Preliminary work on the Statistical Emulation of Regional Climate Models

Antoine Doury, Samuel Somot, Sébastien Gadat, Aurelien Ribes, Lola Corre

CNRM, University of Toulouse, Meteo-France, CNRS
antoine.doury@meteo.fr

European Climate Prediction system

Why do we need statistical emulator ?

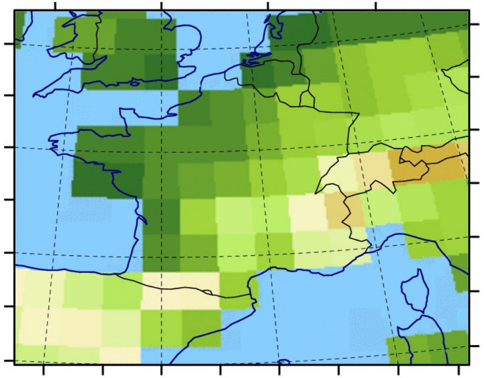
3 kind of climate models :

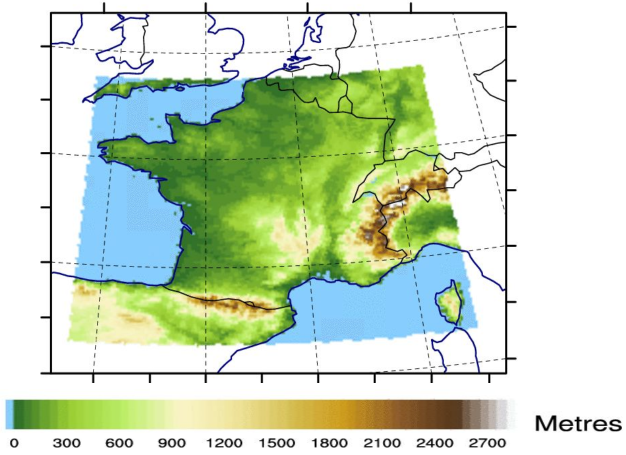
GCM :

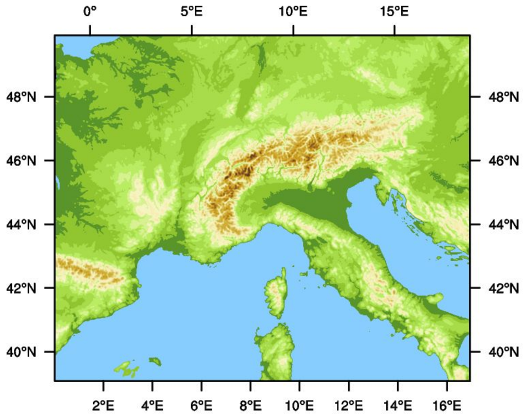
- 100-300 km resolution
- On the whole planet
- Computationally “cheap”

RCM :
- 10-50 km resolution
- Limited area
- Forced at its boundaries by a GCM
- Computationally expensive

CPRCM :
- 1-5 km resolution
- Limited area (smaller than RCMs)
- Often forced by a RCM
- Computationally really expensive







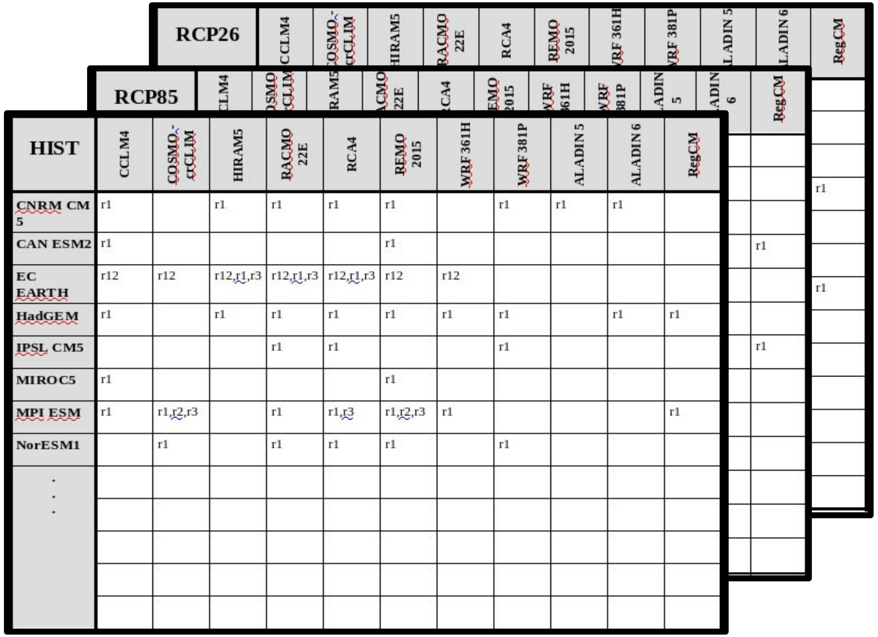
Why do we need statistical emulator ?

To deliver **robust information** on future climate change at **local scales**:

➢ Cover the **full range of uncertainties** about the future climate change signal.

➢ Fill up a [SCENARIO x GCM x RCM] matrix with several members.

➢ Incompatible with the computational costs of Regional Climate Models (even more true with the new generation of CP-RCM).



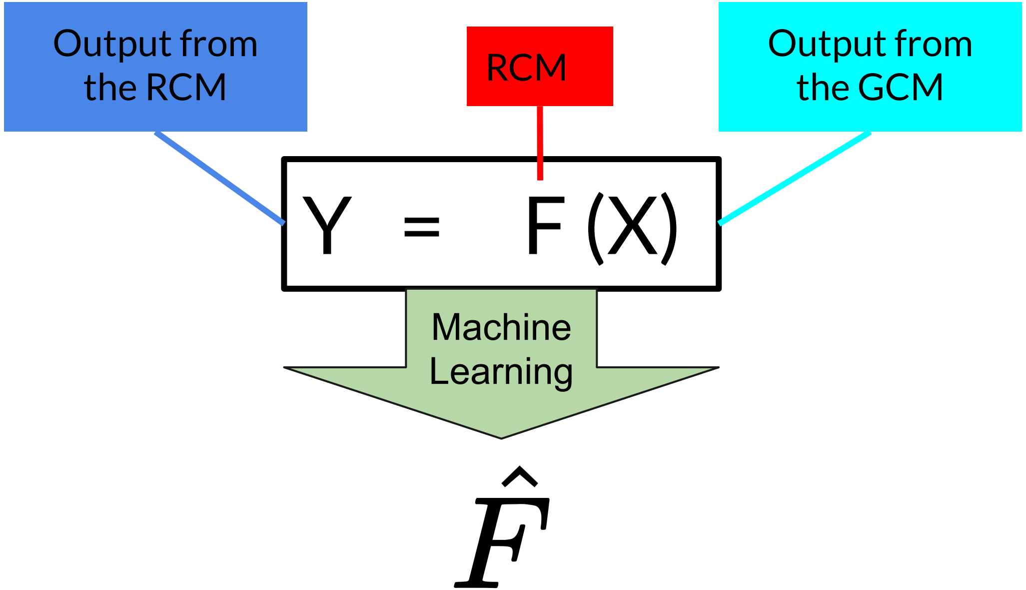
HIST	CCLM4	COSMO-crCLIM	HIRAM5	RACMO 22E	RCA4	REMO 2015	WRF 361H	WRF 381P	ALADIN 5	ALADIN 6	RegCM
CNRM CM 5	r1		r1	r1	r1	r1		r1	r1	r1	
CAN ESM2	r1					r1					
EC EARTH	r12	r12	r12,r1,r3	r12,r1,r3	r12,r1,r3	r12	r12				
HadGEM	r1		r1	r1	r1	r1	r1	r1		r1	r1
IPSL CM5				r1	r1			r1			
MIROC5	r1					r1					
MPI ESM	r1	r1,r2,r3		r1	r1,r3	r1,r2,r3	r1				r1
NorESM1		r1		r1	r1	r1		r1			
. . .											

Euro-Cordex matrix *(Vautard et al 2020, Coppola et al. 2020)*

One solution : Statistical emulation of RCM

Idea : Combine dynamical downscaling and statistical methods (machine learning) to fill up the **[SCENARIO x GCM x RCM]** matrix with new runs and several members.

We **learn** the downscaling function from the existing simulations.



Advantages :

- Learn the future relationship (no question of transferability) and on the whole grid of the RCM (no need of observations).
- Should be able to emulate a new GCM.
- Computationally cheaper than RCMs.

Limitations :

- Reproduce the defaults of the RCMs
- 1 emulator by RCM

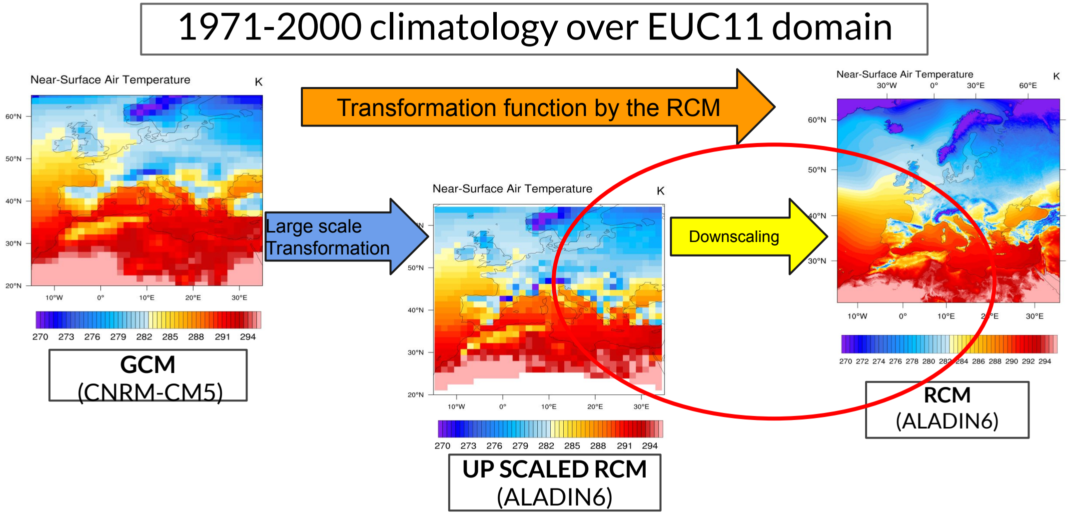
Methods to build the emulator

X coming from an UPSCALED RCM

There are large scale differences between RCM and GCM. Recent papers tend to show that they are due to the lower complexity of the RCM (*Boé et al. 2020, Schwingshackl et al 2019*).

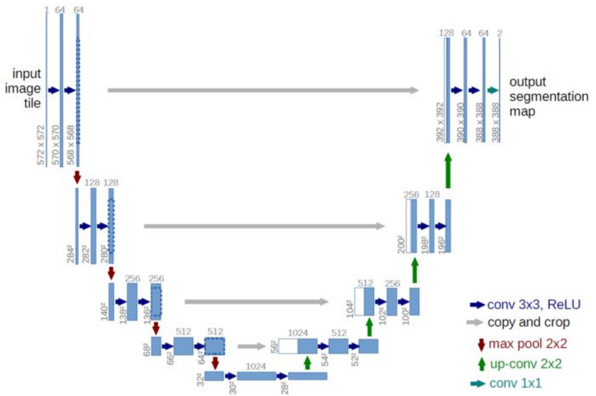
GCM large scale is more trustable than RCM one.

⇒ Focus on the downscaling function of the RCM, train the model with X coming from an upscaled RCM



Machine learning method :Neural Network

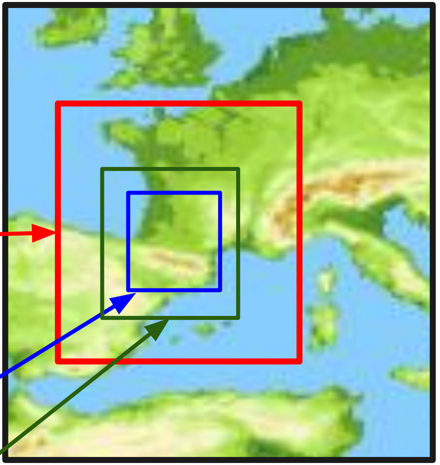
To learn the downscaling function of the RCM we used an adapted U-NET, which is a Neural Network architecture based on Convolutional layers. We selected these kind of architectures for their ability to deal with spatial structures.



U-net architecture, *Ronneberger et al. 2015*

DATA to train the emulator

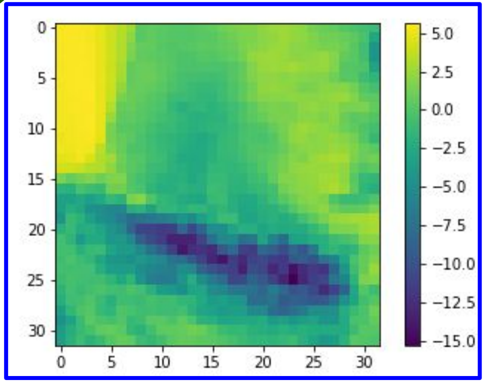
- **X : From the UP RCM on a chosen domain**
  - 4 altitude fields : **ZG, TA, HUS, (UA,VA)** at **850, 700** and **500 hPa**
  - 3 surface variables : **TAS, PR, (UAS,VAS)**
  - on the red domain ([-5,10]E x [35,50]N)
  - Daily frequency



- **Y : RCM output** ⇒ Daily Surface Temperature on box over South West of France including the Pyrenees and the Atlantic coast (blue box on the map).

⇒ **Anomalies mode :** We remove from the local temperature the temperature average on a 5x5 GCM grid point box.
So for a RCM grid point i :

$$Y_{Anom,i} = TAS_{RCM,i} - TAS_{UPS-RCM,5*5box}$$

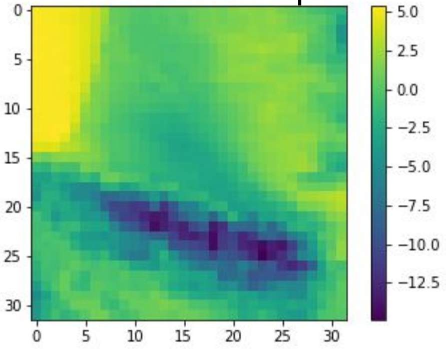


Example of output (Y)

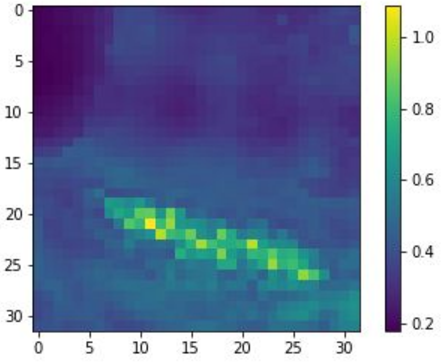
Validation Step

- We first test the ability of the emulator to reproduce the same simulation it was trained on, but different years.
- Simulation : **ALADIN6** 12km forced by **CNRM-CM5** (150km) **RCP4.5**, period 2006-2100, daily timescale
- Training set : 70% of the years, Testing set: 30% of the years
- The emulation presents good results :
  - RMSE, BIAS and correlation are really good on every points.
  - The PDFs and Time Series plots are also really satisfying (see next slide).
  - Some improvement can still be done on highest mountain points.

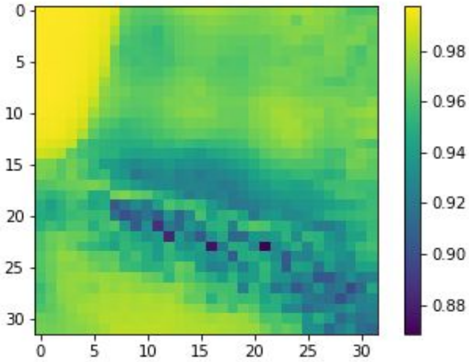
YAnom
Prediction example



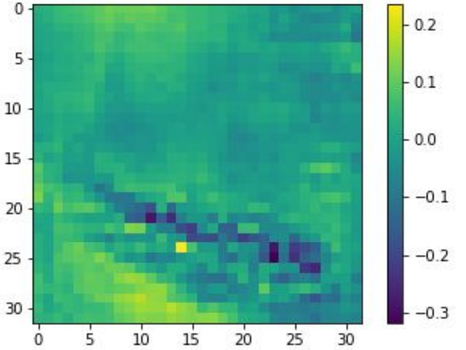
RMSE (°C)



Correlation



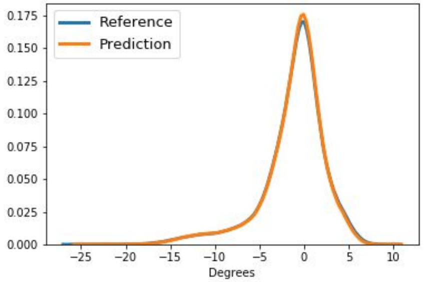
BIAS (°C)

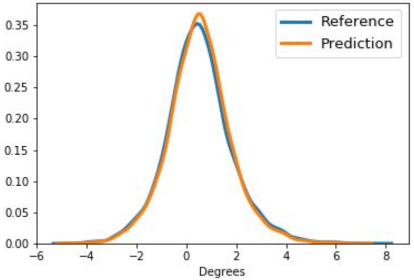


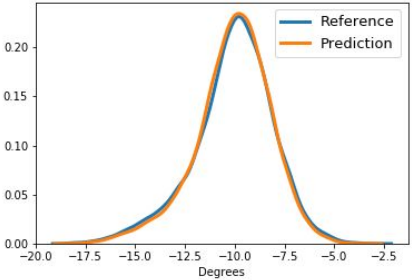
Validation Step

All points pooled | A lowland point | A high mountain point

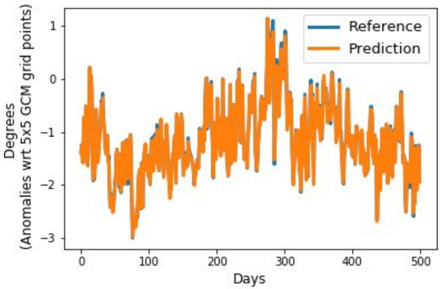
PDFs comparison

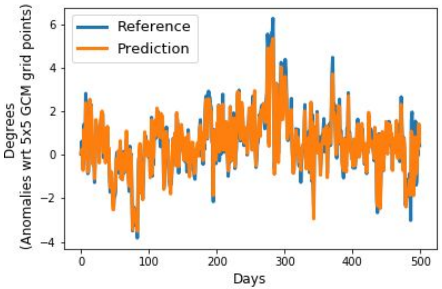


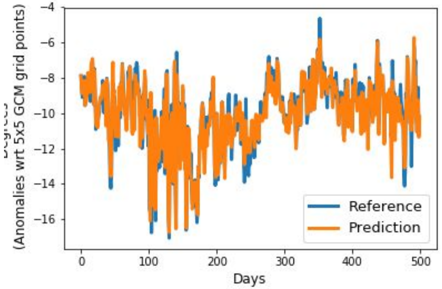




Time series example: 500 consecutive days



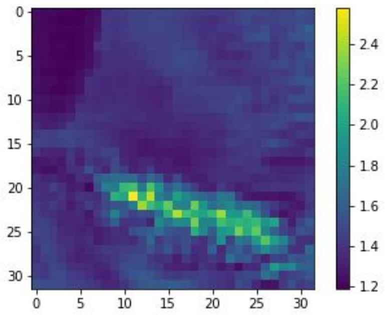




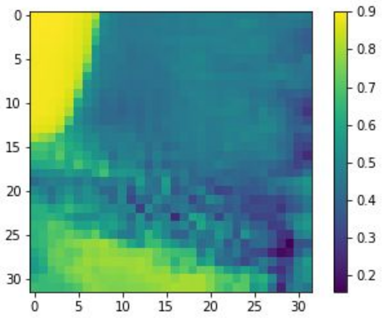
Application : Downscaling of a new run

- Training set : 2006-2100 ALADIN6 simulation forced by CNRM-CM5 with scenario RCP 4.5
- Downscaling of ALADIN6 forced by CNRM-CM5 on 2006-2100 RCP 8.5.
- This simulation exists, so we can verify our emulation
  - ⇒ with for each grid point i, the comparison reference is $TAS_{RCM,i} - TAS_{UPS-RCM,5*5box}$ since we remove the large scale average
- The results are disappointing here but the emulator shows promising results:
  - ⇒ The climatology is respected (reasonable bias and rmse, also visible on PDFs plots on next slide)
  - ⇒ The anomalies seasonality and temporality is respected in most cases (Correlation + Time series plots)

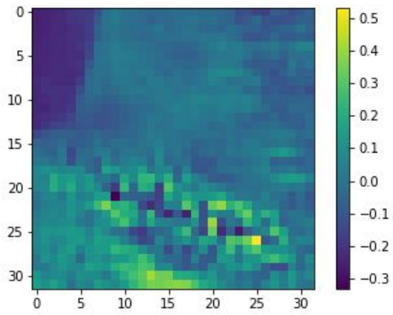
RMSE (°C)



Correlation



BIAS (°C)



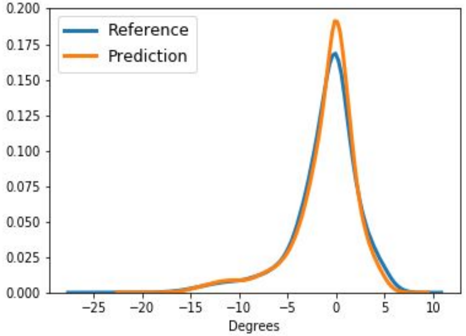
Application : Downscaling of a new run

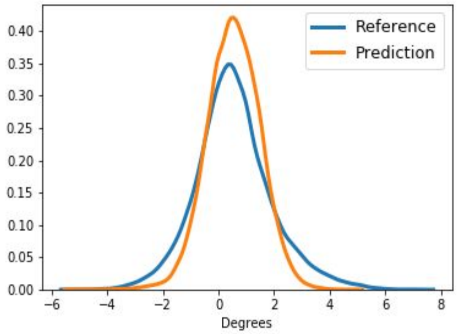
All points pooled

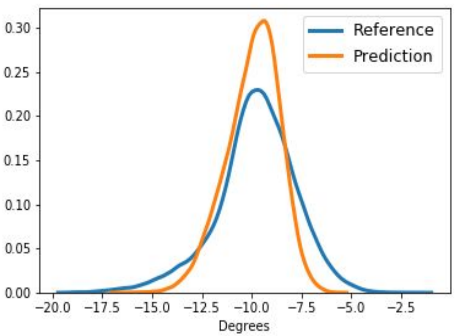
A lowland point

A high mountain point

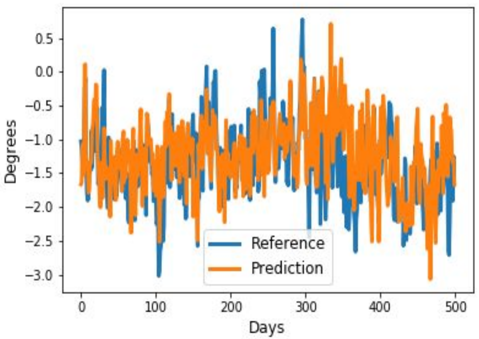
PDFs comparison

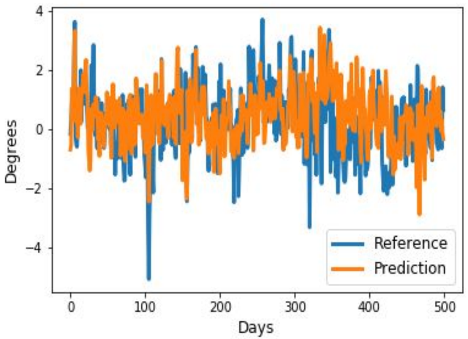


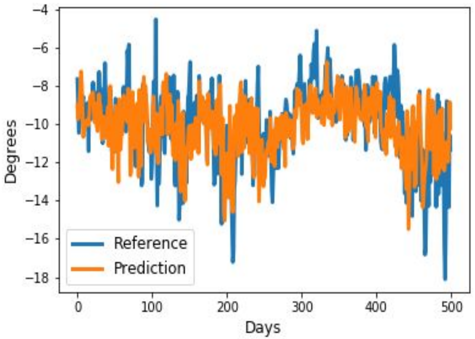




Time series example: 500 consecutive days







Summary and Future work

- We developed the concept of RCM emulators and the methodology.
- We built a satisfying 2D emulator for ALADIN RCM using a Neural Network architecture.
    - The performance of the emulator in the validation step are good but perfectible.
    - The results of the application are promising but we expect more.

- More configurations have to be tested (ex : train on historic and rcp 85, application on rcp 45).
- More test have to be done : downscaling of other GCMs, other variables, larger domain.
- Further work should also be done to define the reference scale for the anomalie mode.