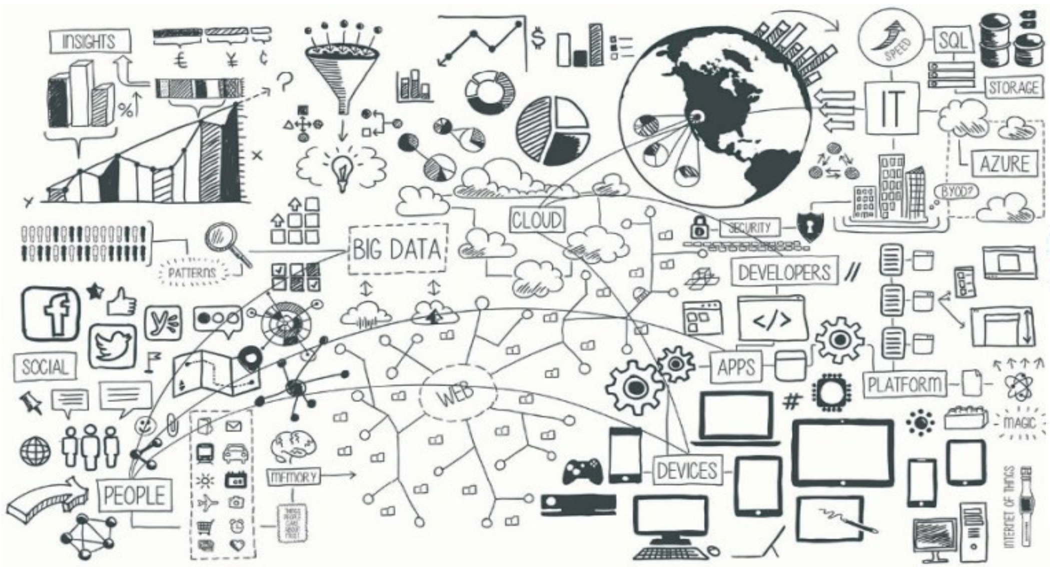
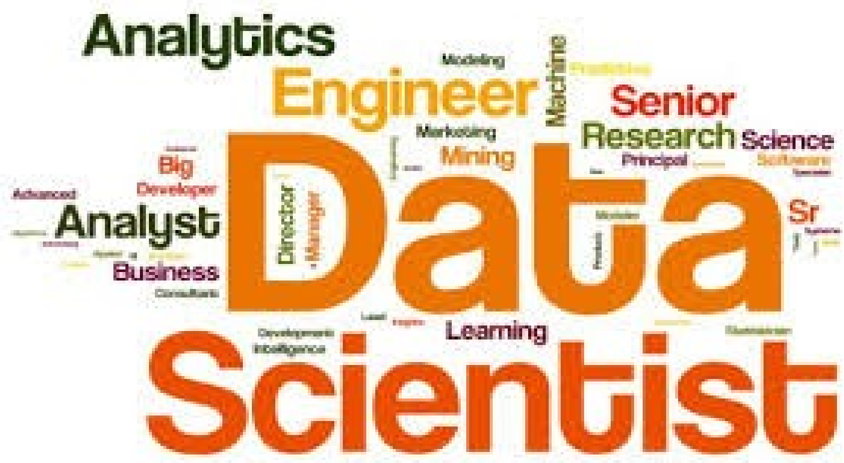
DataLab: Environment and Meteorology

Data Mining and Machine Learning in Climate. Statistical downscaling





Sixto Herrera García

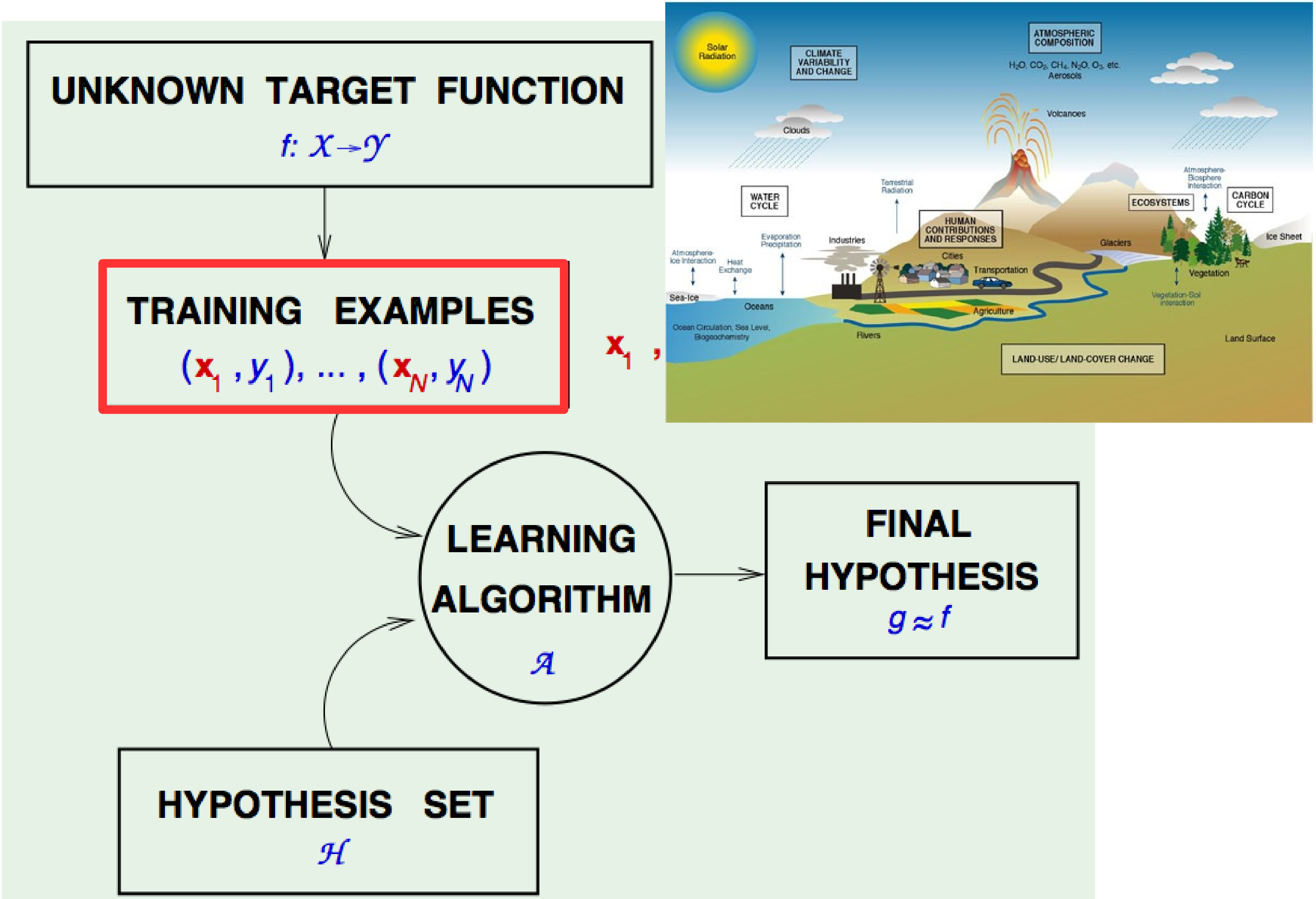
Grupo de Meteorología

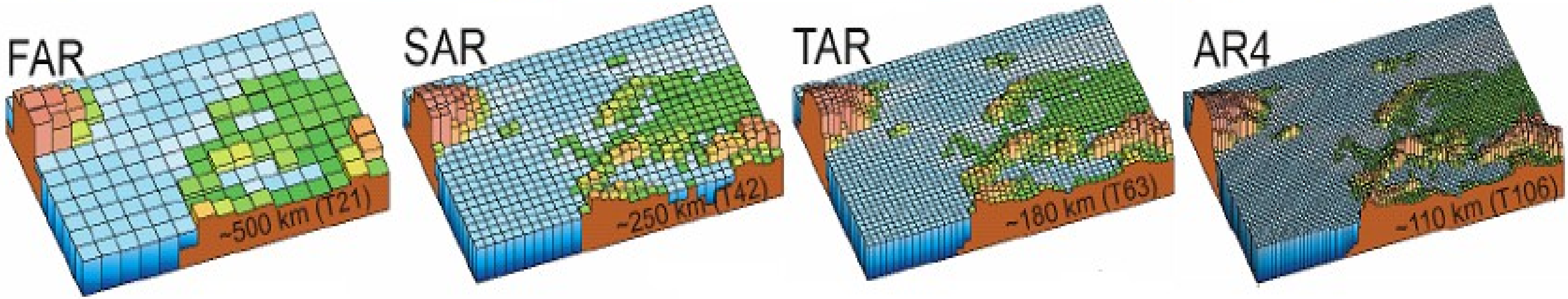
Univ. de Cantabria – CSIC
MACC / IFCA

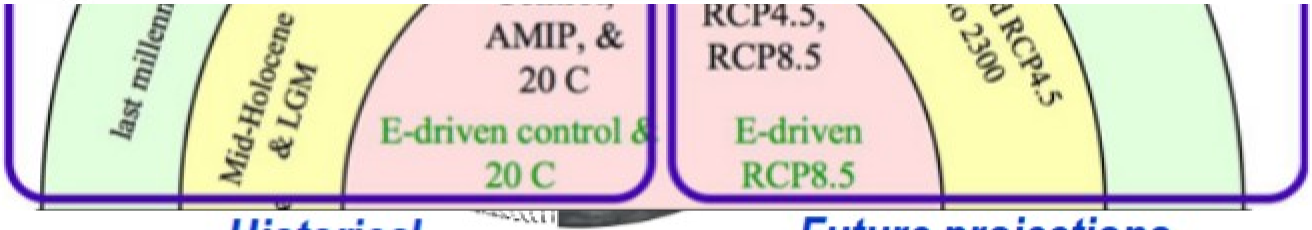
M1980 – Data Laboratory: Environment & Meteorology (16:00-18:00) Virtual classroom: https://meet.jit.si/M1980				
03/25	X	Introduction and Climate4R package	TL	JB
03/26	J	Climatic System & Models (DM & ML in Climate Science)	T	SH
03/27	V	Data Repositories: ESGF & MARS	TL	SH
03/30	L	Data Repositories: ESGF & MARS	TL	SH
03/31	M	Lab: Climate4R – Example 1	L	JB
04/01	X	Lab: Climate4R – Example 2	L	JB
04/02	**J**	**Downscaling: Data Mining in Clime**	**T**	**SH**
04/03	V	Lab: downscaleR	L	JB
04/06	L	Evaluation and Validation	T	SH
04/07	M	Lab: Evaluation and Validation	L	JB
04/08	X	Impacts	L	JB
04/13	L	Impacts	L	JB

SH - Sixto Herrera | **JB** – Joaquín Bedia

Atmosphere + Hydrosphere + Cryosphere + Lithosphere + Biosphere







validation and climate change attribution and projection.

Historical simulations

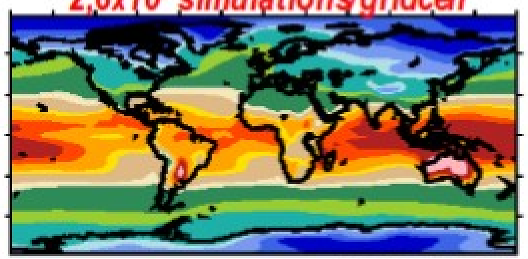
1980 | 1990 | 2000

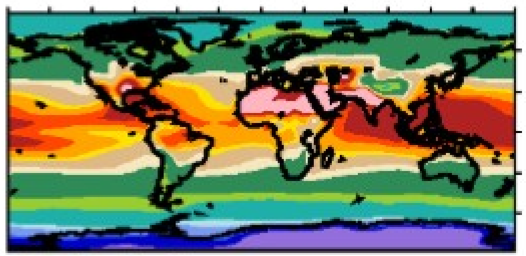
1h step x 30 years

2,6x10^5 simulations/gridcell

Future projections (scenarios)

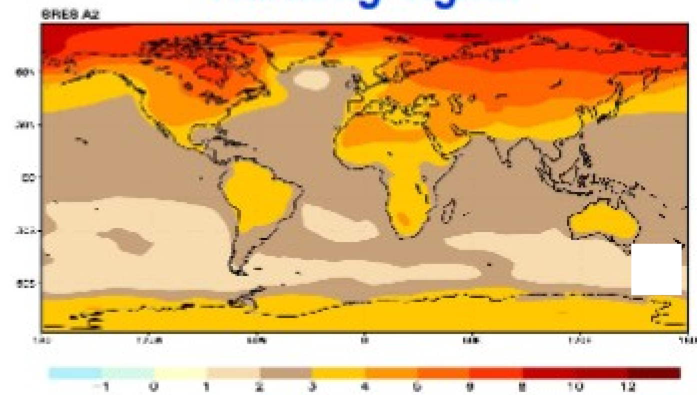
2010 | 2020 | 2030 | 2040 | 2050 | 2060 | 2070 | 2080 | 2090





"delta" method

Warming signal



Computational (and physical) constraints limit the resolution (~100-200 Km)

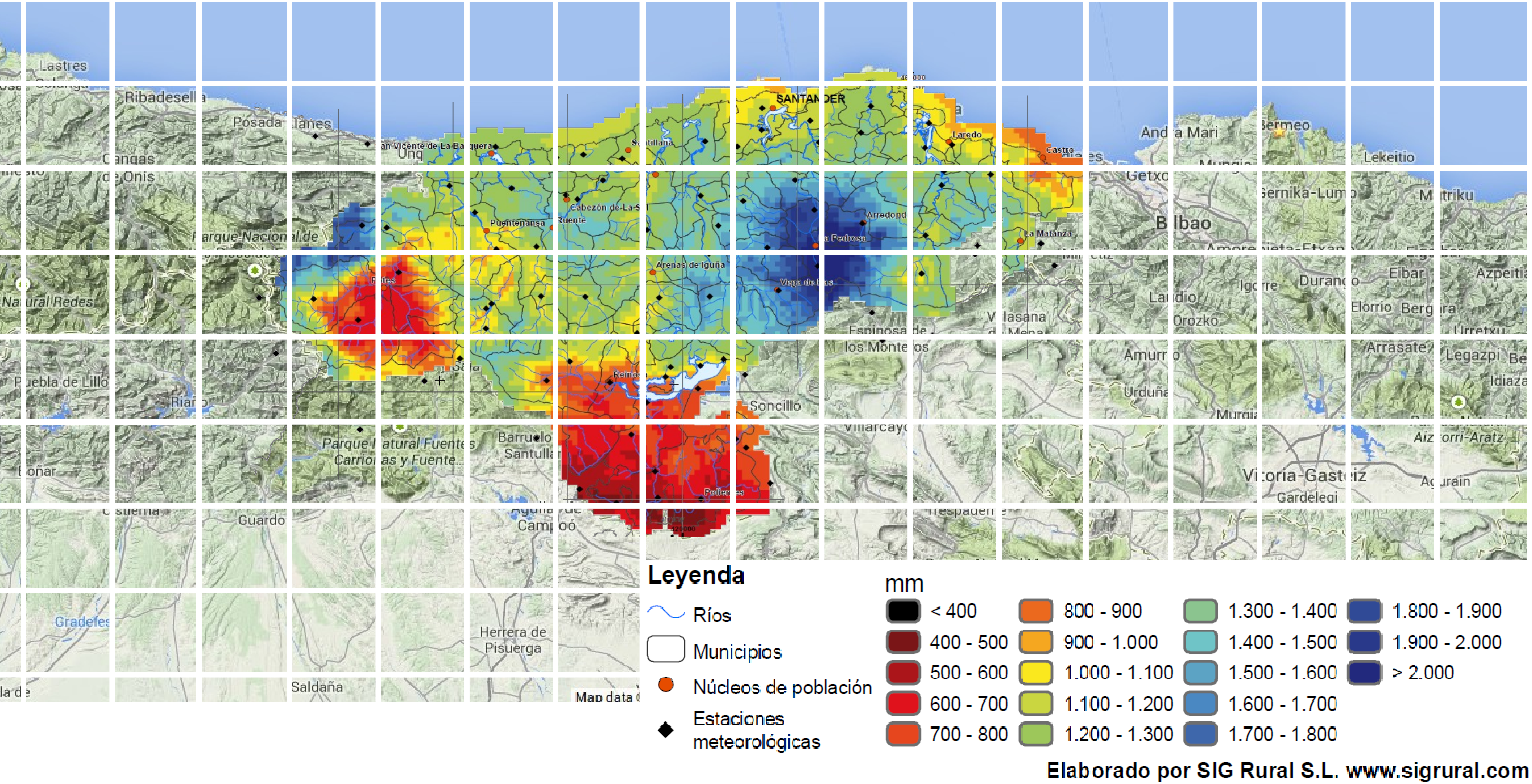
Weather Prediction (short-medium range)

Target → day

Days/weeks

D	L	M	X

15 km (~Operational Forecast ECMWF)



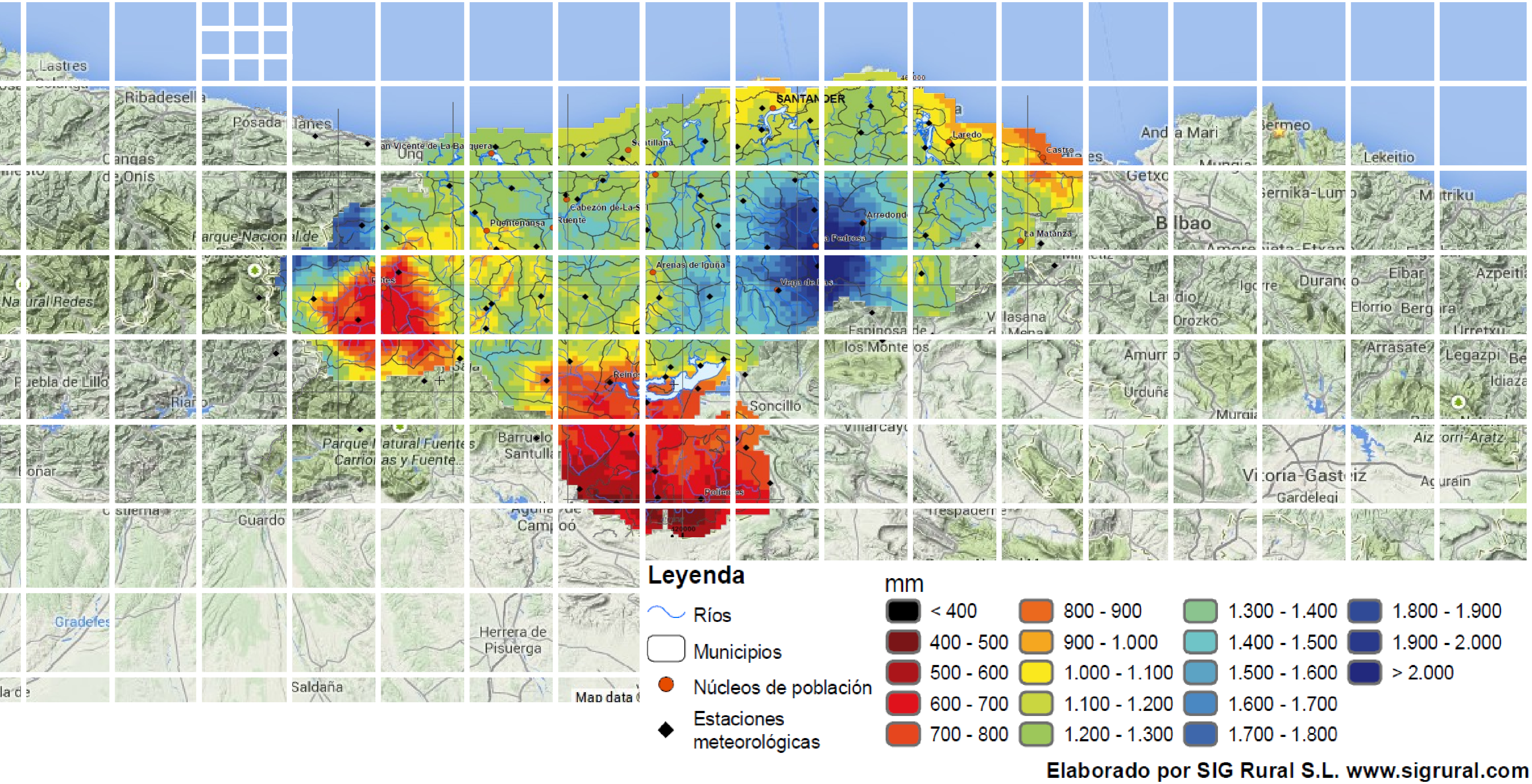
Weather Prediction (short-medium range)

Target → day

Days/weeks

D	L	M	X

5 km (~AEMET HIRLAM) 15 km (~Operational Forecast ECMWF)



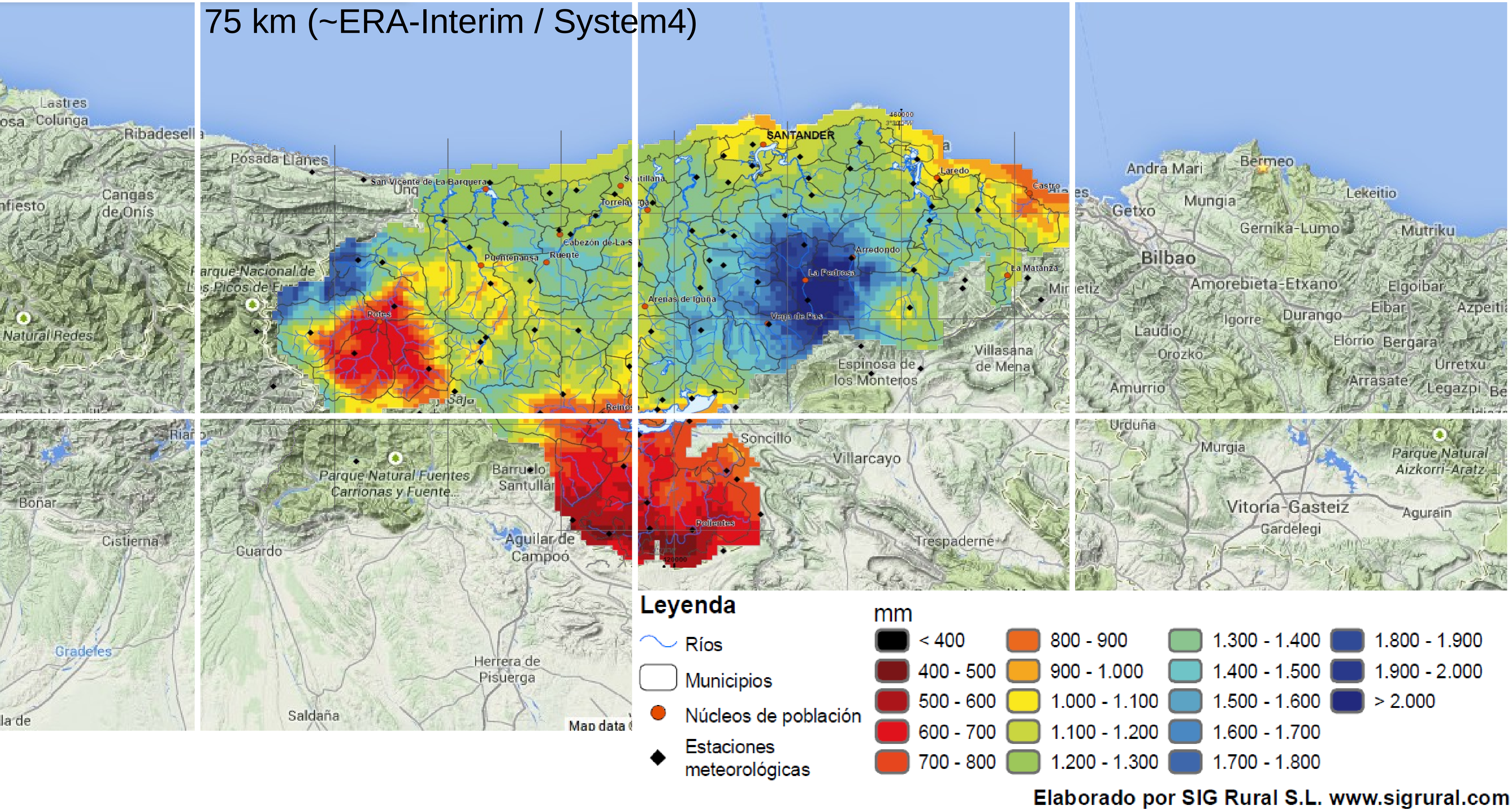
Seasonal Forecast (Prediction)

Target → season

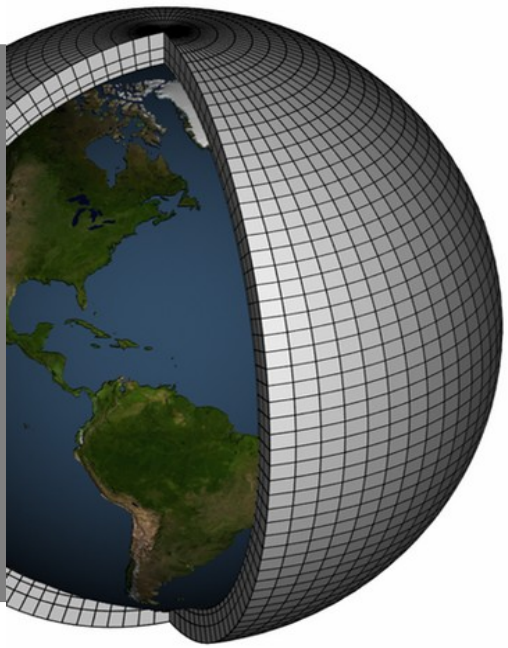
Month/Season

DEF	MAM	JJA	SON

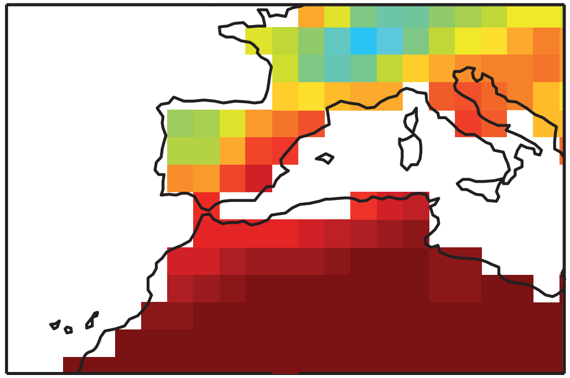
75 km (~ERA-Interim / System4)



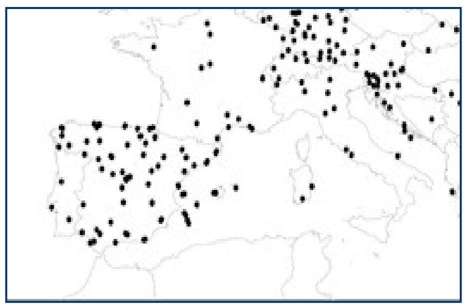
Model's World



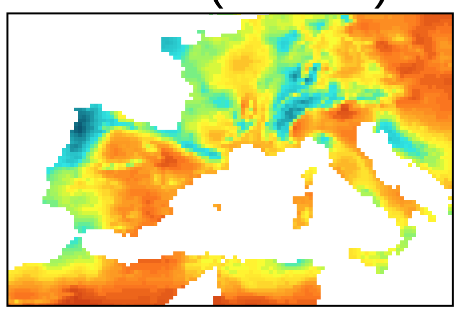
"Real" World

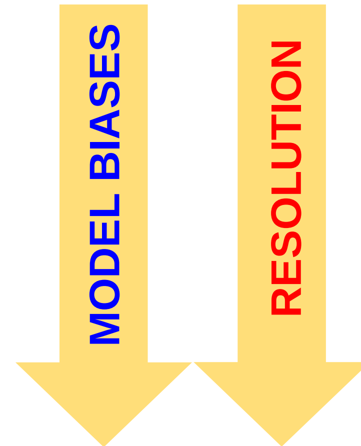


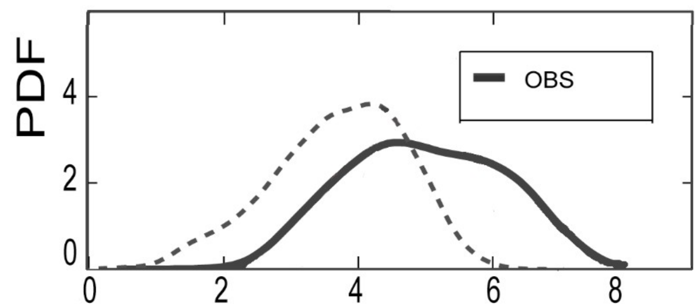
Locations



Gridded (**~10km**)



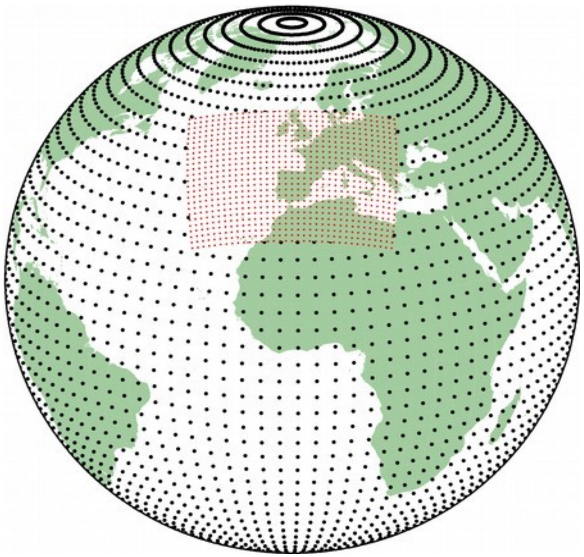


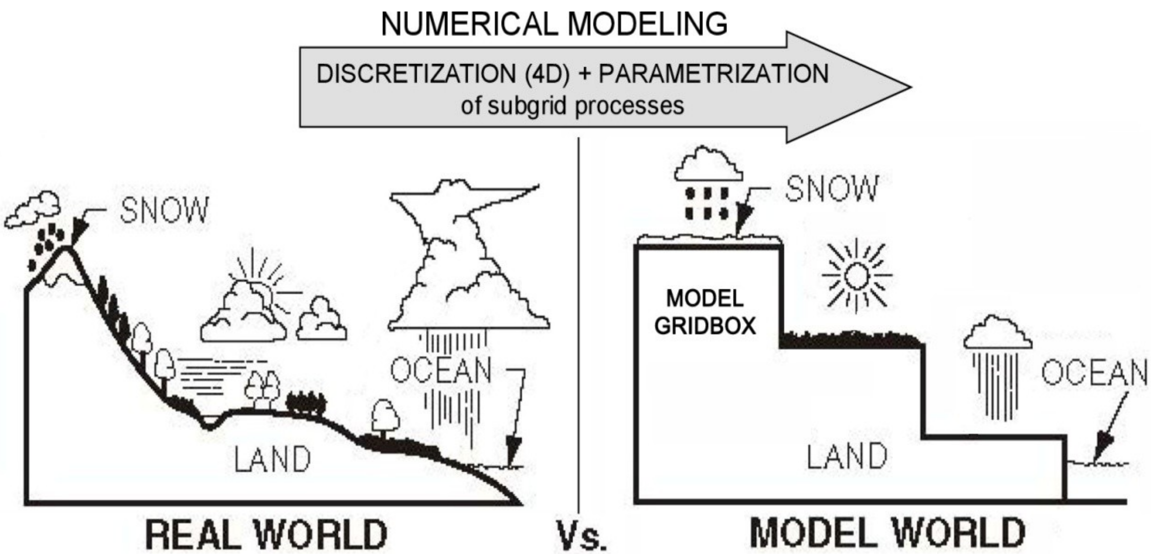


Climate Models have systematic biases w.r.t. observations:

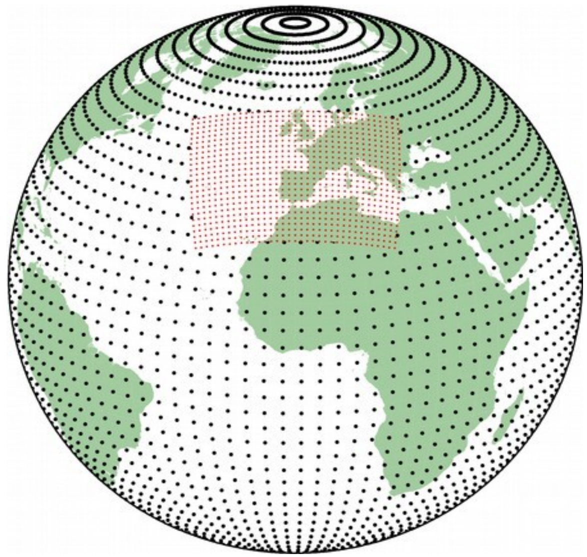
1. **Systematic Bias**
2. **Different resolutions**

Downscaling methods try to bridge the gap between the coarse resolution of the Global Climate Models (GCMs) and the local climate characteristics of a particular region.



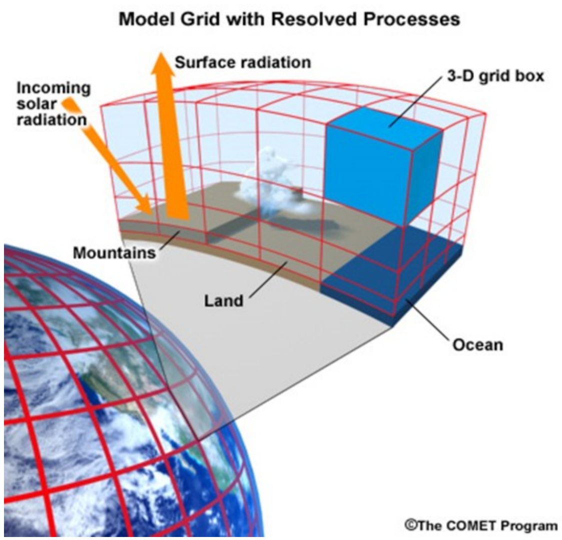


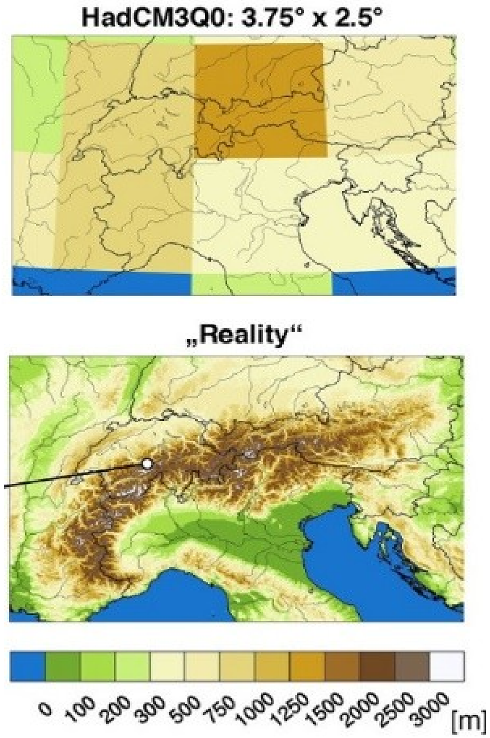
Downscaling methods try to bridge the gap between the coarse resolution of the Global Climate Models (GCMs) and the local climate characteristics of a particular region.

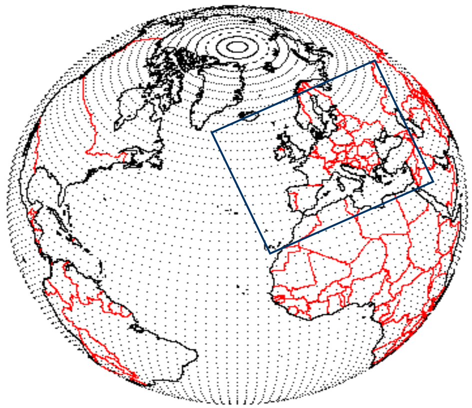


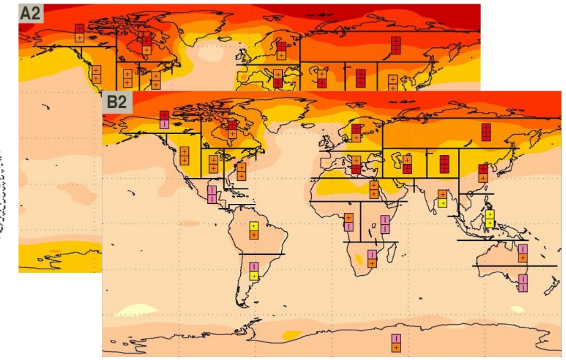
Dynamical Downscaling solves the clymate equation system in a geographical domain covering the target region, letting to increase the final resolution of the simulation.

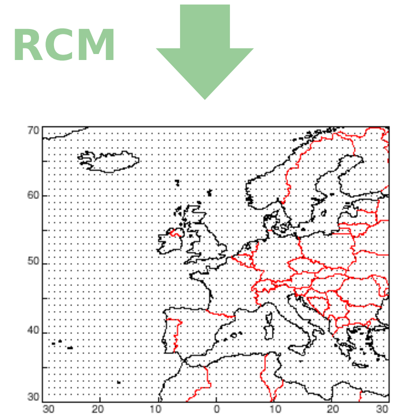
$$
\begin{aligned}
\frac{d\vec{v}}{dt} &= -\frac{1}{\rho}\vec{\nabla}p - \vec{\nabla}\phi + \vec{F} - 2\vec{\Omega}\times\vec{v} \\
\frac{\partial\rho}{\partial t} &= -\vec{\nabla}\cdot(\rho\vec{v}) \\
p &= \rho R T \\
Q &= C_p\frac{dT}{dt} - \frac{1}{\rho}\frac{dp}{dt} \\
\frac{\partial\rho q}{\partial t} &= -\vec{\nabla}\cdot(\rho q\vec{v}) + \rho(E-C)
\end{aligned}
$$

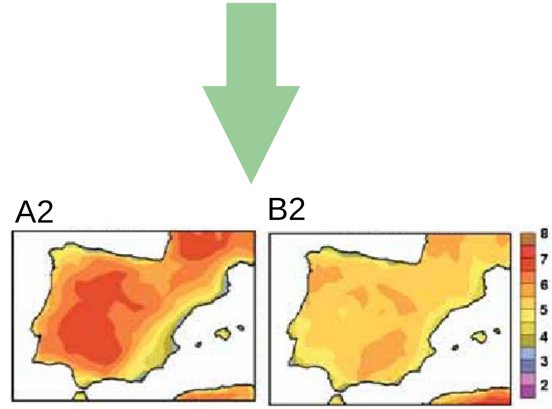




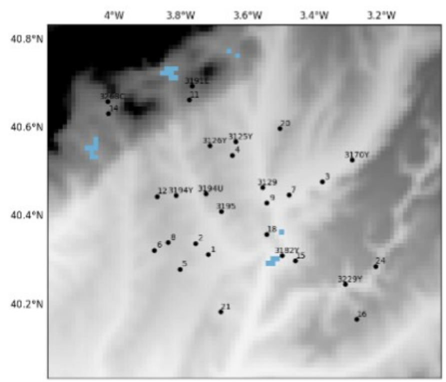


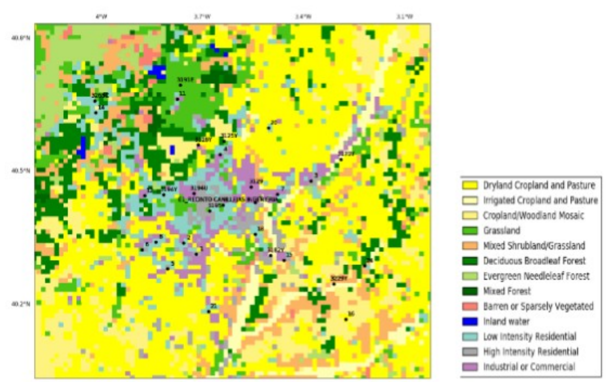






Depending on the target resolution some boundary conditions, as the land use, should be considered and Introduced in the model.





CORINE, en lugar de USGS33

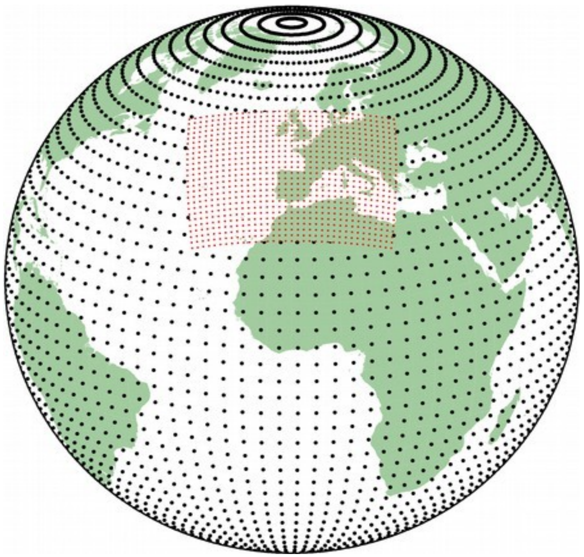
A parameterization is a **statistical** representation of the **net effect** of processes occurring on **spatial scales smaller than the grid** spacing of a dynamical model (GCM, RCM, CRM, LES, …) **over mean variables** at each grid cell.

Parameterizations are based on the **physics** of the processes, plus **simplifying (closure) assumptions** to relate unknown variables to prognostic (mean) model variables.

Parameters are obtained from:

- Theory (e.g. known physical constants)
- Field campaigns
- Higher resolution models (CRM, LES)

Downscaling methods try to bridge the gap between the coarse resolution of the Global Climate Models (GCMs) and the local climate characteristics of a particular region.



Dynamical Downscaling solves the clymate equation system in a geographical domain covering the target region, letting to increase the final resolution of the simulation.

- ☺ Plenty of variables, including 3D
- ☺ Subdaily data available
- ☺ Variables are physically consistent
- 😐 Variables represent areal averages
- ☹ Computationally expensive
- ☹ Biases are commonplace
- ☹ Parameterizations: Stationary hypothesis

Generate a coordinated ensemble of high-resolution, historical/future regional climate projections for land-regions of the globe sampling; multiple GCM/RCP/RCM/ESDs methods. 1st phase based on CMIP5 historical-projection runs and/or ERA-int boundary data

Make data accessible & useable in common format/file structure

Foster coordination between downscaling efforts & encourage local participation, in generating, analysing & communicating potential regional climate change and associated uncertainties & risks

Initial emphasis on African climate & IAV: START/WCRP sponsored 3 analysis/IAV workshops for an Africa-CORDEX team in 2011-12

Similar activities now starting for South Asia, East Asia and South/Central America

WCRP CORDEX

1990	1995	2000	2005	2010

1 mes | 1 año | 1 década | 1 siglo

varios meses | varios años | varias decadas

60 Km | 50 Km | 30 Km | 20 Km | 10 Km

Ejemplos:

- Dickinson et al. **1989**: 20 días (60km) x 3
- Giorgi et al. **1990**: 1 mes (60km) x 6
- Jones et al. **1995**: 10 años (50km) x 2
- Kidson & Thompson **1998**: 5 años (50km)
- Christensen et al. **1998**: 10 años (19km)
- Giorgi et al. **2004**: 30 años (50km) x 4

PRUDENCE several decades (50km)
ENSEMBLES centuries (25km)
EURO-CORDEX centuries (11km)

One Model/One Domain

- Common approach in the application of a RCM

Multi-Model/One Domain

- Approach of the Model Intercomparison Projects (MIPs)
  - **EU:** Regionalization, RACCS, MERCURE, PRUDENCE, ENSEMBLES
  - **USA:** PIRCS, NARCCAP
  - Etc: ARCMIP, SGMIP, ...

One Model/Several Domain

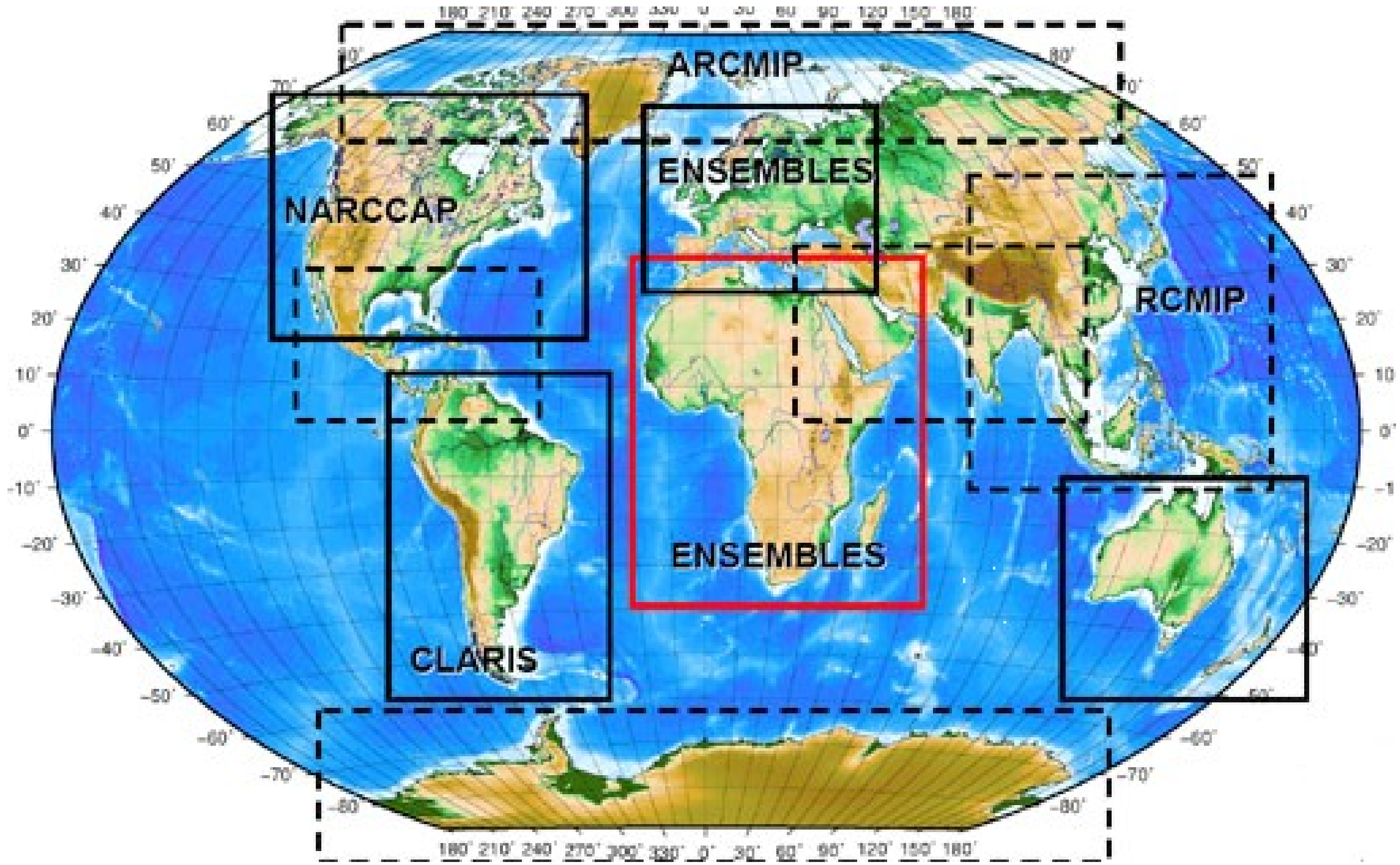
- Used to find problems of the model.

Multi-Model/Several Domain

- RCMs ran out of their "***natural***" domain
  - GEWEX **transferability** experiment
  - **CORDEX** (COordinated Regional Downscaling EXperiment)

CORDEX Phase I experiment design

A WCRP-sponsored initiative for regional climate change projection with dynamical downscaling methods, aimed at providing regional climate change information (with initial focus in Africa) for the IPCC-AR5.

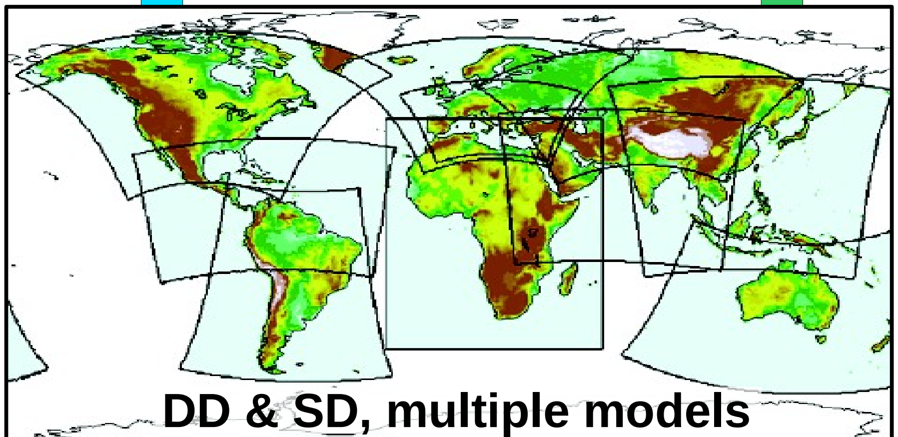


http://wcrp.ipsl.jussieu.fr/SF_RCD_CORDEX.html

CORDEX Phase I experiment design

Model Evaluation Framework

Climate Projection Framework



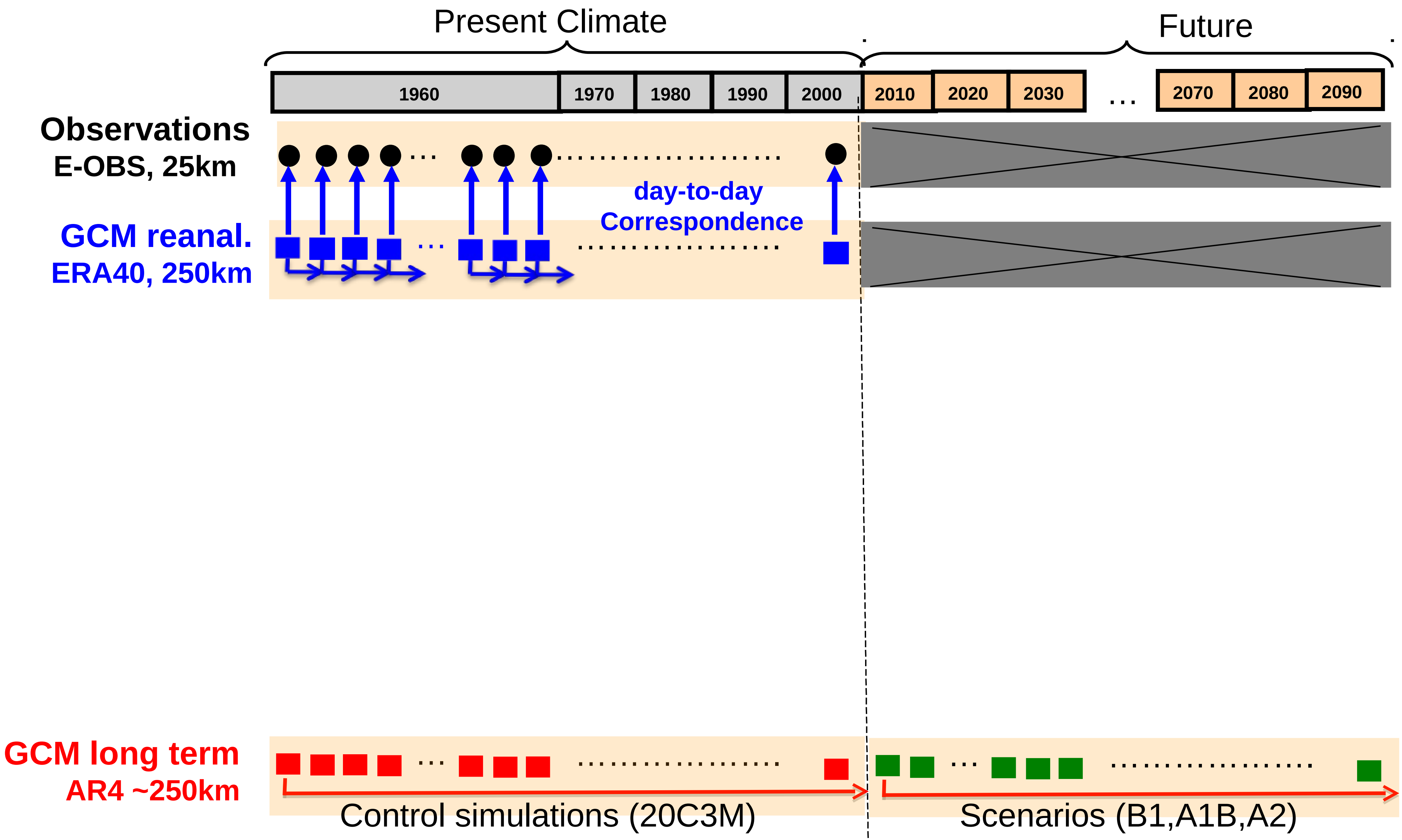
ERA-Interim BC
1989-2008

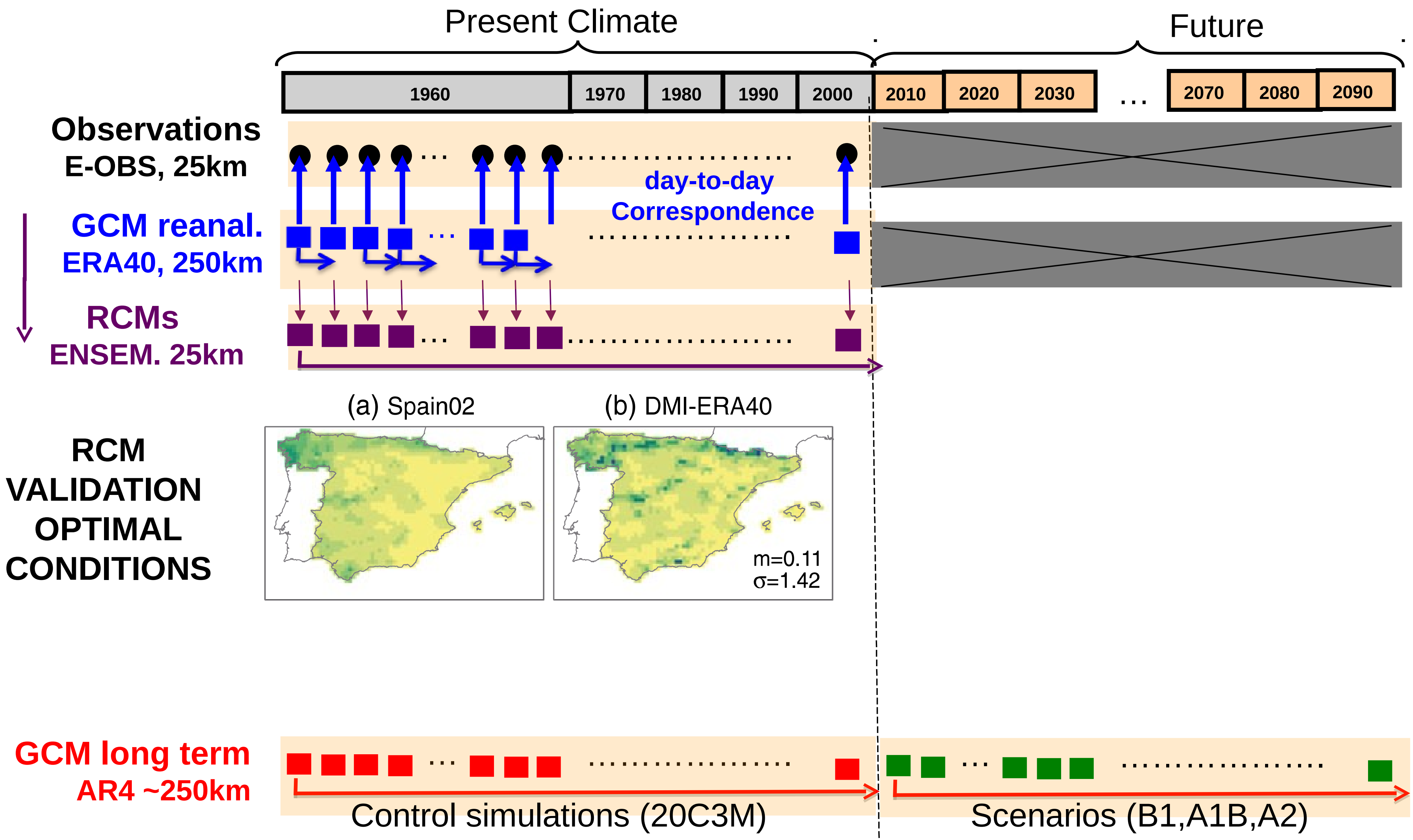
RCP4.5, RCP8.5

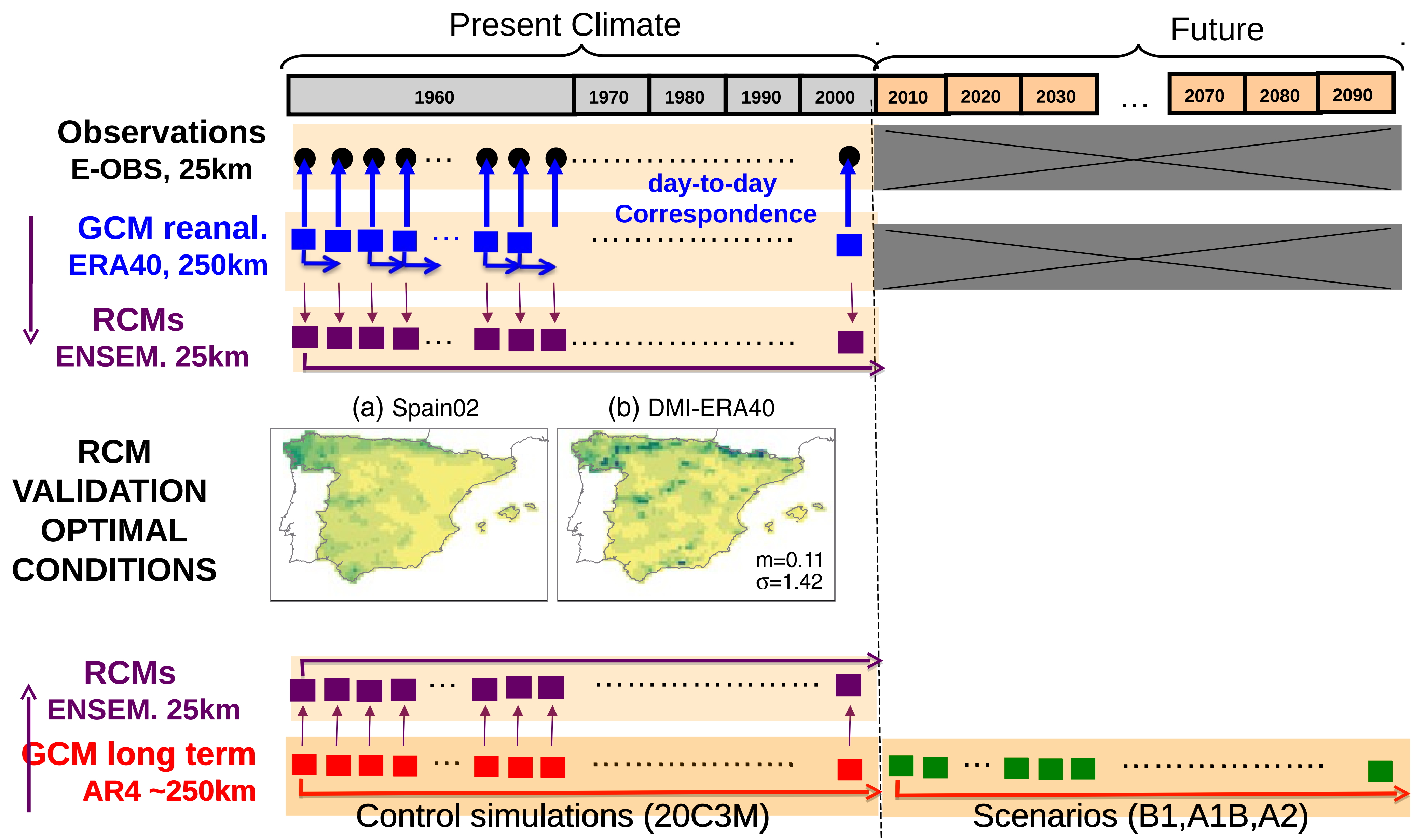
Multiple AOGCMs

Projections
1951-2100

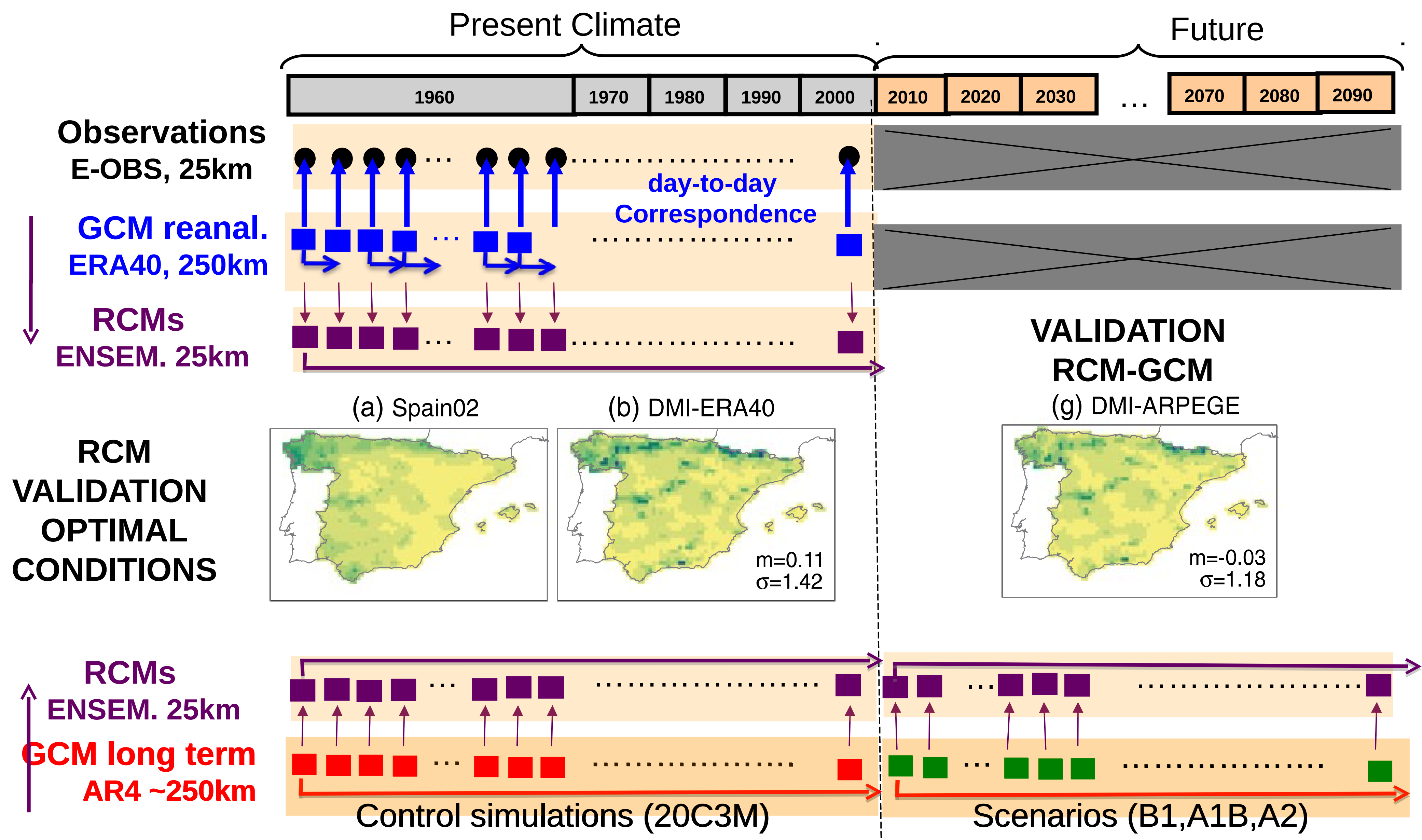
Regional analysis
Regional databanks


Present Climate
Future
1960
1970
1980
1990
2000
2010
2020
2030
...
2070
2080
2090
Observations
E-OBS, 25km
day-to-day
Correspondence
GCM reanal.
ERA40, 250km
GCM long term
AR4 ~250km
Control simulations (20C3M)
Scenarios (B1,A1B,A2)


Present Climate
Future
1960
1970
1980
1990
2000
2010
2020
2030
…
2070
2080
2090
Observations
E-OBS, 25km
day-to-day
Correspondence
GCM reanal.
ERA40, 250km
RCMs
ENSEM. 25km
(a) Spain02
(b) DMI-ERA40
m=0.11
σ=1.42
RCM
VALIDATION
OPTIMAL
CONDITIONS
GCM long term
AR4 ~250km
Control simulations (20C3M)
Scenarios (B1,A1B,A2)

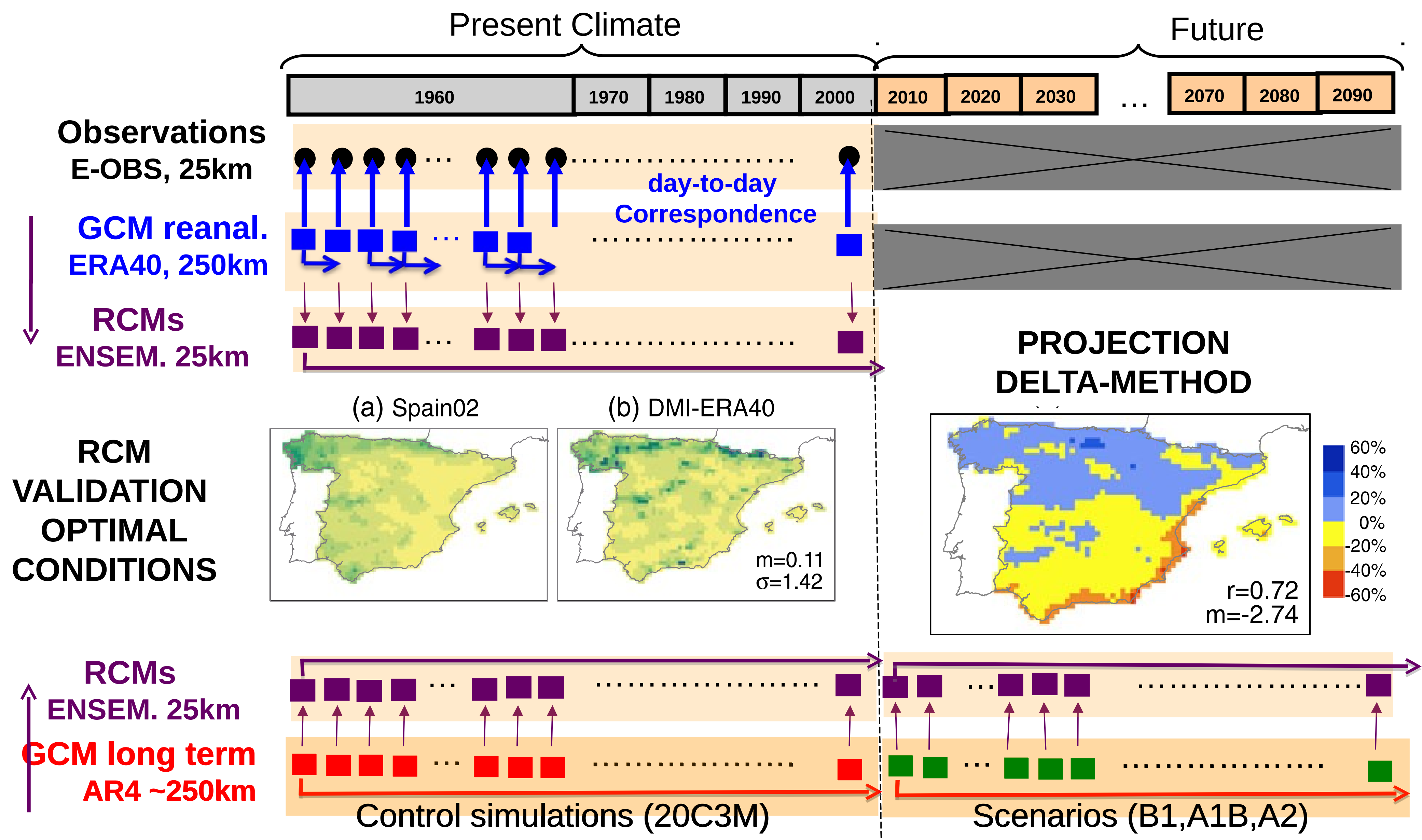


Present Climate
Future
1960
1970
1980
1990
2000
2010
2020
2030
...
2070
2080
2090
Observations
E-OBS, 25km
GCM reanal.
ERA40, 250km
day-to-day
Correspondence
RCMs
ENSEM. 25km
(a) Spain02
(b) DMI-ERA40
m=0.11
σ=1.42
RCM
VALIDATION
OPTIMAL
CONDITIONS
RCMs
ENSEM. 25km
GCM long term
AR4 ~250km
Control simulations (20C3M)
Scenarios (B1,A1B,A2)

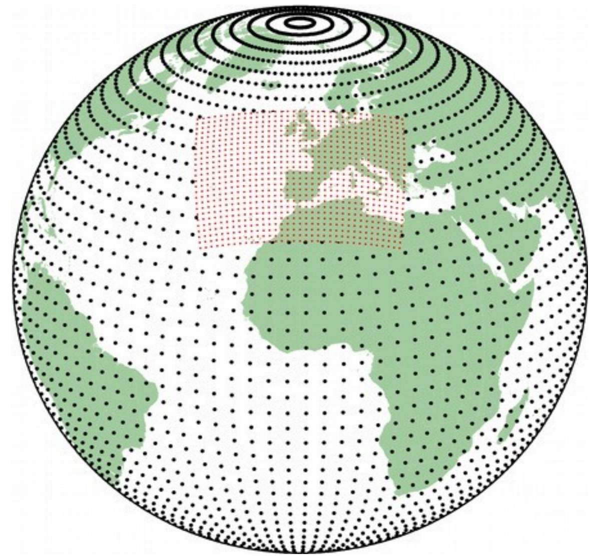


Present Climate
Future
1960
1970
1980
1990
2000
2010
2020
2030
2070
2080
2090
Observations
E-OBS, 25km
GCM reanal.
ERA40, 250km
day-to-day
Correspondence
RCMs
ENSEM. 25km
VALIDATION
RCM-GCM
(a) Spain02
(b) DMI-ERA40
(g) DMI-ARPEGE
RCM
VALIDATION
OPTIMAL
CONDITIONS
m=0.11
σ=1.42
m=-0.03
σ=1.18
RCMs
ENSEM. 25km
GCM long term
AR4 ~250km
Control simulations (20C3M)
Scenarios (B1,A1B,A2)

Turco et al. 2013: https://link.springer.com/article/10.1007%2Fs10584-013-0844-y



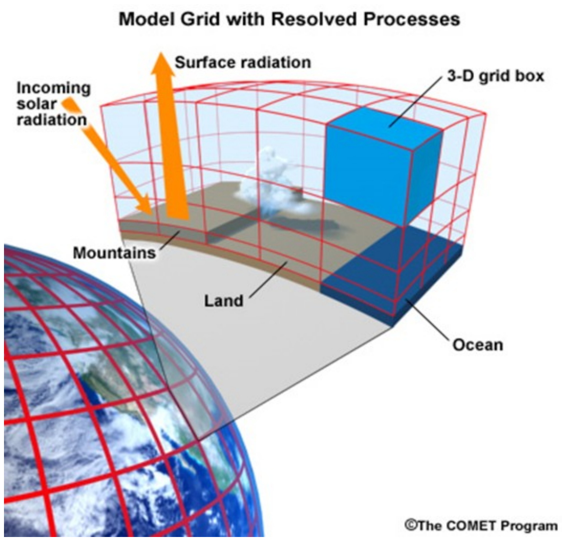
Downscaling methods try to bridge the gap between the coarse resolution of the Global Climate Models (GCMs) and the local climate characteristics of a particular region.

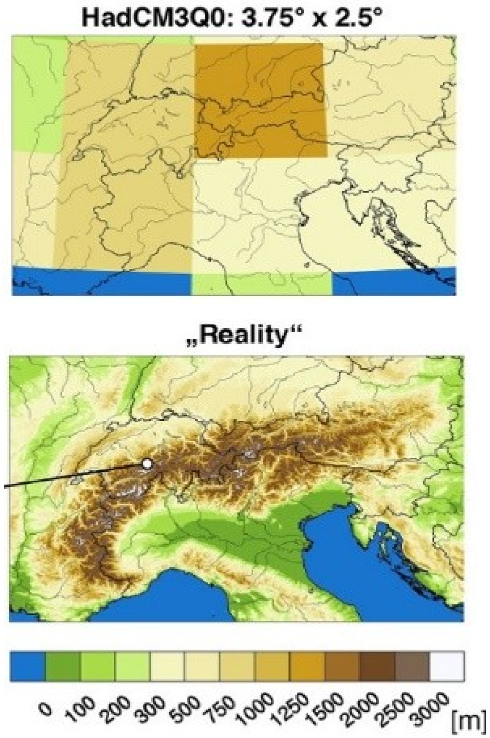


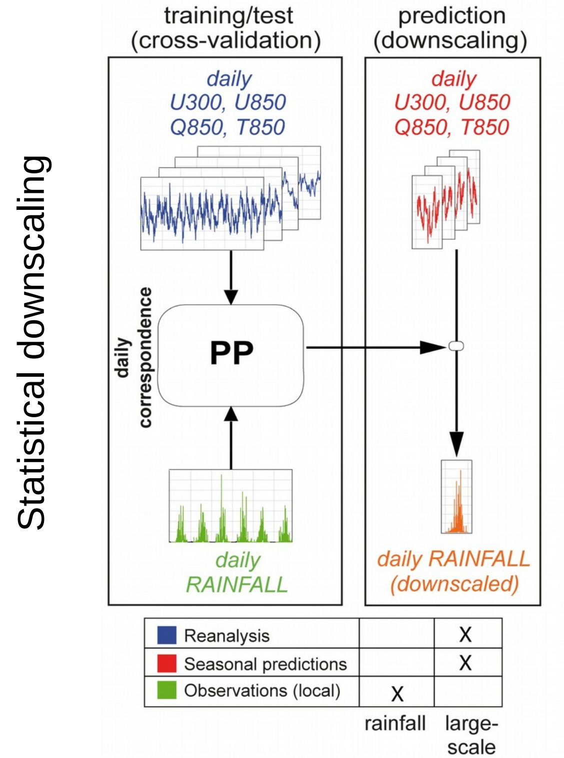
Dynamical Downscaling solves the clymate equation system in a geographical domain covering the target region, letting to increase the final resolution of the simulation.

Statistical Downscaling looks for empirical relations between the local variables (**predictand**) and the given by the models (**predictors**).

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

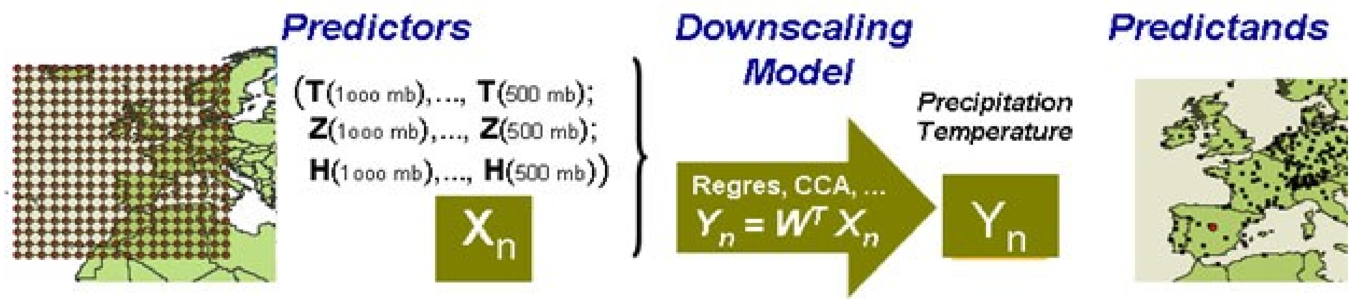




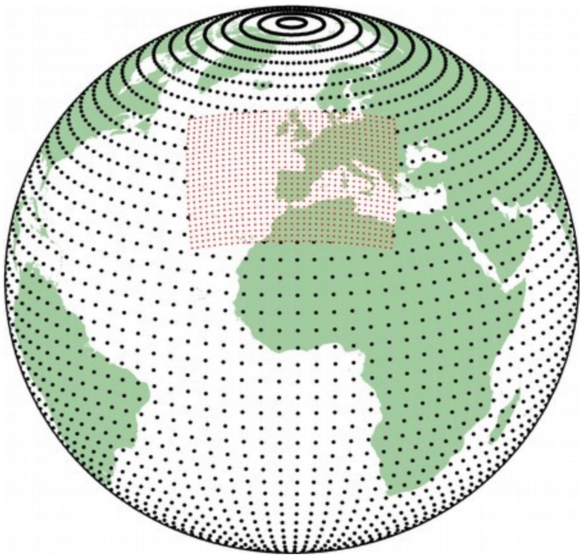


Statistical Downscaling

Statistical methods linking the local observed climate (**predictand *Y***) with the global simulations given by the GCMs (**predictors *X***), through some **function *f*** and/or **parameters** θ : $\mathbf{Y} = f\,(\mathbf{X};\theta)$



Downscaling methods try to bridge the gap between the coarse resolution of the Global Climate Models (GCMs) and the local climate characteristics of a particular region.



Statistical Downscaling looks for empirical relations between the local variables (**predictand**) and the given by the models (**predictors**).

- ☹ Variables available as long as there are obs.
- ☹ Different variables probably do not keep physical or even spatial consistency
- ☹ Stationary hypothesis and extrapolation capabilities.
- 😐 Variables keep the representativity of obs.
- ☺ Computationally cheap
- ☺ Biases are low
- ☺ Non-meteorological variables (e.g. impact indices) could be directly produced.

http://www.cordex.org/domains/cordex-esd/

Cordex ESD

Experiment protocol – Empirical statistical downscaling

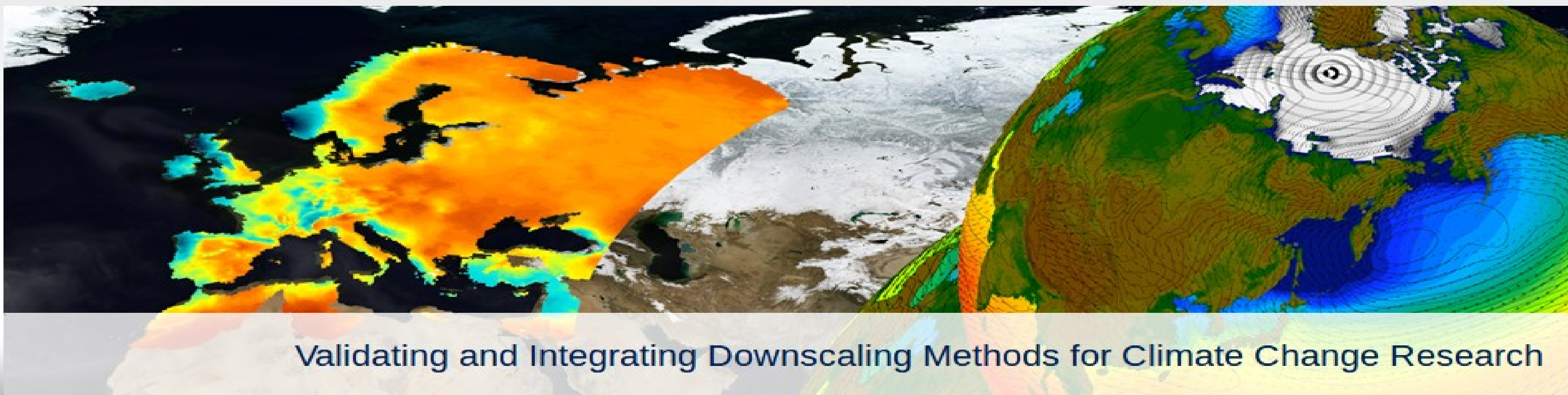
- ESD Overview
- ESD Background
- ESD Experiment1 protocols
- ESD Reference Document
- Register for CORDEX ESD Experiment 1

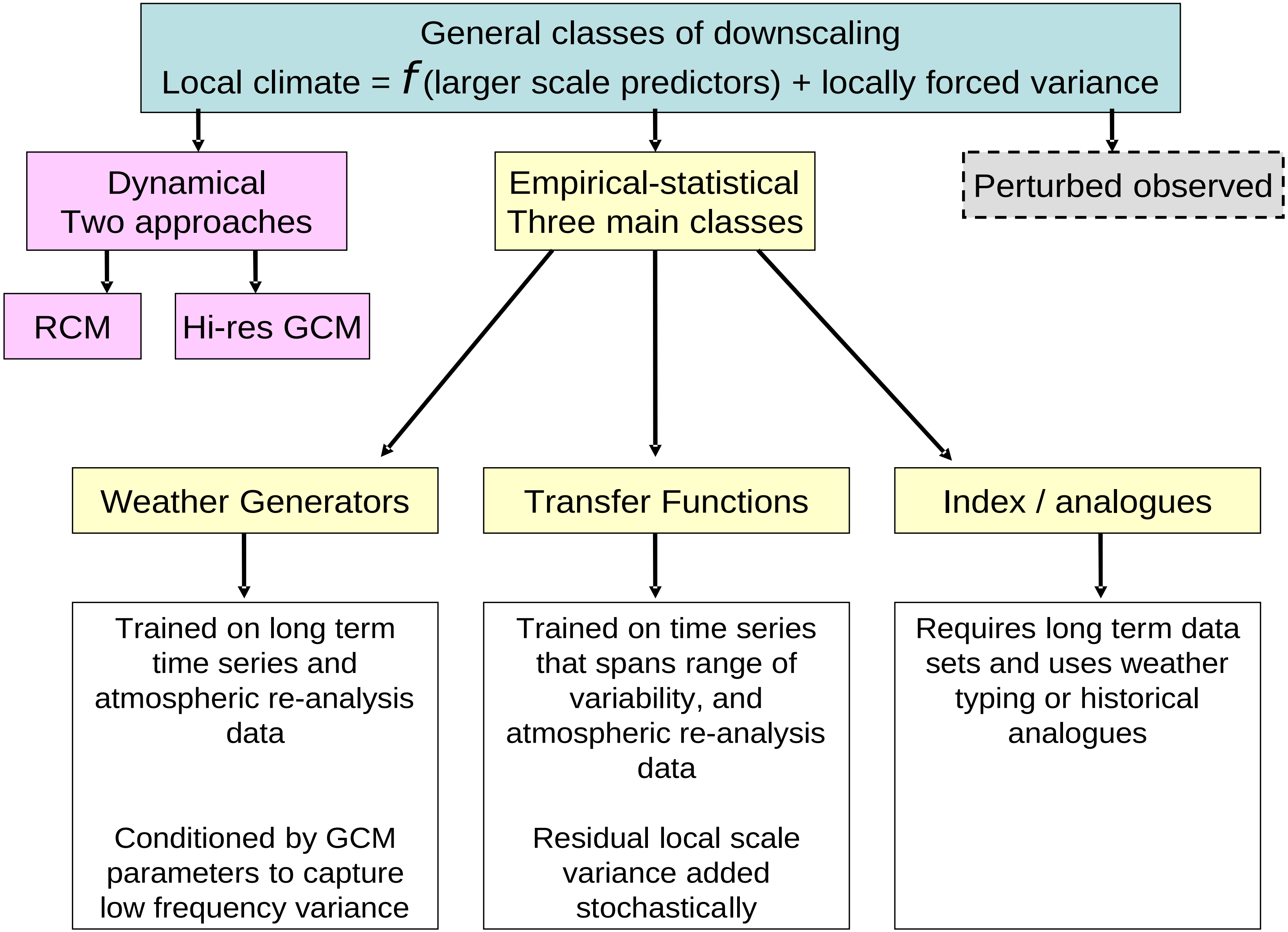
http://www.value-cost.eu/

VALUE: COST Action ES1102 (2012-2015)

Login

CONTRIBUTE TO THE VALIDATION



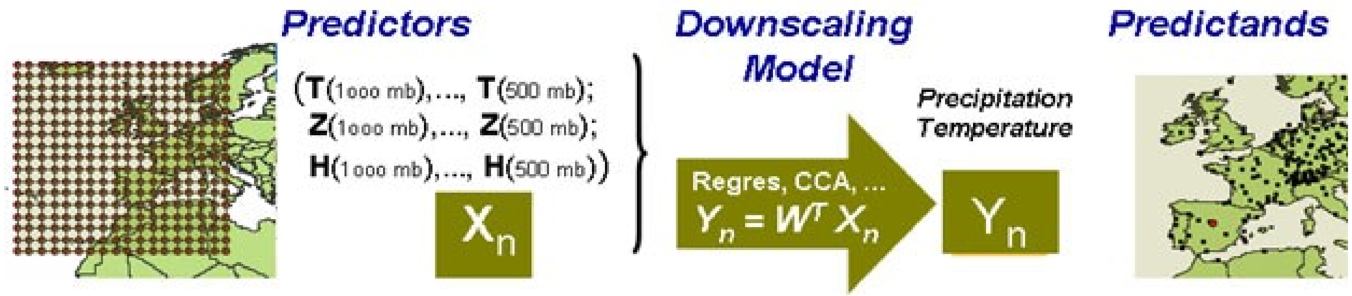


Source:
Bruce Hewitson
(CSAG)

Appro. \ Tech.		Generative		non-Generative	
		Deterministic	Stochastic	Deterministic	Stochastic
PP	Eventwise	Regression, Neural Nets.	GLMs	Analogs, weather types	Analog resampling
	Distribution	Regression on PDF parameters			
MOS	Eventwise	Regression, Neural Nets.	GLMs	Analogs	Analog resampling
	Distribution	Bias correction, parametric q-q map	Nonhomogeneous HMM	q-q map	

Statistical Downscaling

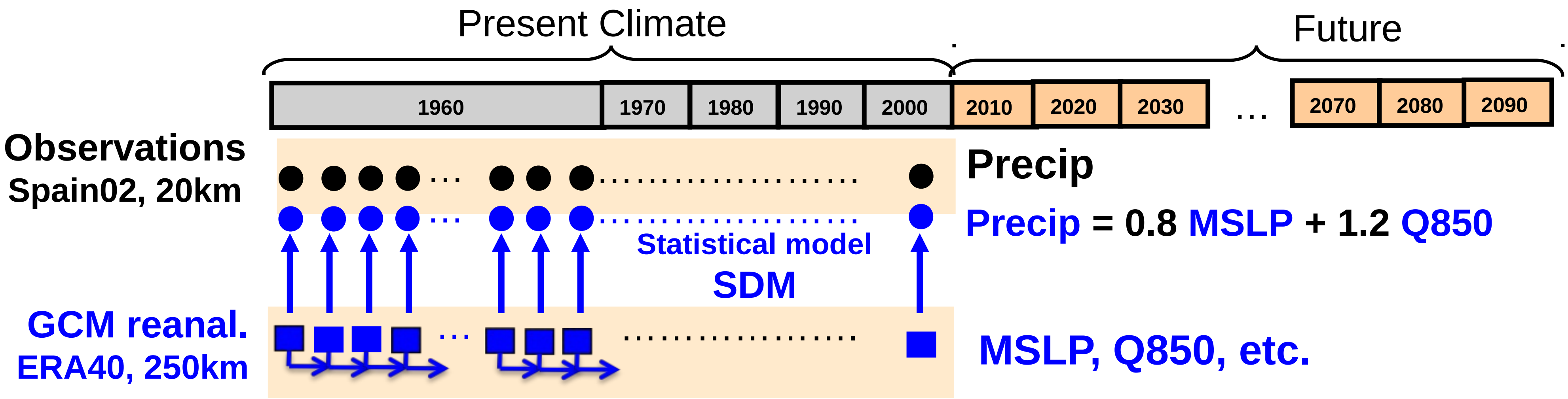
Statistical methods linking the local observed climate (**predictand *Y***) with the global simulations given by the GCMs (**predictors *X***), through some **function *f*** and/or **parameters** θ : $\mathbf{Y} = f(\mathbf{X};\theta)$



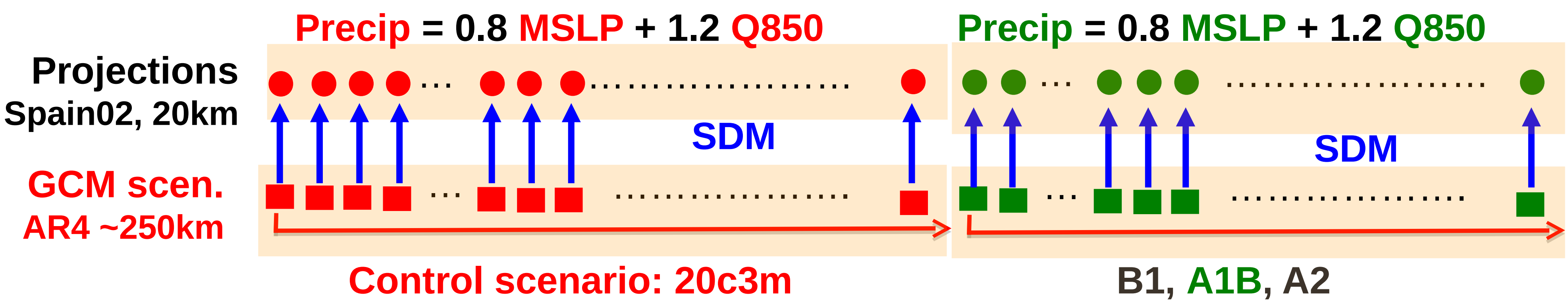
Appro. \ Tech.		Generative		non-Generative	
		Deterministic	Stochastic	Deterministic	Stochastic
PP	Eventwise	Regression, Neural Nets.	GLMs	Analogs, weather types	Analog resampling
	Distribution	Regression on PDF parameters			
MOS	Eventwise	Regression, Neural Nets.	GLMs	Analogs	Analog resampling
	Distribution	Bias correction, parametric q-q map	Nonhomogeneous HMM	q-q map	

Perfect Prognosis (PP): Calibrated in the training phase using observational data for both the predictands and predictors (**reanalysis**). Since different GCMs are used in the training and downscaling phases, large-scale circulation variables well-resolved by the models are typically chosen as predictors in this approach. Variables directly influenced by model parameterizations and orography are not suitable predictors in this approach.

PP	Training	Downscaling
Large scale	Obs.	GCM
Local scale	Obs.	→



- **Assumption 1:** Reanalysis choice
- **Assumption 2:** Choosing consistent predictors:
- **Assumption 3:** Stationarity/robustness: SDM SDM



Present Climate

Journal of Hydrology

A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado

R.L. Wilby[a,*], L.E. Hay[c], G.H. Leavesley[c]

Predictor variable	Abbreviatio
Surface variables	
*Mean sea level pressure	mslp
Zonal velocity component	Us
Meridional velocity component	Vs
Strength of the resultant flow (hPa)	Fs
Vorticity (hPa)	Zs
Divergence (hPa)	Ds
2 m temperatures (°C)	T2m
Relative humidities (%)	RH
*Specific humidity (gm/kg)	SH
Upper-atmosphere variables (500 hPa)	
*500 hPa geopotential heights (m)	H
Zonal velocity component	Uu
Meridional velocity component	Vu
Strength of the resultant flow (hPa)	Fu
Vorticity (hPa)	Zu
Divergence (hPa)	Du

Future

Available online at www.sciencedirect.com
ScienceDirect
Environmental Modelling & Software

Automated regression-based statistical downscaling tool

Masoud Hessami [a,*], Philippe Gachon [b,c], Taha B.M.J. Ouarda [d], André St-Hilaire [d]

No.	Predictors	No.	Predictors
1	Mean sea level pressure	14	500 hPa divergence
2	Surface airflow strength	15	850 hPa airflow strength
3	Surface zonal velocity	16	850 hPa zonal velocity
4	Surface meridional velocity	17	850 hPa meridional velocity
5	Surface vorticity	18	850 hPa vorticity
6	Surface wind direction	19	850 hPa geopotential height
7	Surface divergence	20	850 hPa wind direction
8	500 hPa airflow strength	21	850 hPa divergence
9	500 hPa zonal velocity	22	Relative humidity at 500 hPa
10	500 hPa meridional velocity	23	Relative humidity at 850 hPa
11	500 hPa vorticity	24	Near surface relative humidity
12	500 hPa geopotential height	25	Surface specific humidity
13	500 hPa wind direction	26	Mean temperature at 2 m

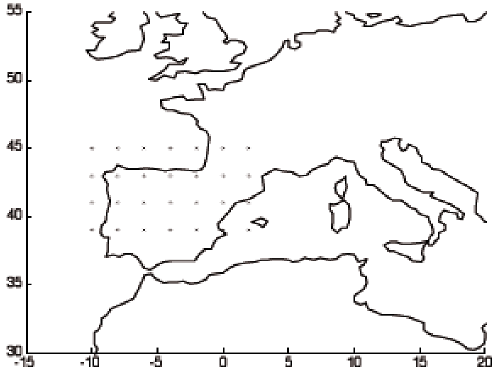


Figura 3. Área de estudio y rejilla utilizada para definir los predictores de los métodos de downscaling estadístico.

Variable	Nivel	Hora
Geopotencial	500	0 UTC
Geopotencial	1000	0 UTC
Temperatura	500	0 UTC
Temperatura	850	0 UTC
Humedad Relativa	850	0 UTC

140 parameters (5 variables, 28 gridboxes), n=16434

Redundancy (correlation):

- Principal Components *(domain selection)*
- Nearest grid-boxes
- Both (e.g. 15 CPs + values in 1 gridbox)

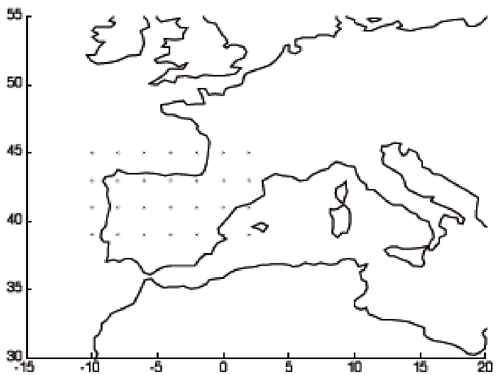


Figura 3. Área de estudio y rejilla utilizada para definir los predictores de los métodos de downscaling estadístico.

- ☹ PP approach strongly depends on the region and variable
- ☹ It is difficult to apply PP methods at continental scale
- ☹ Predictors' screening is a very time consuming task.

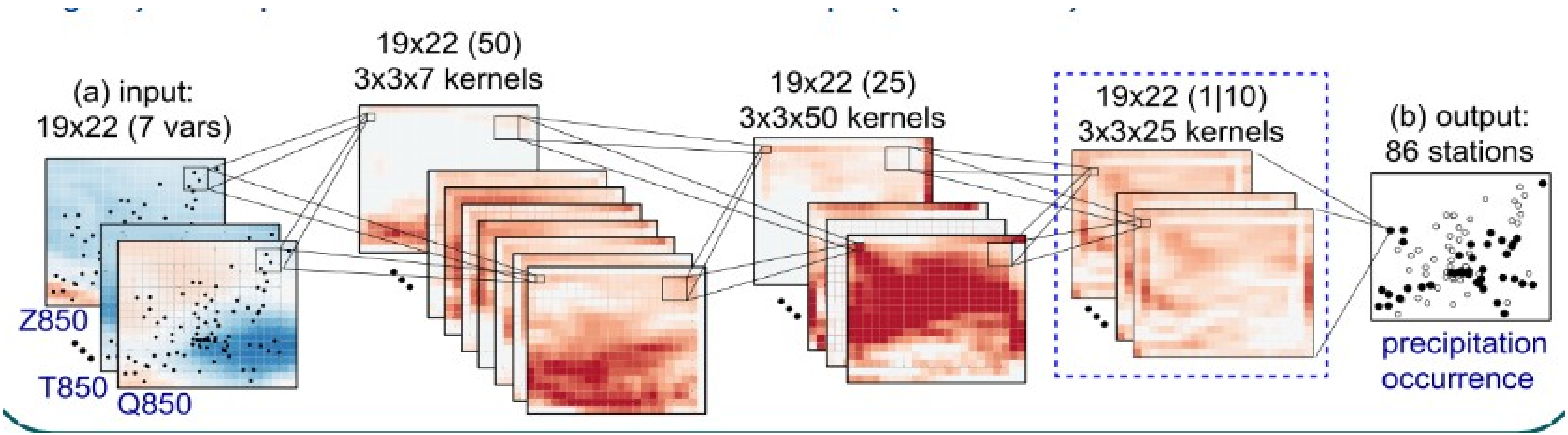
140 parameters (5 variables, 28 gridboxes), n=16434

Redundancy (correlation):

Principal Components *(domain selection)*
Nearest grid-boxes
Both (e.g. 15 CPs + values in 1 gridbox)

Convolution Neural Networks:

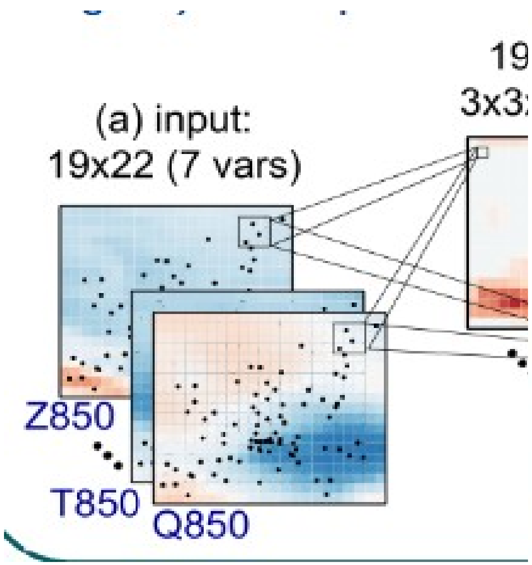
- Simplify the predictor’s screening
- Can be applied to multi-variate problems (variables & location).
-

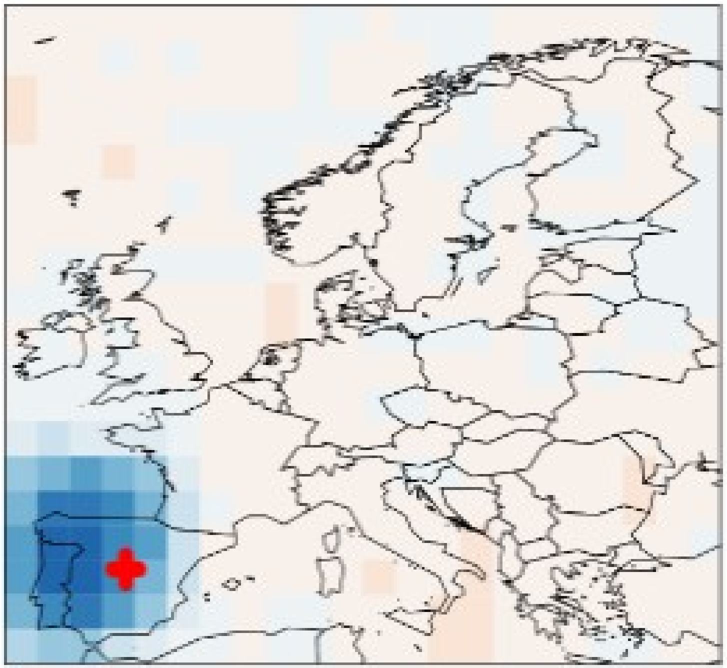


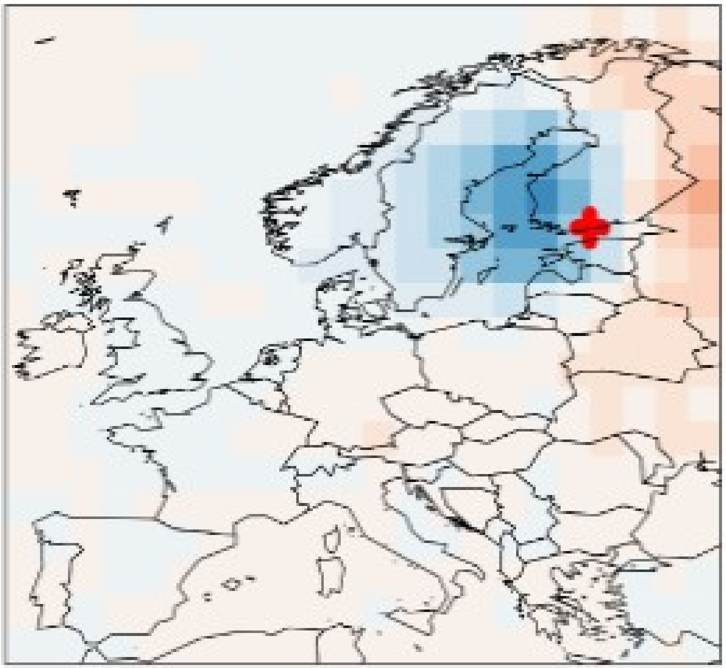
Baño-Medina y Gutiérrez (2018)

Convolution Neural Networks:

- Simplify the predictor's screening
- Can be applied to multi-variate problems (variables & location).
- Can be applied at continental scale





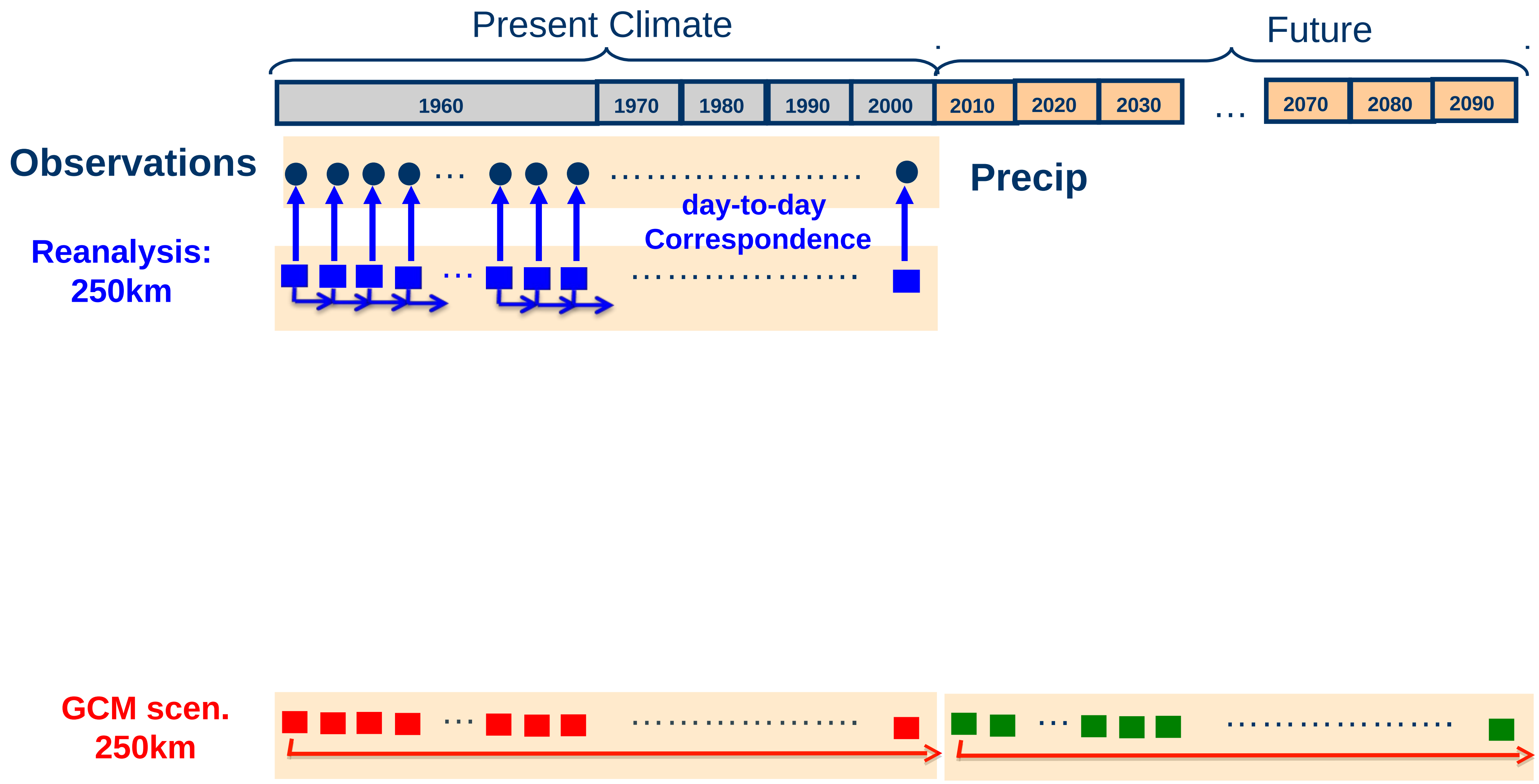


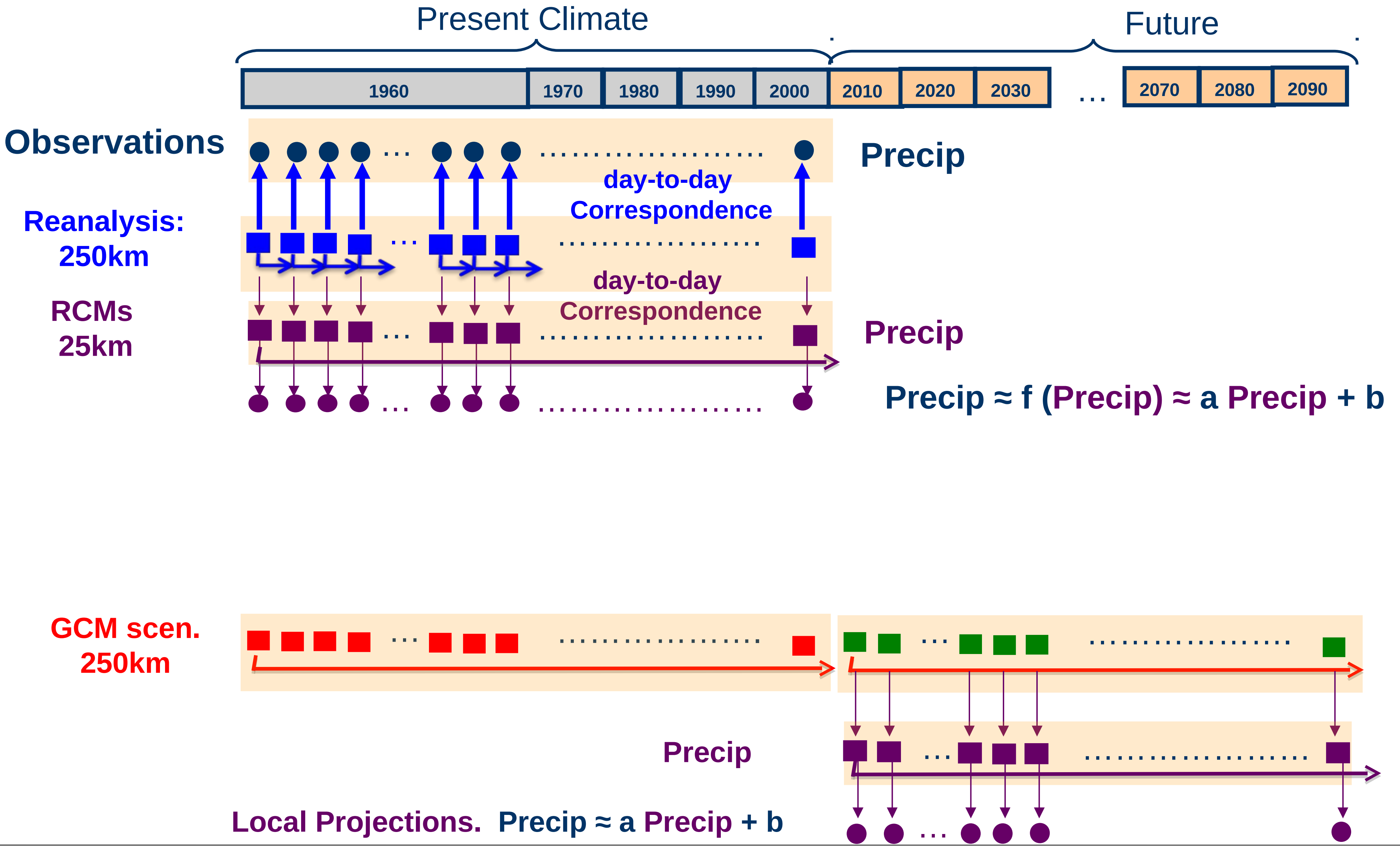
Baño-Medina y Gutiérrez (2018)

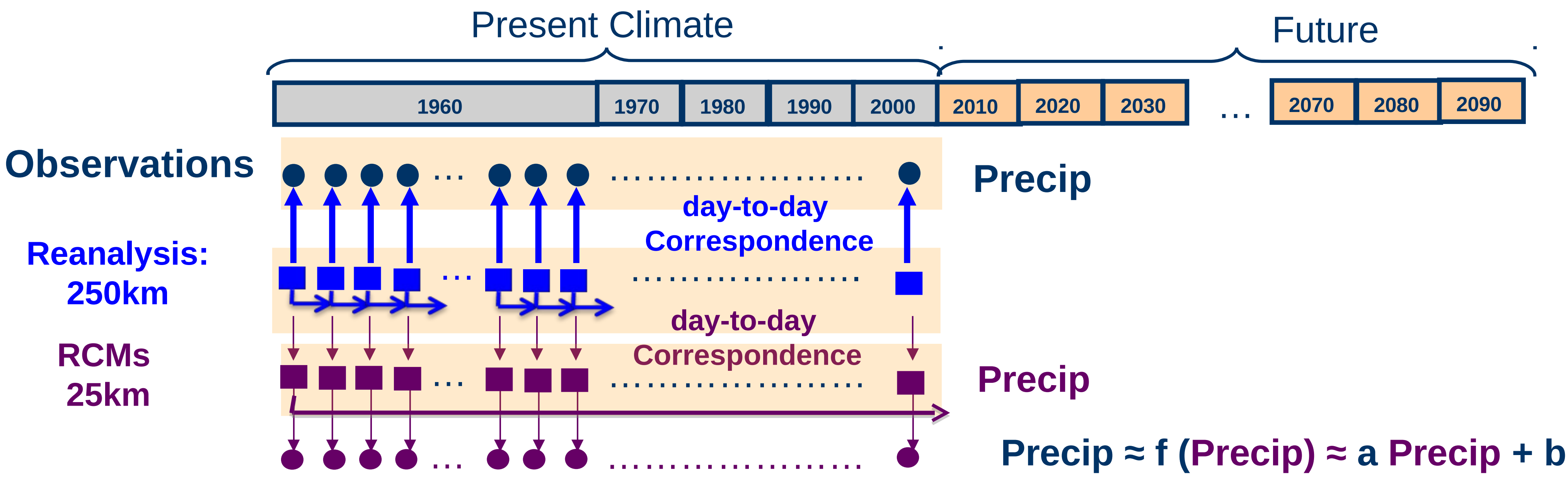
Appro. \ Tech.		Generative		non-Generative	
		Deterministic	Stochastic	Deterministic	Stochastic
PP	Eventwise	Regression, Neural Nets.	GLMs	Analogs, weather types	Analog resampling
	Distribution	Regression on PDF parameters			
MOS	Eventwise	Regression, Neural Nets.	GLMs	Analogs	Analog resampling
	Distribution	Bias correction, parametric q-q map	Nonhomogeneous HMM	q-q map	

Model Output Statistics (MOS): Predictors are taken from the global (or regional) model for both training and downscaling phases. They require the model output to have day-to-day correspondence with observations. These methods can work with the variable of interest as predictor. For instance, local precipitation can be derived from the direct model precipitation forecasts.

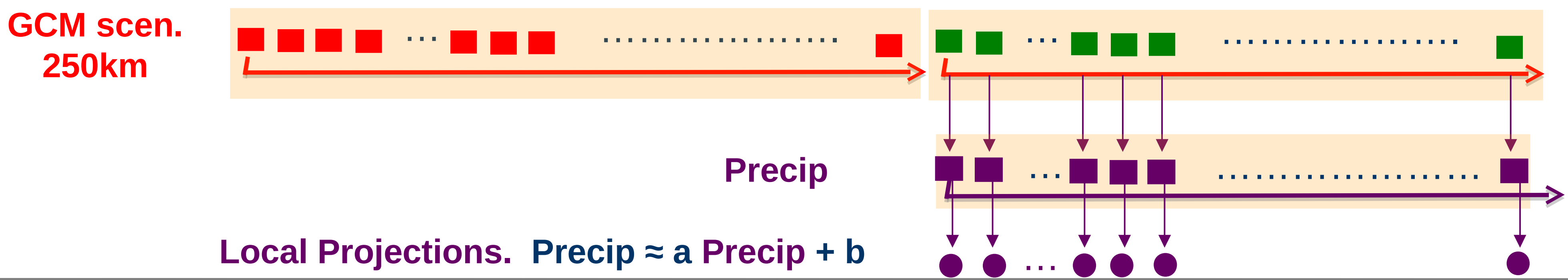
MOS	Training	Downscaling
Large scale	GCM	GCM
Local scale	Obs.	→


Present Climate
Future
1960
1970
1980
1990
2000
2010
2020
2030
…
2070
2080
2090
Observations
Precip
day-to-day
Correspondence
Reanalysis:
250km
GCM scen.
250km


Present Climate
Future
1960
1970
1980
1990
2000
2010
2020
2030
2070
2080
2090
Observations
Precip
Reanalysis: 250km
day-to-day Correspondence
RCMs 25km
day-to-day Correspondence
Precip
Precip ≈ f (Precip) ≈ a Precip + b
GCM scen. 250km
Precip
Local Projections. Precip ≈ a Precip + b



- **Limitation 1:** The Reanalysis-RCM coupling can be done with nudging (perfect correspondence) or in climate mode (only boundaries; partial correspondence). The latter is the typical situation in most databases.



- **Model Output Statistics (MOS):** The model is trained using observations and GCM outputs (which include biases/errors).

$$\mathbf{precip}_{obs}[d] = \mathbf{f}\ (\mathbf{precip}_{gcm}[d])$$

First introduced in weather forecast (Glahn and Lowry, 1972),
but problematic for climate projection.

Adapted for climate projection under the name **"bias-correction"** in a PDF-wise approach:

$$\mathbf{PDF(precip}_{obs}) = \mathbf{F}\ (\mathbf{PDF(precip}_{gcm}))$$

- **Perfect Prognosis (PP):** The model is trained using observations and reanalysis (quasi-observations)**. Predictors are large-scale variables well represented by GCMs**.

$$\mathbf{preci}_{obs}[d] = \mathbf{f}\ (\mathbf{SLP}_{rea}[d], \mathbf{Q850}_{rea}[d])$$

BIASES RESOLUTION

Model Output Statistics (MOS): The model is trained using observations and GCM outputs (which include biases/errors).

$$\mathbf{precip}_{obs}[d] = \mathbf{f}\ (\mathbf{precip}_{gcm}[d])$$

First introduced in weather forecast (Glahn and Lowry, 1972),
but problematic for climate projection.

RCMs

BIASES

Adapted for climate projection under the name **"bias-correction"** in a PDF-wise approach:

$$\mathbf{PDF(precip}_{obs}\mathbf{)} = \mathbf{F}\ (\mathbf{PDF(precip}_{gcm}\mathbf{)})$$

BIASES RESOLUTION

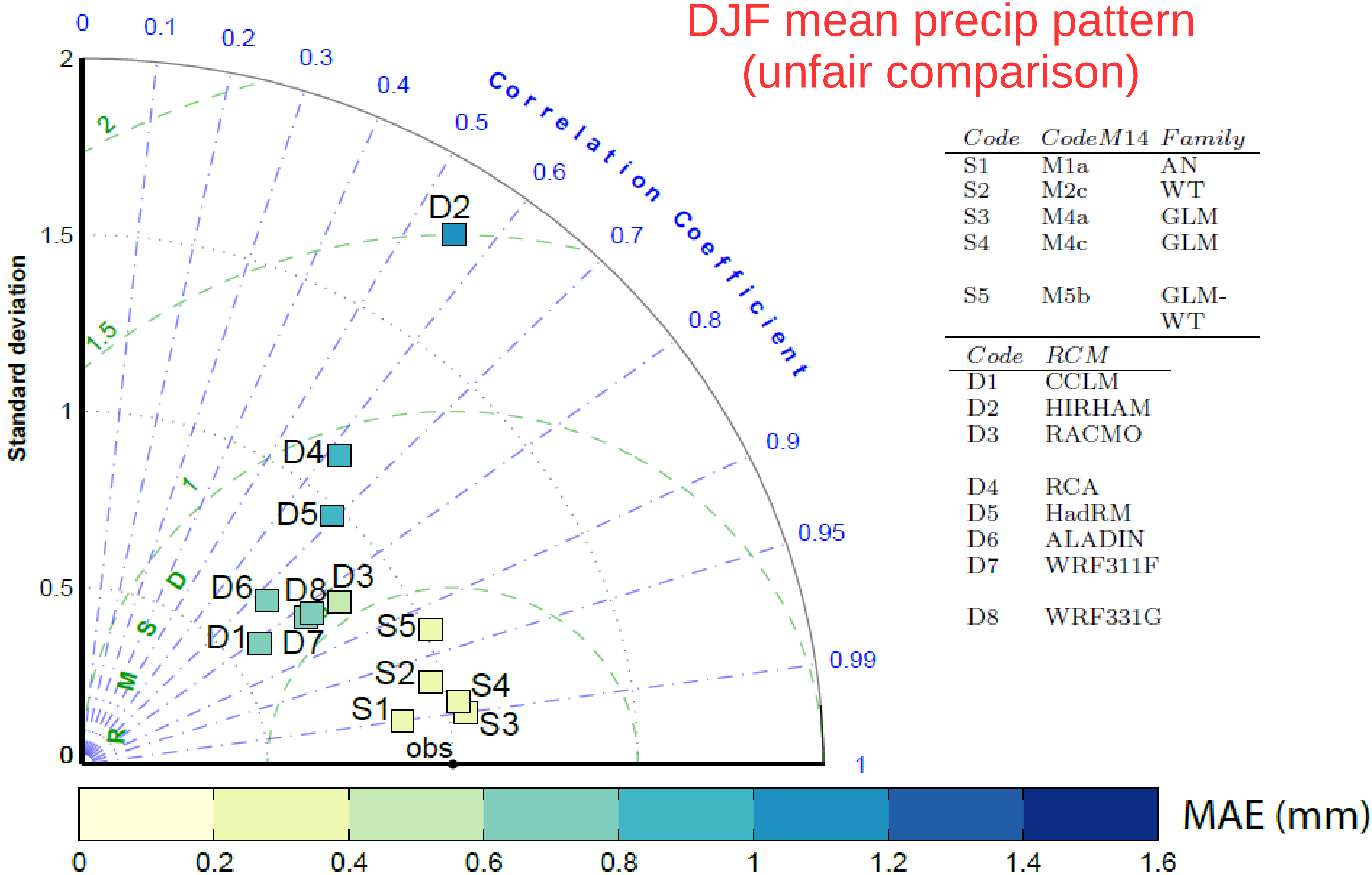
Perfect Prognosis (PP): The model is trained using observations and reanalysis (quasi-observations)**. Predictors are large-scale variables well represented by GCMs**.

$$\mathbf{preci}_{obs}[d] = \mathbf{f}\ (\mathbf{SLP}_{rea}[d], \mathbf{Q850}_{rea}[d])$$

GCMs

Source: Ana Casanueva, UC

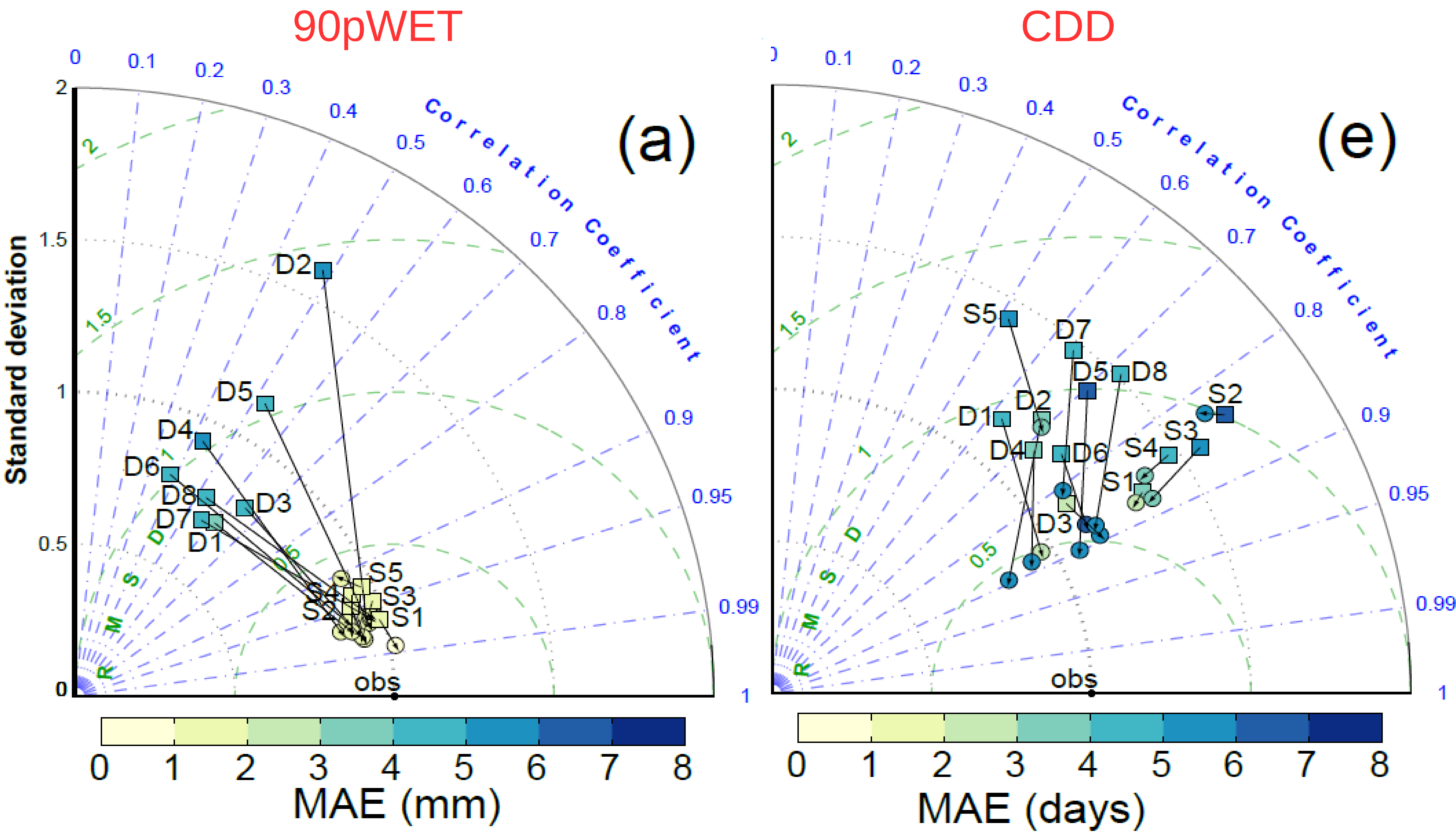
DJF mean precip pattern (unfair comparison)



Code	*CodeM14*	*Family*
S1	M1a	AN
S2	M2c	WT
S3	M4a	GLM
S4	M4c	GLM
S5	M5b	GLM-WT

Code	*RCM*
D1	CCLM
D2	HIRHAM
D3	RACMO
D4	RCA
D5	HadRM
D6	ALADIN
D7	WRF311F
D8	WRF331G

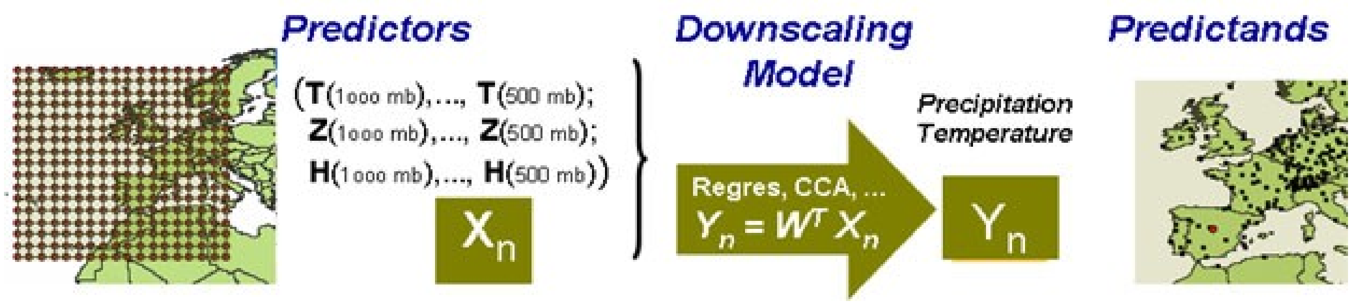
Source: Ana Casanueva, UC



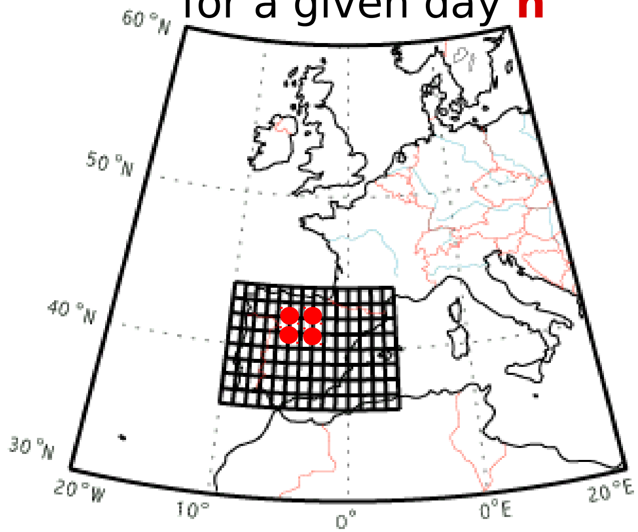
Appro. \ Tech.		Generative		non-Generative	
		Deterministic	Stochastic	Deterministic	Stochastic
PP	Eventwise	Regression, Neural Nets.	GLMs	Analogs, weather types	Analog resampling
	Distribution	Regression on PDF parameters			
MOS	Eventwise	Regression, Neural Nets.	GLMs	Analogs	Analog resampling
	Distribution	Bias correction, parametric q-q map	Nonhomogeneous HMM	q-q map	

Statistical Downscaling

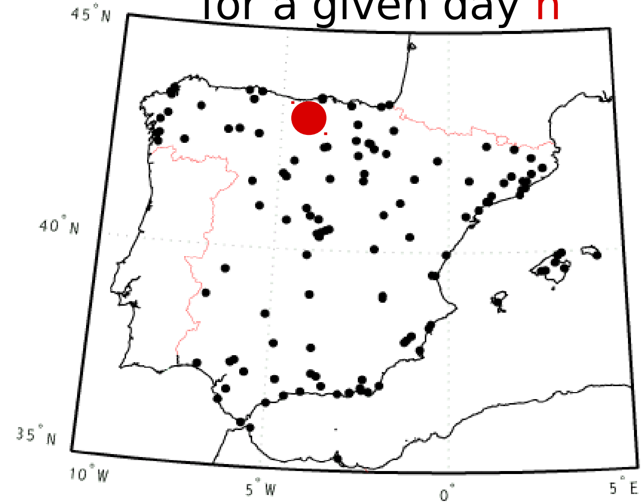
Statistical methods linking the local observed climate (**predictand *Y***) with the global simulations given by the GCMs (**predictors *X***), through some **function *f*** and/or **parameters** θ : $\mathbf{Y} = f(\mathbf{X}; \theta)$



Grids of atmospheric patterns
for a given day **n**



Predictands: *precip*., etc.
for a given day n

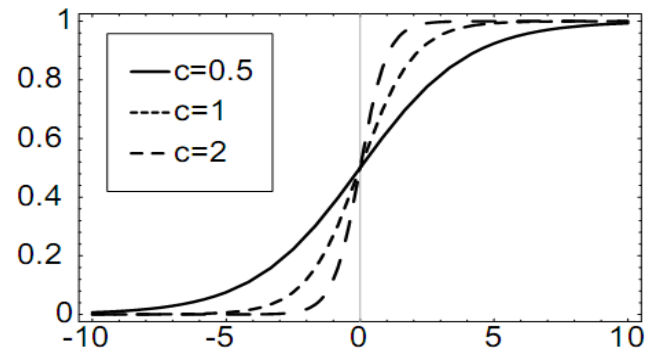


(**T**(1ooo mb),..., **T**(500 mb);
Z(1ooo mb),..., **Z**(500 mb);
.......;
H(1ooo mb),..., **H**(500 mb)) = X_n

Y_n

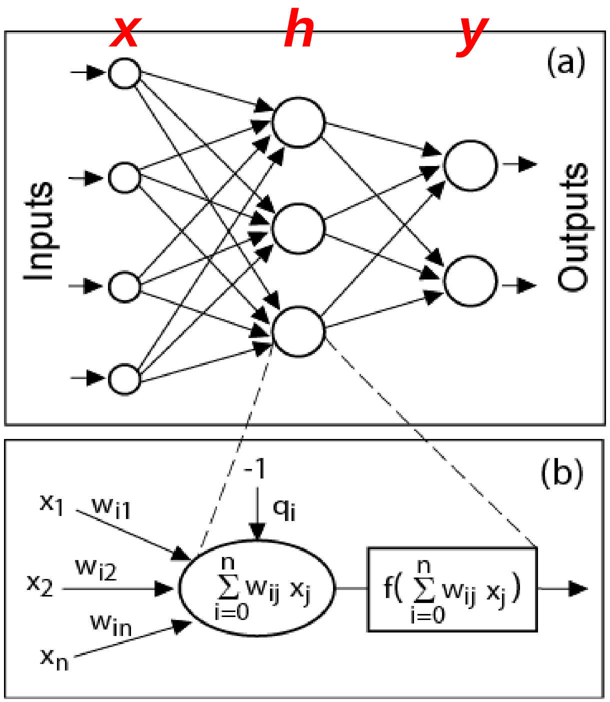
Linear regression:

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



Logistic regression
Probabilistic prediction

$\hat{Y}_n = F(a\, X_n + b)$



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

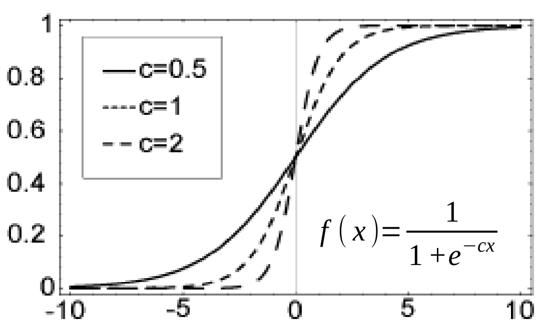
Inputs $\{x_{1p}, \ldots, x_{mp}\}$

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

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

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

nonlinear function



1. Init the neural weight with random values

2. Select the input and output data and train it

3. Compute the error associate with the output

$$\delta_{jp} = (y_{jp} - \hat{y}_{jp})f'(\boldsymbol{\beta}_j^T \hat{\mathbf{h}}_p)$$

4. Compute the error associate with the hidden neurons

$$\psi_{jp} = \sum_k \delta_{jp}\, \beta_{jk}\, f'(\boldsymbol{\alpha}_k^T \mathbf{x}_p)$$

5. Comput

$$\Delta\beta_{jk} = \eta\, \hat{h}_k\, \delta_{jp},\ \Delta\alpha_{ki} = \eta \sum_j x_{ip}\, \delta_{jp}\, \psi_{jp},$$

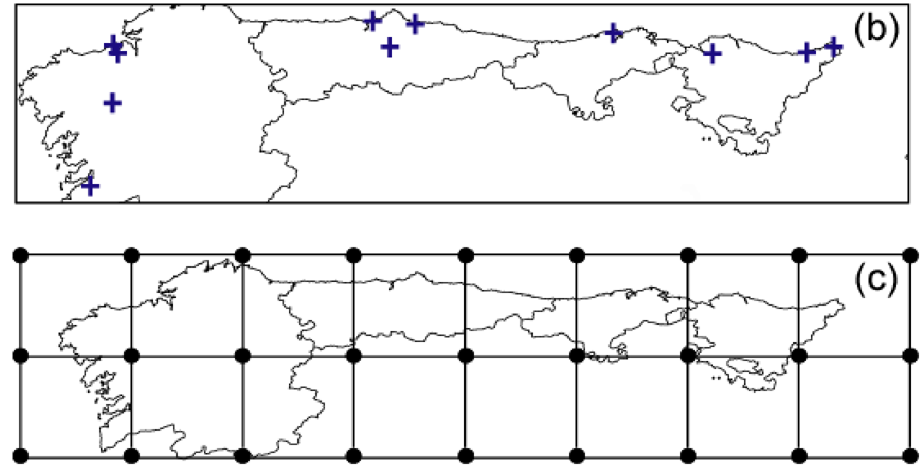
and update the neural weight according to these values

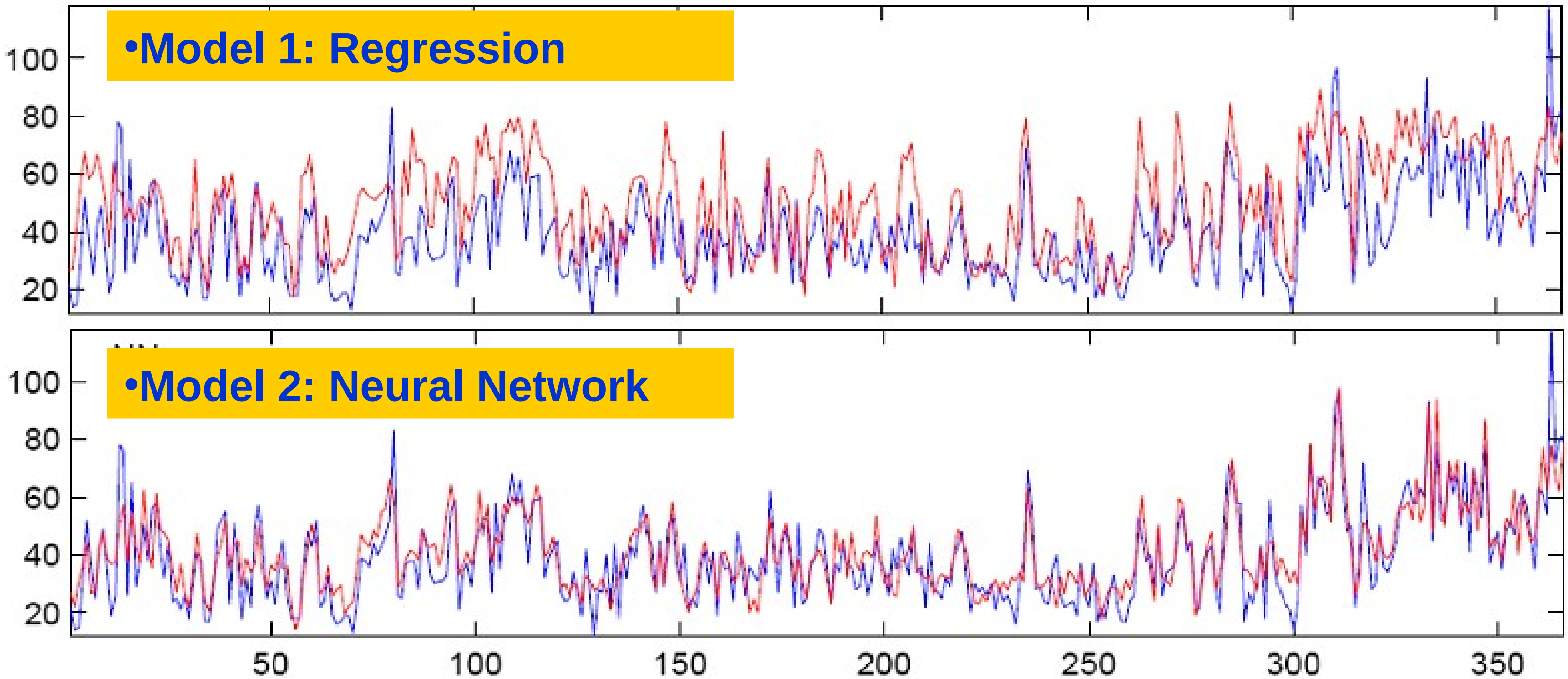
Wind Speed ➔ [0,∞)

Observations from **1977- 2002.**

ERA40 over 27 grid points for the same period

60% for trainning and
40% for validation

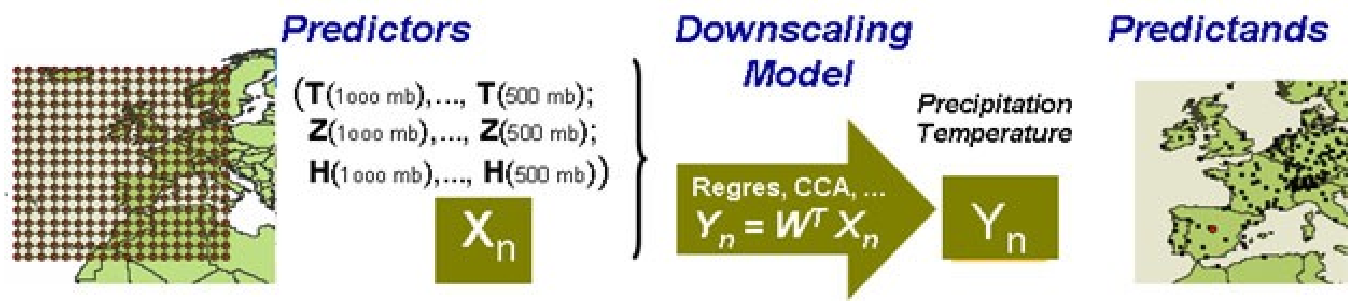




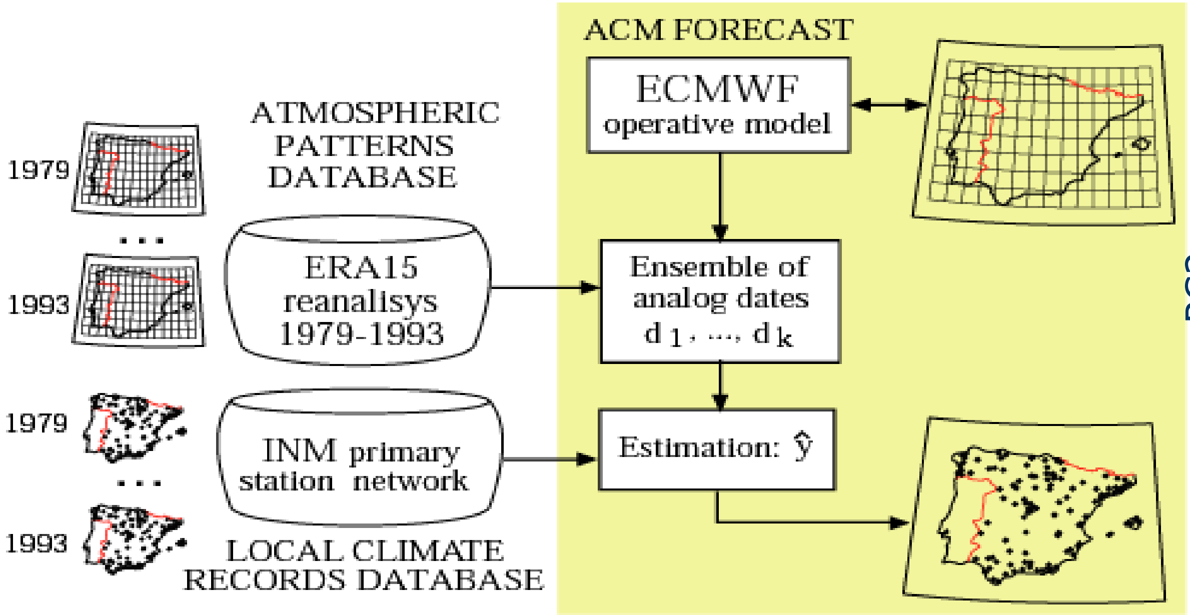
Appro. \ Tech.		Generative		non-Generative	
		Deterministic	Stochastic	Deterministic	Stochastic
PP	Eventwise	Regression, Neural Nets.	GLMs	Analogs, weather types	Analog resampling
	Distribution	Regression on PDF parameters			
MOS	Eventwise	Regression, Neural Nets.	GLMs	Analogs	Analog resampling
	Distribution	Bias correction, parametric q-q map	Nonhomogeneous HMM	q-q map	

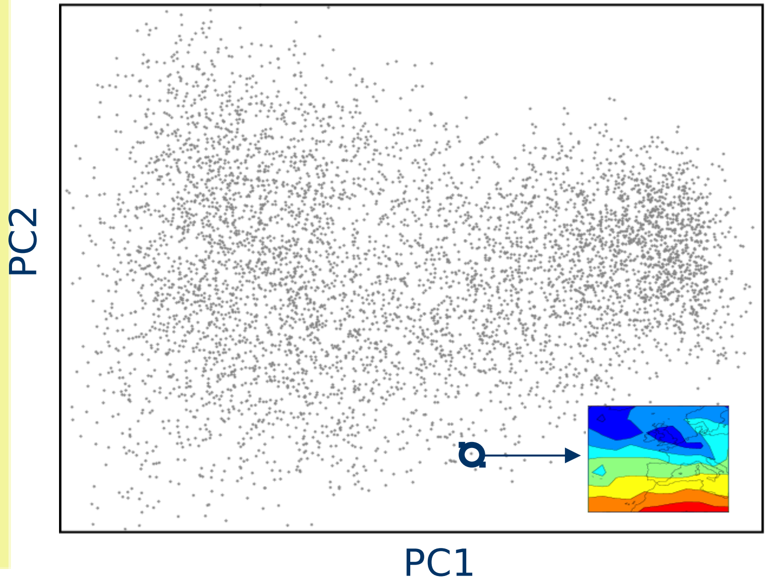
Statistical Downscaling

Statistical methods linking the local observed climate (**predictand *Y***) with the global simulations given by the GCMs (**predictors *X***), through some **function *f*** and/or **parameters** θ : $\mathbf{Y} = f(\mathbf{X};\theta)$

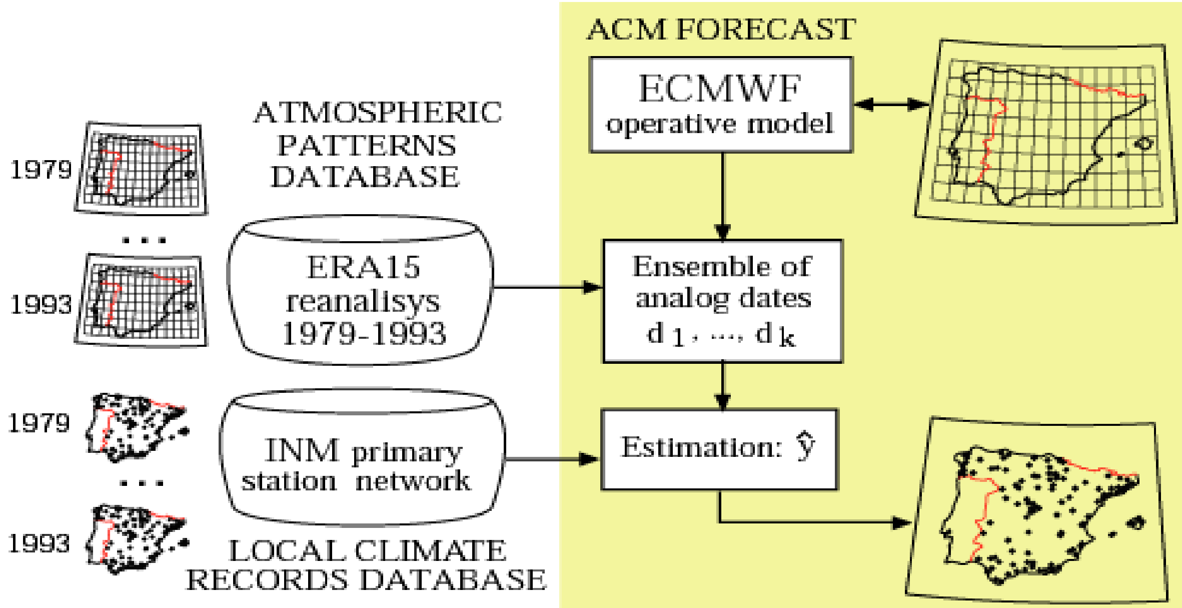


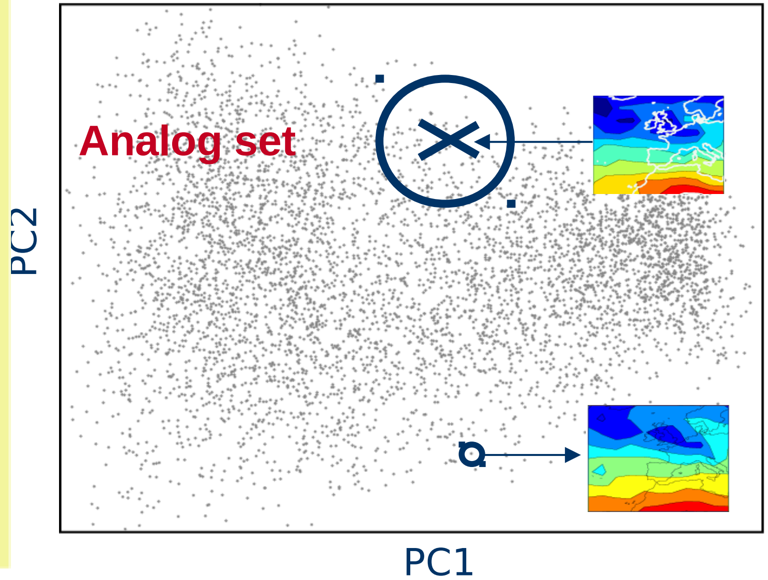
The method of analogs (k-nearest neighbors) is one of the most popular techniques in statistical downscaling, introduced by E. Lorenz (1969).



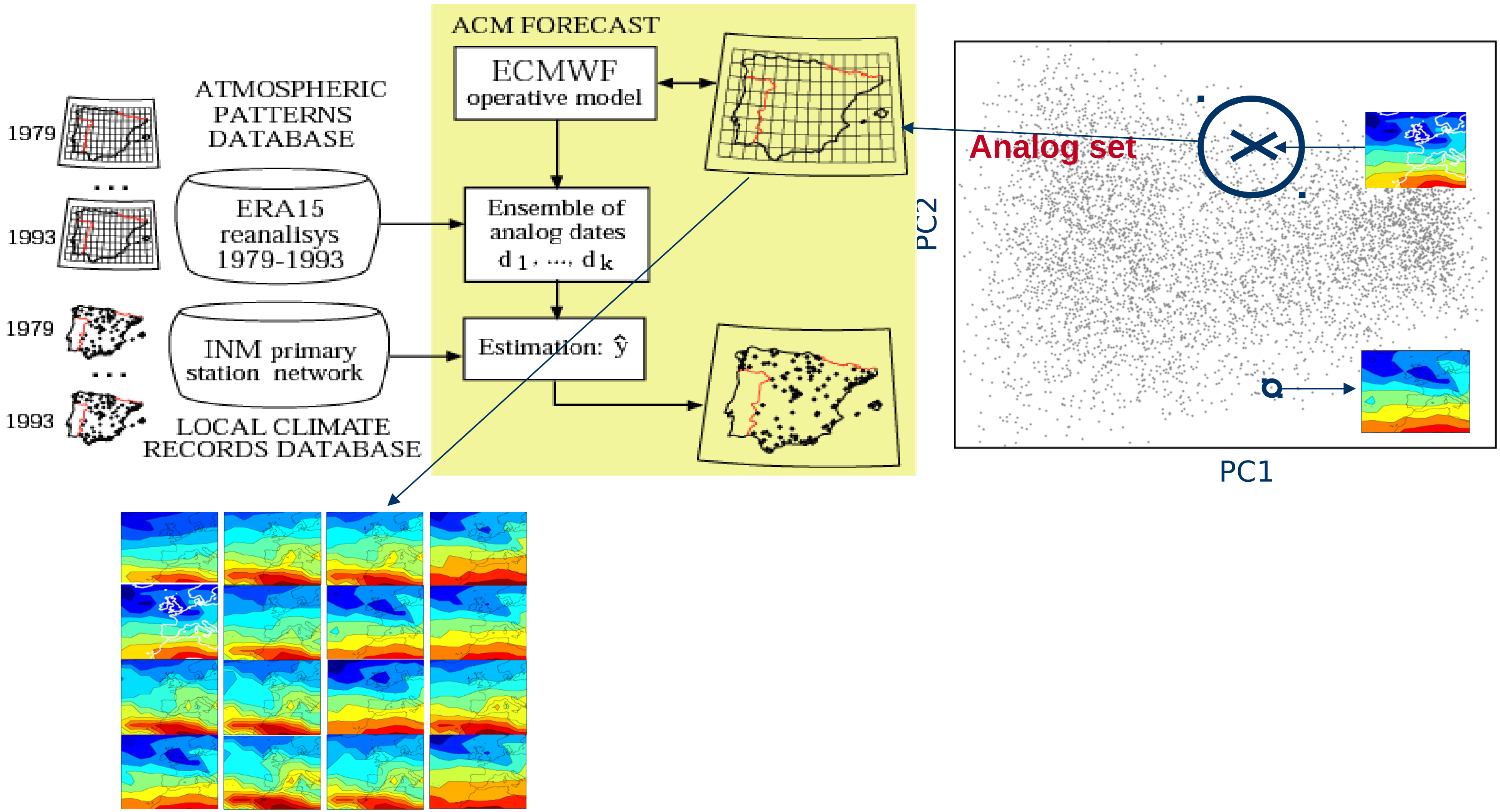


The method of analogs (k-nearest neighbors) is one of the most popular techniques in statistical downscaling, introduced by E. Lorenz (1969).

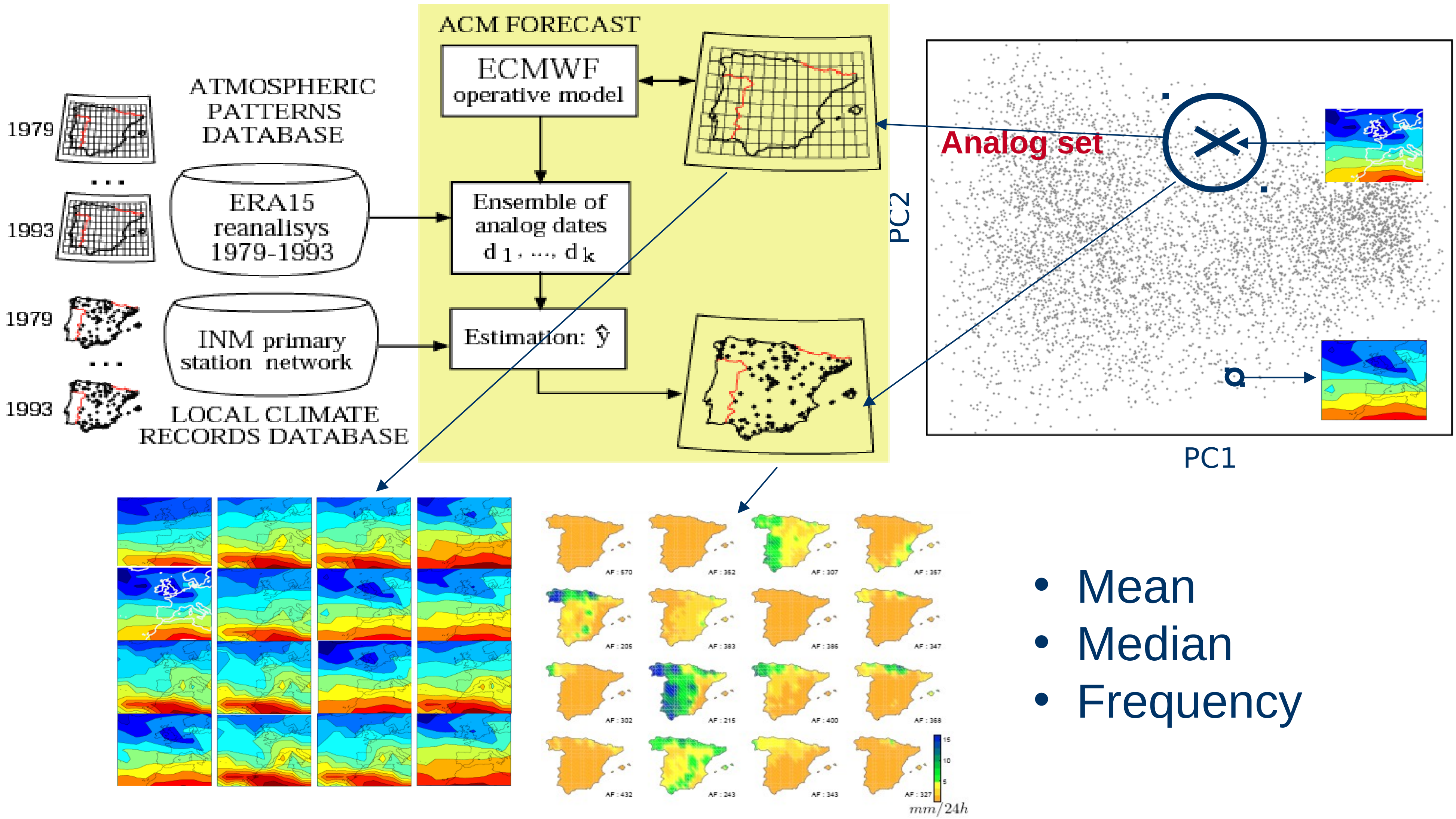




The method of analogs (k-nearest neighbors) is one of the most popular techniques in statistical downscaling, introduced by E. Lorenz (1969).



The method of analogs (k-nearest neighbors) is one of the most popular techniques in statistical downscaling, introduced by E. Lorenz (1969).



- Mean
- Median
- Frequency

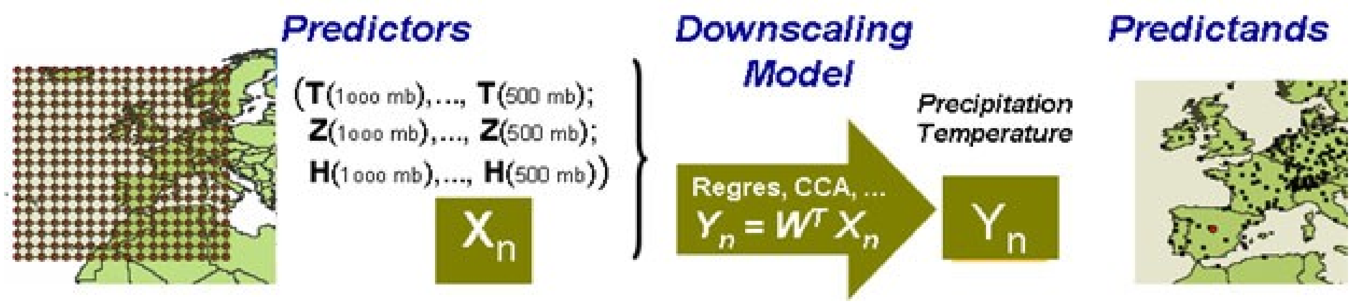
Appro. \ Tech.		Generative		non-Generative	
		Deterministic	Stochastic	Deterministic	Stochastic
PP	Eventwise	Regression, Neural Nets.	GLMs	Analogs, weather types	Analog resampling
	Distribution	Regression on PDF parameters			
MOS	Eventwise	Regression, Neural Nets.	GLMs	Analogs	Analog resampling
	Distribution	Bias correction, parametric q-q map	Nonhomogeneous HMM	q-q map	

There is **day-to-day** correspondence.

There is **not day-to-day** correspondence.

Statistical Downscaling

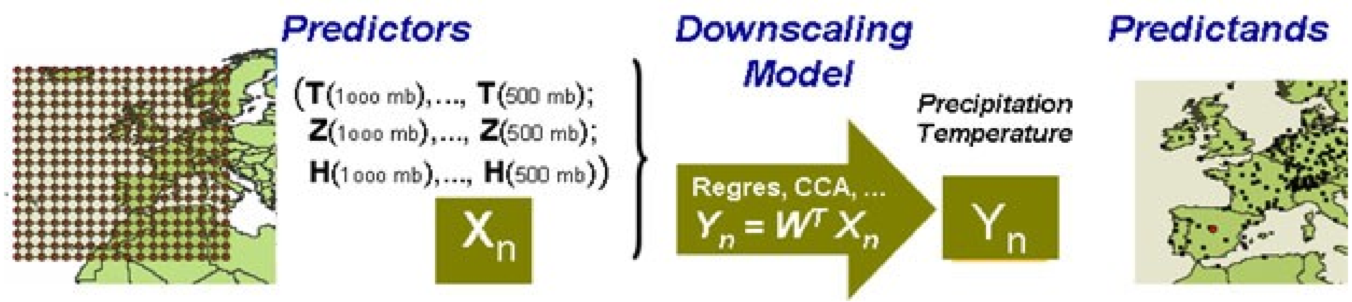
Statistical methods linking the local observed climate (**predictand *Y***) with the global simulations given by the GCMs (**predictors *X***), through some **function *f*** and/or **parameters** θ : $\mathbf{Y} = f\,(\mathbf{X};\theta)$



Appro. \ Tech.		Generative		non-Generative	
		Deterministic	Stochastic	Deterministic	Stochastic
PP	Eventwise	Regression, Neural Nets.	GLMs	Analogs, weather types	Analog resampling
	Distribution	Regression on PDF parameters			
MOS	Eventwise	Regression, Neural Nets.	GLMs	Analogs	Analog resampling
	Distribution	Bias correction, parametric q-q map	Nonhomogeneous HMM	q-q map	

Statistical Downscaling

Statistical methods linking the local observed climate (**predictand *Y***) with the global simulations given by the GCMs (**predictors *X***), through some **function *f*** and/or **parameters** θ : $\mathbf{Y} = f\,(\mathbf{X};\theta)$



1. **Linear Transformation (ax+b):** unbiasing/ scaling (Déqué. 2007; global or local, Widmann et al. 2003; Schmidli et al. 2006), delta change (Hay et al. 2000). These methods were proposed to adjust the arithmetic mean and /or the standard deviation.

2. **Empirical CDF correction:** quantile-quantile mapping and variations (Themeßl et al., 2010; Mengual et al. 2011; Wilcke et al. 2013)

3. **CDF correction:** parametric adjust of the probability/cumulative distribution function (Vidal and Wade, 2008a, 2008b; Piani et al. 2010).

4. **Other approaches:** In the framework of the first Inter-Sectoral Impact Model Intercomparison Project, ISI-MIP, was developed a trend-preserving bias correction method (Hempel et al. 2013) designed to synthesise impact projections in the agriculture, water, biome, health, and infrastructure sectors at different levels of global warming.

Additive Correction

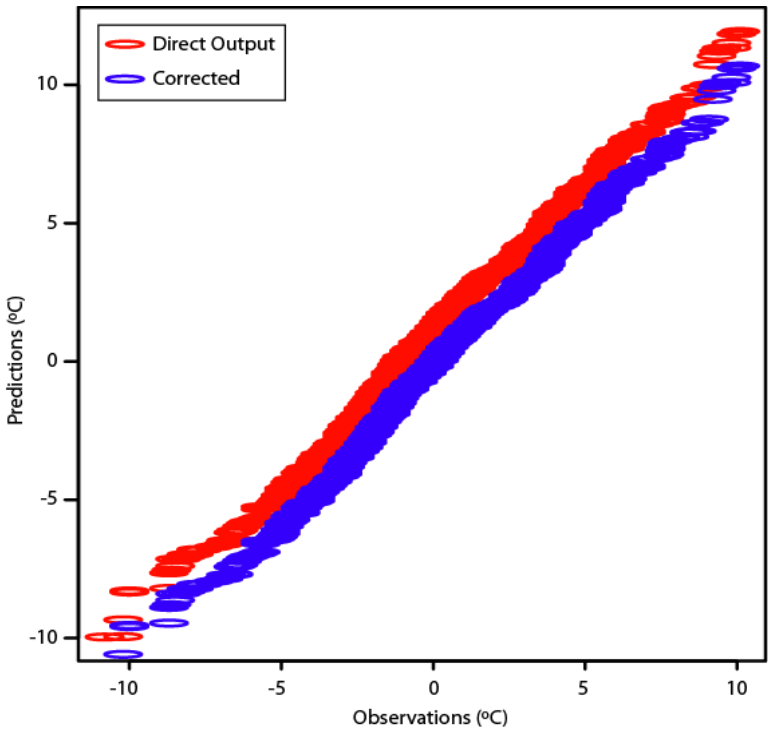
This method has the advantage of simplicity. Two straight-forward corrections consist of adding the climatological difference between future and control climate scenario simulations to an observed baseline (the so-called **delta method**) or removing the bias from future simulation by applying the climatological difference between the observed and control data (the **unbiasing method**; Déqué 2007).

1. The **delta approach** assumes that the variability in the test period remains unchanged.

$$Y_{sim(bc)} = Obs + mean(Y_{test} - Y_{prd})$$

1. The **unbiasing method** assumes that the GCM/RCM variability is perfect.

$$Y_{sim(bc)} = Y_{sim} + mean(Obs - Y_{prd})$$



Multiplicative Correction

This method is the equivalent to the linear correction but for precipitation-like variables. In this case both delta and scaling methods assume that the variability in the climate scenario changes in the same proportion than the mean.

1. **Delta approach**:

$$Y_{sim(bc)} = Obs * mean(Y_{sim})/mean(Y_{prd})$$

2. Scaling method:

$$Y_{sim(bc)} = Y_{sim} * mean(Obs)/mean(Y_{prd})$$

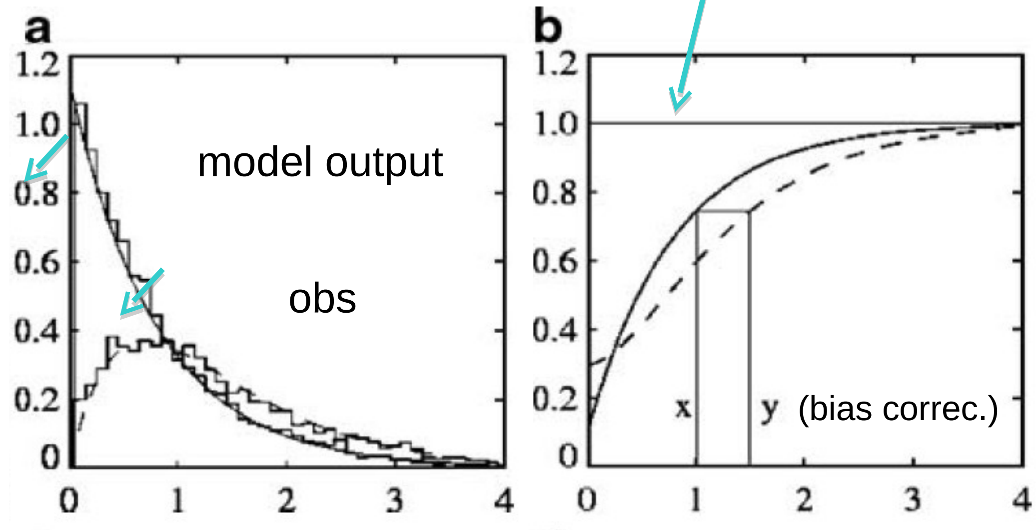
Bias correction: QQ Mapping

Fig. 1 Statistical correction applied to a synthetic dataset. **a** Synthetic pdf of simulated daily precipitation (*solid line*), synthetic pdf of observed daily precipitation (*dashed line*). **b** cdfs obtained by integrating the corresponding pdfs in **a**. **c** Transfer function obtained graphically from **b** by solving: $\mathrm{cdf}_{\mathrm{obs}}(y)=\mathrm{cdf}_{\mathrm{sim}}(x)$ (*thick solid line*). **d**

Source Piani et al. 2010

$$\boxed{\mathrm{cdf}(x) = \int_0^x \frac{e^{\left(-\frac{x}{\theta}\right)} x'^{(k-1)}}{\Gamma(k)\theta^k} dx' + \mathrm{cdf}(0)}$$

$$\boxed{\mathrm{pdf}(x) = \frac{e^{\left(-\frac{x}{\theta}\right)} x^{(k-1)}}{\Gamma(k)\theta^k}}$$



Requires a "reference" PDF !!!!

More sensitive to non-stationarity issues !!!!

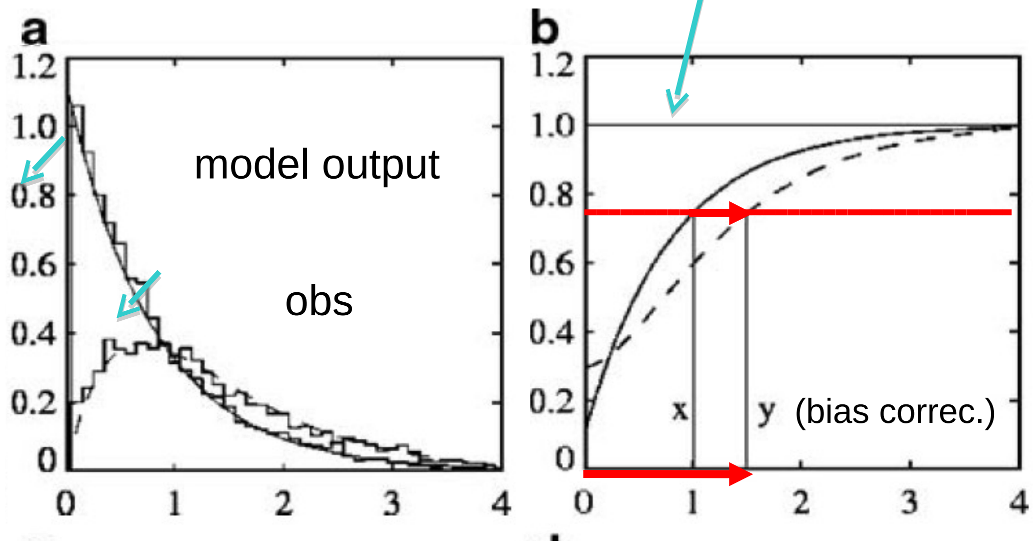
Requires cross-validation (e.g. Gutiérrez et al. 2013)

Fig. 1 Statistical correction applied to a synthetic dataset. **a** Synthetic pdf of simulated daily precipitation (*solid line*), synthetic pdf of observed daily precipitation (*dashed line*). **b** cdfs obtained by integrating the corresponding pdfs in **a**. **c** Transfer function obtained graphically from **b** by solving: $\mathrm{cdf}_{obs}(y)=\mathrm{cdf}_{sim}(x)$ (*thick solid line*). **d**

Source Piani et al. 2010

$$\mathrm{cdf}(x) = \int_0^x \frac{e^{\left(-\frac{x}{\theta}\right)} x'^{(k-1)}}{\Gamma(k)\theta^k} dx' + \mathrm{cdf}(0)$$

$$\mathrm{pdf}(x) = \frac{e^{\left(-\frac{x}{\theta}\right)} x^{(k-1)}}{\Gamma(k)\theta^k}$$



Requires a "reference" PDF !!!!

More sensitive to non-stationarity issues !!!!

Requires cross-validation (e.g. Gutiérrez et al. 2013)

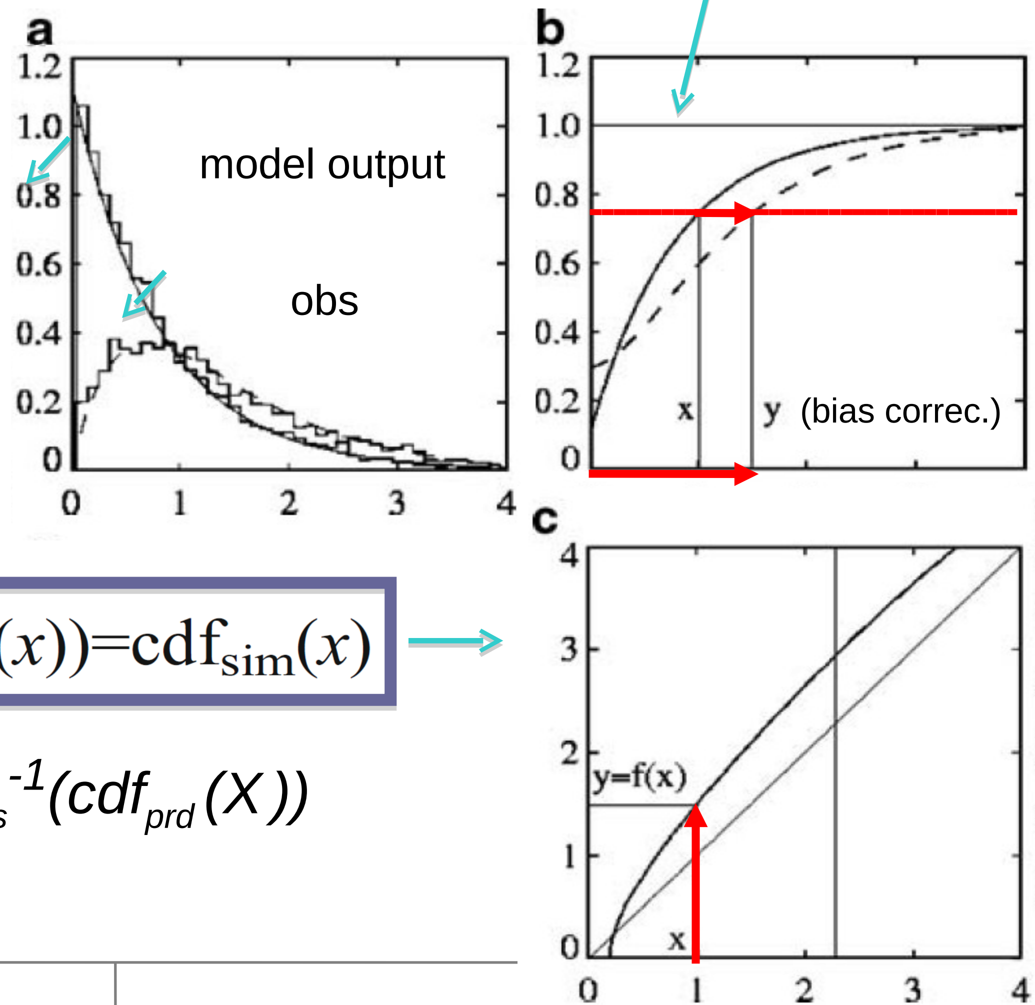
Bias correction: QQ Mapping

Fig. 1 Statistical correction applied to a synthetic dataset. **a** Synthetic pdf of simulated daily precipitation (*solid line*), synthetic pdf of observed daily precipitation (*dashed line*). **b** cdfs obtained by integrating the corresponding pdfs in **a**. **c** Transfer function obtained graphically from **b** by solving: $\mathrm{cdf}_{obs}(y)=\mathrm{cdf}_{sim}(x)$ (*thick solid line*). **d**

Source Piani et al. 2010

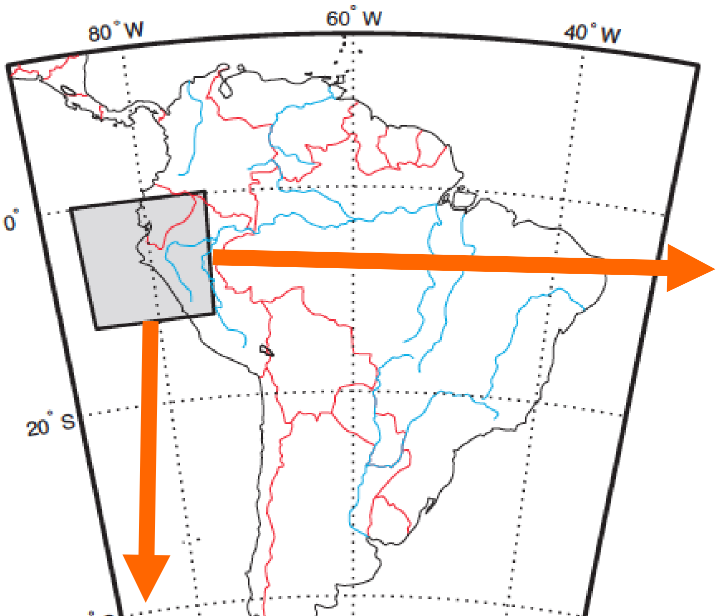
$$\mathrm{cdf}(x) = \int_0^x \frac{e^{\left(-\frac{x}{\theta}\right)} x'^{(k-1)}}{\Gamma(k)\theta^k} dx' + \mathrm{cdf}(0)$$

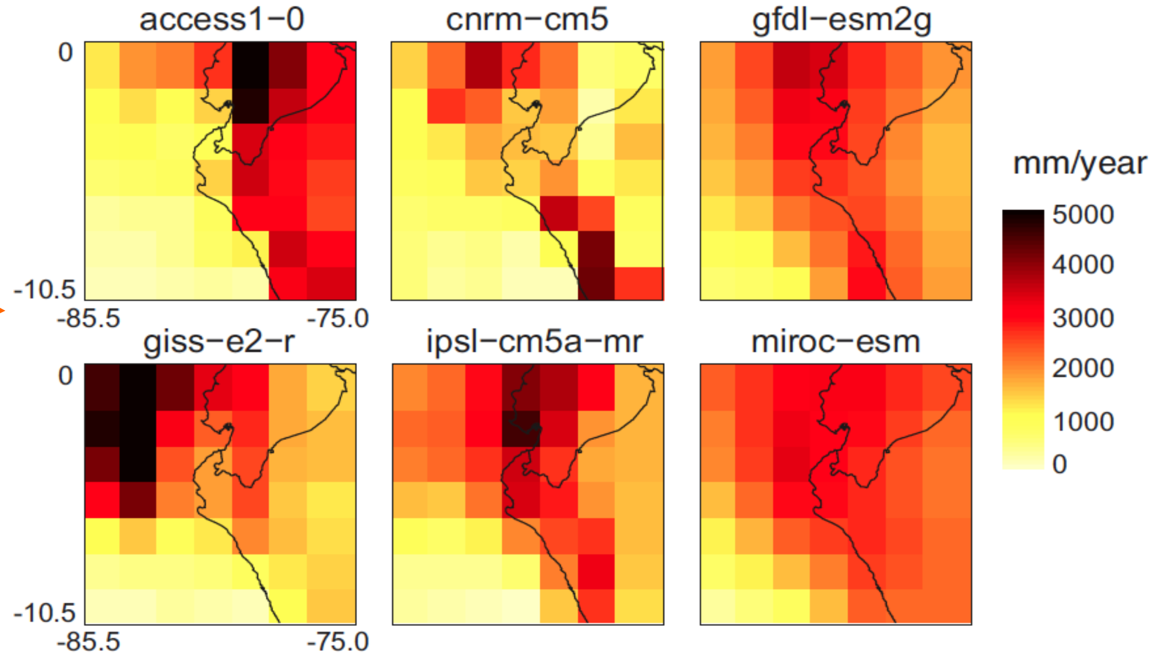
$$\mathrm{pdf}(x) = \frac{e^{\left(-\frac{x}{\theta}\right)} x^{(k-1)}}{\Gamma(k)\theta^k}$$

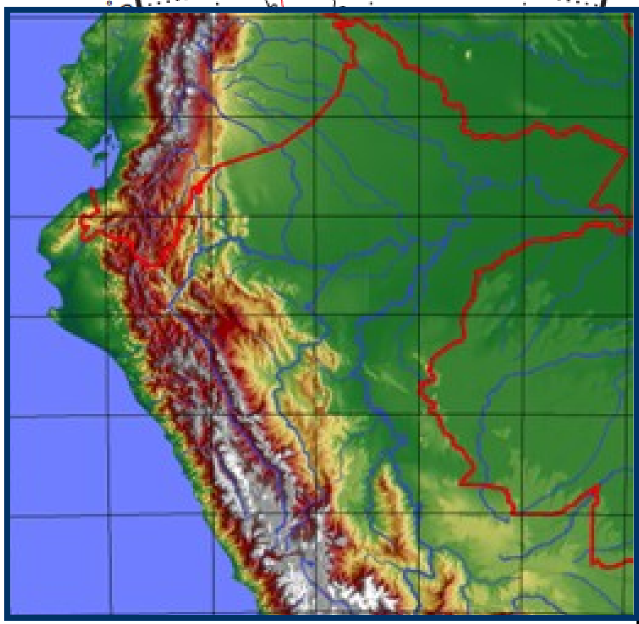


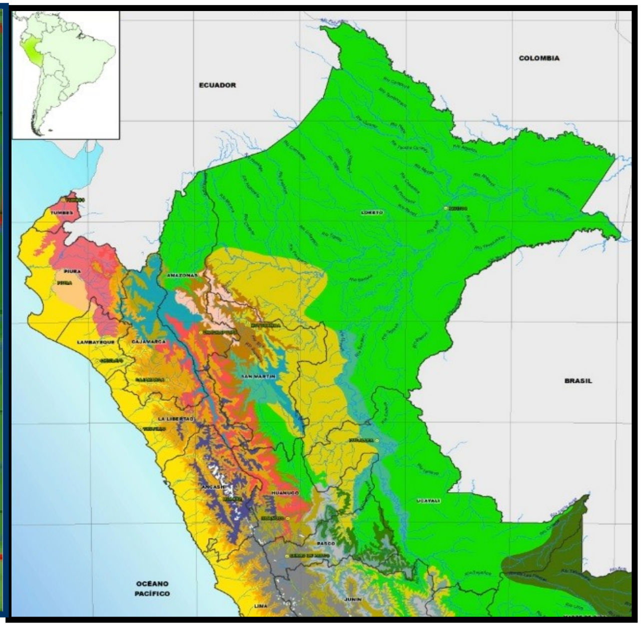
$$\mathrm{cdf}_{obs}(\mathrm{f}(x))=\mathrm{cdf}_{sim}(x)$$

$$y = f(x) = cdf_{obs}^{\;-1}(cdf_{prd}(X))$$





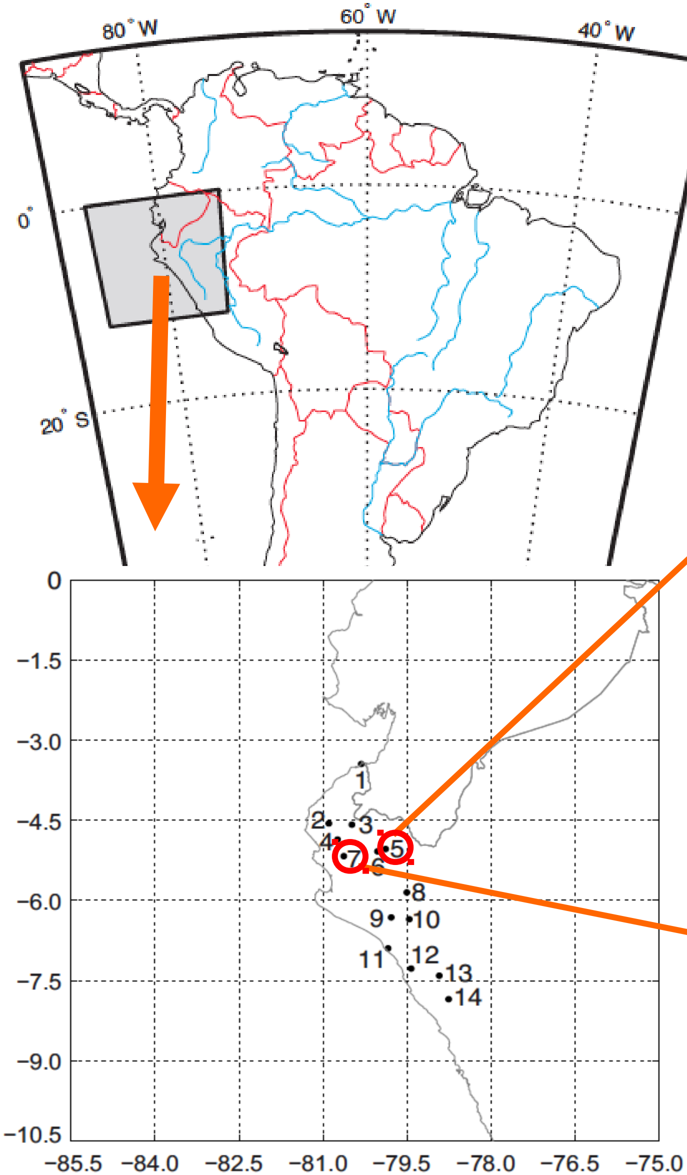


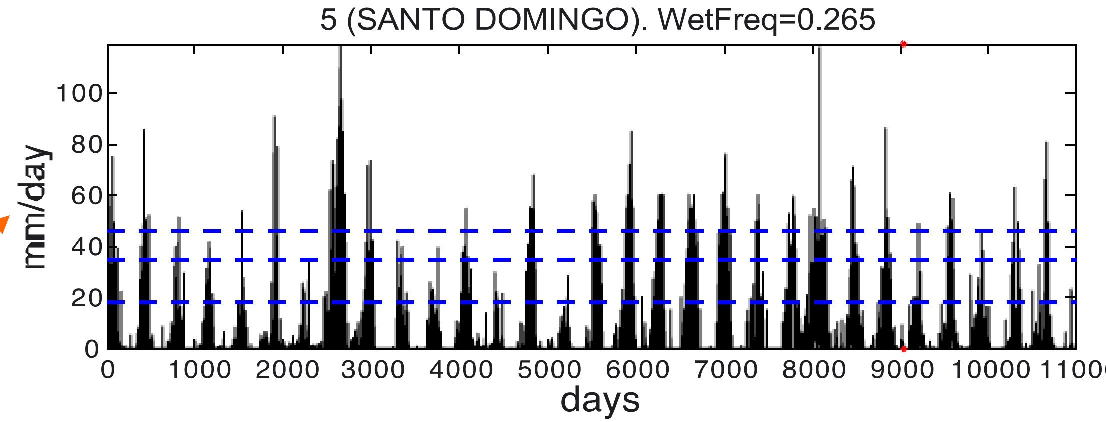


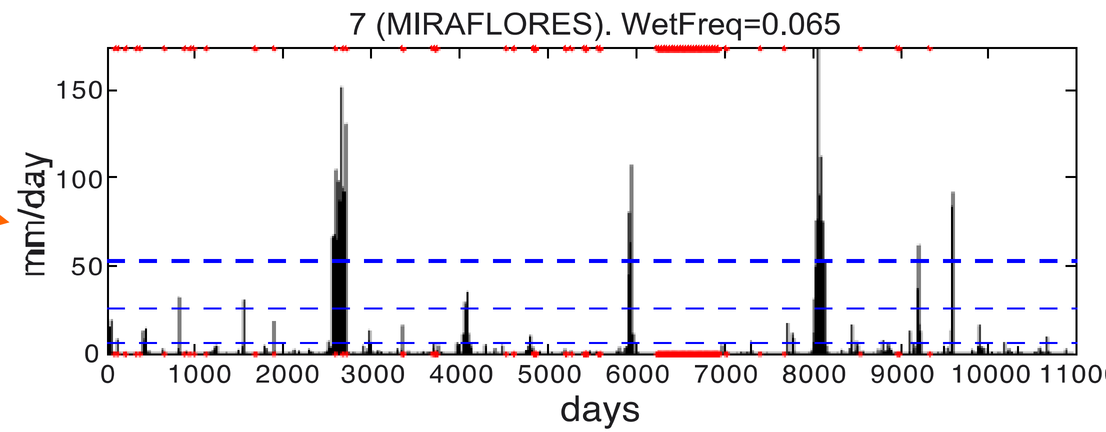
Mean yearly accumulated precipitation from six CMIP5 models over a 1.5º grid (*historical scenario 1976-2005*) ➔ **Different *precip* patterns.**

🡐 elevation and climates

14 stations from SENAMHI with less than 10% of missing data in 1976-2005.
Two main different regimes (ENSO).







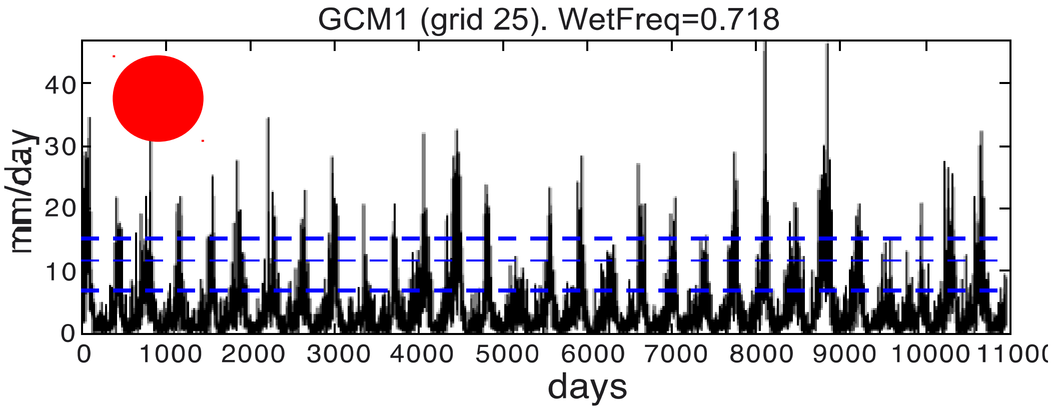
Values over P90 only in El Niño years

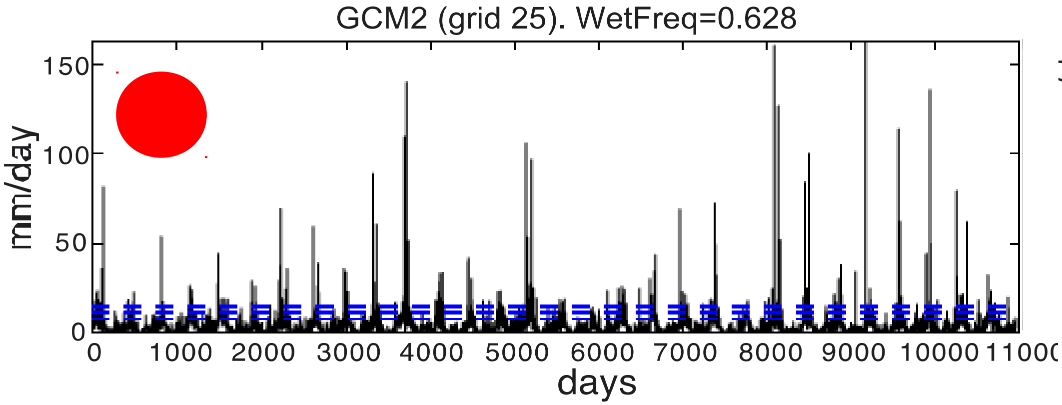
percentiles 75, 90 and 95 of wet days

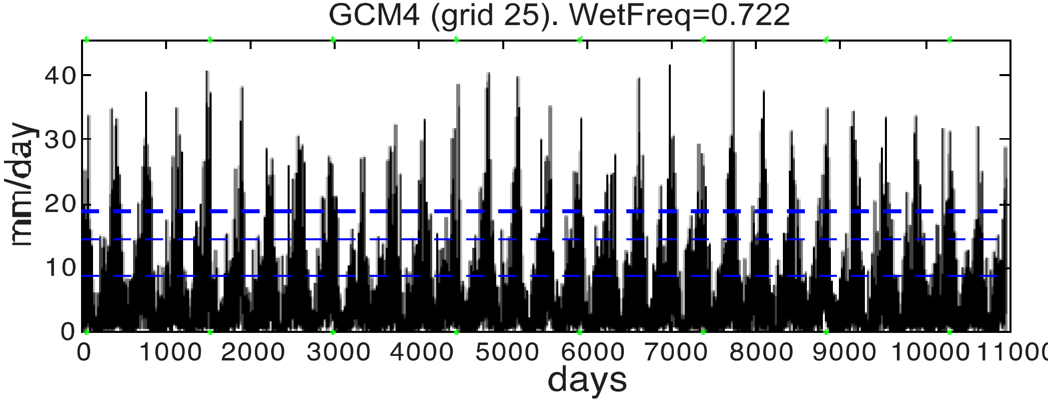
Obs and Sim in the same gridbox

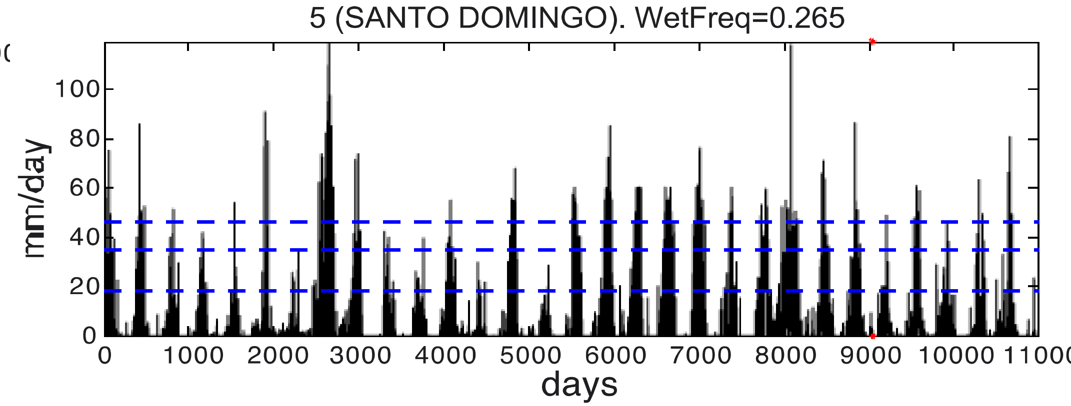
None of the six GCMs exhibits "values over P90 only in El Niño years"

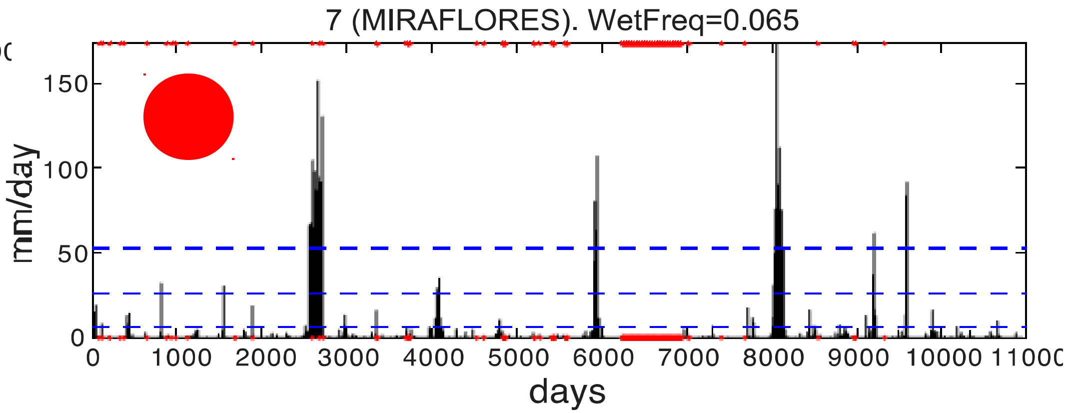
Thus, bias corrected data using *qqmap* will not exhibit El Niño regimes.



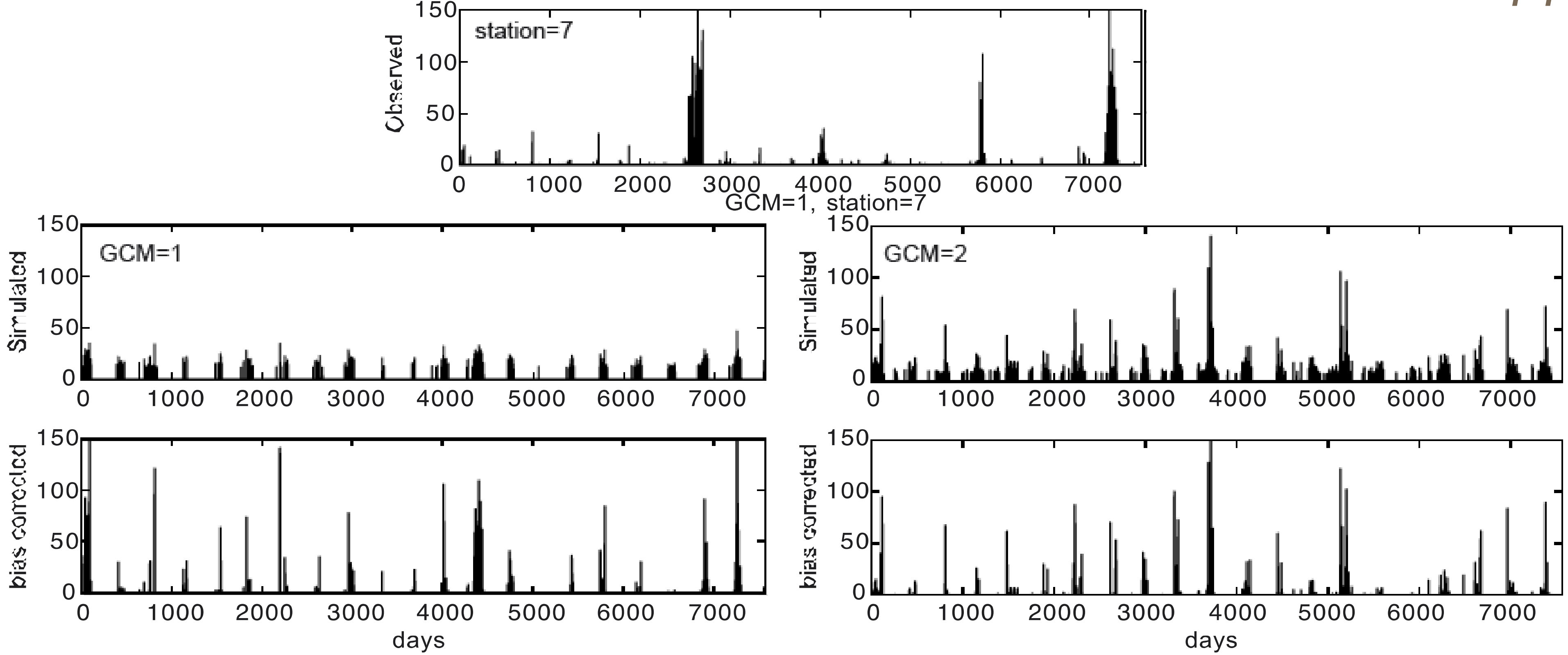


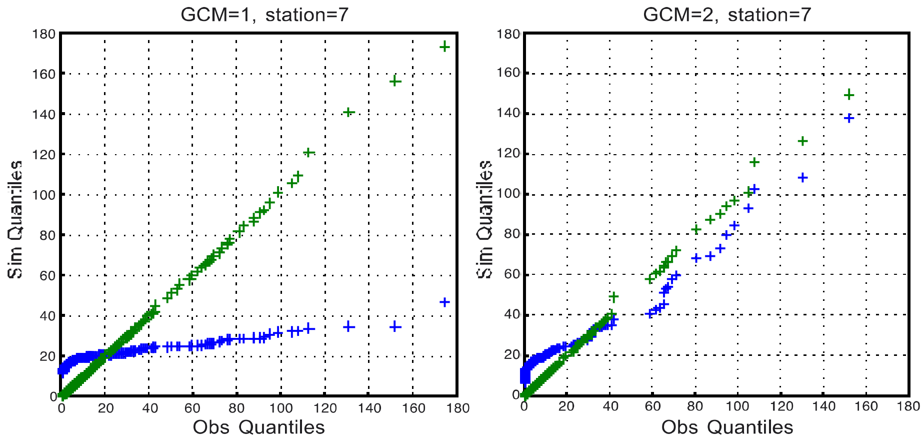






percentiles 75, 90 and 95 of wet days





Bias correction techniques do not add regional weather information. This is given by the GCM.

ISI-MIP

precipitation, radiation, wind speed and surface pressure

Temperature

snowfall, max. and min. temperature,wind components

Multiplicative[2]

Additive[1]

Dependent Variables[3]

Monthly Mean Correction

Daily Variability Correction

Freq. adjust and redistribute drizzle

Correction of precip. amount of wet days

Linear regression of the rank ordered daily anomalies.

variable name
average temperature[1]
minimum temperature[3]
maximum temperature[3]
total precipitation[2]
snowfall[3]
shortwave radiation[2]
longwave radiation[2]
near-surface wind speed[2]
near-surface eastward wind[3]
near-surface northward wind[3]
surface pressure[2]

Earth Syst. Dynam., 4, 219–236, 2013
www.earth-syst-dynam.net/4/219/2013/
doi:10.5194/esd-4-219-2013

A trend-preserving bias correction – the ISI-MIP approach

S. Hempel, K. Frieler, L. Warszawski, J. Schewe, and F. Piontek

1. **Calibration's quality** strongly depends and is limited by the **observations' quality**.
2. **Stationary Hypothesis:** it is implicitly assumed that the bias doesn't change when applied to a non-observed period, introducing aditional uncertainties (Raisanen and Raty, 2012; Maraun, 2012).