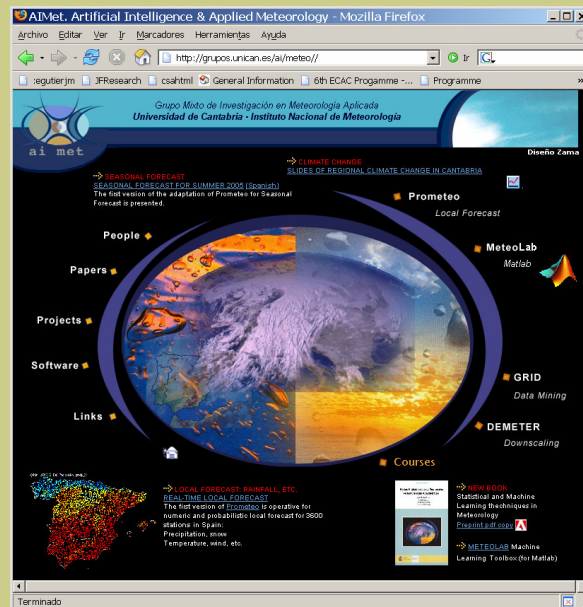


Introducción a los métodos estadísticos de regionalización

Madrid, 18-20 octubre de 2006



Rafael Ancell Trueba
Instituto Nacional de Meteorología
Centro Meteorológico Regional de
Cantabria y Asturias

rct@inm.es

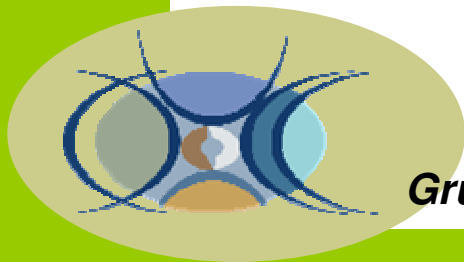


Grupo de Meteorología Aplicada

Introducción a los métodos estadísticos de regionalización

Madrid, 18-20 octubre de 2006

En este documento se pretende dar una **síntesis cualitativa** de los principales métodos estadísticos de regionalización que se utilizan en el estudio del cambio climático (CC). Primero se explica el problema de la regionalización posicionándolo en el marco general del estudio del CC, después se muestra una **clasificación descriptiva** donde se delimitan los diferentes **ámbitos de aplicación**, con sus ventajas e inconvenientes para finalizar con un interesante apunte acerca del manejo de las **incertidumbres en la declaración** de resultados.

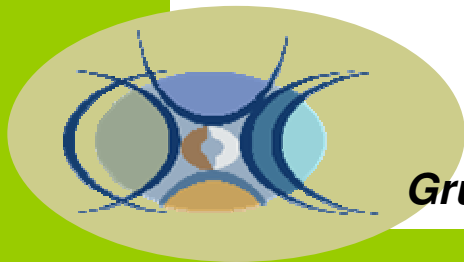


Grupo de Meteorología Aplicada

Introducción a los métodos estadísticos de regionalización

Madrid, 18-20 octubre de 2006

1. Introducción
2. Justificación de los métodos de regionalización
3. Downscaling estadístico en CC
4. Hipótesis de la regionalización estadística
5. Métodos de regionalización estadística
6. Regionalización estadística versus dinámica
7. Incertidumbre del método estadístico
8. Bibliografía y enlaces

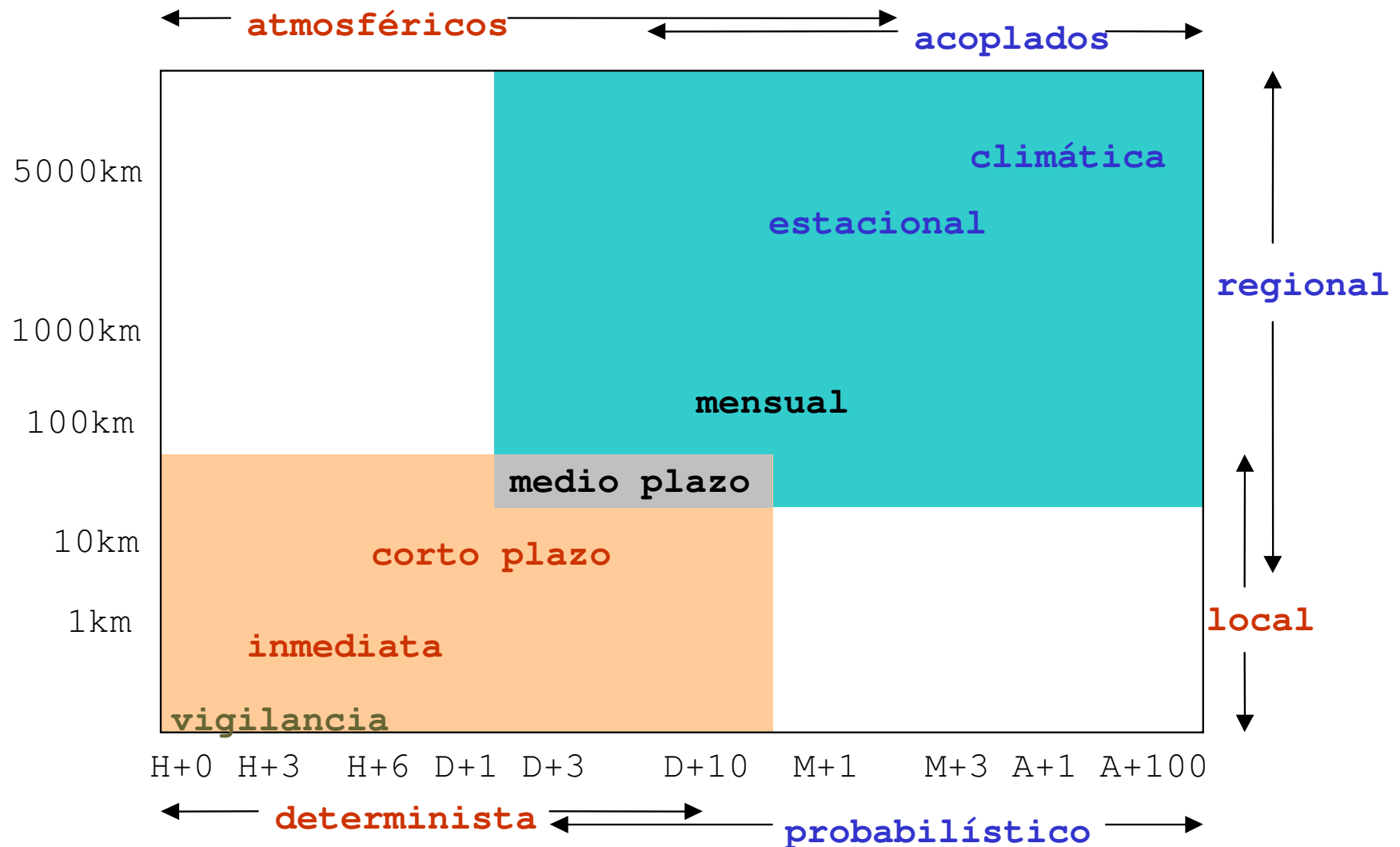


Grupo de Meteorología Aplicada



Modelos y predicciones

En el corto y medio plazo, se analiza y predice la dinámica 'rápida' del sistema (atmósfera); para horizontes más lejanos hay que añadir los procesos 'lentos' (suelo y océano), por lo que la incertidumbre es muy superior: **los predictandos son muestrales (media, percentil, etc) y se predice en términos probabilísticos.**

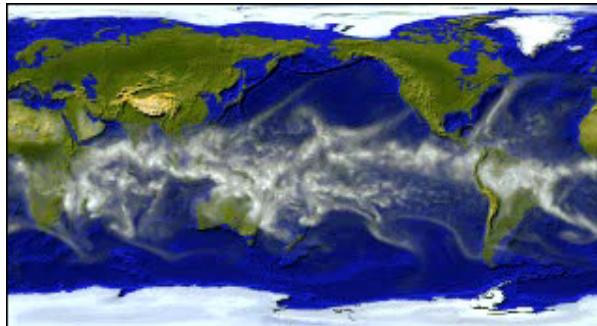




Modelos y predicciones

Actualmente, los supercomputadores paralelos permiten simular con gran realismo los fenómenos más significativos de la circulación atmosférica.

ECMWF. Thunderstorm (>2000 procesadores)



Sin imagen



Actores del estudio del Cambio Climático

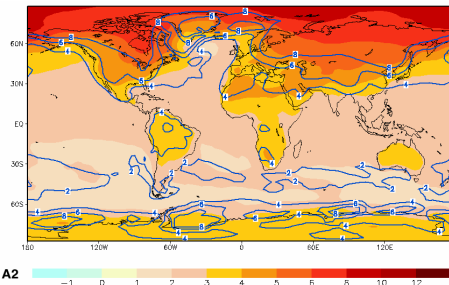
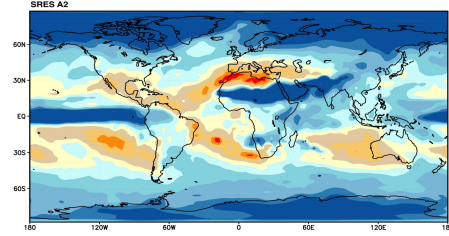
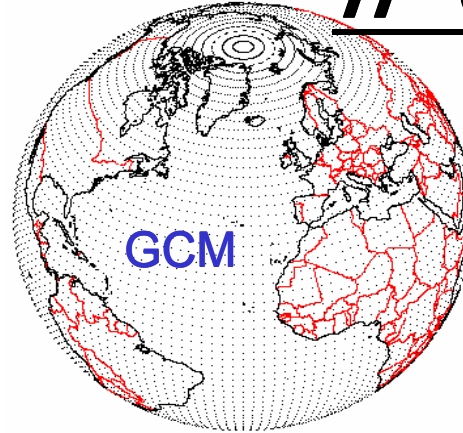
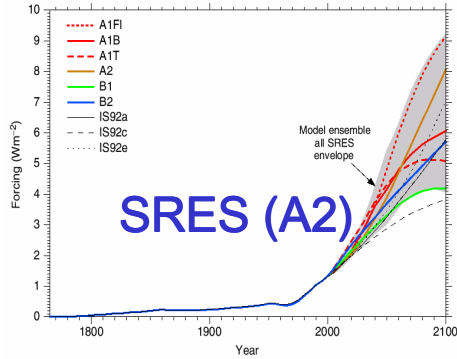
Son los principales elementos utilizados para simular el pasado presente y futuro del sistema climático. Todos y cada uno tienen su ámbito de aplicación delimitado por las hipótesis que de forma explícita o implícita llevan asociadas. Cada uno de ellos supone una fuente de incertidumbre.

1. Escenarios de Emisión (**SRES**: A2, B2, etc.)
2. Modelos Globales (**GCMs**);
3. Modelos Regionales (**RCMs**);
4. Series de datos (observaciones, proxy, paleo,..)
5. **Técnicas Estadísticas**
6. Proyecciones
7.impactos, políticas, etc.

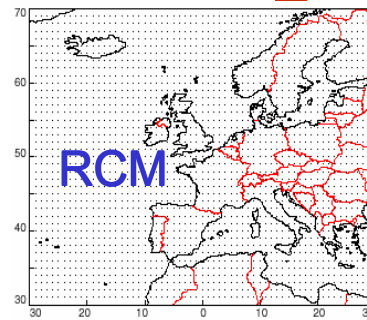
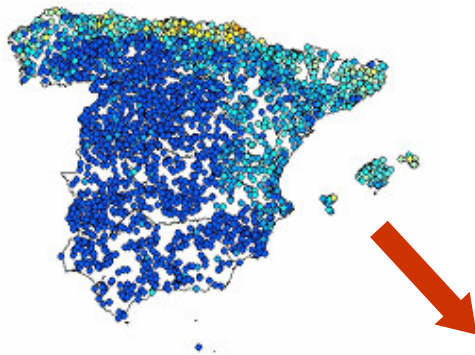
Actores del estudio del Cambio Climático

IPCC

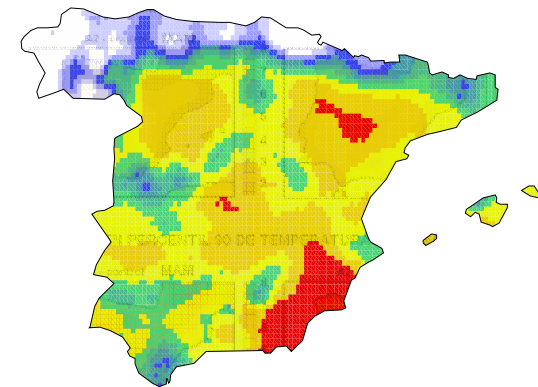
Proyecciones globales



Series de datos



Proyecciones regionales

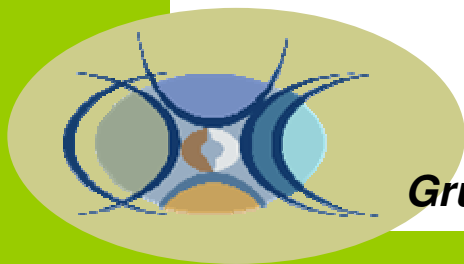


$$Y = f(X; \theta)$$

Introducción a los métodos estadísticos de regionalización

Madrid, 18-20 octubre de 2006

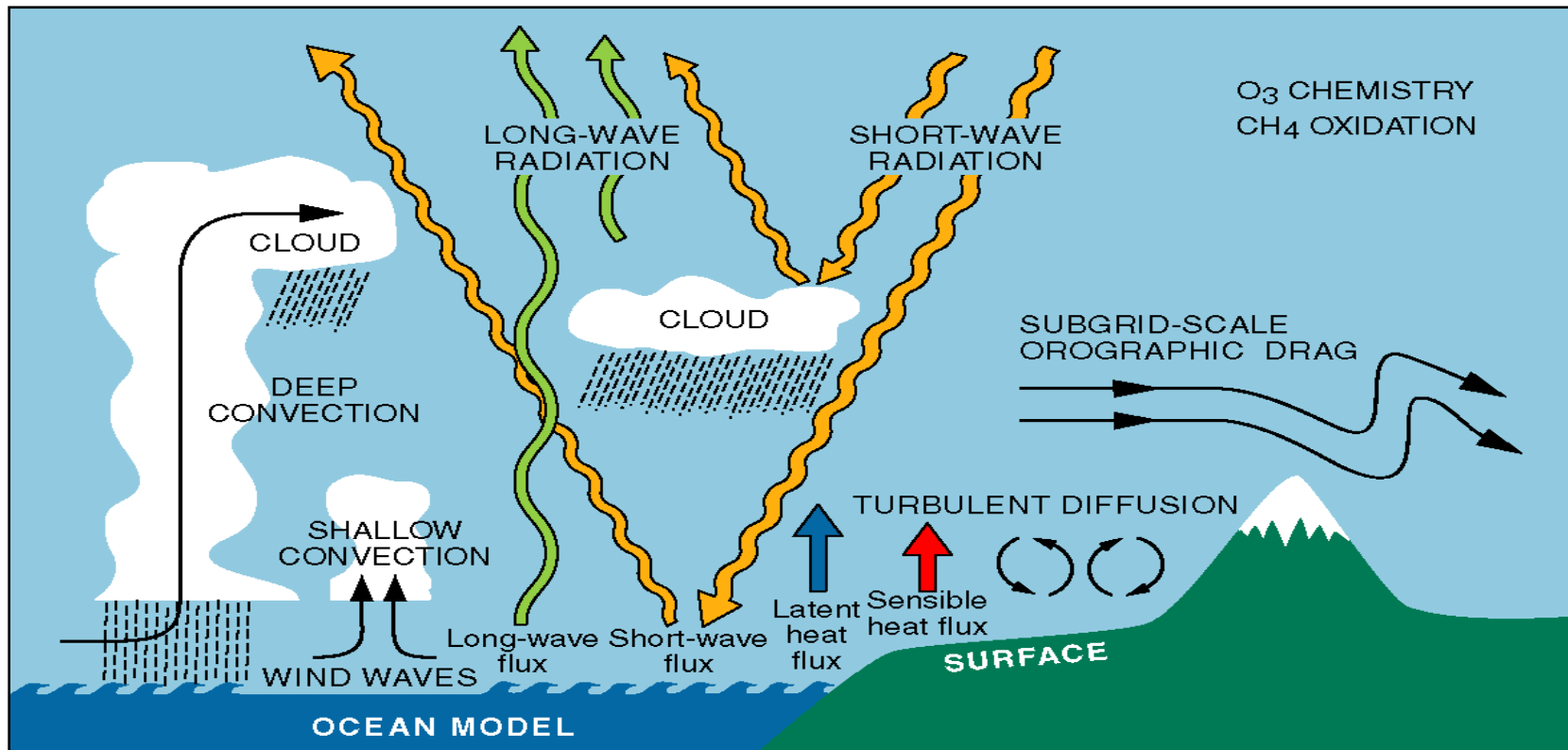
1. Introducción
- 2. Justificación de los métodos de regionalización**
3. Downscaling estadístico en CC
4. Hipótesis de la regionalización estadística
5. Métodos de regionalización estadística
6. Regionalización estadística versus dinámica
7. Incertidumbre del método estadístico
8. Bibliografía y enlaces



Grupo de Meteorología Aplicada

Justificación de los métodos de regionalización

El sistema climático se caracteriza por una gran heterogeneidad superficial que se traduce en una dinámica con estructuras cuya escala espacio-temporal es muy variada (brisas, borrascas, huracanes, ENSO, etc..) y que además se encuentra **sometida a forzamientos**, tanto naturales como antropogénicos que alteran su variabilidad natural. Surge de aquí una primera realidad: **no existe operativo ningún modelo** numérico de circulación global (GCM) capaz de manejar simultáneamente toda la complejidad del SC, de manera que los actuales modelos climáticos parametrizan las escalas pequeñas y como es lógico sus productos **carecen del detalle que corresponde a dichas escalas**.





Modelos Globales

Dinámica oceánica, suelo, criosfera, biosfera, etc...

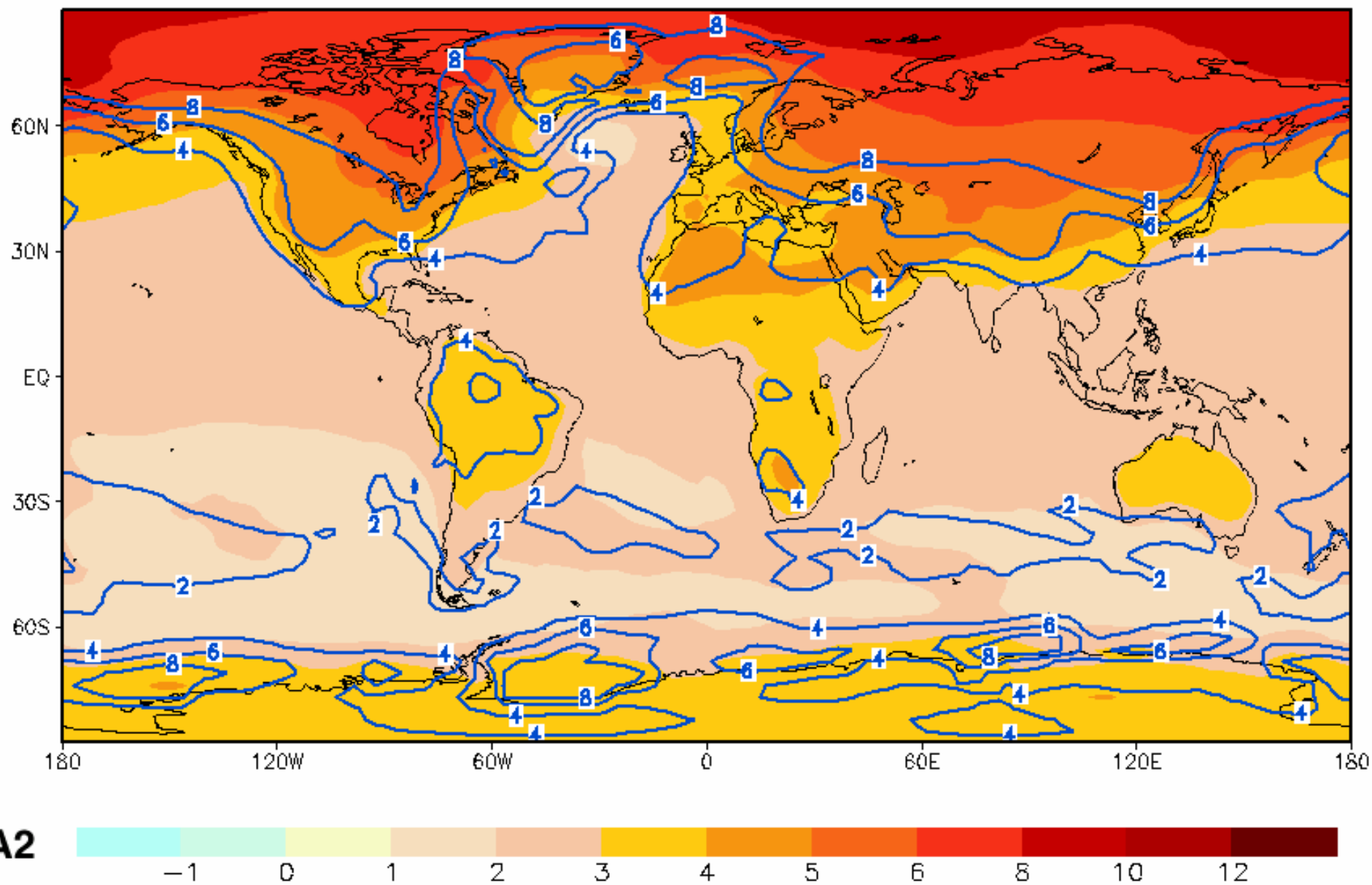
GCMs

Macro-α	15.000 km	1 mes	Circulación general, Ondas largas.
Macro-β	5.000 km	1 semana	Depresiones y Anticiclones.
Macro-γ	500 km	2 días	Frentes, Huracanes.
Meso-β	100 km	6 horas	Brisas, Ondas de montaña, Chorros de bajo nivel, Complejos convectivos, Isla térmica.
Meso-γ	5 km	1 hora	Tormentas, TAC.
Meso-δ	500 m	30 mín.	Cúmulos, Tornados, Vientos catabáticos.
Micro-β	50 m	5 mín.	Penachos, Estelas, Trombas, Tolvaneras.
Micro-γ	2 m	1 seg.	Ondas sonoras, Turbulencia.

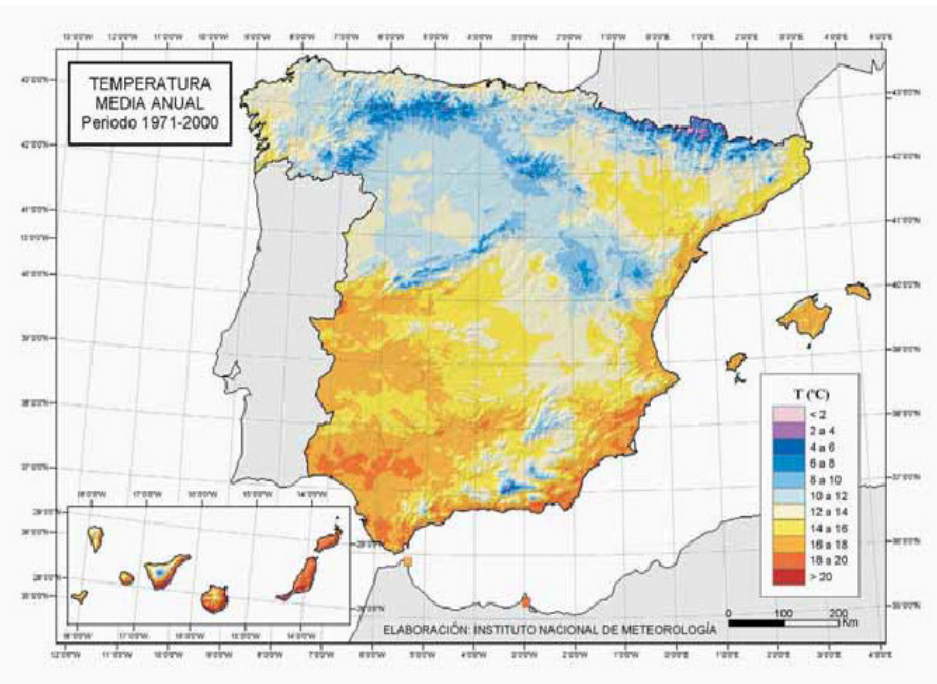
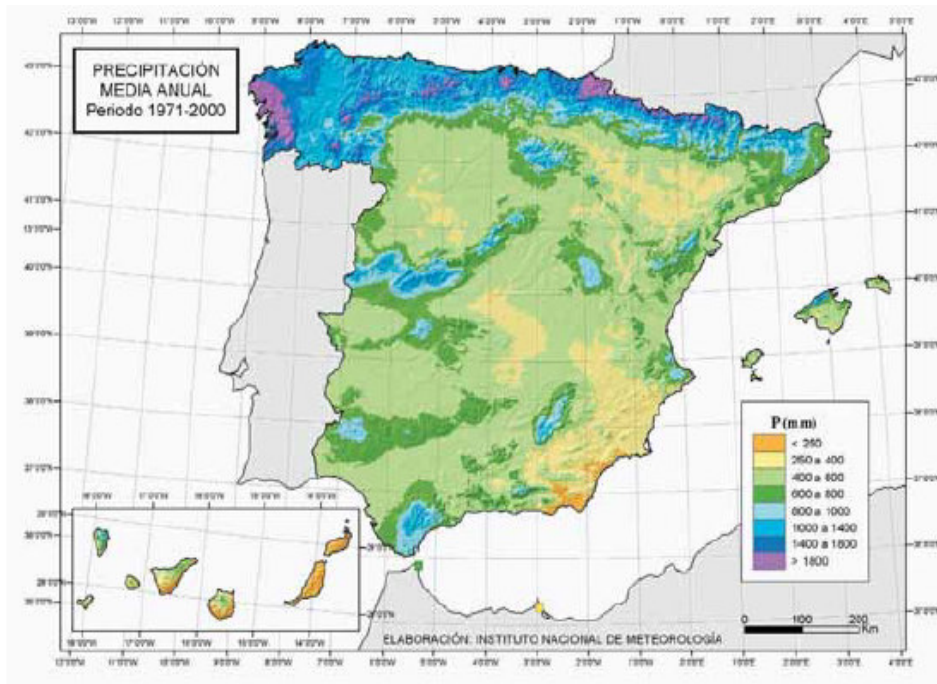
Nombre Escala Duración Fenómenos Atmosféricos

La descripción macroscópica del sistema climático es una primera aproximación a una realidad mucho más heterogénea. El 90% del intercambio energético ocurre en la escala pequeña (100 millones de cúmulos, todos los días, intercambian mucha más energía que un huracán).

Justificación de los métodos de regionalización

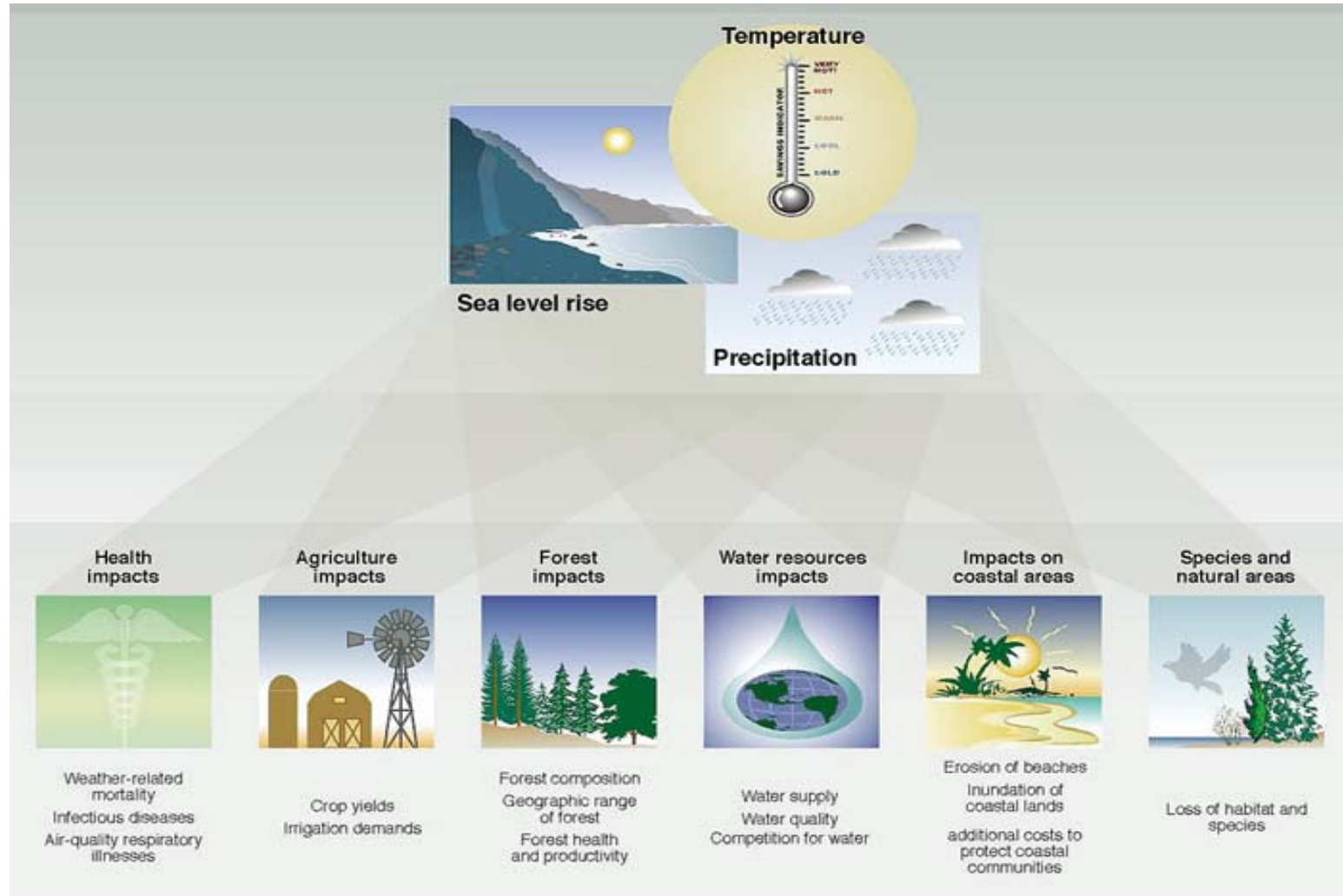


Justificación de los métodos de regionalización



Justificación de los métodos de regionalización

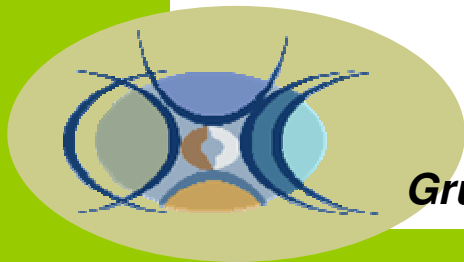
Los estudios de impacto se focalizan sobre áreas muy concretas: demandantes y principales usuarios de la regionalización.



Introducción a los métodos estadísticos de regionalización

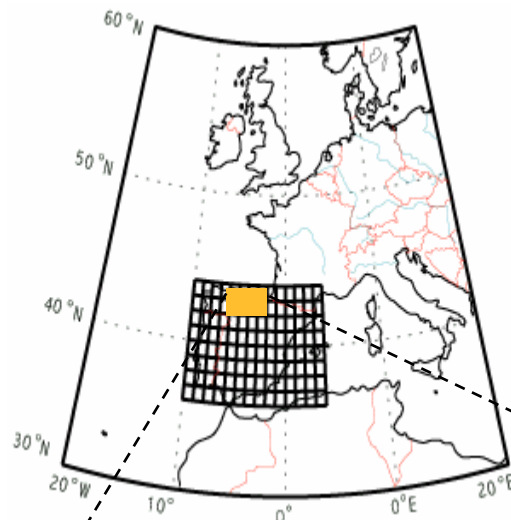
Madrid, 18-20 octubre de 2006

1. Introducción
2. Justificación de los métodos de regionalización
- 3. Downscaling estadístico en CC**
4. Hipótesis de la regionalización estadística
5. Métodos de regionalización estadística
6. Regionalización estadística versus dinámica
7. Incertidumbre del método estadístico
8. Bibliografía y enlaces



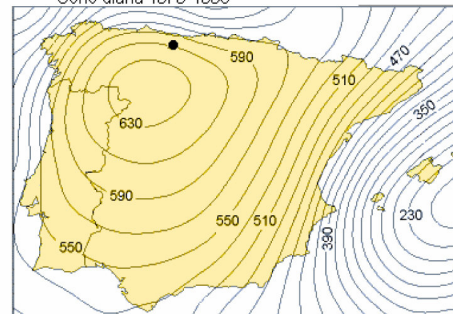
Grupo de Meteorología Aplicada

Modelización Regional. Downscaling

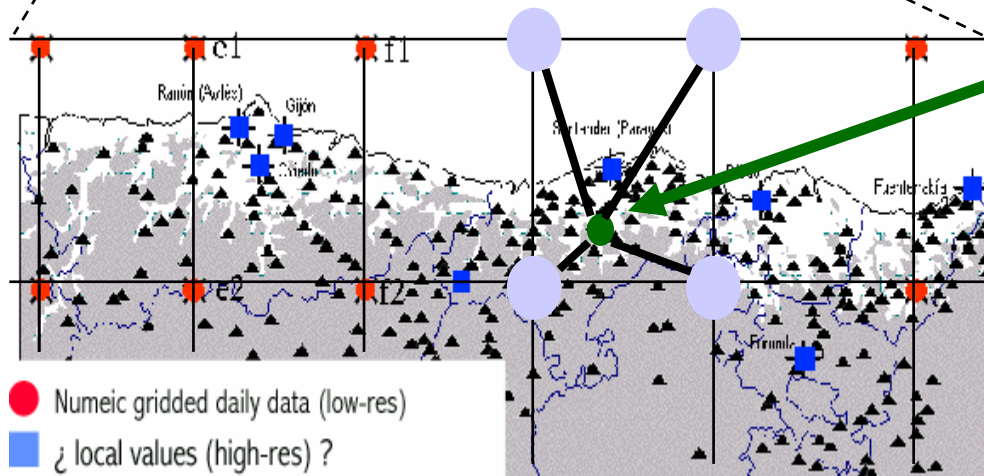
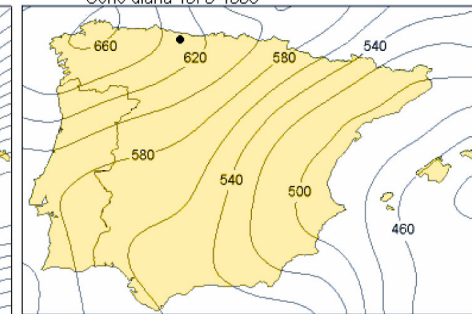


Localidades muy cercanas pueden presentar climatologías muy distintas. Por ejemplo, Oviedo y Ranón están a una distancia de 25 km, pero la climatología de ambas es de influencia distinta (continental y oceanica, respectivamente).

Temperatura máxima en Oviedo
Serie diaria 1979-1993



Temperatura máxima en Ranón
Serie diaria 1979-1993



● Numeic gridded daily data (low-res)
■ ¿ local values (high-res) ?

¿Cómo se puede estimar un valor en un punto de interés?

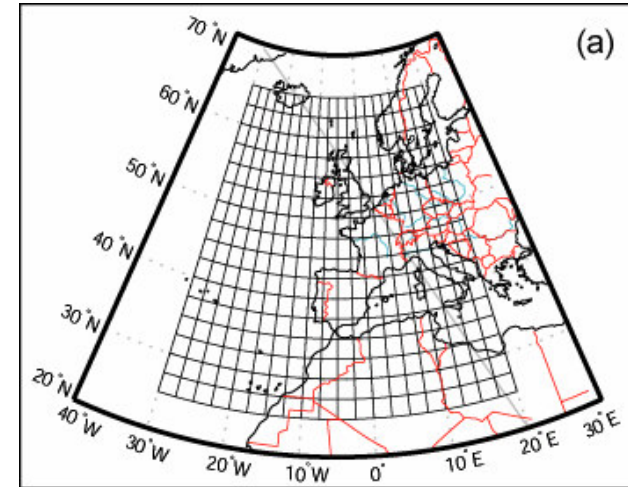
- ✓ **Interpolando las predicciones**
weather.yahoo.com, etc
- ✓ **Con técnicas estadísticas.**
- ✓ **Downscaling dinámico.**
- ✓ **Con técnicas híbridas (Downscaling estadístico).**

Aspectos experimentales del downscaling estadístico

- predictando
 - discreta
 - continua
 - diaria
 - mensual
 - local
 - regional
 - gaussiana*
- predicción
 - determinista
 - probabilística
 - estacional
 - climática



Dominio



Predictandos: **Y**

Predictores: **X**

Técnica **f**

El modelo $\mathbf{Y} = f(\mathbf{X}; \boldsymbol{\theta})$ se obtiene estimando los parámetros $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ a partir de registros históricos de **X** e **Y**.

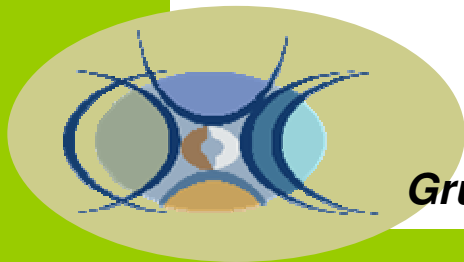
Dada una predicción: \mathbf{X}_τ , se puede obtener una estimación:

$$\hat{\mathbf{Y}}_\tau = f(\mathbf{X}_\tau; \boldsymbol{\theta})$$

Introducción a los métodos estadísticos de regionalización

Madrid, 18-20 octubre de 2006

1. Introducción
2. Justificación de los métodos de regionalización
3. Downscaling estadístico en CC
- 4. Hipótesis de la regionalización estadística**
5. Métodos de regionalización estadística
6. Regionalización estadística versus dinámica
7. Incertidumbre del método estadístico
8. Bibliografía y enlaces



Grupo de Meteorología Aplicada



Hipótesis de la regionalización estadística

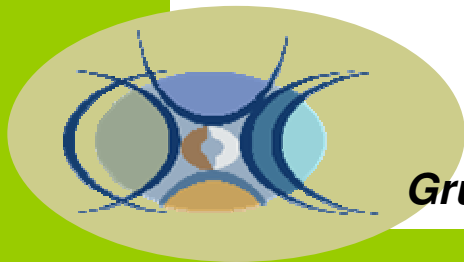
$$Y = f(X; \theta)$$

1. La regionalización estadística se realiza bajo la hipótesis *Perfect Prog* –la alternativa *MOS* contempla modificaciones en los modelos y sólo se utiliza en predicción a corto y medio plazo-; esto quiere decir que se utiliza un periodo de referencia estático para ajustar los métodos estadísticos, no teniendo en cuenta las desviaciones que puedan tener los modelos en el futuro.
2. En segundo lugar, la regionalización estadística es *estacionaria*, lo cual significa que la relación predictor-predictando obtenida en el periodo de ajuste o calibración es aplicada durante el periodo de predicción, a pesar de que los forzamientos del SC son diferentes en este periodo.
3. La hipótesis de *linealidad* consiste en suponer que el modelo estadístico, o lo que es lo mismo, la relación entre predictor y predictando, es lineal, como ocurre entre variables cuya distribución de probabilidad (PDF) es *gaussiana*. Si se tiene en cuenta que los predictores generalmente son variables *gaussianas*, esta hipótesis sólo es apropiada para los predictandos *gaussianos*.
4. Otra hipótesis consiste en suponer que los predictandos son *independientes* entre sí y por tanto solo dependen de los predictores, permitiendo construir un modelo conjunto por la simple agregación de modelos individuales.
 - Estas dos últimas hipótesis implican una enorme eficacia y robustez en las técnicas correspondientes, pero sólo deben ser utilizadas cuando la naturaleza del problema lo permita.

Introducción a los métodos estadísticos de regionalización

Madrid, 18-20 octubre de 2006

1. Introducción
2. Justificación de los métodos de regionalización
3. Downscaling estadístico en CC
4. Hipótesis de la regionalización estadística
- 5. Métodos de regionalización estadística**
6. Regionalización estadística versus dinámica
7. Incertidumbre del método estadístico
8. Bibliografía y enlaces



Grupo de Meteorología Aplicada



Métodos de regionalización estadística (catálogo del IPCC)

Para seleccionar un método de regionalización estadística hay que tener en cuenta la naturaleza de las variables y de la propia predicción:

- Los predictores y predictandos pueden ser discretos, continuos, diarios, mensuales, locales, regionales, *gaussianos*, no *gaussianos*, etc.
- La predicción puede ser local, regional, diaria, mensual, determinista, probabilística, estacional, climática, etc.

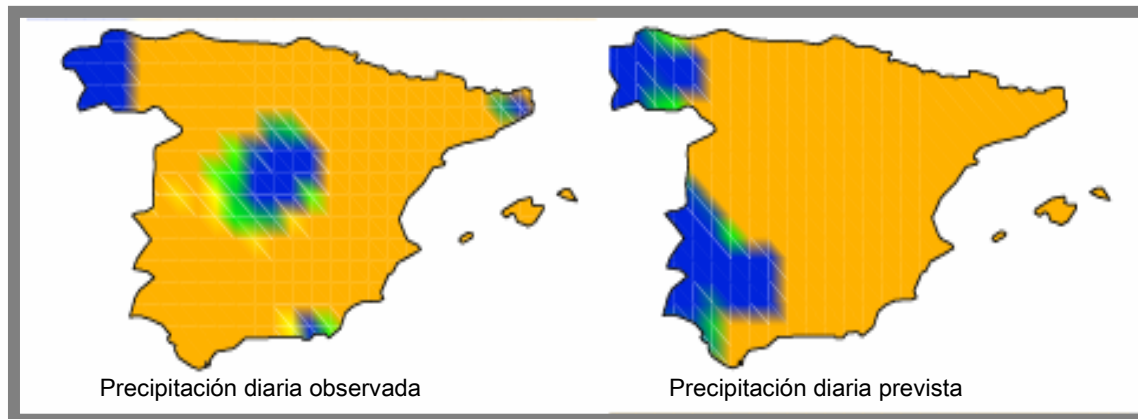
Hay varias formas de catalogar los diferentes métodos de regionalización estadística, aquí se utilizará la propuesta por el IPCC en la cual se distinguen tres categorías:

deterministas
probabilísticos
generadores

Métodos de regionalización estadística (catálogo del IPCC)

Deterministas

Obtienen sus modelos mediante **regresión**/ajuste de parámetros minimizando el error cuadrático medio (**RMSE**), razón por la cual recortan mucho la varianza del predictando. De este tipo son la **regresión múltiple**, el **análisis de correlación canónica**, los **filtros de Kalman** y las **redes neuronales** por citar algunos de los más conocidos. La mayoría de estos métodos son **lineales** –excepto la redes neuronales- y aplican la hipótesis de predictandos **independientes**; solo la correlación canónica considera de forma directa la dependencia -solo la lineal- entre predictandos. En el resto de técnicas se puede considerar esta dependencia indirectamente, pero presentan problemas de **sobreajuste** si se combinan demasiados predictandos.



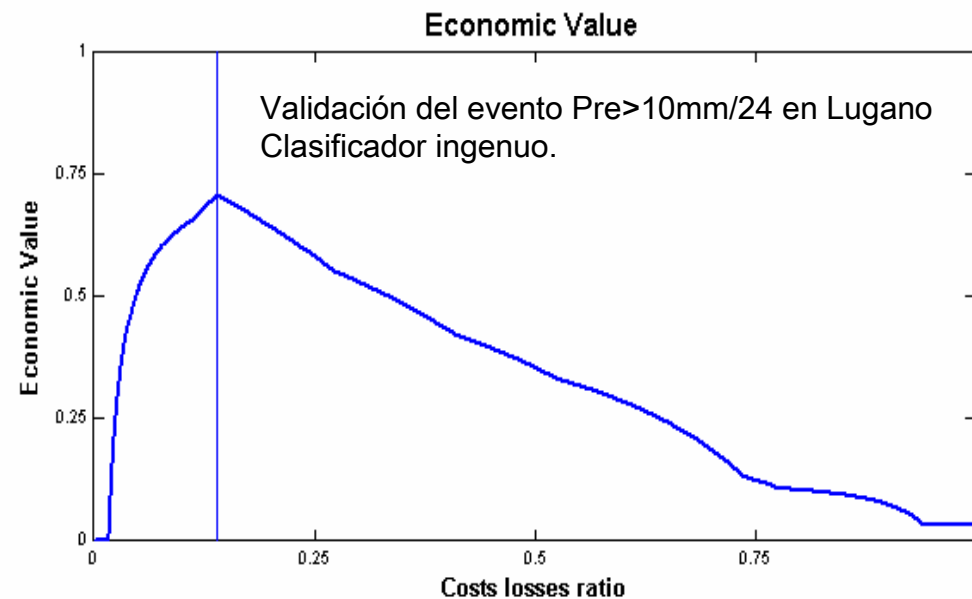
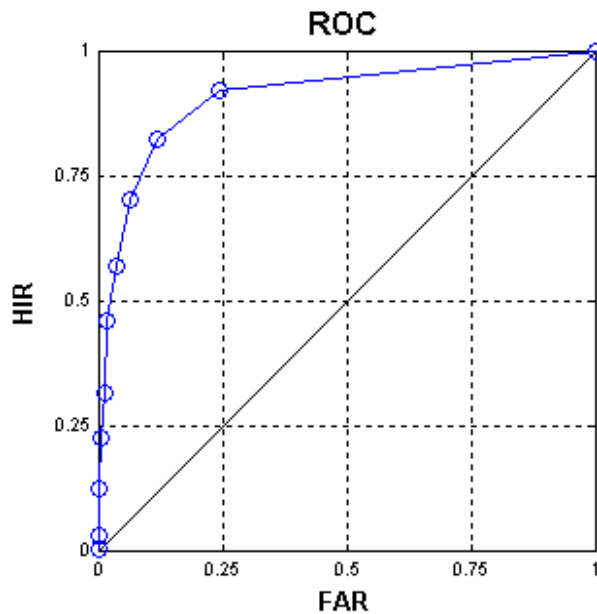
La correlación canónica es una generalización de la regresión múltiple para varios predictandos con un único modelo.



Métodos de regionalización estadística (catálogo del IPCC)

Probabilísticos

Estiman la **PDF conjunta** predictor-predictando obteniendo un modelo capaz de estimar la PDF del predictando dado el predictor. Se basan en dependencia (no sólo en correlación lineal), además son no-lineales, y capaces de considerar la dependencia espacial entre predictandos. Las más generales son las redes probabilísticas (en particular las **redes Bayesianas**). En la práctica existen dos versiones: cuando el condicionamiento se basa en **analogías** entre predictores, se tiene el conocido método de **Análogos de Lorenz**, y si se basa en la **clasificación** de predictores se tienen los **Clasificadores** o *Weather Typing*. **Los clasificadores son más eficientes que los métodos de análogos**, aunque no tan precisos, sin embargo cuando se utilizan en predicción por conjuntos son igual de precisos y mucho más eficientes.





Métodos de regionalización estadística (catálogo del IPCC)

Weather Generators

Estos métodos simulan series *sintéticas* a partir de una predicción (p.e. la precipitación mensual prevista). Utilizan cadenas de Markov para simular series de predictandos secundarios a partir de los predictandos generados por los métodos espaciales de regionalización anteriormente descritos; éstas series simuladas sólo se deben emplear para extraer sus propiedades estadísticas. Son utilizados en problemas de desagregación (p.e. mensual a diaria), en la obtención de nuevos predictandos (p.e. radiación y temperatura a partir de precipitación) y en la simulación de series largas de eventos extremos. Actualmente su principal problema es que no consideran la dependencia espacial entre predictandos.

Por ejemplo para un modelo de orden 1 bastará con conocer:

$$P_{10} = P(r_t = 1 | r_{t-1} = 0)$$

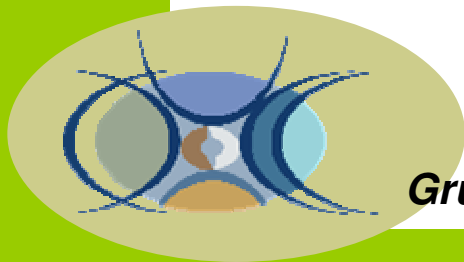
$$P_{11} = P(r_t = 1 | r_{t-1} = 1)$$

Así utilizando números aleatorios se pueden generar series, estudiar las rachas secas, periodos de retorno, etc.

Introducción a los métodos estadísticos de regionalización

Madrid, 18-20 octubre de 2006

1. Introducción
2. Justificación de los métodos de regionalización
3. Downscaling estadístico en CC
4. Hipótesis de la regionalización estadística
5. Métodos de regionalización estadística
- 6. Regionalización estadística versus dinámica**
7. Incertidumbre del método estadístico
8. Bibliografía y enlaces



Grupo de Meteorología Aplicada



Regionalización estadística versus dinámica (STARDEX)

Method provides:	Dynamical downscaling	Statistical downscaling
Station-scale information	No	Yes
Grid-box information	Yes	Yes
European-wide information	Yes	Some methods
Daily time series	Yes	Yes – for 'daily' methods
Temperature and rainfall values which are physically consistent with each other on a daily/seasonal basis	Yes, in theory	Some methods available
Physically and spatially consistent values for multiple sites	Yes, in theory	A few methods available
Information at sites with no observations	Yes	No
Method requirements:		
Computing resources	High	Medium/low
Volume of data inputs	High	Medium/low
Availability of input data	Currently restricted to a few global climate models	Medium/low for observed data



Regionalización estadística versus dinámica

Sopesando las prestaciones y los requerimientos de ambos métodos, quizás la mayor ventaja del método estadístico sea su eficiencia computacional y la facilidad para conseguir series de observaciones casi para cualquier dominio del planeta, lo cual lo convierte en una **alternativa sencilla**, barata y hasta cierto punto competitiva si se compara con los métodos dinámicos.

Además el método estadístico es el único capaz de obtener resultados específicos para **localidades** -esto está limitado a la existencia previa de un registro histórico- evitando el problema que supone considerar datos observacionales interpolados en rejillas que pueden suavizar y hasta eliminar características regionales de los datos.

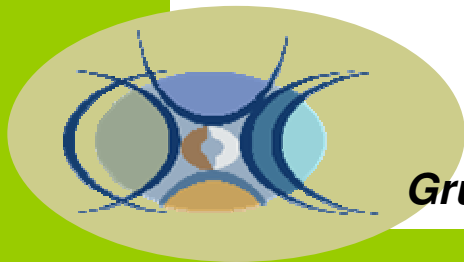
Por otra parte, los métodos estadísticos que suponen independencia entre predictandos no garantizan la **consistencia espacial** ni física de la regionalización mientras que los dinámicos, al menos en teoría, sí.

Finalmente la desventaja más clara del método estadístico es que aplica un modelo **estacionario** a un sistema que, a escalas de tiempo climáticas, no lo es, con el agravante de que no se puede estimar de forma directa la validez de dicha aproximación ya que **el periodo de observación suele ser demasiado homogéneo en términos de forzamiento** –a veces se evalúa de forma indirecta utilizando series de observaciones con forzamientos diferentes generadas de forma sintética por RCMs-.

Introducción a los métodos estadísticos de regionalización

Madrid, 18-20 octubre de 2006

1. Introducción
2. Justificación de los métodos de regionalización
3. Downscaling estadístico en CC
4. Hipótesis de la regionalización estadística
5. Métodos de regionalización estadística
6. Regionalización estadística versus dinámica
- 7. Incertidumbre del método estadístico**
8. Bibliografía y enlaces



Grupo de Meteorología Aplicada



Incertidumbre del método estadístico

Para terminar, un apunte acerca de la interpretación de los resultados. La progresiva disponibilidad de diferentes métodos de regionalización estadística brinda la posibilidad de utilizar un amplio abanico de soluciones para cada problema concreto; como hemos visto, las posibilidades de diseño de un método de regionalización estadístico son muy variadas ya que dependiendo del dominio y naturaleza del predictando, se pueden combinar métodos y predictores de muchas maneras, suministrando resultados diferentes para el mismo GCM. Esta incertidumbre, presente en cada método, debe ser **puesta en evidencia y añadida** a las restantes -la del GCM, la del forzamiento y la natural del SC-; tampoco hay que olvidar que la incertidumbre presenta una distribución espacial heterogénea a pesar de la relativa pequeñez del dominio de aplicación de los métodos estadísticos; por ejemplo, los métodos de clasificación y de análogos obtienen significativamente peores resultados en zonas donde el carácter de la precipitación es más convectivo. En este sentido se recomienda, a iniciativa de STARDEX, realizar, siempre que sea posible, varios experimentos de regionalización estadística, utilizando diferentes técnicas, para conseguir una mejor representación de la incertidumbre de los resultados. También se debe separar claramente la incertidumbre asociada al método empleado (confianza) de la asociada a resultados específicos (verosimilitud); por ejemplo, un buen método puede dar una predicción muy ambigua y un mal método dar una predicción muy concreta.



Fuentes principales de incertidumbre

Dado que las predicciones a estas escalas tan largas son muy difíciles de evaluar directamente, existe un gran riesgo de fallo derivado de un manejo inadecuado de la incertidumbre. Por ello es preceptivo, especialmente en predicción climática, identificar todas y cada una de las fuentes de incertidumbre; que básicamente son de tres tipos:

- **Variabilidad natural:** Al ser un sistema caótico, es muy sensible a pequeñas diferencias en las condiciones iniciales. Esta variabilidad se puede simular considerando una distribución de condiciones iniciales de temperatura, viento, humedad, etc.
- **Forzamientos:** El clima es perturbado por agentes externos naturales, como pueden ser la actividad volcánica o solar, y antropogénicos, como las emisiones de gases de efecto invernadero. Para ello se hacen simulaciones para diferentes escenarios de forzamiento.
- **Modelos y métodos:** Nuestra capacidad de simulación es limitada por lo tanto cada modelo y/o método produce resultados diferentes.



Guía terminológica de la incertidumbre.

método empleado: (confianza)

resultados específicos: (verosimilitud)

Terminology Degree of confidence in being correct

Very High confidence At least 9 out of 10 chance of being correct

High confidence About 8 out of 10 chance

Medium confidence About 5 out of 10 chance

Low confidence About 2 out of 10 chance

Very low confidence Less than 1 out of 10 chance

Terminology Likelihood of the occurrence/ outcome

Virtually certain > 99% probability of occurrence

Very likely > 90% probability

Likely > 66% probability

About as likely as not 33 to 66% probability

Unlikely < 33% probability

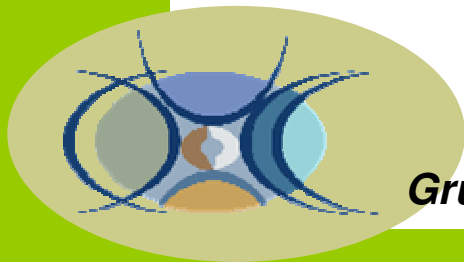
Very unlikely < 10% probability

Exceptionally unlikely < 1% probability

Introducción a los métodos estadísticos de regionalización

Madrid, 18-20 octubre de 2006

1. Introducción
2. Justificación de los métodos de regionalización
3. Downscaling estadístico en CC
4. Hipótesis de la regionalización estadística
5. Métodos de regionalización estadística
6. Regionalización estadística versus dinámica
7. Incertidumbre del método estadístico
8. **Bibliografía y enlaces**



Grupo de Meteorología Aplicada



Bibliografía y enlaces empleados

J.M. Gutiérrez, R. Cano, A.S. Cofiño, and C. Sordo

Redes Probabilísticas y Neuronales en las Ciencias Atmosféricas

Monografías del Instituto Nacional de Meteorología, Ministerio de Medio Ambiente, Madrid
(2004) *ISBN: 84-8320-281-6*

Wilby, R.L., S.P. Charles, E. Zorita, B. Timbal, P. Whetton, and L.O. Mearns, (2004):
Guidelines for use of climate scenarios developed from statistical downscaling methods,
available from the DDC of IPCC TGCIA, 27 pp.

http://www.cru.uea.ac.uk/projects/stardex/reports/STARDEX_FINAL_REPORT.pdf

<http://www.ipcc.ch/activity/uncertaintyguidancenote.pdf>

<http://www.ipcc.ch/activity/ar.htm>

<http://www.clivar.org/publications/reports/reports.php>

<http://www.usclivar.org/Pubs/report.html>



Modelización Regional. Downscaling

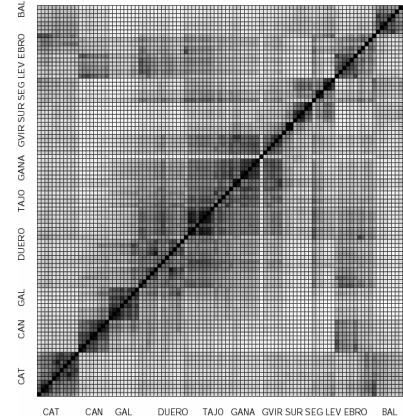


Macro-α	15.000 km	1 mes	Circulación general, Ondas largas.
Macro-β	5.000 km	1 semana	Depresiones y Anticiclones.
Macro-γ	500 km	2 días	Frentes, Huracanes.
Meso-β	100 km	6 horas	Brisas, Ondas de montaña, Chorros de bajo nivel, Complejos convectivos, Isla térmica.
Meso-γ	5 km	1 hora	Tormentas, TAC.
Meso-δ	500 m	30 mín.	Cúmulos, Tornados, Vientos catabáticos.
Micro-β	50 m	5 mín.	Penachos, Estelas, Trombas, Tolvaneras.
Micro-γ	2 m	1 seg.	Ondas sonoras, Turbulencia.

Nombre Escala Duración Fenómenos Atmosféricos

La descripción macroscópica del estado de la atmósfera es una primera aproximación a una realidad mucho más heterogénea. La investigación de las diferentes técnicas de adaptación entre lo previsto a una escala y lo observado a escalas inferiores es uno de los grandes retos de la meteorología y se conoce como **downscaling**.

Correlation and dependency



Evident questions:

Independence implies no correlation?

Correlation implies dependence?

} True Anycase? Yes.

Not so evident questions:

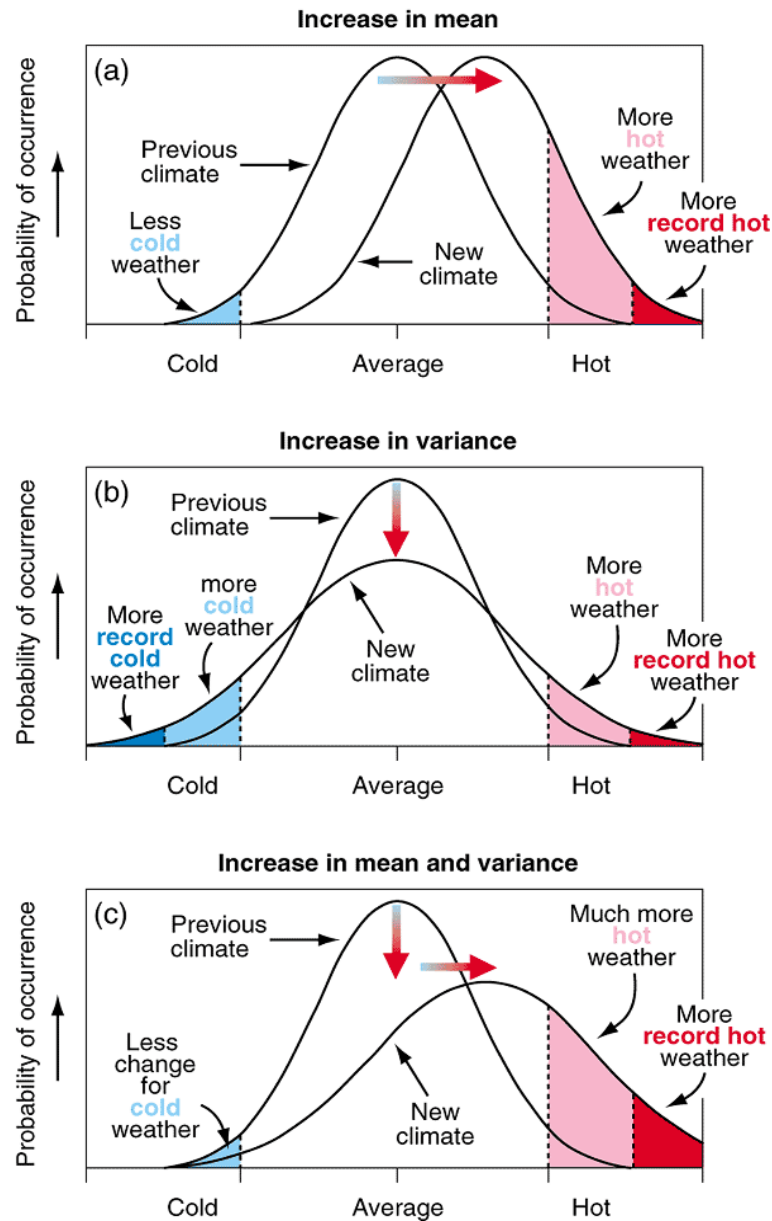
No correlation implies independence?

Dependence implies correlation?

} True if linear
False other case

Dependency is more general than correlation

Modelización Regional. Downscaling estadístico.



There is a growing recognition that changes in the frequency and intensity of extreme events are likely to have more of an impact on the environment and human activities than changes in mean climate.

Schematic showing the effect on **extreme temperatures** when (a) the mean temperature increases, (b) the variance increases, and (c) when both the mean and variance increase for a normal distribution of temperature. For variables that are not well approximated by normal distributions, like **precipitation**, the situation is even more complex, especially for dry climates. For precipitation, for example, changes in the mean total precipitation can be accompanied by other changes like the frequency of precipitation or the shape of the distribution including its variability. All these changes can affect the various aspects of precipitation extremes including the intensity of precipitation (amount per unit time).

