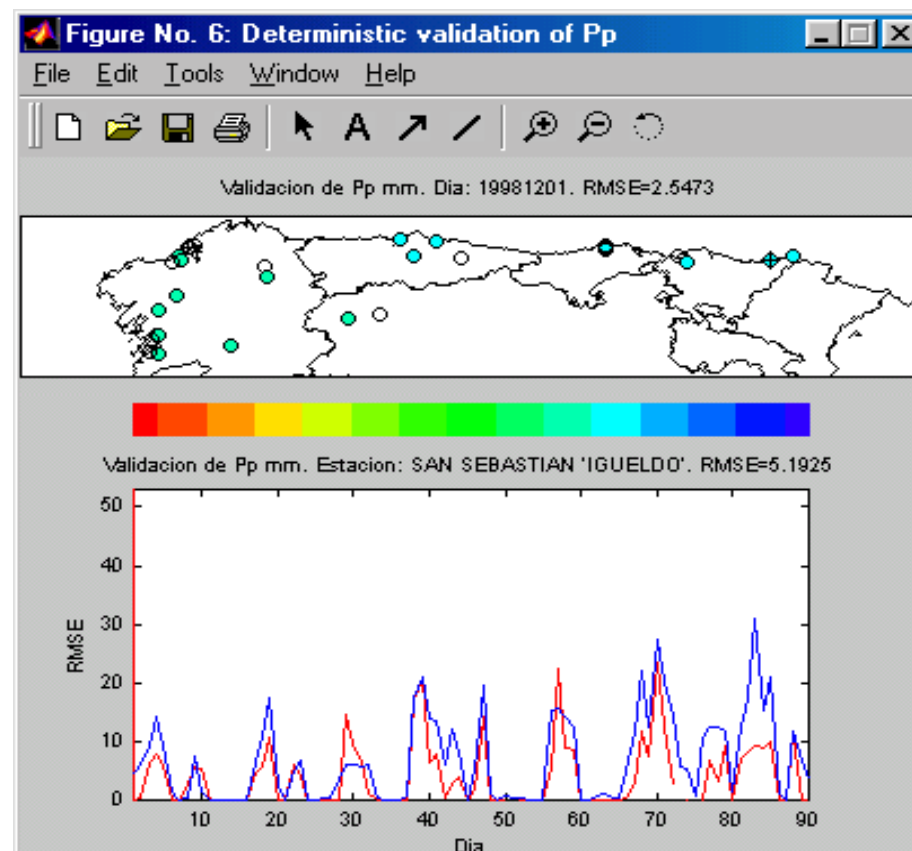
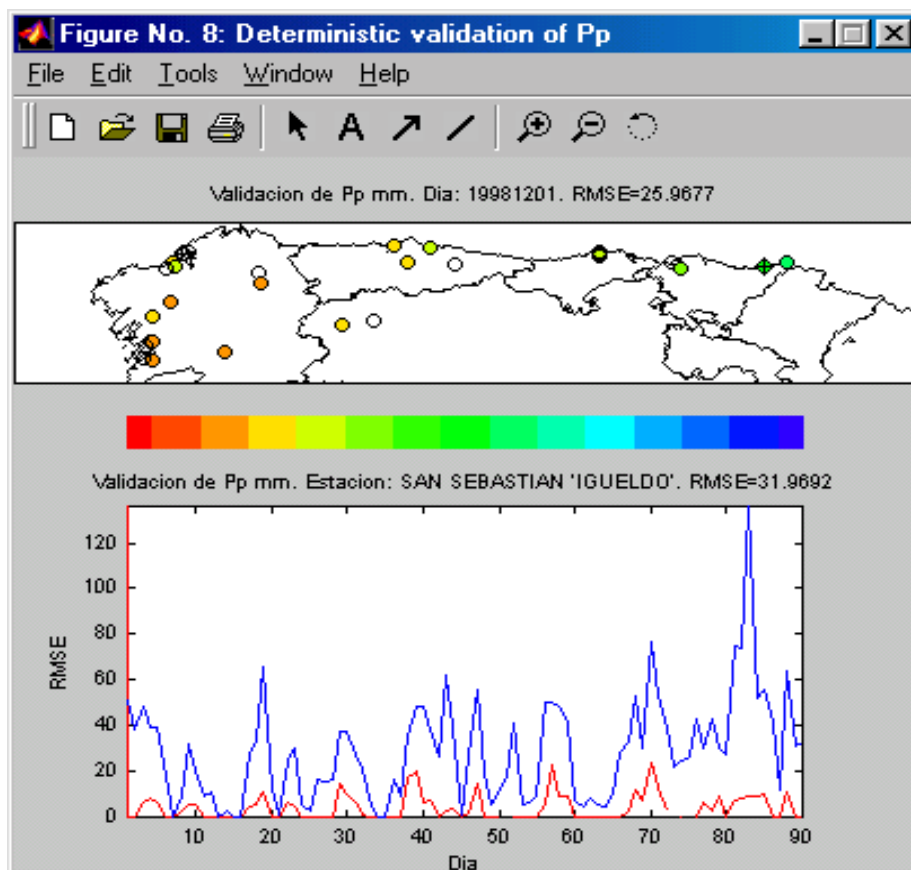


Precipitación y Temperaturas en la Red Principal

Precipitación

Predicción determinista.

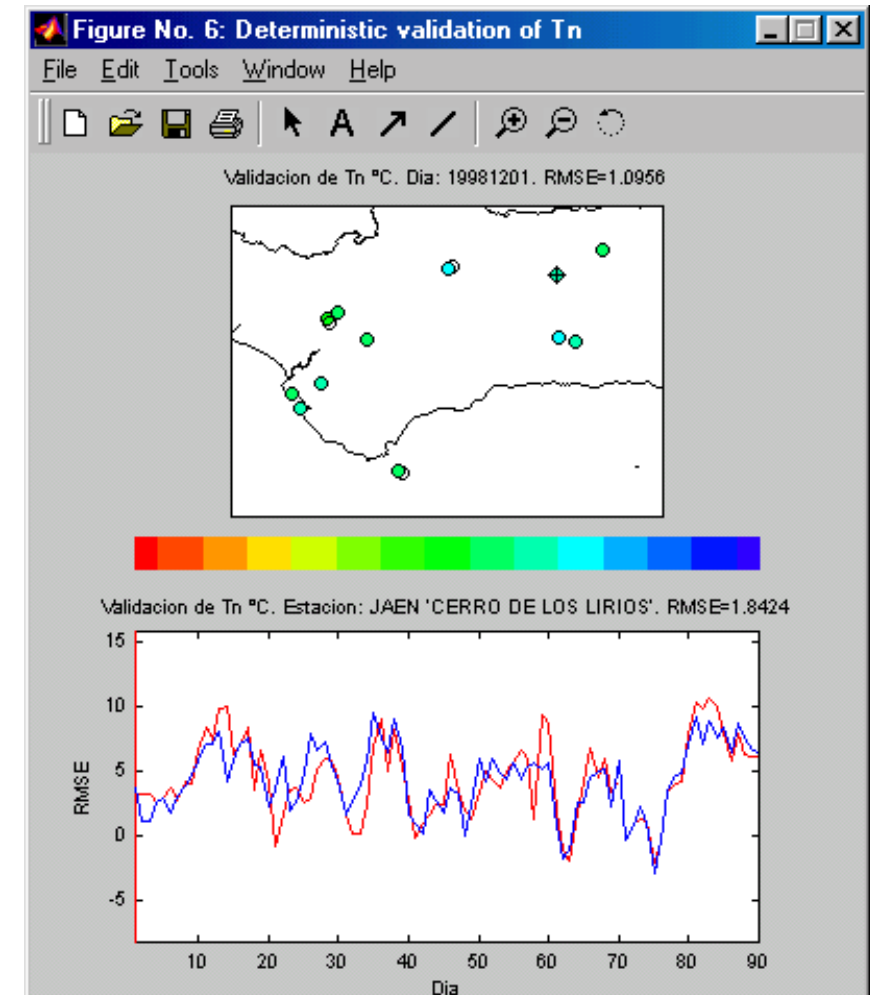
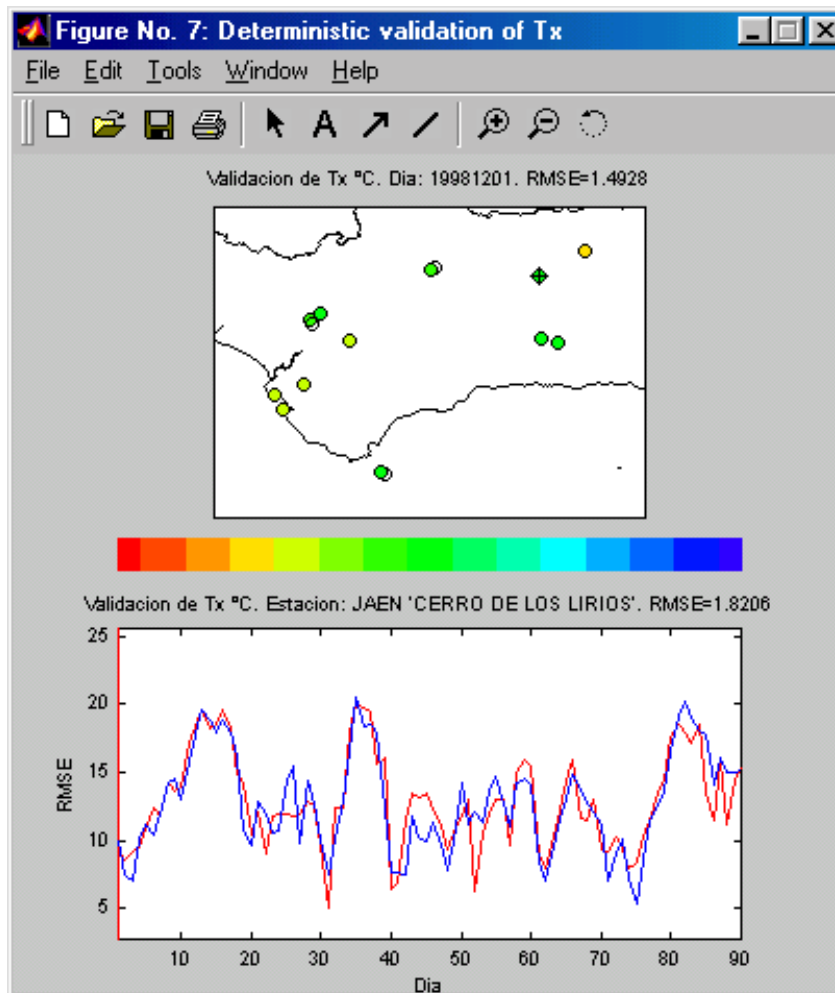
Precipitación = ?



Temperaturas máxima y mínima

Predicción determinista. $T_{max} = ?$, $T_{min} = ?$

Ejemplo: Cuenca del Guadalquivir



VALIDACIÓN. Cuencas Norte y del Guadalquivir

	Pp								Rx				In			
	>0.5mm		>2mm		>10mm		>20mm		>50km/h		>80km/h		>20%		>80%	
	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
DJF	0,45	0,57	0,44	0,55	0,28	0,40	0,18	0,29	0,33	0,37	0,33	0,44	0,46	0,47	0,24	0,36
MAM	0,45	0,54	0,41	0,48	0,27	0,34	0,16	0,16	0,32	0,36	0,32	0,28	0,20	0,22	0,33	0,29
JJA	0,33	0,38	0,32	0,36	0,20	0,24	0,18	0,29	0,19	0,23	0,85	0,67	0,15	0,16	0,28	0,29
SON	0,42	0,54	0,35	0,48	0,29	0,36	0,23	0,25	0,33	0,36	0,39	0,47	0,28	0,27	0,24	0,25

Tabla 1: Resultados de la validación (D+1) en la *cuenca Norte* para los umbrales de precipitación, racha máxima de viento, e insolación indicados en el texto.

	Pp								Rx				In			
	>0.5mm		>2mm		>10mm		>20mm		>50km/h		>80km/h		>20%		>80%	
	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2	M1	M2
DJF	0,64	0,70	0,56	0,60	0,32	0,33	0,31	0,29	0,37	0,37	0,37	0,42	0,34	0,40	0,43	0,46
MAM	0,31	0,41	0,14	0,27	0,12	0,28	-	-	0,19	0,26	-	-	0,21	0,23	0,34	0,42
JJA	0,18	0,23	0,42	0,72	0,57	0,90	0,32	0,91	0,14	0,15	0,91	0,80	0,45	0,66	0,32	0,40
SON	0,29	0,47	0,38	0,48	0,36	0,44	0,19	0,3	0,35	0,35	0,56	0,54	0,18	0,23	0,35	0,37

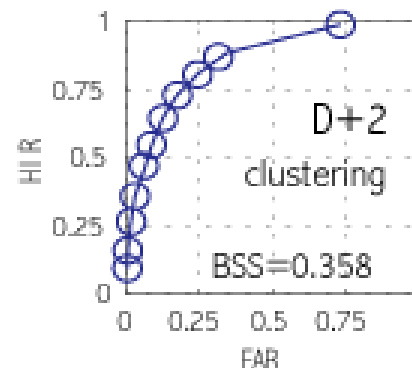
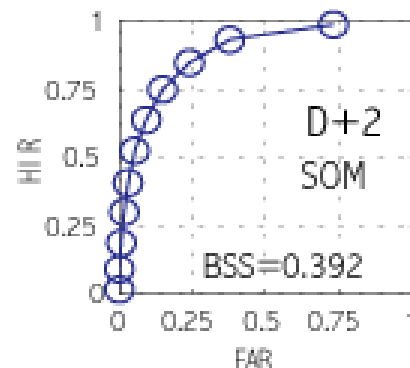
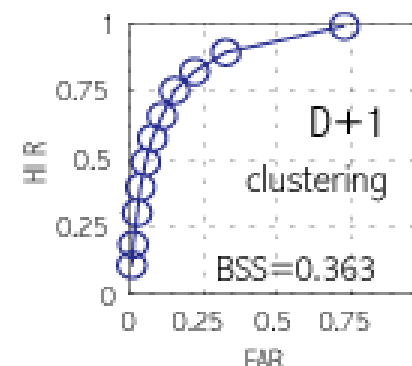
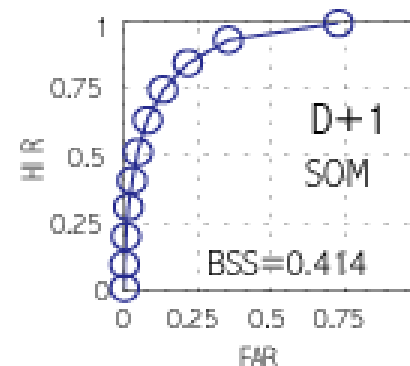
Tabla 2: Resultados de la validación (D+1) en la *cuenca del Guadalquivir* para los umbrales de precipitación, racha máxima de viento, e insolación indicados en el texto.

Validación en toda la Península

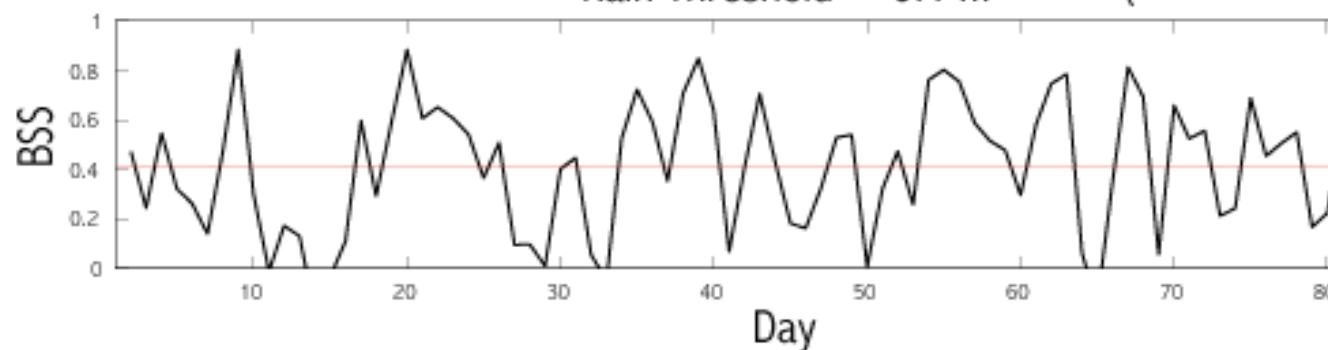
*122 estaciones completas
del INM*



Precipitación > 2mm



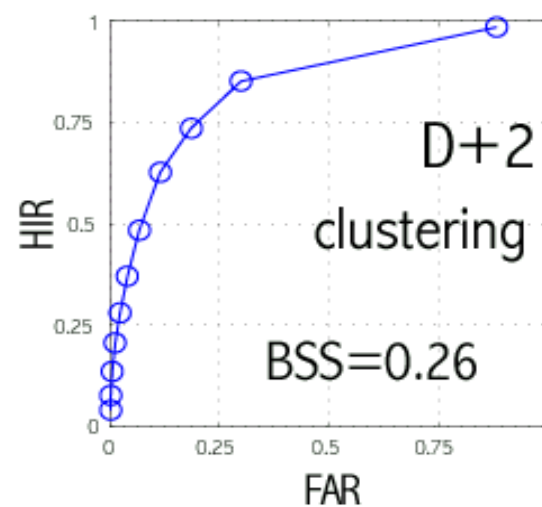
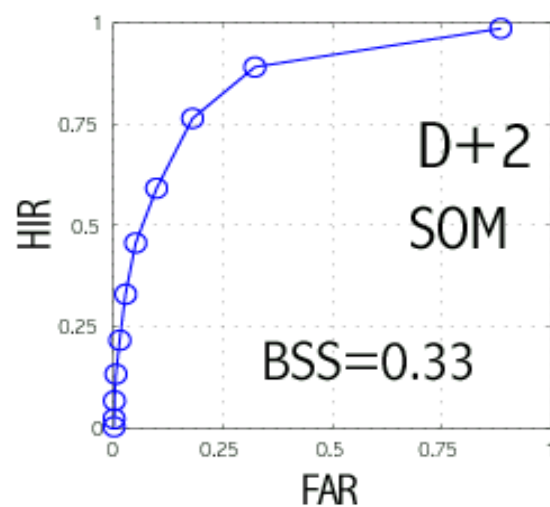
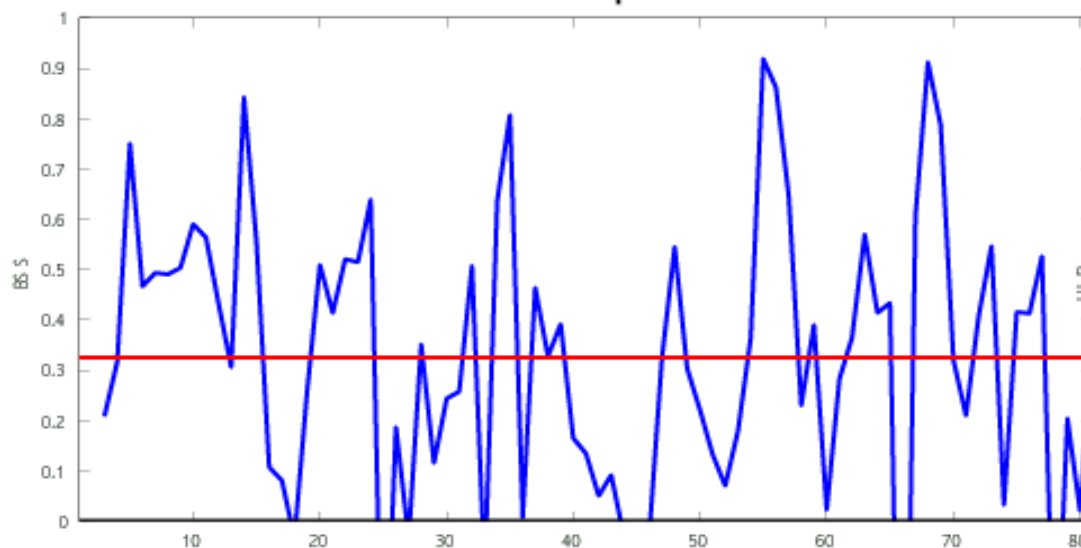
Rain Threshold = 0.1 m (BSS=0.414)



Validación en toda la Península



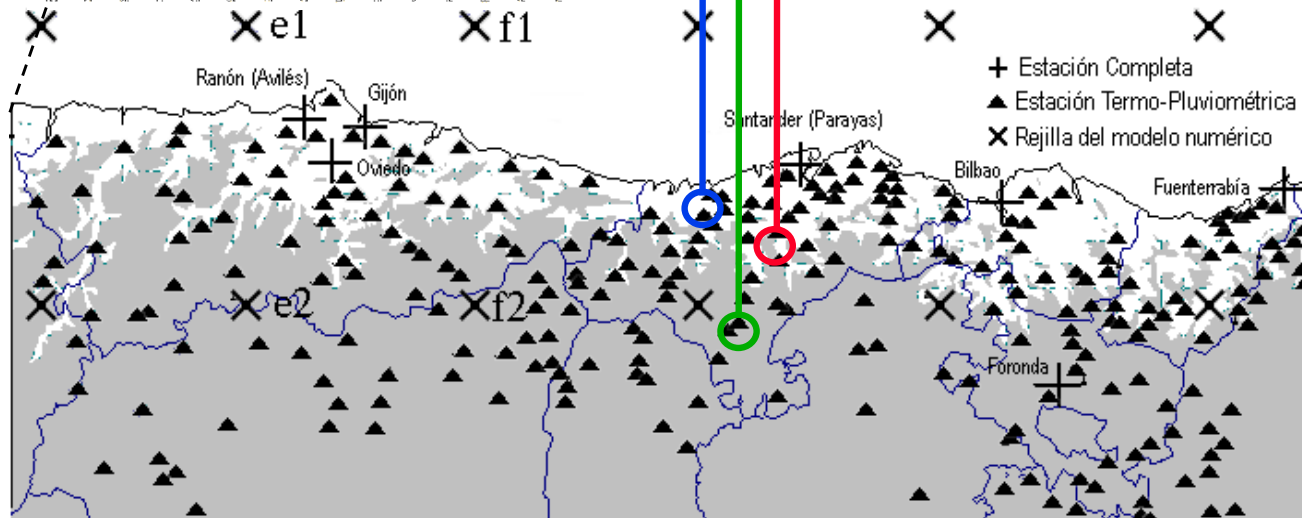
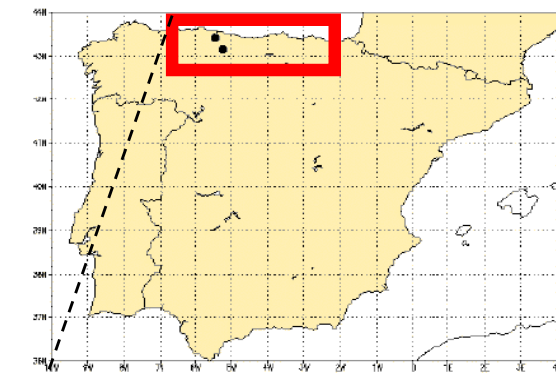
Maximum wind speed > 50Km/h



Validación en la Red Secundaria (Modelo 2)

~ 2600 estaciones

Precipitación
Meteoros (nieve, niebla, ...)



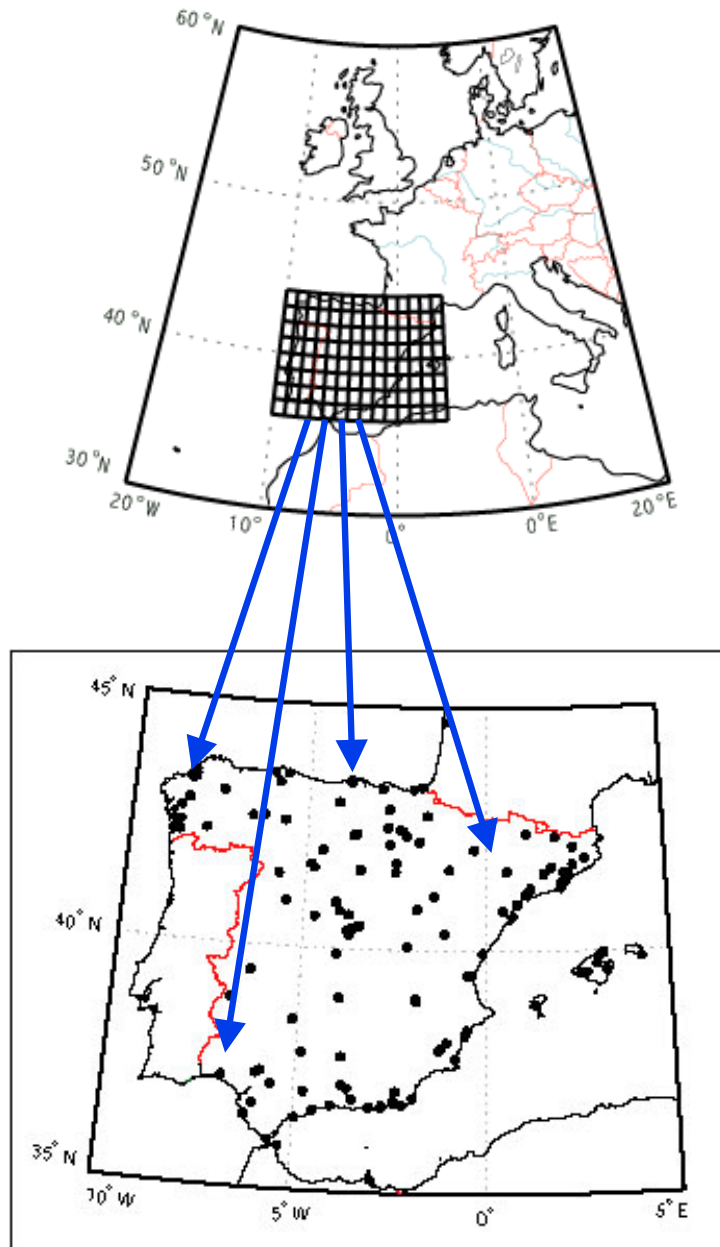
BSS para D+1

Precipitación	Nieve	Niebla
0.16	----	-0.22
0.43	0.52	0.09
0.58	0.36	0.27

*Hay una gran
variabilidad !!!!!*

*Calidad
inhomogénea
de los datos.*

Redes Probabilísticas



¿Qué es una Red Bayesiana?

Un conjunto de variables mas la probabilidad de todas sus combinaciones (PDF conjunta) forman una Red Probabilística. Cuando las relaciones son causales y acíclicas, la Red es Bayesiana.

Cada combinación de variables define un estado del sistema. Obviamente la PDF conjunta la aprende de una Base de Datos.

¿Para qué sirve?

Para inferir las variables desconocidas a partir de las conocidas (EVIDENCIAS), como en una regresión pero con argumentos probabilísticos. Admite evidencias categóricas o en forma de PDF (EPSs).

Dependencia entre Meteoros (Física)

Algunos problemas involucran gran número de variables y se conocen ciertas relaciones de independencia entre ellas.

Obtener un modelo probabilístico

Cto. de relaciones
 $I(X,Y|Z)_M$

Cto. de variables aleatorias
 $\{X_1, X_2, \dots, X_n\}$

Función de probabilidad conjunta
 $P(X_1, X_2, \dots, X_n)$

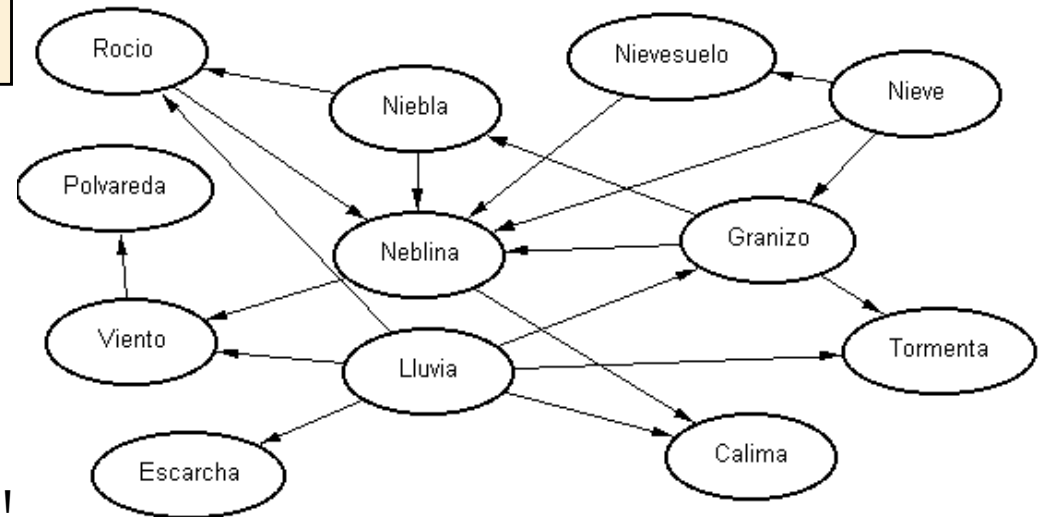
Factorización de la probabilidad !!

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P_i(X_i | \pi_i)$$

Lluvia	Nieve	Granizo	Tormenta	Niebla ...
5	0	0	0	0 ...
1	0	0	0	0 ...
5	0	0	1	0 ...

Relaciones de dependencia

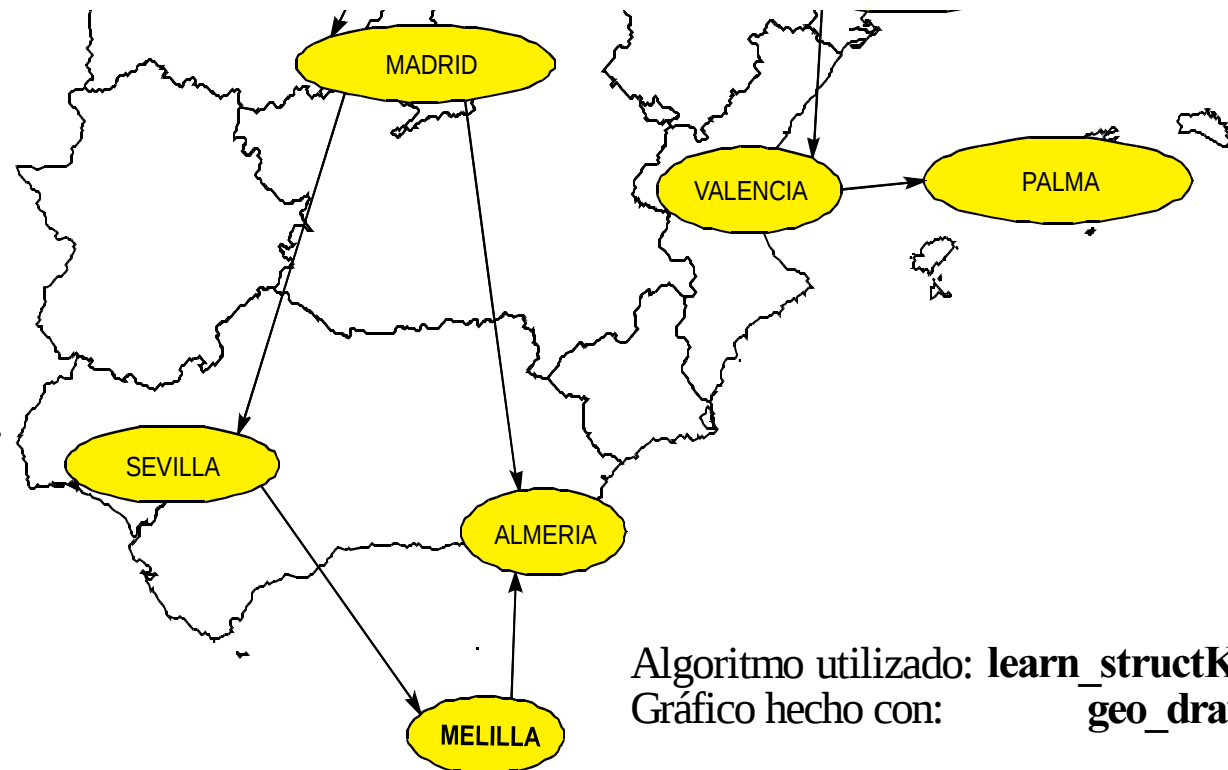
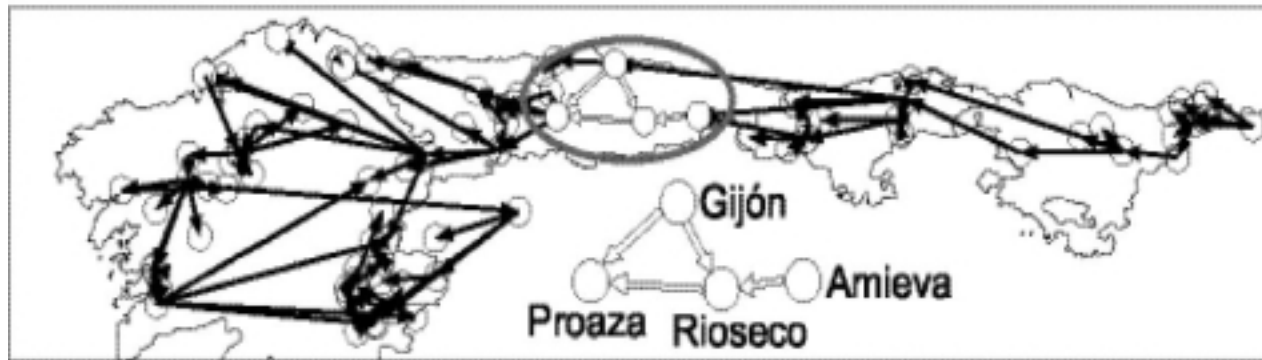
Mediante un grafo dirigido donde cada variable tiene sus antecedentes.



Escarcha		
Lluvia	0.0	1.0
5.0	0.998	0.002
1.0	0.996	0.004
3.0	0.990	0.010
0.0	0.963	0.037
4.0	0.917	0.083

Cuantificación
Funciones de prob. condicionada.

Dependencia entre localidades (espacial).



La existencia de Modos Normales o situaciones típicas, indica que existe un acoplamiento espacial importante.

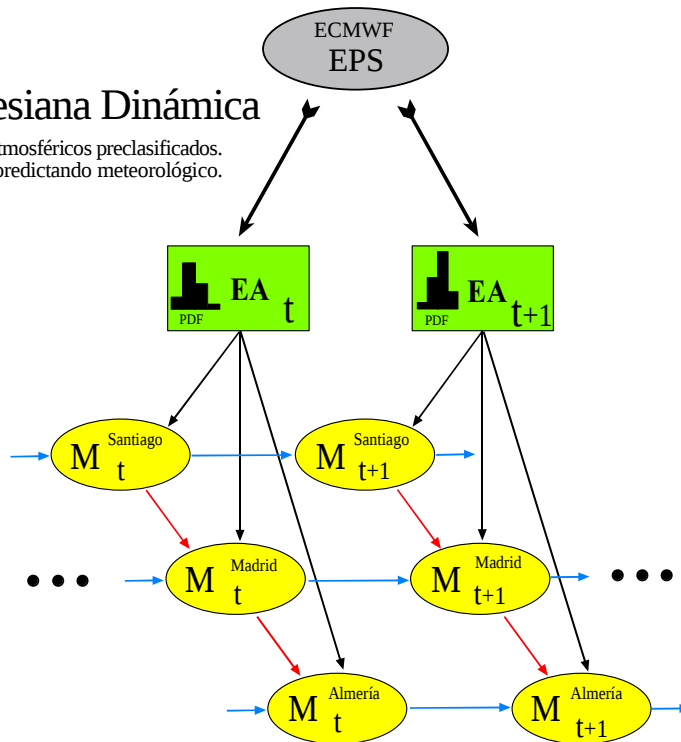
Algoritmo utilizado: **learn_structK2**
Gráfico hecho con: **geo_draw**

Dependencia temporal. Modelo Propuesto

Modelo de Análogos Acoplado

Red Bayesiana Dinámica

EA: Estados Atmosféricos preclasificados.
M: Cualquier predictando meteorológico.



$$P(m_t^l) \approx P(m_t^l | ea_t, m_{t-1}^l, m_t^j)$$

Ventajas del planteamiento probabilístico

EPS: es el *input* ideal por su naturaleza probabilística.

Actualización: asimila datos aislados *on line*.

Causalidad: estudio *off line* de modelos conceptuales.

Ventajas específicas del CAM

Coherencia: no genera pronósticos incompatibles.

Genérico: es apto para predecir cualquier predictando.

Realismo: el acoplamiento compensa el suavizado.



Gracias por la atención !!!!!!!!!!!!!

<http://grupos.unican.es/ai/meteo>