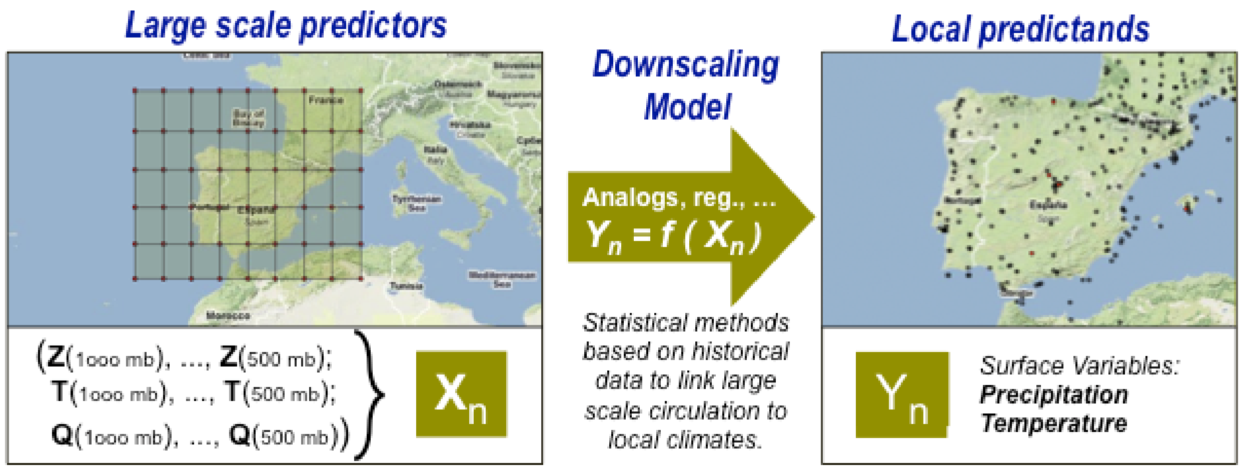
María Dolores Frías
friasmd@unican.es

http://www.meteo.unican.es

Statistical Downscaling:
A user friendly portal



Thanks to: Ana Casanueva
Jose Manuel Gutiérrez
Sixto Herrera
Daniel San Martín
Max Tuni

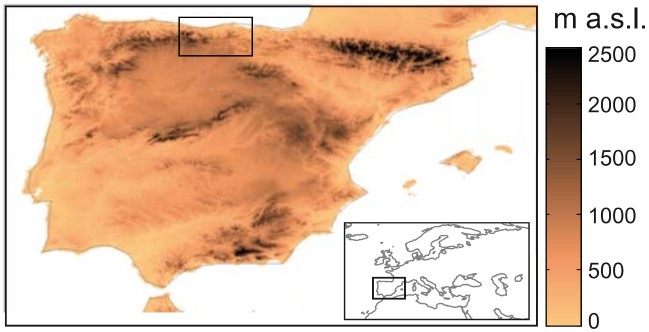
Santander Meteorology Group:
Universidad de Cantabria
CSIC
Dpto. Matemática Aplicada y Ciencias de la Computación
Instituto de Física de Cantabria

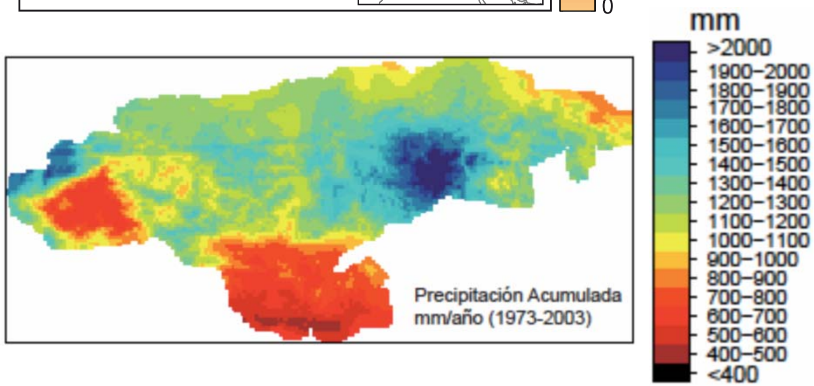
First CLIM-RUN Workshop on Climate Services. Trieste, 15-19 October 2012

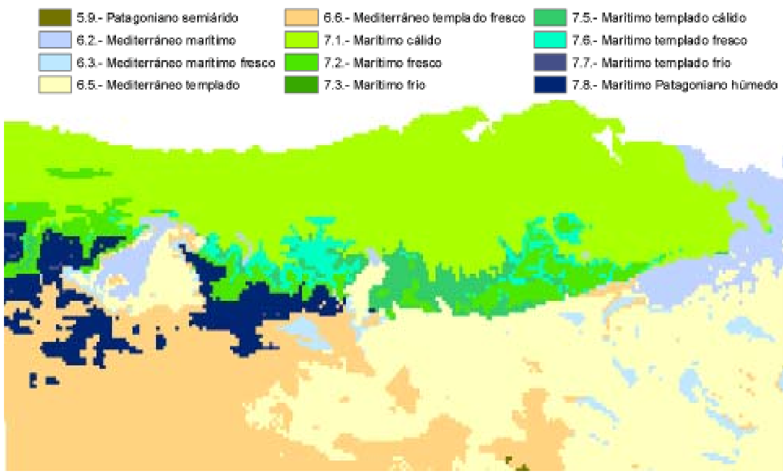
Outline

- ***Introduction to statistical downscaling***
- ***Techniques:*** ***Weather typing, transfer functions and weather generators.***
- ***Validation in perfect model conditions***
  - **Accuracy**
  - **Observed-simulated distributional consistency.**
  - **Stationarity/robustness under climate change conditions.**
- ***The statistical downscaling portal***

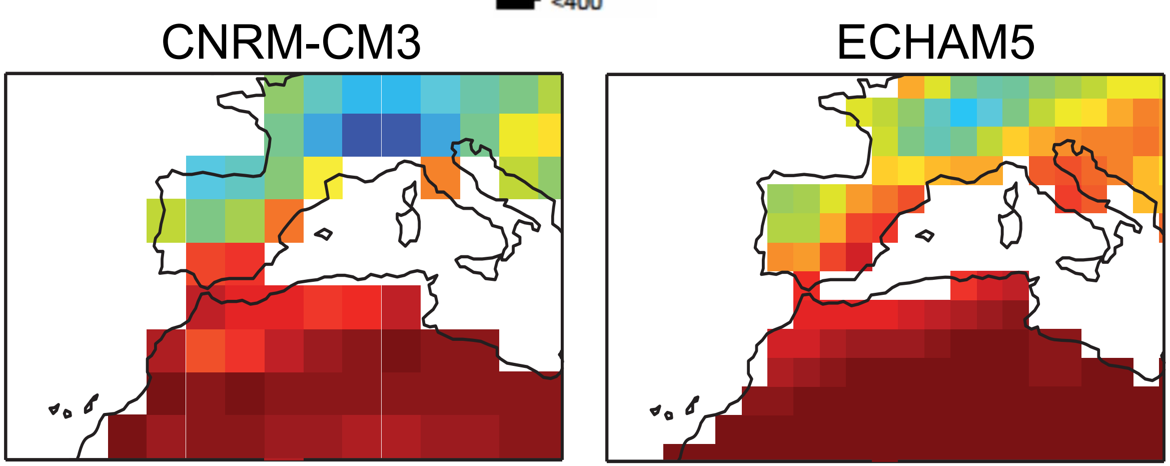
Introduction

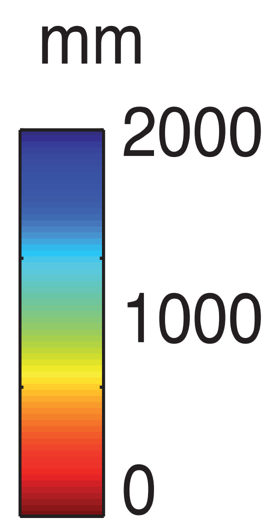






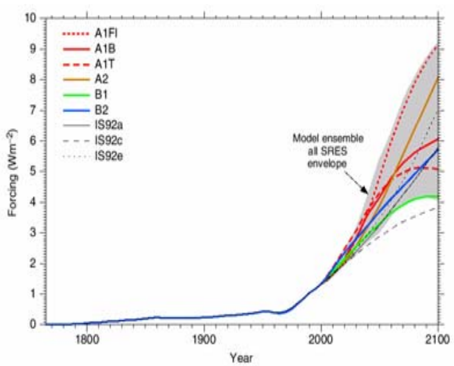
The spatial scales at which GCMs produce useful information do not match the scales that many users require.

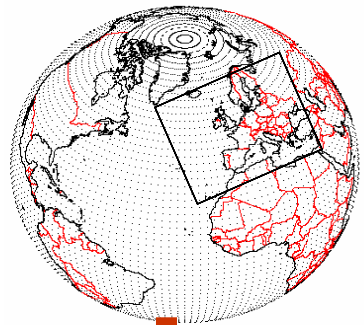




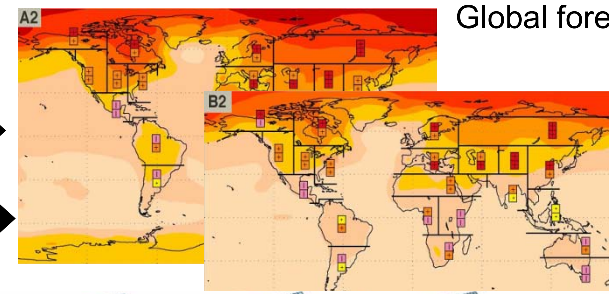
Introduction

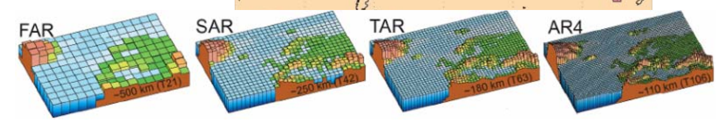
Scenarios



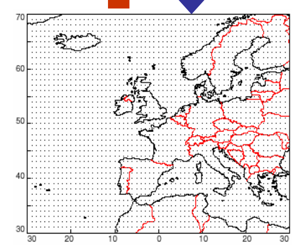


Global forecasts

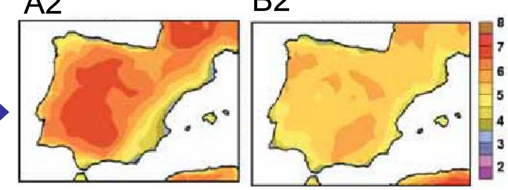




RCM



Dynamical Downscaling:
based on regional climate models (RCMs)



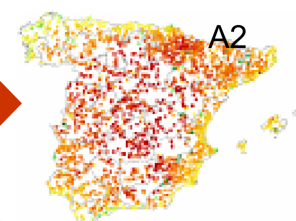
Historical library

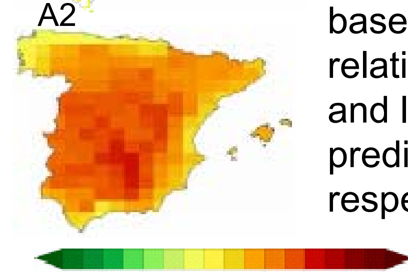
Station data

High resolution gridded data

$Y = f(X;\theta)$

The statistical model parameters are fitted from observed or modeled data.

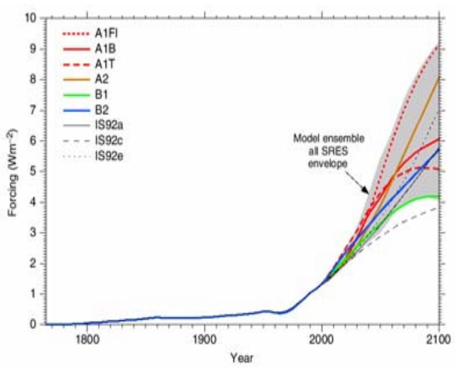


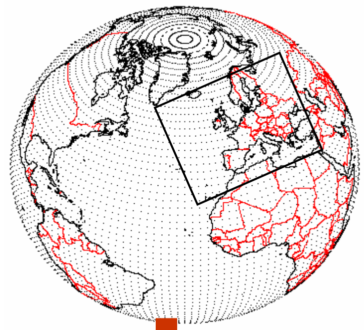


Statistical Downscaling:
based on empirical relationships between large and local scale variables, predictors and predictands respectively.

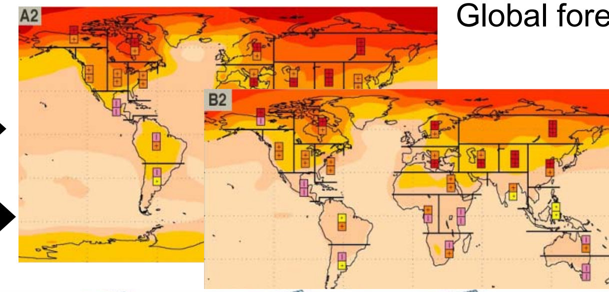
Introduction

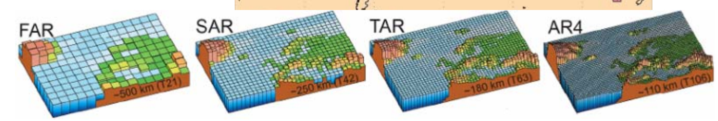
Scenarios





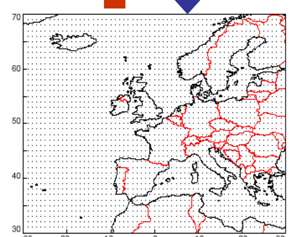
Global forecasts



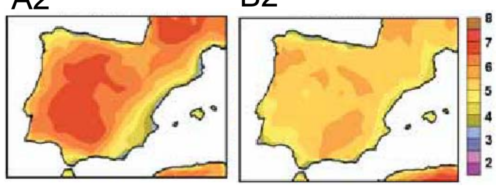


RCM

Dynamical Downscaling:
based on regional climate models (RCMs)



A2 B2



Historical library

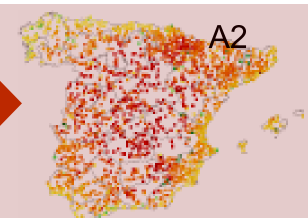
Station data

Gridded data (**20 km**)

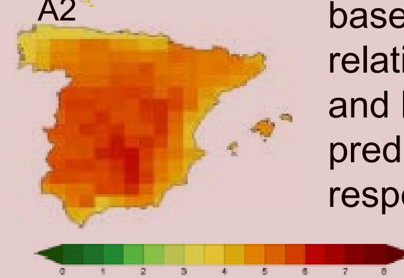
$\mathbf{Y} = f\,(\mathbf{X};\theta)$

The statistical model parameters are fitted from observed or modeled data.

A2



A2



Statistical Downscaling:
based on empirical relationships between large and local scale variables, predictors and predictands respectively.

Main advantages:

- ***Less computationally intensive***
- ***SD can be applied to non-climate predictands (e.g. FWI)***

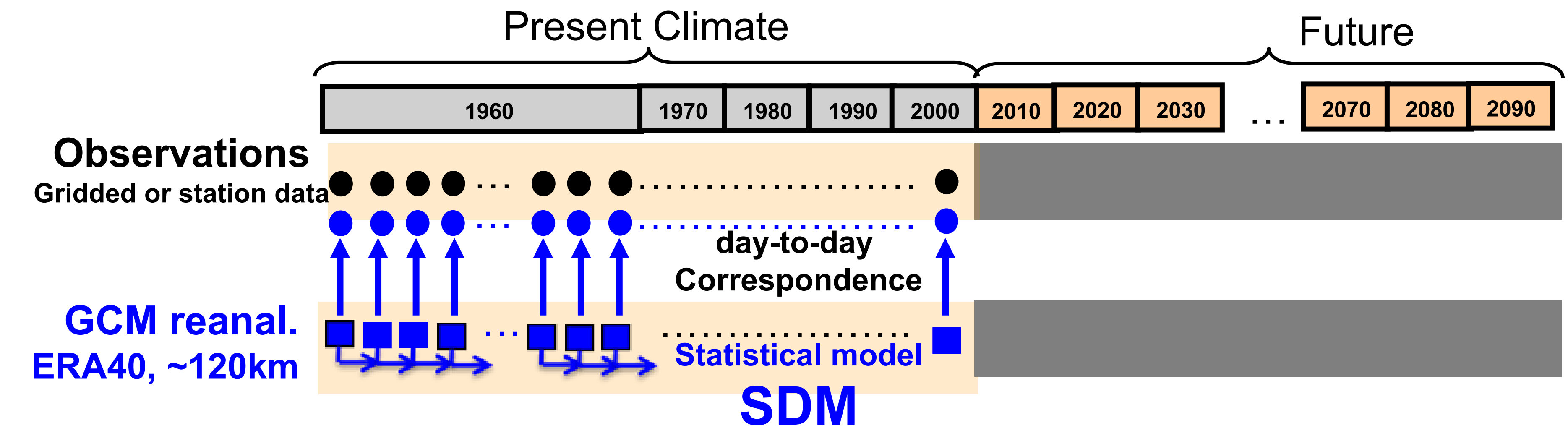
$$\mathbf{Y} = f\,(\mathbf{X};\theta)$$

Perfect Prognosis (Perfect Prog): uses only observations (large scale and local) to train the statistical model, which is later applied to the GCM output, assuming it provides perfect (observed-like) large scale fields.

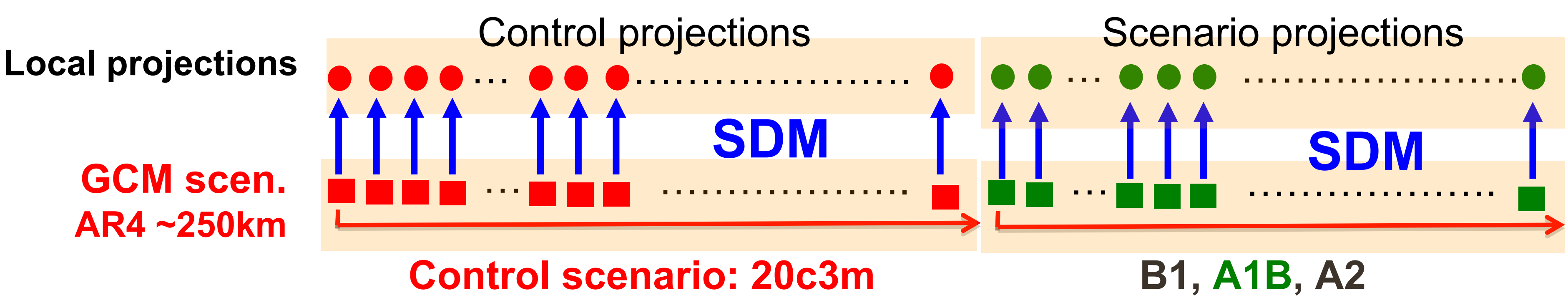
Model Output Statistics (MOS): GCM output is directly related to locally observed variables to train a statistical model, which is later applied to future GCM forecasts. Common in weather forecasting.

The statistical downscaling portal is based on the Perfect Prog idea of developing a statistical model independent of model forecasts.

Statistical Downscaling: Perfect Prog.



- **PROBLEM 1: Choosing consistent predictors:**
- **PROBLEM 2: Stationarity/robustness: SDM SDM**



Outline

- *Introduction to statistical downscaling*
- ***Techniques:*** ***Weather typing, transfer functions and weather generators.***
- ***Validation in perfect model conditions***
  - **Accuracy**
  - **Observed-simulated distributional consistency.**
  - **Stationarity/robustness under climate change conditions.**
- ***The statistical downscaling portal***

Empirical-statistical
Three main classes

Weather Generators

Trained on long term time series and atmospheric re-analysis data

Conditioned by GCM parameters to capture low frequency variance

Transfer Functions

Trained on time series that spans range of variability, and atmospheric re-analysis data

Residual local scale variance added stochastically

Index / analogues

Requires long term data sets and uses weather typing or historical analogues

Source:
Bruce Hewitson

Climate System Analysis Group
University of Cape Town
www.csag.uct.ac.za

- **Transfer-Function Approaches**
- **Weather typing**

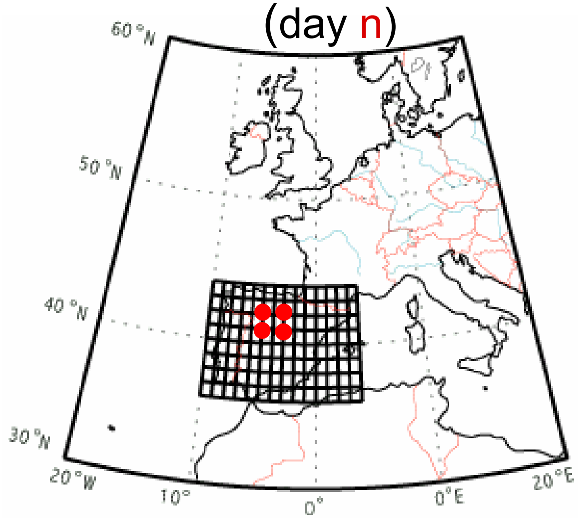
	Advantages	Shortcomings
Linear Regression GLMs		
Neural Networks		
Analogs		
Weather Typing (k-means, SOM, etc.)		

1. Transfer functions: Linear regression

Huth (2002 and 2004)

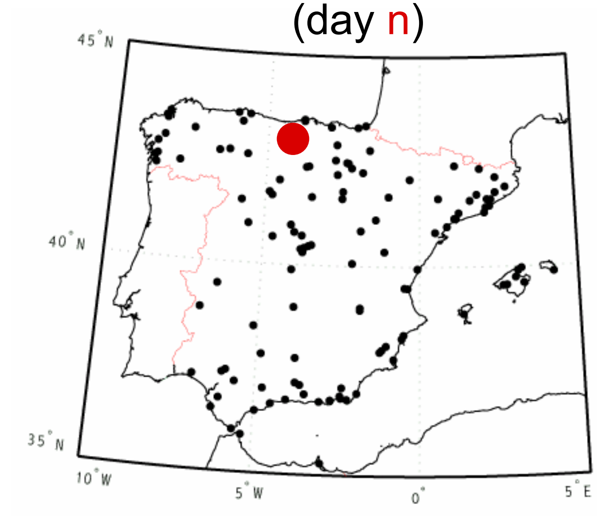
Large scale pattern or predictor

(day n)



Local data or predictand

(day n)



$$\left.\begin{array}{l}(\mathbf{T}(1000\text{ mb}),\ldots,\ \mathbf{T}(500\text{ mb});\\ \mathbf{Z}(1000\text{ mb}),\ldots,\ \mathbf{Z}(500\text{ mb});\\ \quad\ldots\ldots;\\ \mathbf{H}(1000\text{ mb}),\ldots,\ \mathbf{H}(500\text{ mb})) = \boxed{X_n}\end{array}\right\}$$

$$\boxed{Y_n}$$

Linear Regression

$$\boxed{\hat{Y}_n = a\,X_n + b}$$

Redundancy: EOF & Clustering

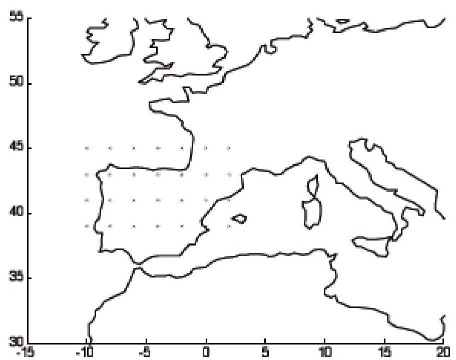
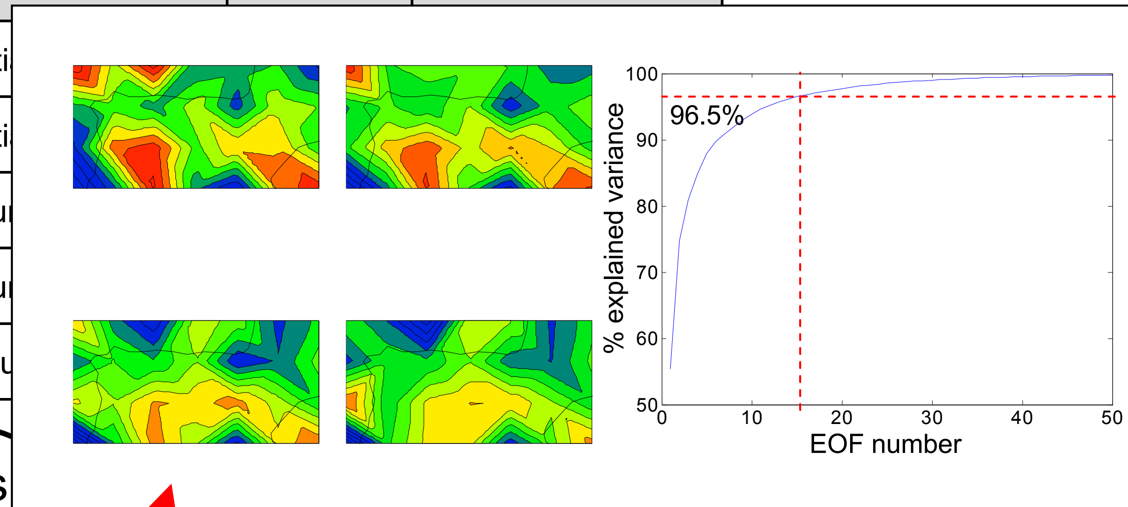


Figura 3. Área de estudio y rejilla utilizada para definir los predictores de los métodos de downscaling estadístico.

Variable	Level	Time
Geopotenti		
Geopotenti		
Temperatu		
Temperatu		
Relative Hu		

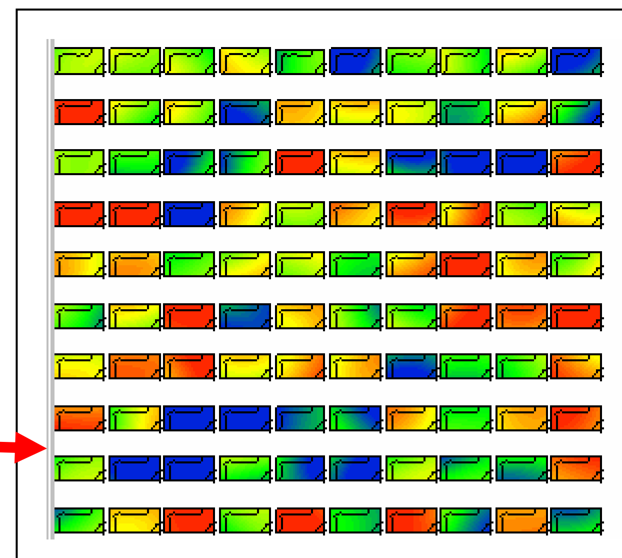
140 parameters (5 variables



Redundancy (correlation):

- Principal Components

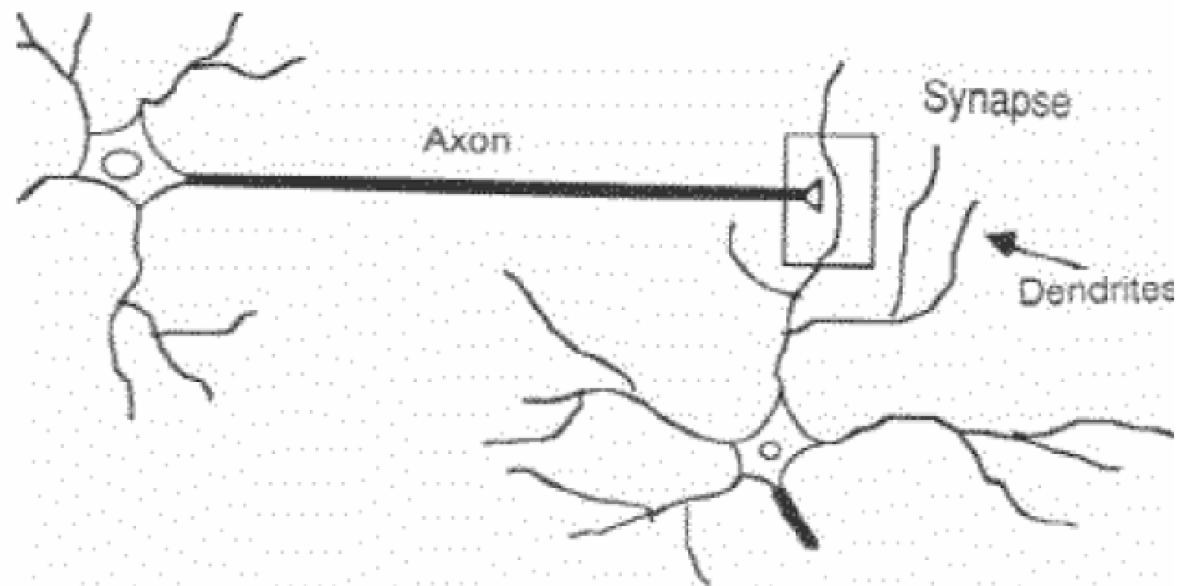
- Clustering



1. Transfer functions: Neural Networks

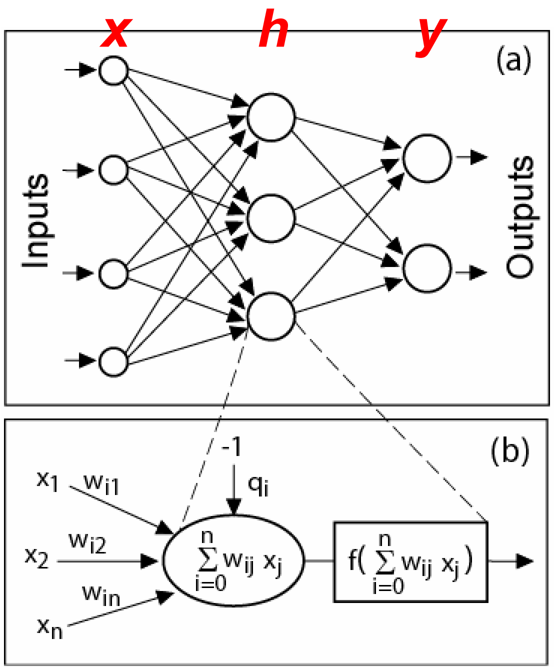
Artificial Neural Networks are inspired in the structure and functioning of the **brain**, which is a collection of **interconnected neurons** (the simplest computing elements performing information processing):

- ✓ Each neuron consists of a cell body, that contains a cell **nucleus**.
- ✓ There are number of fibers, called **dendrites**, and a single long fiber called **axon** branching out from the cell body.
- ✓ The axon connects one neuron to others (through the dendrites).
- ✓ The connecting junction is called **synapse**.

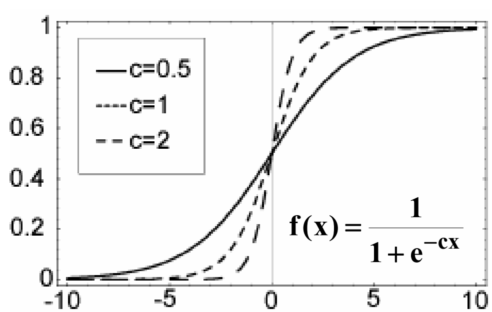


1. Transfer functions: Neural Networks

Trigo and Palutikov (1999); Huang et al (2004)



The neural activity (output) is given by a non-linear function.



$$y_j = f\Big(\sum_i \beta_{ji} \underbrace{f\Big(\sum_k \alpha_{ik}\, x_{kp}\Big)}_{h_i}\Big)$$

Inputs $\{x_{1p}, \ldots, x_{mp}\}$

Outputs $\{y_{1p}, \ldots, y_{np}\}$

$$\begin{aligned} E(\alpha,\beta) &= \frac{1}{2}\sum_{j,p}\Big(y_{jp} - f\Big(\sum_i \beta_{ji} f\Big(\sum_k \alpha_{ik}\, x_{kp}\Big)\Big)\Big)^2 \\ &= \sum_p \|\mathbf{y}_p - f(\beta^T f(\alpha^T\, \mathbf{x}_p))\| \end{aligned}$$

Gradient descent

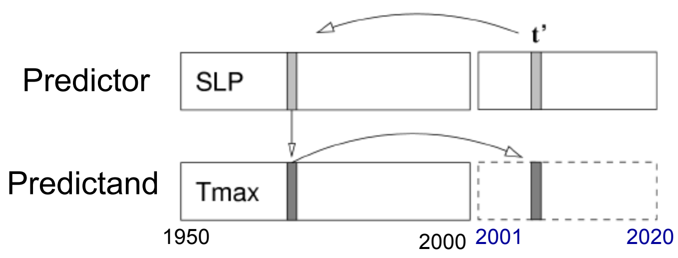
$$\Delta\beta_{ik} = -\eta \frac{\partial E}{\partial \beta_{ik}};\quad \Delta\alpha_{kj} = -\eta \frac{\partial E}{\partial \alpha_{kj}},$$

1. Init the neural weights with random values
2. Select the input and output data and train it
3. Compute the error associated with the output and update the neural weight according to these values.

2. Weather typing: Analogs

Zorita and von Storch (1999)

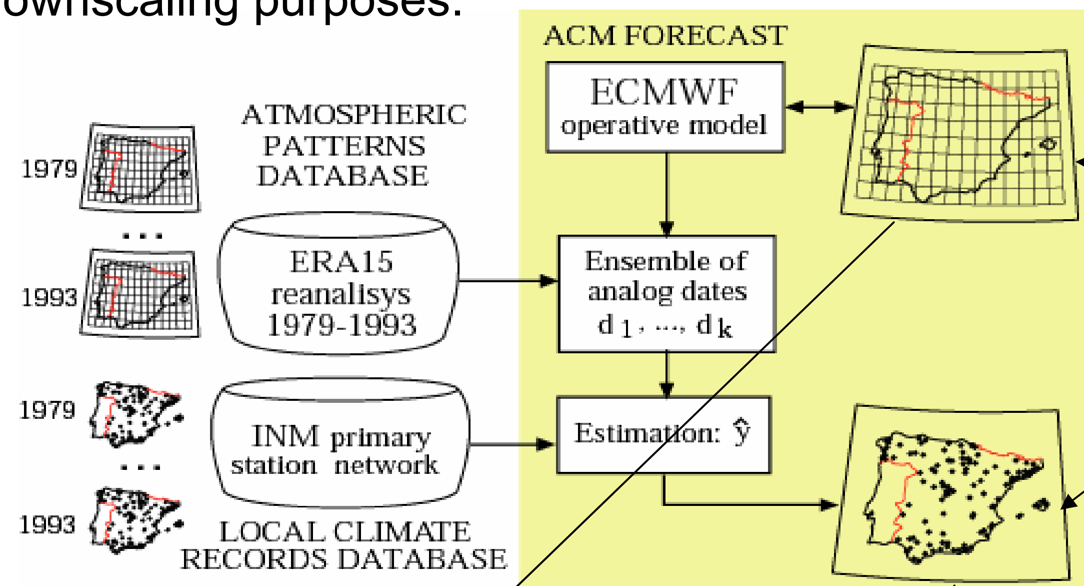
The analog method (nearest neighbour) was introduced by E. Lorenz (1969) and has been considered in different applications, in particular in statistical downscaling purposes.

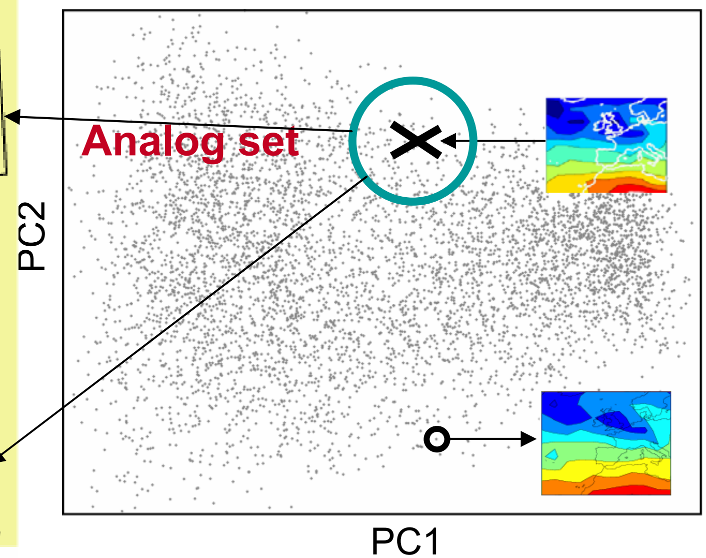


2. Weather typing: Analogs

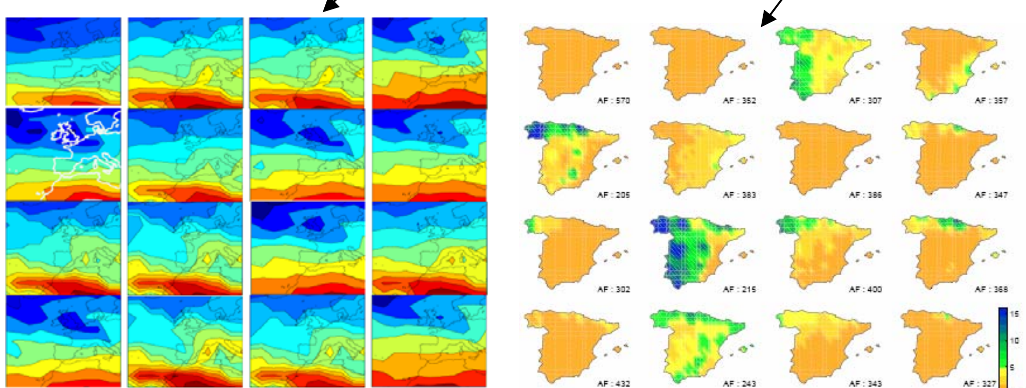
Zorita and von Storch (1999)

The analog method (nearest neighbour) was introduced by E. Lorenz (1969) and has been considered in different applications, in particular in statistical downscaling purposes.





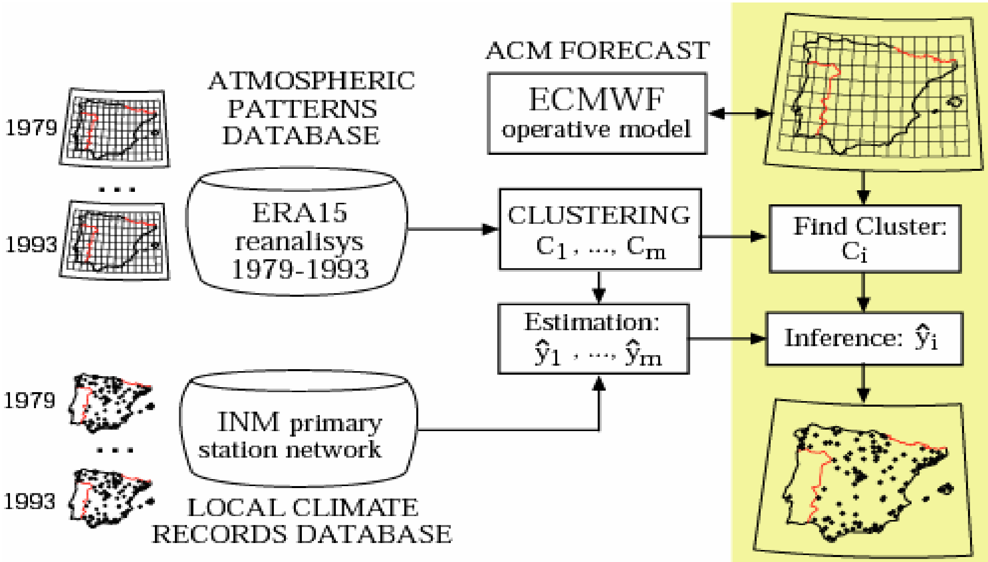
Euclidean distance



- Mean
- Median
- Frequency

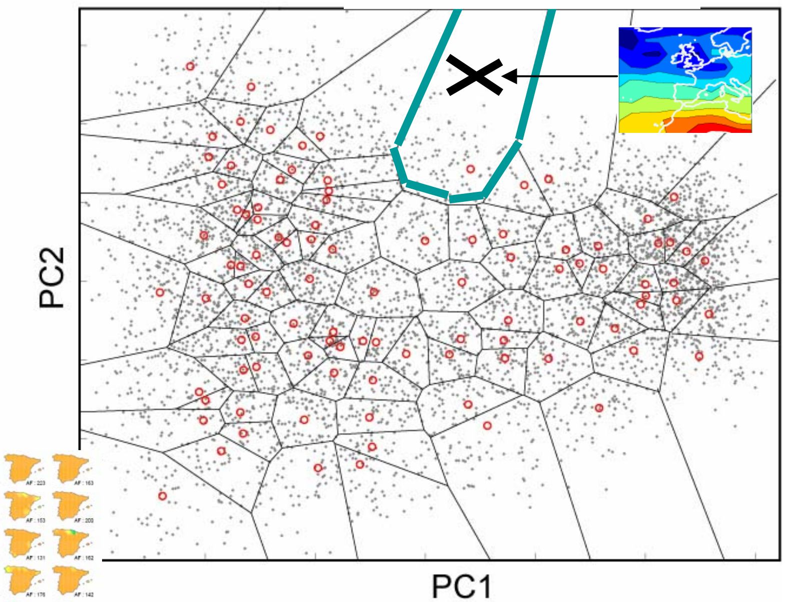
2. Weather typing: K-means

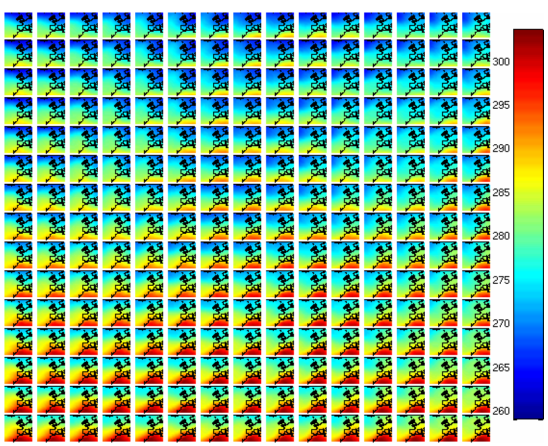
Gutierrez et al (2004), Huth (2010)

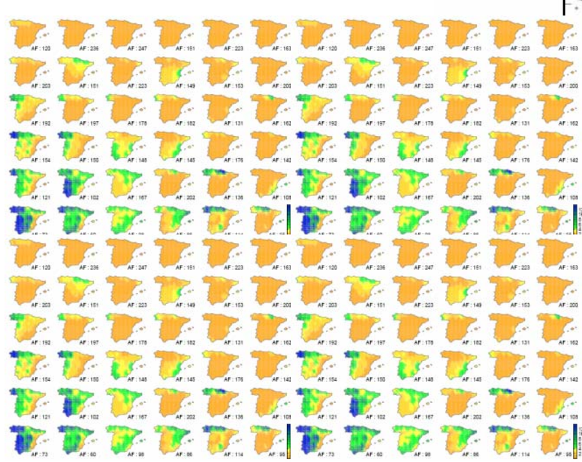


K-means clustering algorithm

centroides: $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$

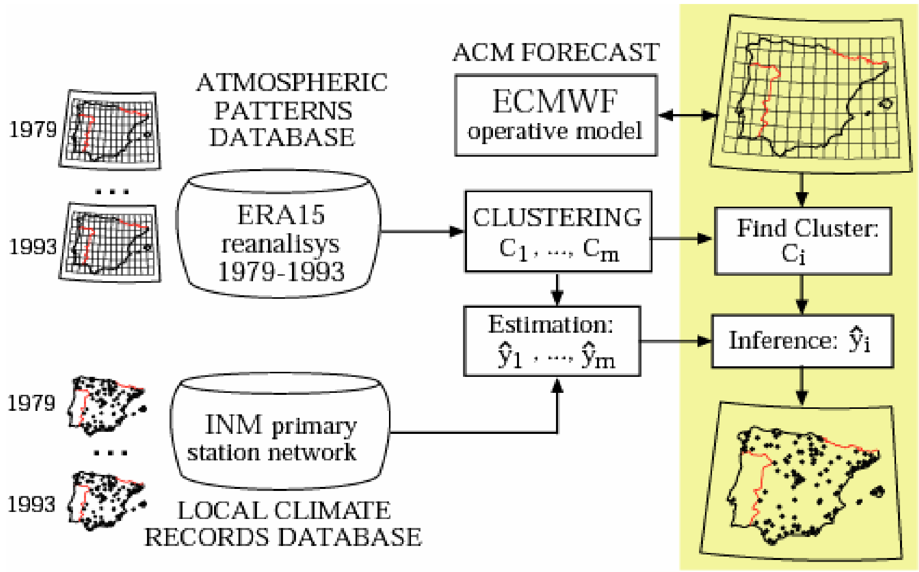


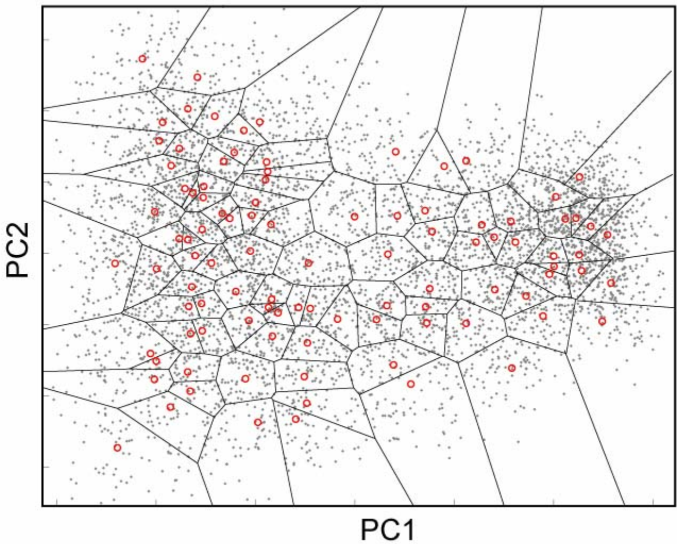




A particular day (X) is assigned to the closest cluster, Ck P(y>u|Ck)

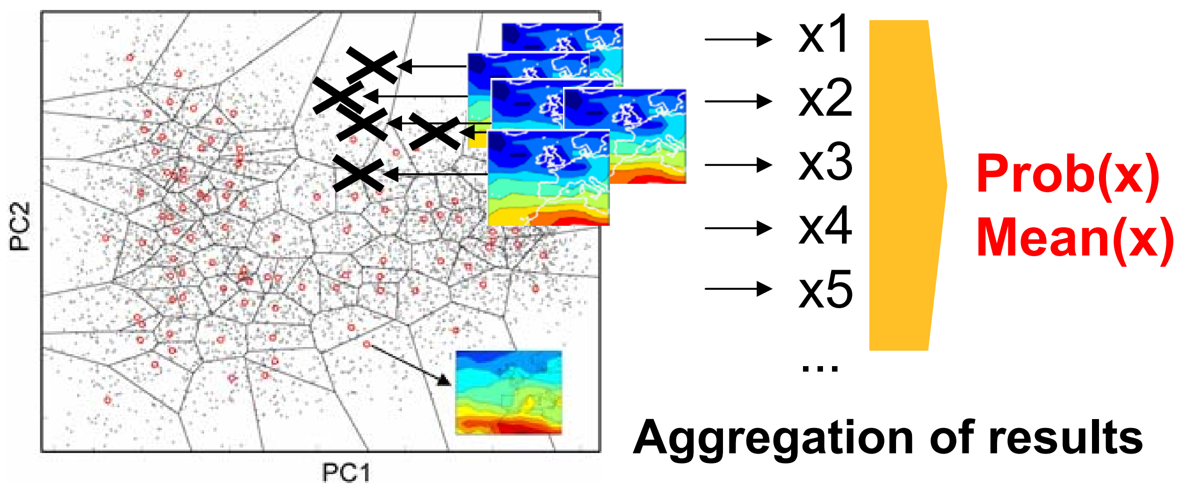
2. Weather typing: K-means





$$P_{forecast}(\text{precip} > u) = \Sigma_{Ck}\, P(\text{precip} > u \mid Ck)\ P_{forecast}(Ck)$$

The application to an EPS requires applying the method to each of the ensemble members:



- **Transfer-Function Approaches**
- **Weather typing**

	Advantages	Shortcomings
Linear Regression **GLMs**	Simple Easy to interpret	Linear assumption Spatially inconsistent Selection of predictors
Neural Networks	Nonlinear “Universal” interpolator	Complex blackbox-like Optimization required Selection of predictors
Analogs	Nonlinear Spatial consistency	Algorithmic. No model. Difficult to interpret
Weather Typing **(k-means, SOM, etc.)**	Nonlinear Easy to interpret Spatial consistency	Loss of variance Problem with borders

Domain selection

Monthly Weather Review (2004)

Clustering Methods for Statistical Downscaling in Short-Range Weather Forecasts

J. M. Gutiérrez and A. S. Cofiño
Department of Applied Mathematics, E.T.S.I. Caminos, University of Cantabria, Santander, Spain
R. Cano
Instituto Nacioal de Meteorología, CMT/CAS, Santander, Spain
M. A. Rodríguez
Instituto de Física de Cantabria, CSIC, Santander, Spain

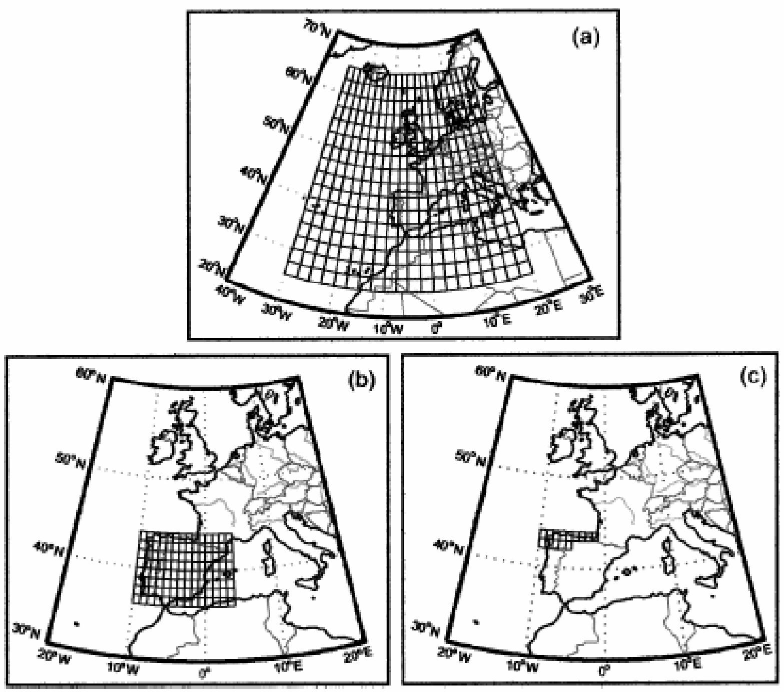


Fig. 2. Maps of the model grid domains used in this study: (a) large-scale macro-β grid considered for model 1, (b) meso-α grid covering the peninsula for model 2, and (c) meso-β model 3 grid for the northern basin. (Twelve different grids were considered, one for each basin of the Iberian Peninsula. For the sake of clarity only the north basin is shown.)

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

$$\mathbf{x} = (\mathbf{x}_{06}, \mathbf{x}_{30}).$$

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

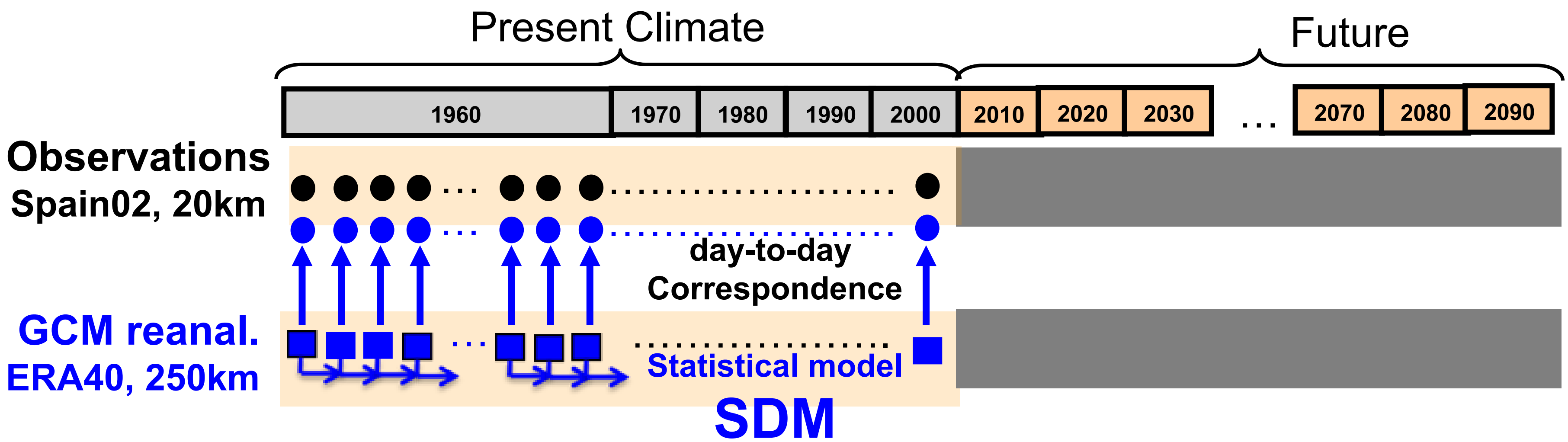
Annual spatial averaged RSA for precipitation

Forecast	Method	>0.1 mm 1	2	3
D + 1	Analog	0.647	0.750	0.791
	Cluster	0.538	0.682	0.744
	WCluster	0.597	0.733	0.783
D + 2	Analog	0.633	0.737	0.771
	Cluster	0.523	0.669	0.716
	WCluster	0.588	0.711	0.763
D + 3	Analog	0.572	0.693	0.734
	Cluster	0.449	0.640	0.678
	WCluster	0.542	0.680	0.726

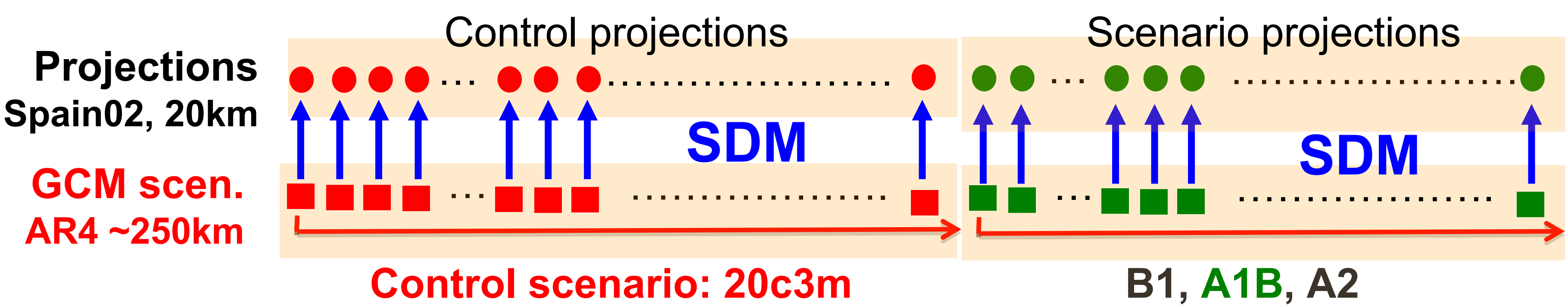
Outline

- ***Introduction to statistical downscaling***
- ***Techniques:*** ***Weather typing, transfer functions and weather generators.***
- ***Validation in perfect model conditions***
  - **Accuracy**
  - **Observed-simulated distributional consistency.**
  - **Stationarity/robustness under climate change conditions.**
- ***The statistical downscaling portal***

Statistical Downscaling: Perfect Prog.



- **PROBLEM 1: Choosing consistent predictors:**
- **PROBLEM 2: Stationarity/robustness: SDM SDM**



Spanish Regional Climate Change Program: PNACC 2012

Journal of Climate 2012 ; e-View
doi: http://dx.doi.org/10.1175/JCLI-D-11-00687.1

Reassessing statistical downscaling techniques for their robust application under climate change conditions

J. M. Gutiérrez,* D. San-Martín, S. Brands, R. Manzanas, and S. Herrera

Instituto de Fisisca de Cantabria (UNICAN-CSIC), Santander, Spain

INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (2010)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.2256

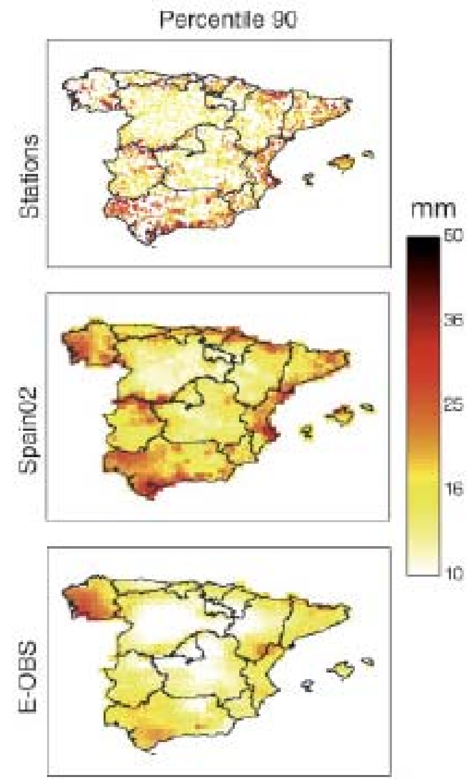
RMetS
Royal Meteorological Society

Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02)

S. Herrera,[a]* J. M. Gutiérrez,[a] R. Ancell,[b] M. R. Pons,[b] M. D. Frías[c] and J. Fernández[c]
[a] *Instituto de Física de Cantabria, CSIC-University of Cantabria, Avenida de los Castros s/n, Santander, Spain*
[b] *Agencia Estatal de Meteorología (AEMET), Santander, Spain*
[c] *Department of Applied Mathematics and Computer Science, Universidad de Cantabria, Santander, Spain*

Precipitación: 2756 Estaciones

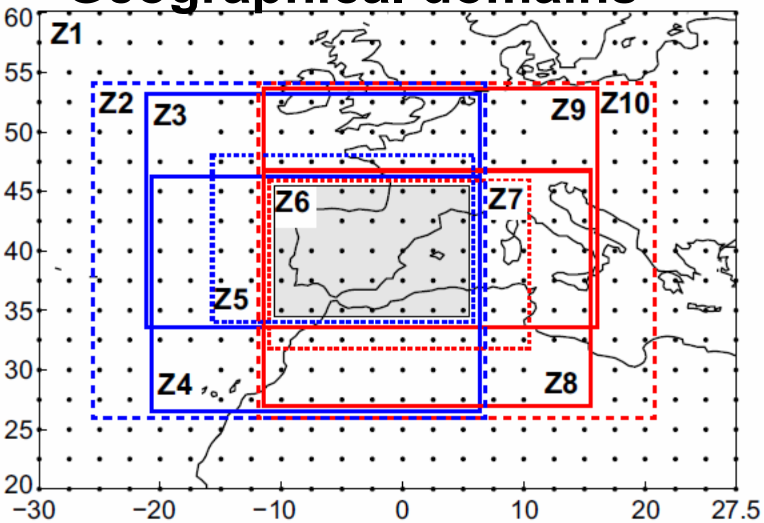
Temperatura: 864 Estaciones



Precipitation, min. and max. temperatures

Calibration and Selection of SD Methods

Geographical domains



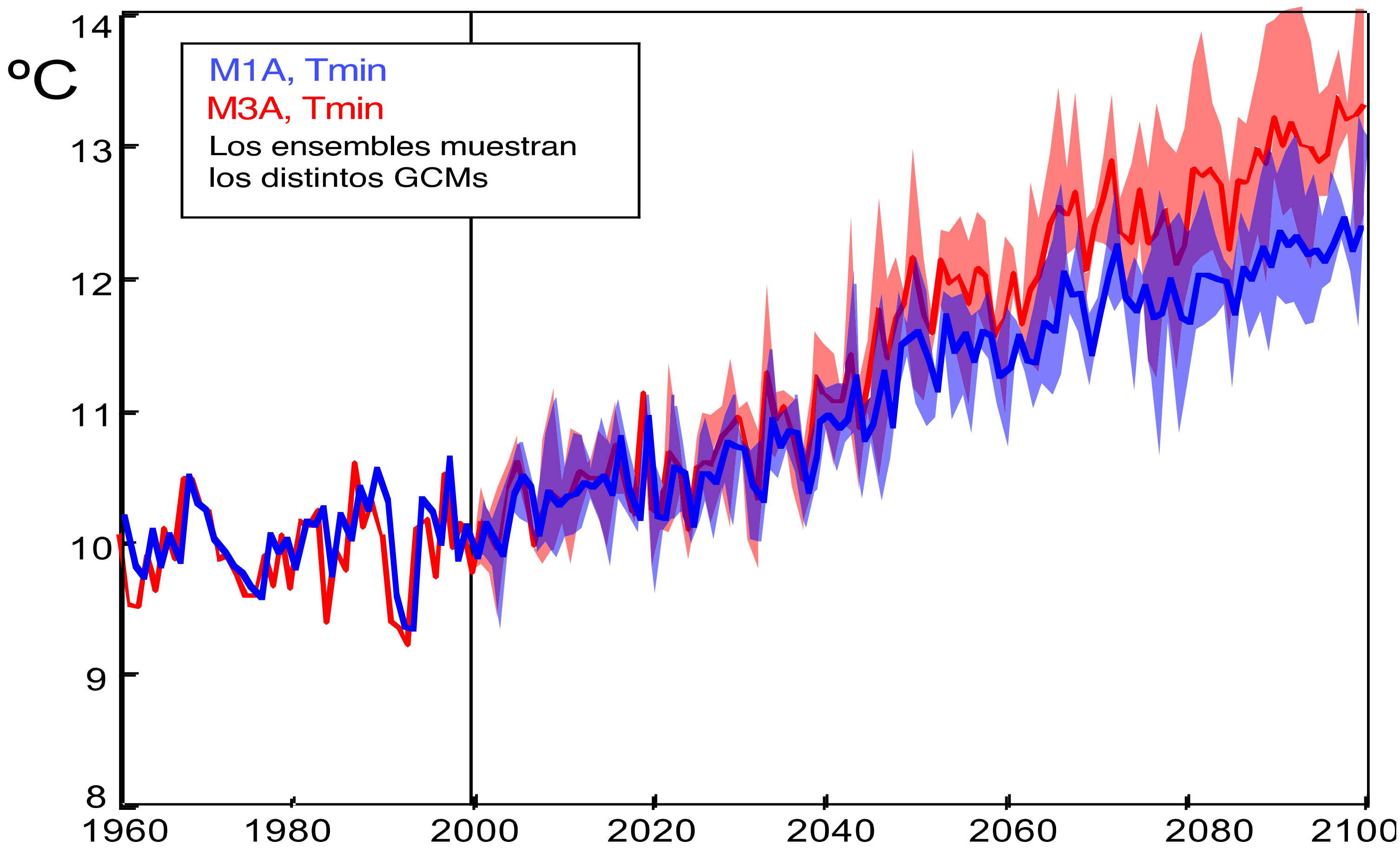
Consistent Predictors

Code	*Predictor variables*
P1–P1d	SLP, T850, Q850, U500, V500
P2–P2d	SLP, T850, Q850, Z500
P3–P3d	SLP, T850, Q850
P4–P4d	SLP, T850
P5	SLP, T2m
P6–P6d	T850
P7	T2m
P8	Tx
P9	Tn

SD Methods

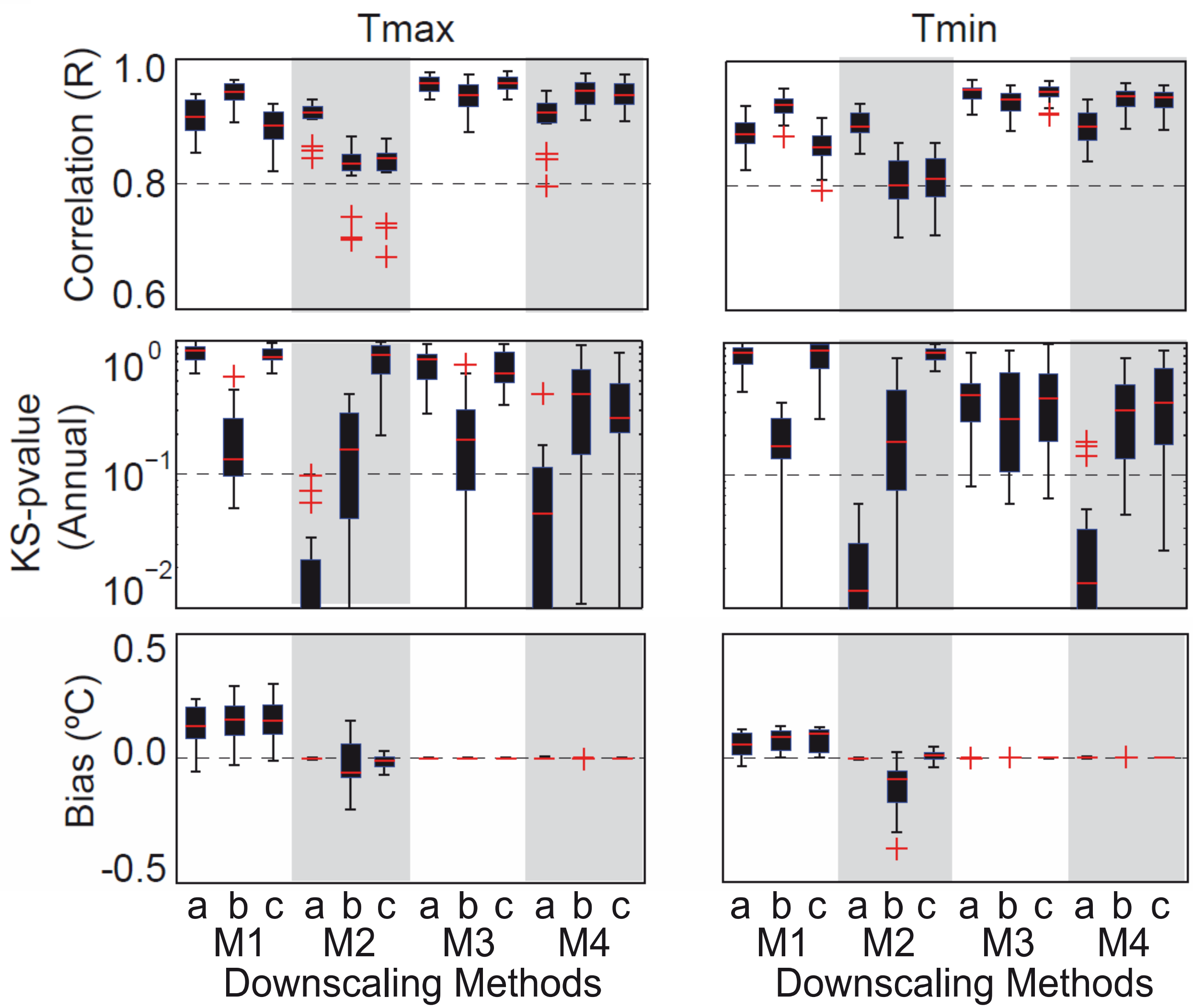
Code	*Type*	*Method and Predictor Field*
M1a	AM	Nearest neighbour (1 analogue)
M1b	AM	Mean of 5 neighbours
M1c	AM	One out of 15 neighbors, random selection
M2a	WT	100 WTs (k-means), mean of the observations
M2b	WT	100 WTs (k-means), random selection
M2c	WT	100 WTs (k-means), simulation from gaussian distribution
M3a	LR	Linear regression with n PCs (95% variance)
M3b	LR	Local predictor values in the nearest grid box
M3c	LR	15 PCs + Nearest grid box
M4a	LR-WT	M3c conditioned on 10 WTs (k-means)
M4b	LR-WT	M3b conditioned on 10 WTs (k-means)
M4c	LR-WT	M3b (T,Q) conditioned on 10 WTs (SLP)

The lack of robustness can lead to wrong future projections. In the example below the difference between two SD methods is much larger than inter-GCM variability.



Calibration and Selection of SD Methods

Code	Type	Method and Predictor Field
M1a	AM	Nearest neighbour (1 analogue)
M1b	AM	Mean of 5 neighbours
M1c	AM	One out of 15 neighbors, random selection
M2a	WT	100 WTs (k-means), mean of the observations
M2b	WT	100 WTs (k-means), random selection
M2c	WT	100 WTs (k-means), simulation from gaussian distribution
M3a	LR	Linear regression with n PCs (95% variance)
M3b	LR	Local predictor values in the nearest grid box
M3c	LR	15 PCs + Nearest grid box
M4a	LR-WT	M3c conditioned on 10 WTs (k-means)
M4b	LR-WT	M3b conditioned on 10 WTs (k-means)
M4c	LR-WT	M3b (T,Q) conditioned on 10 WTs (SLP)



Differences in terms of accuracy, distributional similarity and robustness are much larger among the predictors than among the geographical domains, with optimum results for 2T, as compared to T850, and small geographical domains.

With an appropriate predictor selection, all methods exhibits a good performance in terms of correlation. However some of them suffers from significant distributional inconsitencies indicating that could not be suitable for climate change applications.

Geography Compass 5/6 (2011): 275–300, 10.1111/j.1749-8198.2011.00425.x

Climate Scenario Development and Applications for Local/Regional Climate Change Impact Assessments: An Overview for the Non-Climate Scientist

Part I: Scenario Development Using Downscaling Methods

Julie A. Winkler[1]*, Galina S. Guentchev[2], Perdinan[1], Pang-Ning Tan[3], Sharon Zhong[1], Malgorzata Liszewska[4], Zubin Abraham[3], Tadeusz Niedźwiedź[5] and Zbigniew Ustrnul[6]

[1]*Department of Geography, Michigan State University*
[2]*UCAR CLIVAR Postdocs Applying Climate Expertise (PACE) Program*
[3]*Department of Computer Science and Engineering, Michigan State University*
[4]*Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw*
[5]*Department of Climatology, University of Silesia*
[6]*Department of Climatology, Jagiellonian University*

Part II: Considerations When Using Climate Change Scenarios

Julie A. Winkler[1]*, Galina S. Guentchev[2], Malgorzata Liszewska[3], Perdinan[1] and Pang-Ning Tan[4]

[1]*Department of Geography, Michigan State University*
[2]*UCAR CLIVAR Postdocs Applying Climate Expertise (PACE) Program*
[3]*Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw*
[4]*Department of Computer Science and Engineering, Michigan State University*

Statistical vs. Dynamical Downscaling

Consideration	Why important?	Regional climate models (dynamic downscaling)	Analogs (empirical-dynamic downscaling)	Perfect Prog transfer functions with circulation	Transfer functions (disaggregation downscaling)	Spatial interpolation (disaggregation downscaling)	Weather generators
Spatial resolution	For some impact assessments, it is essential to capture the influence of local site characteristics on climate, whereas for other applications regional-scale variations in climate are sufficient	Not appropriate for point (station) scale; multiple-nested RCMs can be used to obtain scenarios on a fine (1–10 km) mesh grid; single nested RCMs are usually used to obtain scenarios on a 25–50 km grid	Can be used for a range of spatial scales, but most often used to obtain scenarios at a point (station) scale	Can be used for a range of spatial scales, but most often used to obtain scenarios at the point (station) scale	Can be used for a range of spatial scales. Frequently used to obtain scenarios for a station or for grid points (the latter requires that observed gridded fields of the climate variable are available)	Not appropriate for point (station) scale; frequently used to obtain scenarios at a fine (1–10 km) resolution grid	Point (station) scale
Temporal resolution	Scenarios often serve as input to ecological, process, or activity models. The time step used in these models often determines the temporal resolution required for the climate scenarios	Sub-daily	Typically daily	Scenarios can be generated at sub-daily, daily, monthly and longer temporal aggregations	Most appropriate for monthly or longer temporal aggregations	Most appropriate for monthly or longer temporal aggregations	Daily (some attempts have been made to develop weather generators for sub-daily time steps)

Outline

- ***Introduction to statistical downscaling***
- ***Techniques:*** ***Weather typing, transfer functions and weather generators.***
- ***Validation in perfect model conditions***
  - **Accuracy**
  - **Observed-simulated distributional consistency.**
  - **Stationarity/robustness under climate change conditions.**
- ***The statistical downscaling portal***

ENSEMBLES Downscaling Portal (version 2)

http://ensembles-eu.metoffice.com

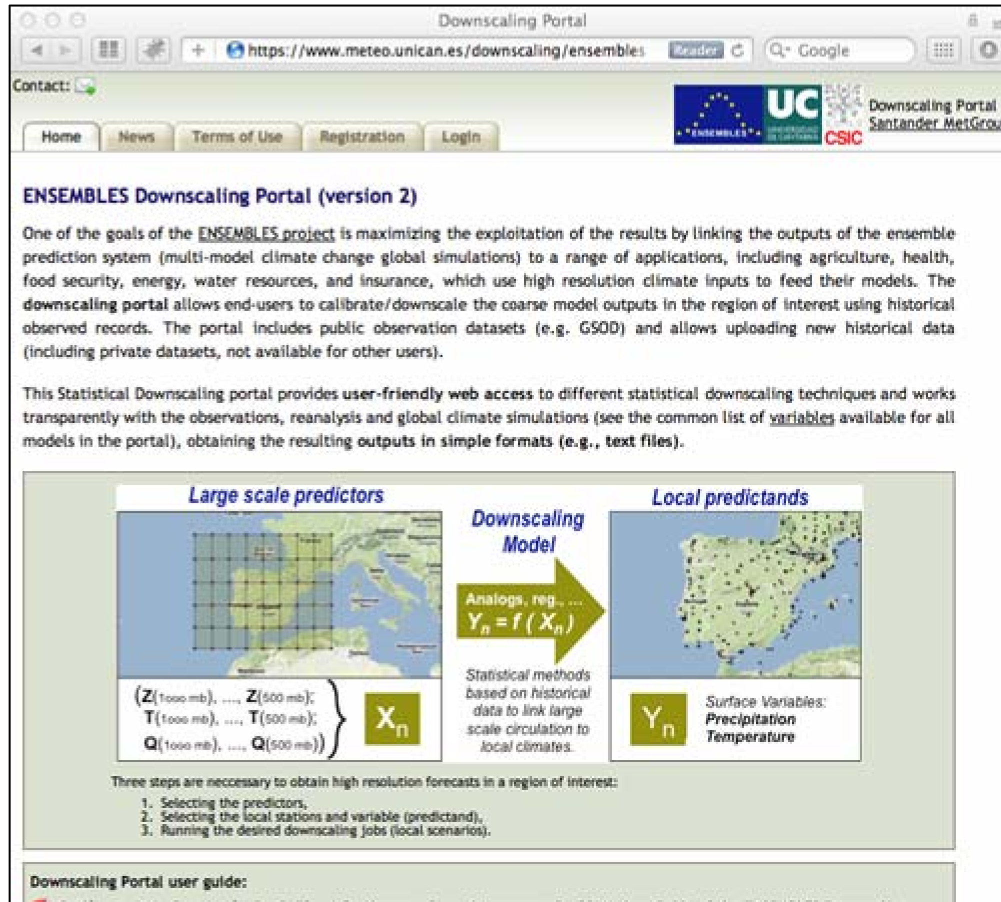
ENSEMBLES Project (2004-2009)

Develop an ensemble prediction system for climate change and linking the outputs to a range of applications.

- **RCM simulations.**
- **Statistical Downscaling.**
- **Gridded observations: E-OBS**

The statistical downscaling portal is a free tool for user-friendly downscaling.

http://www.meteo.unican.es/ensembles



Meteolab: **an open-source Matlab toolbox**

http://www.meteo.unican.es/en/software/meteolab

Currently, the ENSEMBLES datasets included in the portal contain only Climate Change Scenarios data. Data from seasonal experiments (multi-model simulations) will be included soon.

Observations:

- ECA stations + GSOD
- E-OBS 50km + Spain02
- E-OBS 25km

Reanalysis (global coverage):

- ERA40
- NCEP

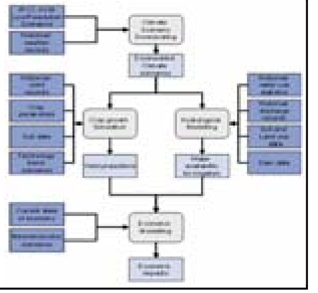
GCM scenarios (global coverage):

- ENSEMBLES Stream1 (CMIP3):
    - BCM2.0, CNRM-CM3, ECHAM5, ECHO-G, HADGEM, IPCM4
- ENSEMBLES Stream2:
    - CNRM-CM33, ECHAM5c, HADCM3C, HADGEM2, IPCMv2

Downscaling Portal: Follow-on Activities

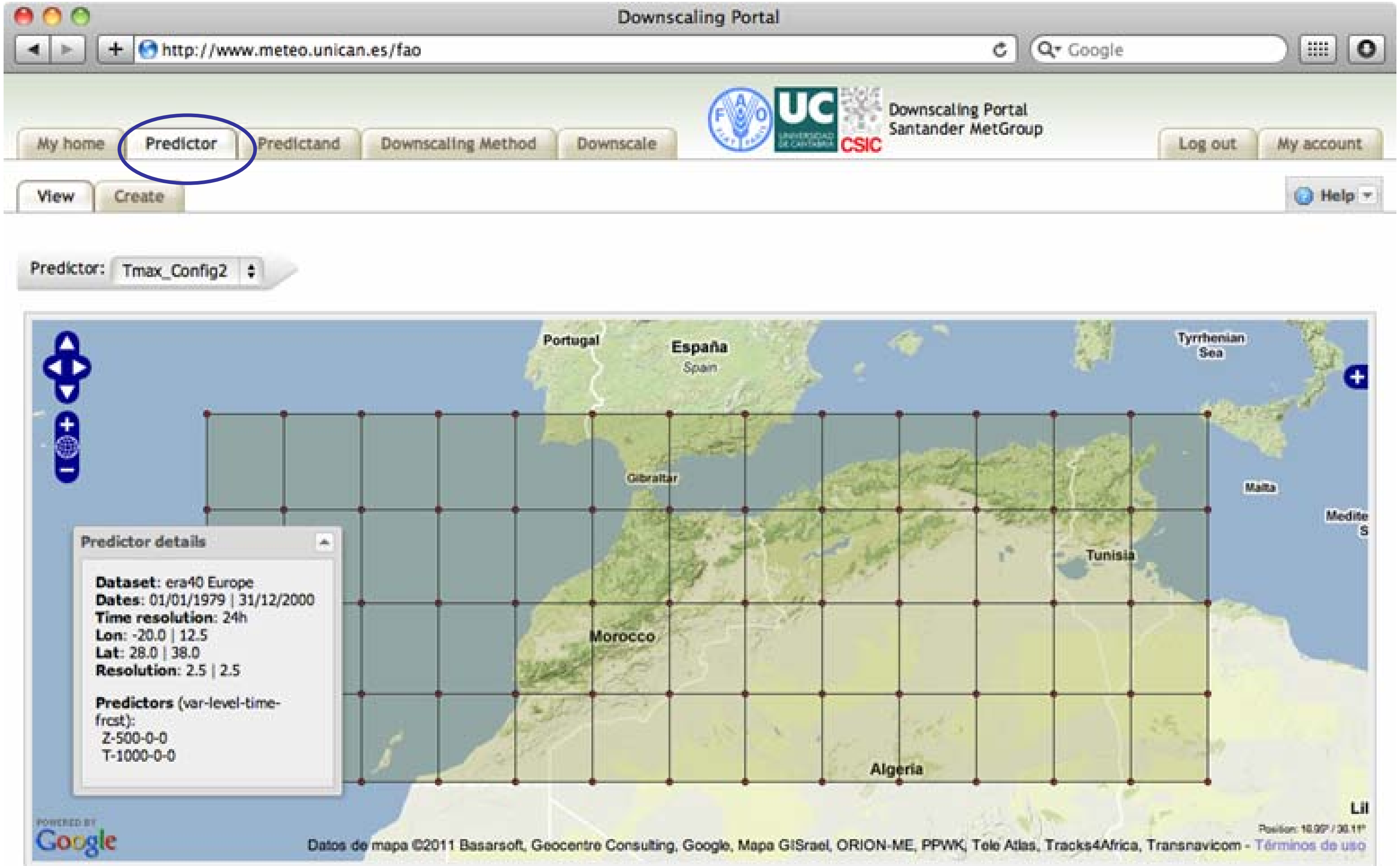
The activities started in ENSEMBLES have a follow on in several EU-funded and international projects, involving different impact communities, and dealing with different CORDEX-related activities.

- FUME: Impacts in forest **fires**
- QWeCI (Quantifying Weather & Climate Impacts on health in developing countries): Impacts in **health**
- CLIM-RUN: **Impacts** in tourism, energy, and natural hazards
- metafor: Appropriate **metadata** for GCMs and downscaling.
- MOSAICC: **Integration** with impact tools: crop + hydrology + economy



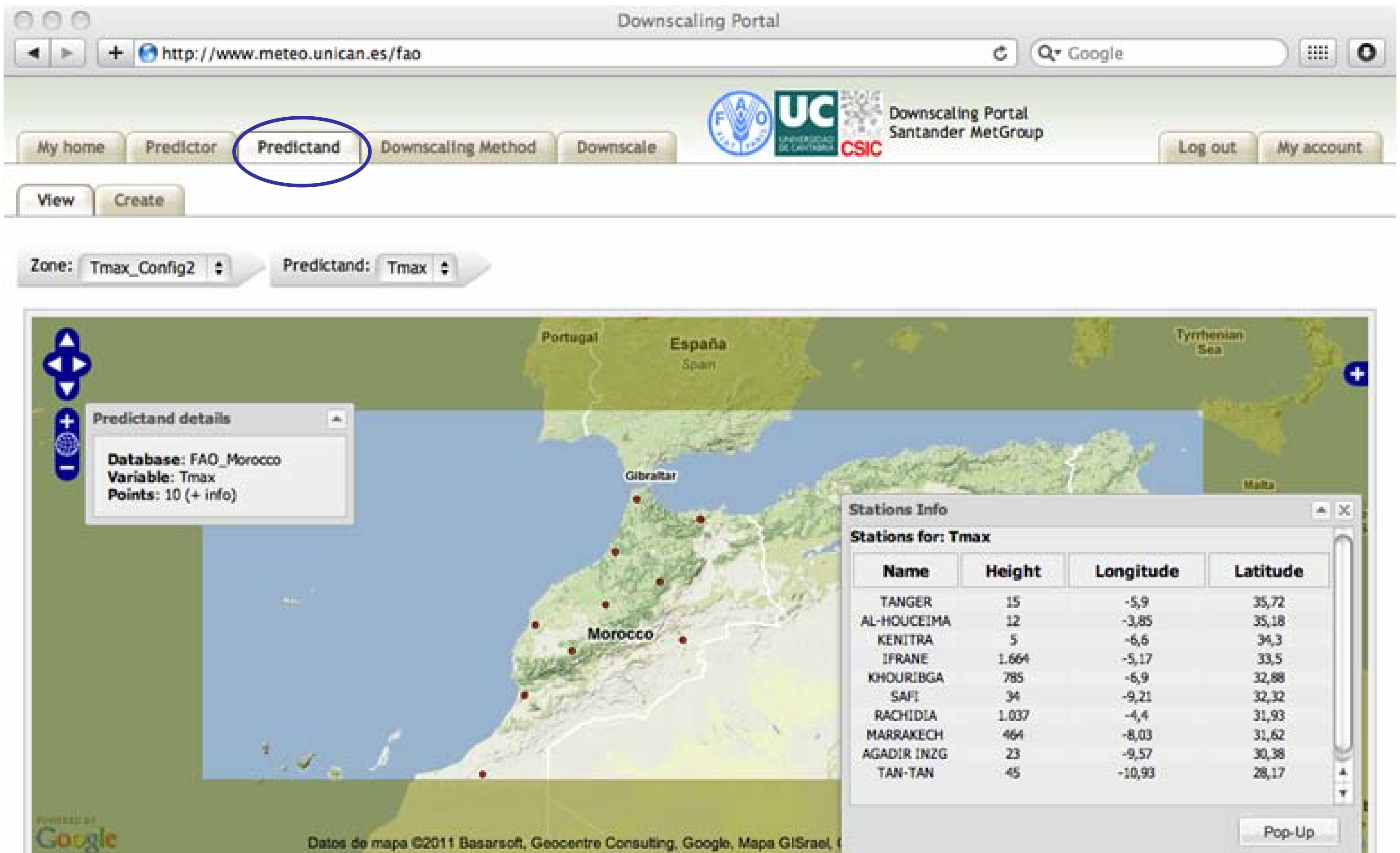
SD Portal: Predictors

The SDS Portal allows creating downscaling experiments selecting a region of interest and the predictors to be used (Z500 and T1000 in this example).



SD Portal: Local predictands

It also allows selecting a local variable of interest (e.g. max. Temp.) in a number of stations from any of the available historical datasets (in this case a dataset developed for the project *FAO_Marocco*).



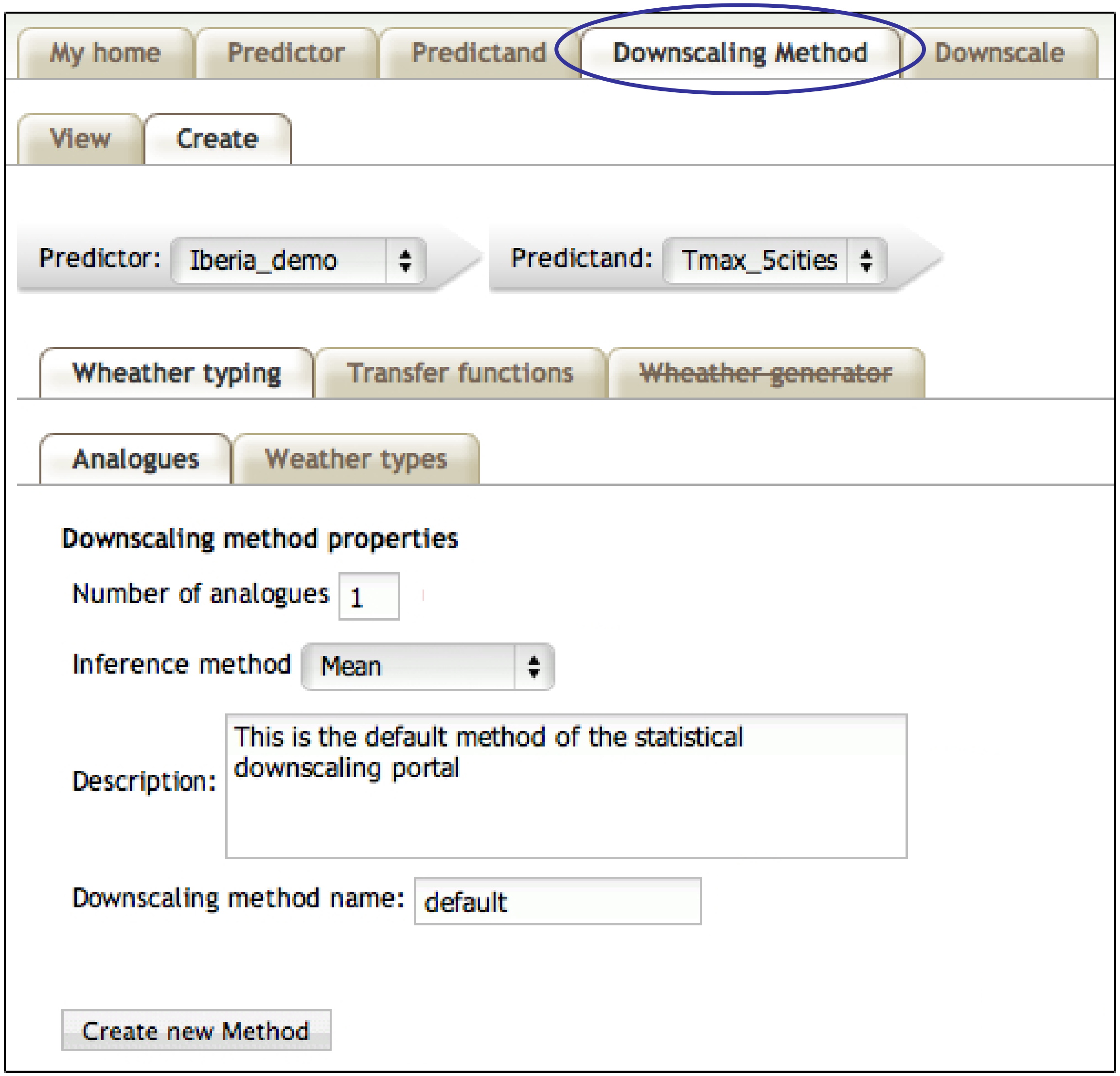
SD Portal: Downscaling Method

It also allows selecting a particular downscaling algorithm from the different families of methods:

- **Analogs**
- **Regression + GLMs**
  - From CPs
  - From grid-points
- **Neural Network**
- **K-means weather types**
- Weather generators

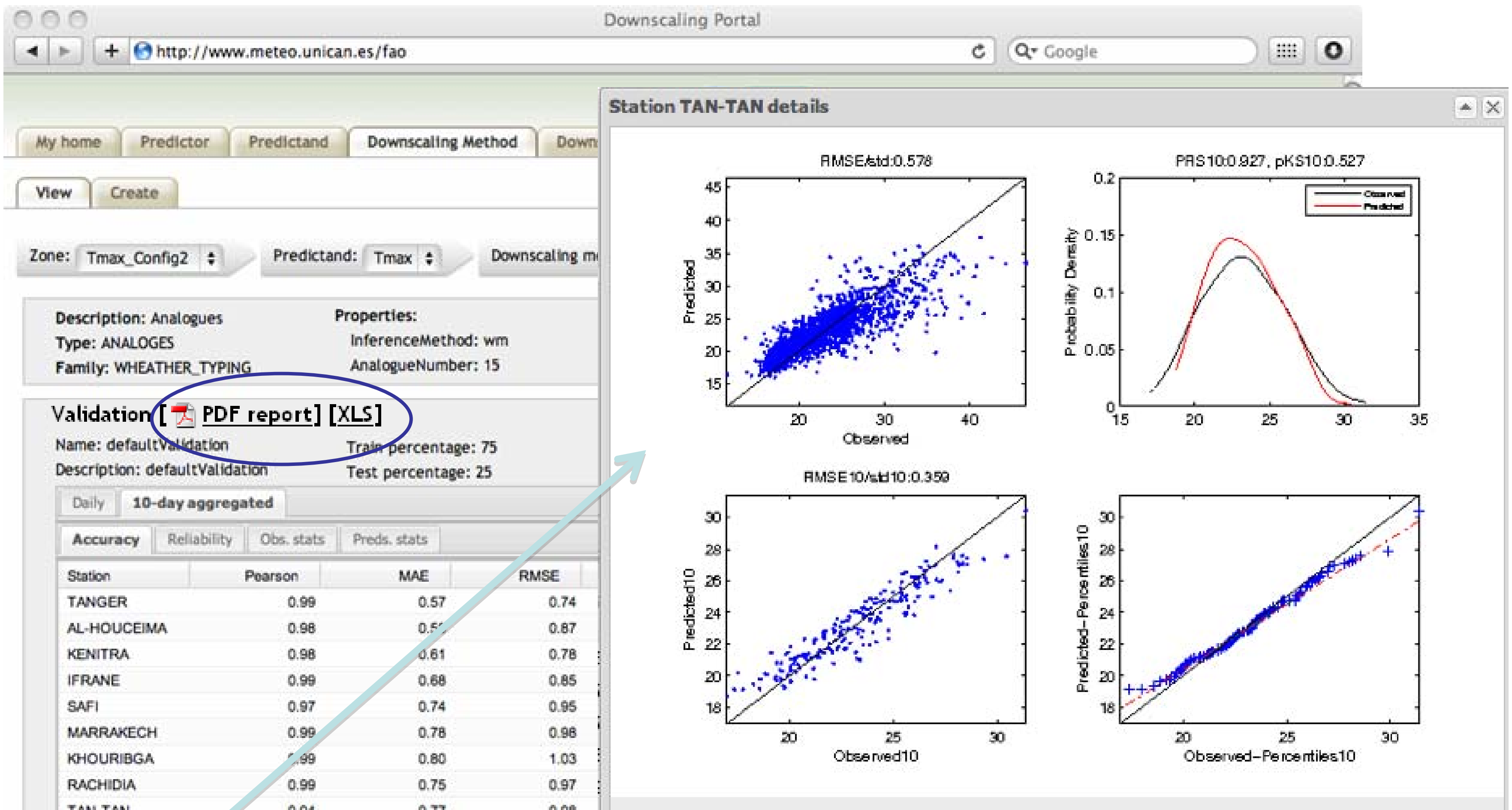
and defining a particular configuration:

• Number of analogs
• Number of CPs.
• etc.



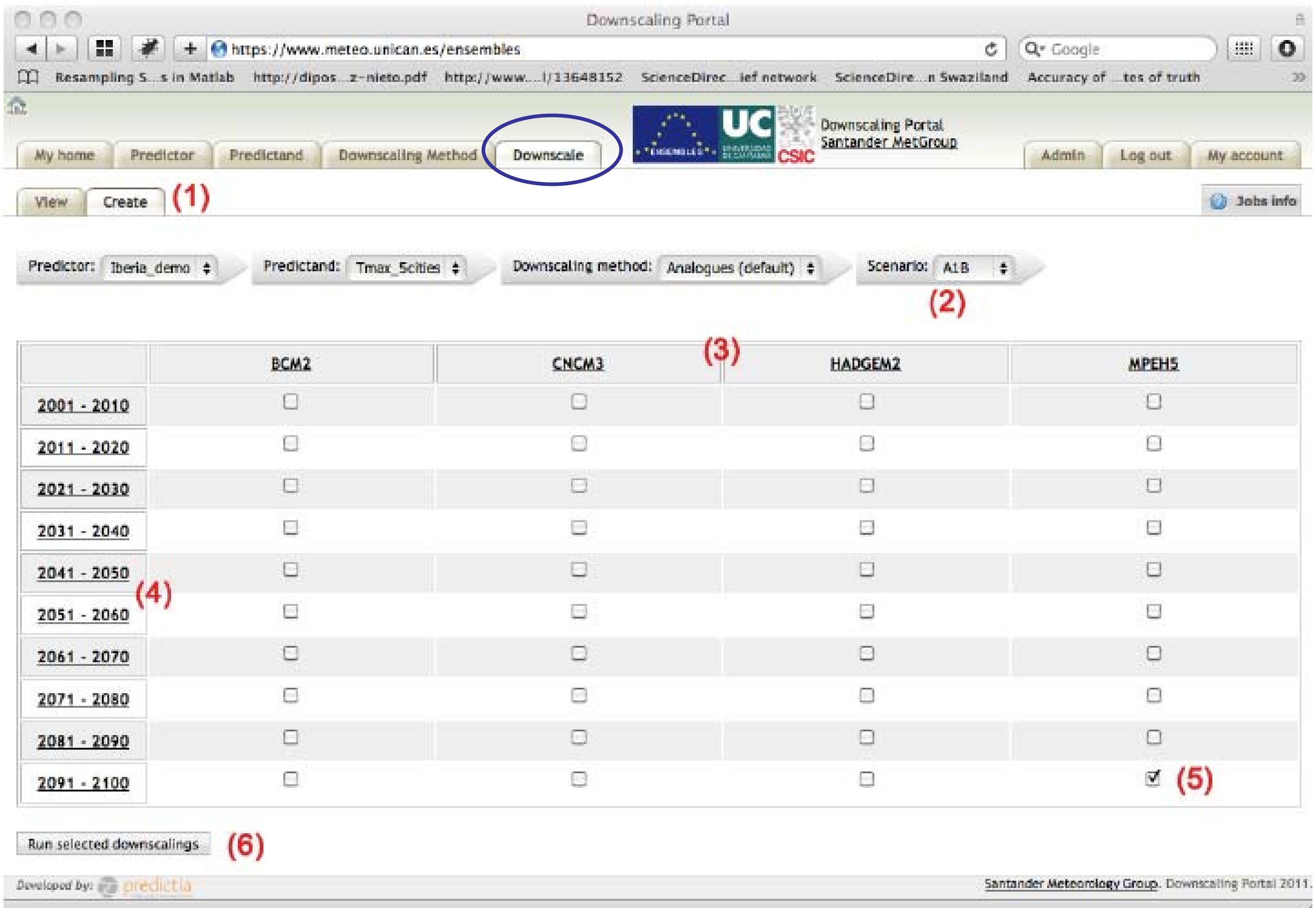
SD Portal: Calibration & validation

Finally, it allows selecting a downscaling method (from the list of available ones, including regression, analogs, weather typing, etc.) and obtaining a cross-validation in present climate using renalysis data.



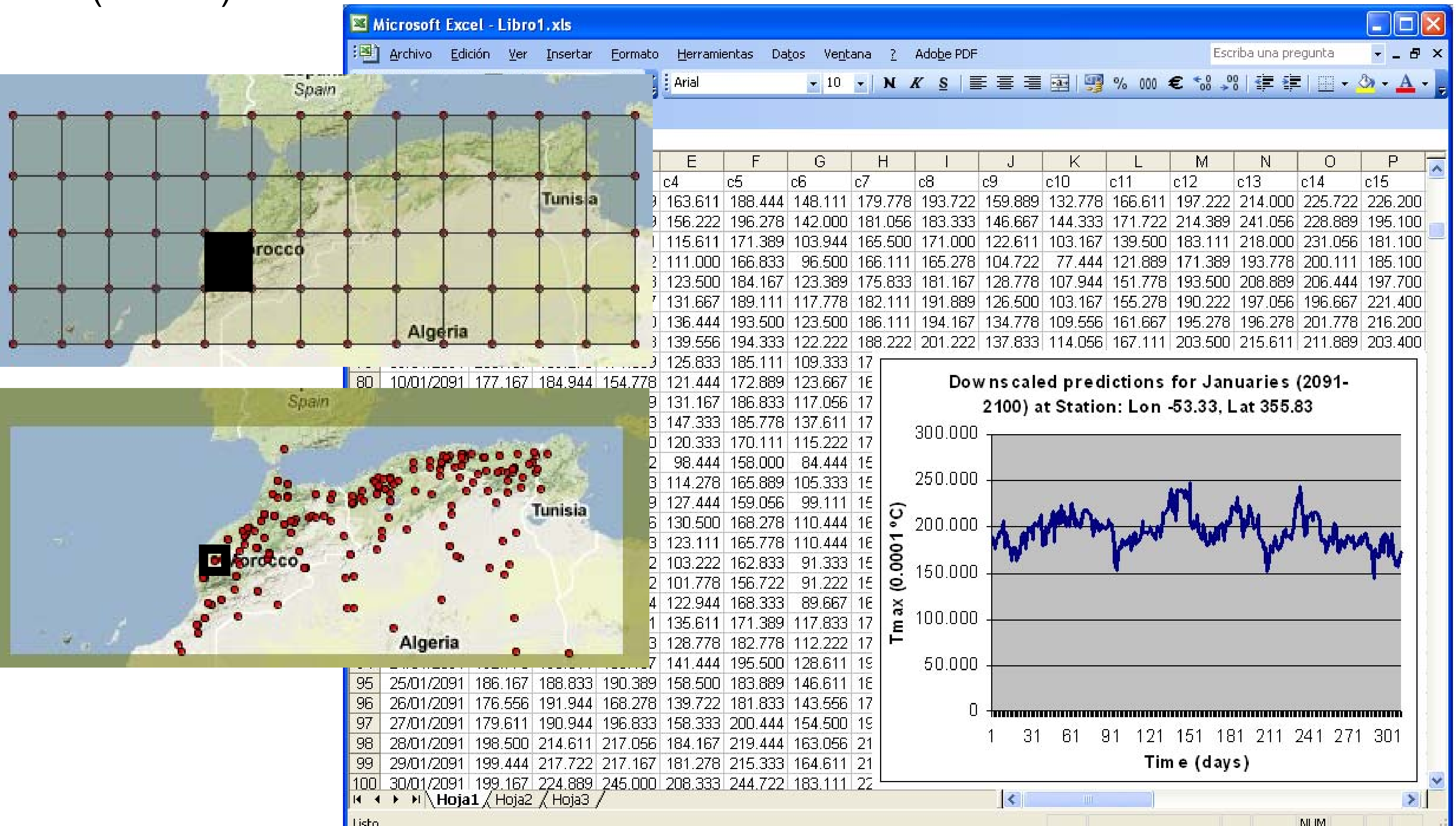
SD Portal: PRODUCTION

Once the method is defined and validated it can be used to downscale GCM models for future scenarios decade by decade.



SD Portal: Friendly Output

The resulting daily locally projected simulations can be downloaded as Excel (or ascii) files.



Recommendations and Support for End-Users

These portals **should not be used as a black-box tool (particularly the downscaling portal)** to avoid wrong applications and errors. Some background knowledge is required and the limitations should be known (e.g. the different assumptions of the statistical downscaling methodology). **The users are requested to collaborate with downscaling experts. In some cases of mutual interest we provide support and/or training.**

User tutorials, indications and **recommendations for downscaling** are provided and referred to, e.g. in the ENSEMBLES web site.

Technical Notes
Santander Meteorology Group (CSIC-UC)
SMG:2.2011

Santander Meteorology Group
A multidisciplinary approach for weather & climate

User Guide of the ENSEMBLES Downscaling Portal (version 2)

J.M. Gutiérrez[1], D. San Martín[1,2], A.S. Cofiño[3], S. Herrera[1,2], R. Manzanas[1], M.D. Frías[3]
[1] *Instituto de Física de Cantabria, CSIC-Universidad de Cantabria, Santander, Spain.*
[2] *Predictia Intelligent Data Solutions, Santander, Spain.*
[3] *Dpto. Matemática Aplicada y C.C. Universidad de Cantabria. Santander, Spain*
correspondence: gutierjm@ifca.unican.es, daniel@predictia.es, cofinoa@unican.es

version: v3–*November 2012*