RCM-EMULATOR: EXAMPLE OF APPLICATION TO A LARGE ENSEMBLE.

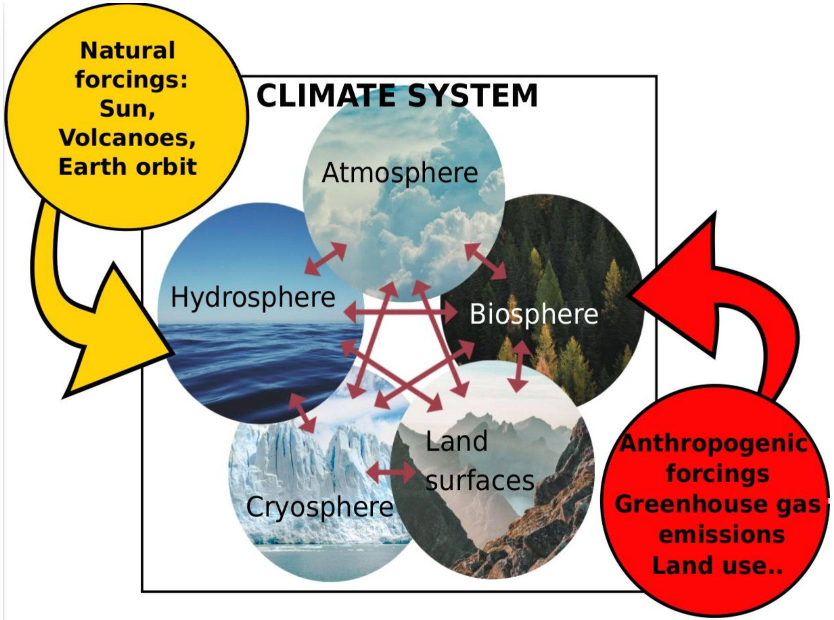
Antoine Doury, Samuel Somot, Elizabeth Harader-Coustau, Christophe Cassou

Journées IA Météo-France, 14/02/2025

antoine.doury@meteo.fr

Climate models

- Numerical representation of the climate system
  - 5 components plus interactions
  - External forcings (anthropogenic & natural)
- Climate simulations :
  - Time evolution of diagnostic variables (Temperature, Humidity, ...)
- Long & multiple simulations :
  - to reach system equilibrium
  - to study the reaction of the system to different scenario of external forcings
  - to take into account the uncertainty range
- Expensive tools : compromise between complexity, resolution, and the length or the number of simulations

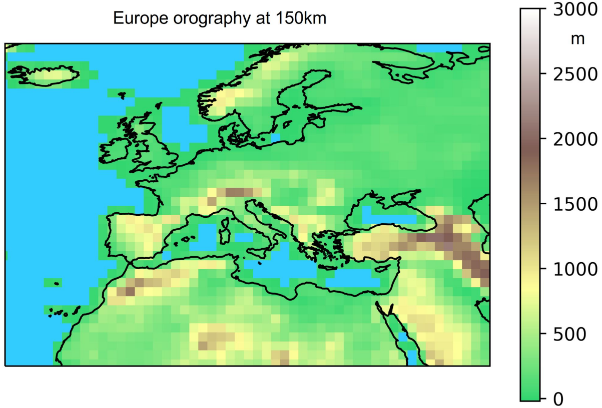


Regional Climate Models

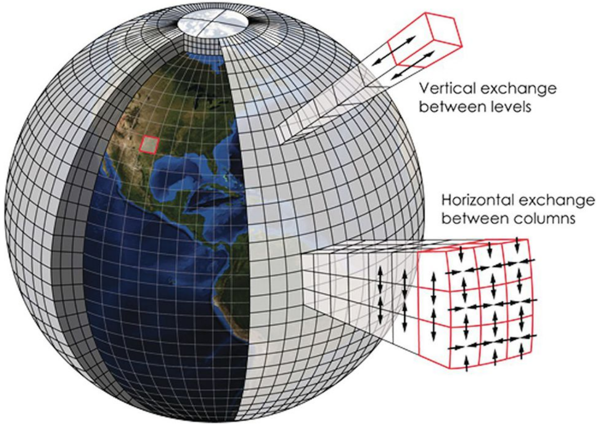
Global Climate Models (GCM):

- Driven by human activities scenarios
- Generally high complexity :
  Coupled model with various components
  Atmosphere, ocean, sea-ice, surface, vegetation, rivers…
- Horizontal resolution ~ 50km to 200km

⇒ Possible to run long and large ensemble of simulations, but too coarse to study the local impacts of climate change.



SOCIO-ECONOMICS SCENARIOS



Regional Climate Models

Global Climate Models (GCM):

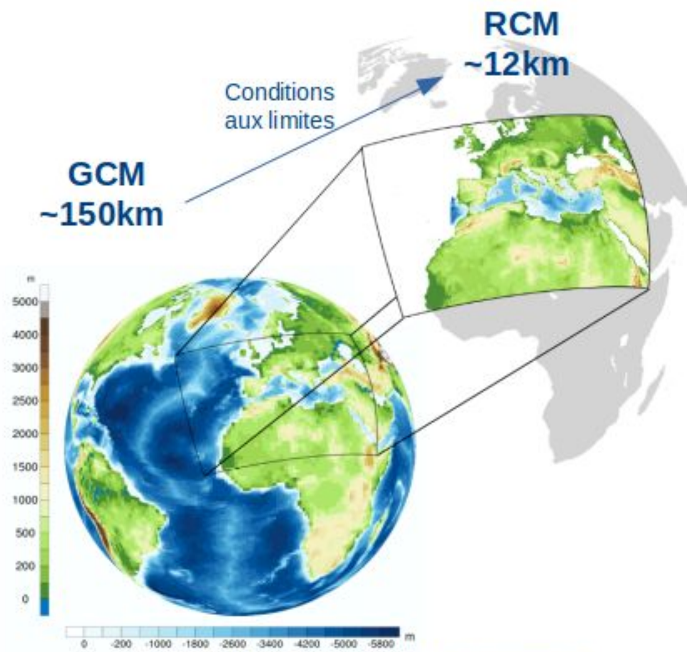
- Driven by human activities scenarios
- Generally high complexity :
  Coupled model with various components
  Atmosphere, ocean, sea-ice, surface, vegetation, rivers…
- Horizontal resolution ~ 50km to 200km

⇒ Possible to run long and large ensemble of simulations, but too coarse resolution to study the local impacts of climate change.

Regional Climate Models (RCM):

- Limited Area models → **driven by a GCM at the border of the domain**
- Horizontal resolution 50km to 1km
- Simpler model: generally only the atmosphere

⇒ Better representation of the local/extremes events, but too expensive to cover the range of uncertainties



Source : l'Encyclopédie de l'environnement - P. Nabat et A. Valdoire

Regional Climate Models

Global Climate Models (GCM):

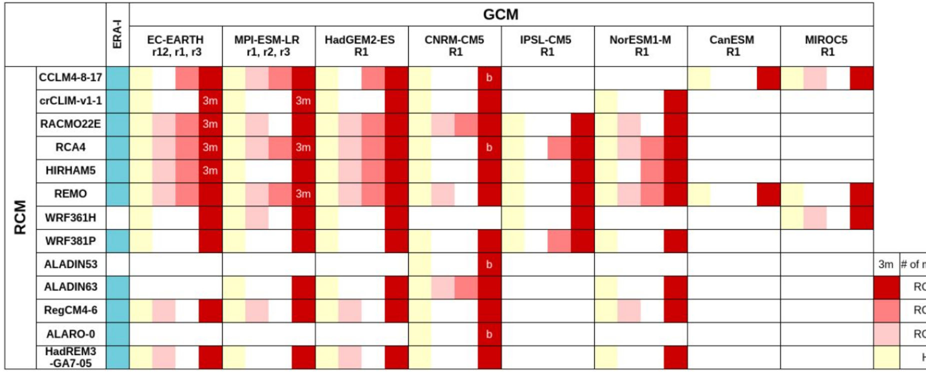
- Driven by human activities scenarios
- Generally high complexity :
  Coupled model with various components
  Atmosphere, ocean, sea-ice, surface, vegetation, rivers…
- Horizontal resolution ~ 50km to 200km

⇒ Possible to run long and large ensemble of simulations, but too coarse resolution to study the local impacts of climate change.

Regional Climate Models (RCM):

- Limited Area models → driven by a GCM at the border of the domain
- Horizontal resolution 50km to 1km
- Simpler model: generally only the atmosphere

⇒ Better representation of the local/extremes events, but too expensive to cover the range of uncertainties



Euro-Cordex matrice, (Vautard et al, 2020; Coppola et al, 2021)

RCM-Emulators

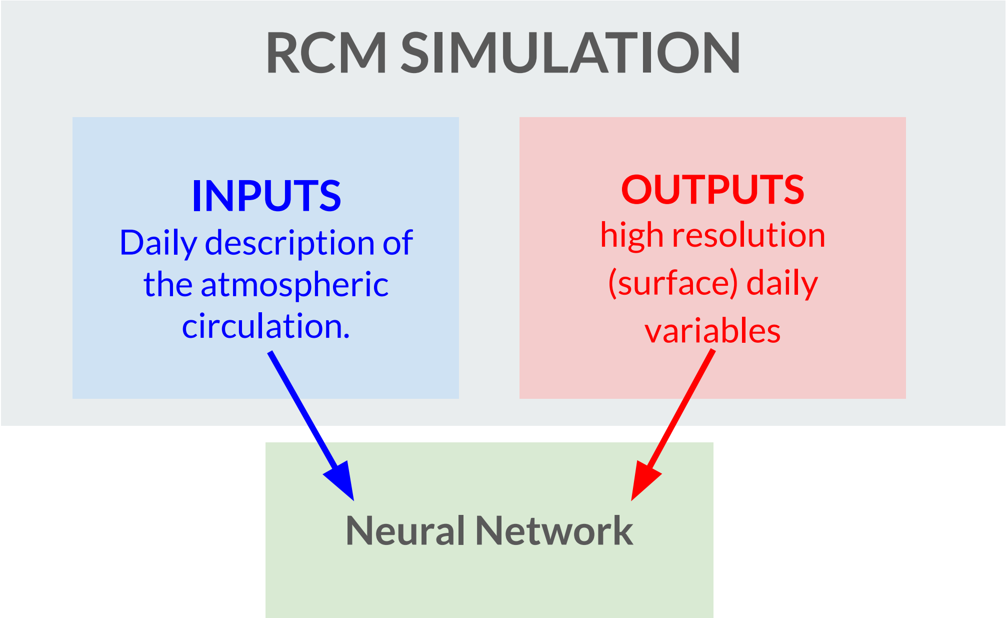
Use RCM simulations to train a machine learning algorithm to capture the relationship between low resolution variables (INPUTS) and high resolution variables of interest (OUTPUT).

Interest : Once the relationship is captured, it can be applied to any new GCM at low cost.

- Build large ensembles by downscaling various GCMs, and multiple members.
- Similar approach tan Empirical/Statistical Downscaling (learns the same relationship in observational data)
  - Advantages: No need for observations: more regions of the world, more variables, explore future climate.
  - Disadvantages : learns an imperfect relationship (learns the defaults of the RCM)

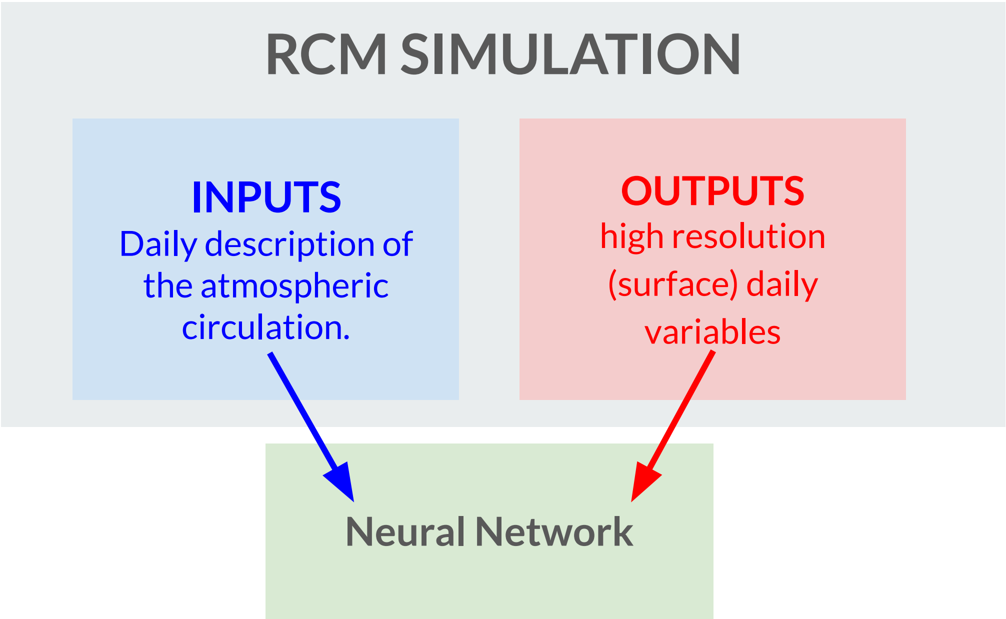
Training : Perfect model strategy

⇒ The relationship is learned INSIDE the RCM simulations used for training



Training : Perfect model strategy

⇒ The relationship is learned INSIDE the RCM simulations used for training



RCM = LARGE SCALE modification + Downscaling

Ensure a PERFECT relationship between inputs and outputs.

RCM, forced only at the boundaries, modifies the large scale of its driving GCM:

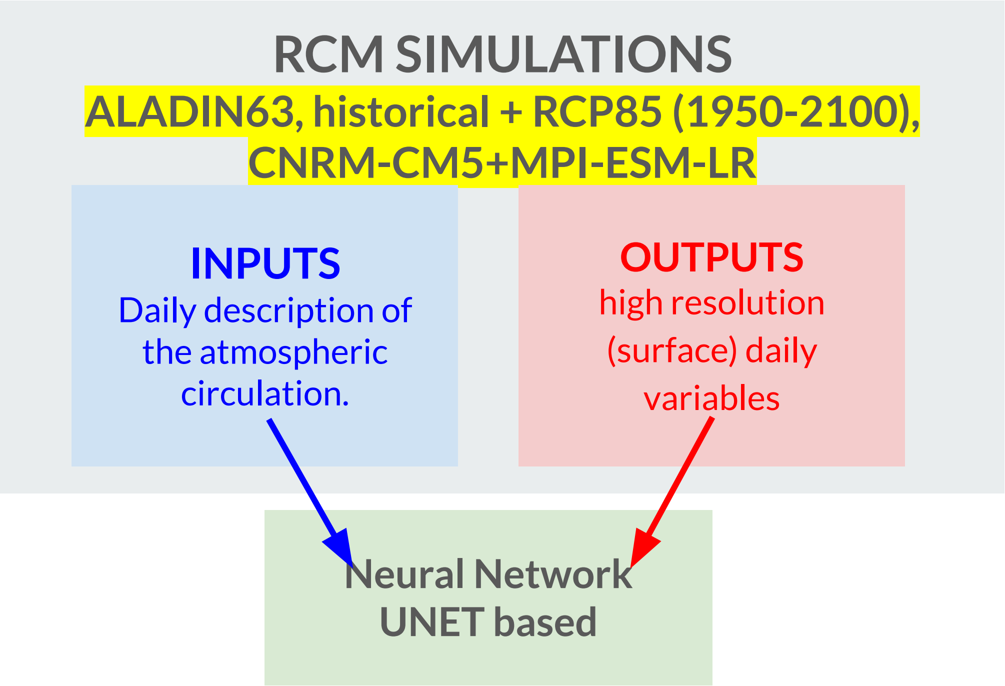
- Day to day chronology,
- But also at the climatological scale

⇒ We do not learn this LS modification..
→ Probably and partially for bad reasons
→ GCM-dependent.

... and **focus only on the *downscaling function*** included in the RCM.

Training : Perfect model strategy

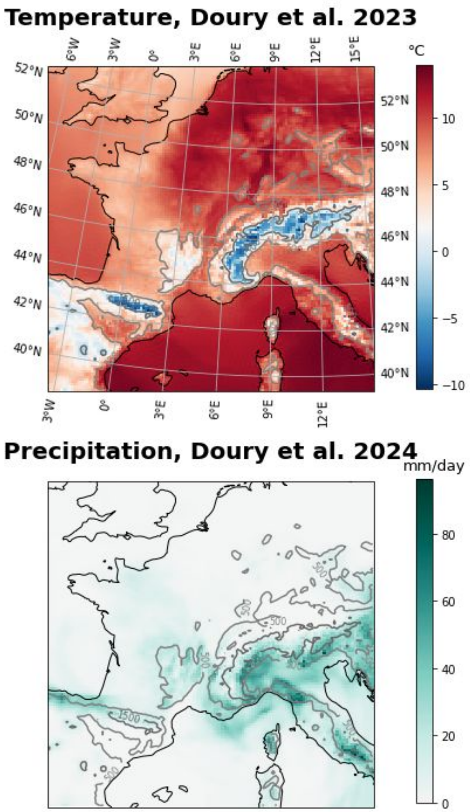
⇒ The relationship is learned INSIDE the RCM simulations used for training



Emulator conception

RCM: ALADIN63 (12km, driven by CMIP5 runs)

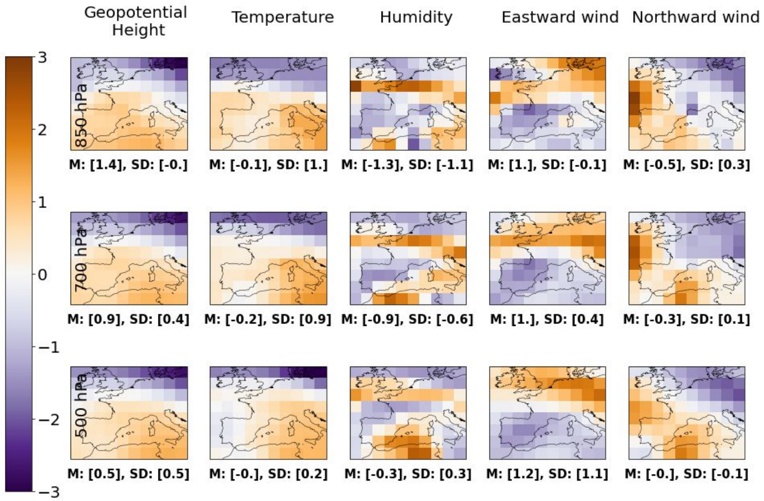
Target variables : Daily Temperature & Precipitations



Emulator conception

RCM: ALADIN63 (12km, driven by CMIP5 runs)

- **Target variables :** Daily Temperature & Precipitations
- Inputs : **Daily description of the atmospheric conditions**
  - ➢ Geopotential, temperature, wind components, humidity at 3 vertical levels + external forcing (aerosols, Greenhouse gases)
  - ➢ 2 steps standardization : Temporal and spatial information given separately



Each map has mean 0 and variance 1

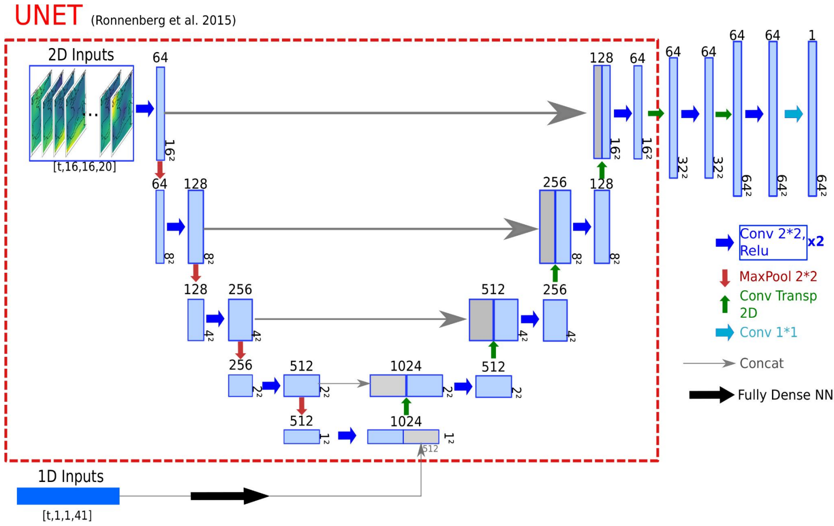
Emulator conception

RCM: ALADIN63 (12km, driven by CMIP5 runs)

- **Target variables :** Daily Temperature & Precipitations

- Inputs : **Daily description of the atmospheric conditions**
  - ➢ Geopotential, temperature, wind components, humidity at 3 vertical levels + external forcing (aerosols, Greenhouse gases)
  - ➢ 2 steps standardization : Temporal and spatial information given separately

- **Neural network** architecture : **UNet** based
  - ➢ Efficient management of multidimensional data
  - ➢ **Fully convolutional** : helps the network to better capture the spatial relationship (Gonzalez-Abbad et al. 2023)

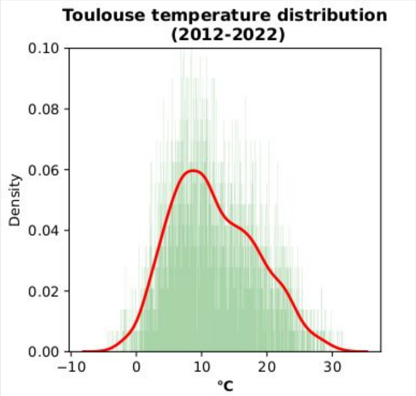


✤ ~ 25 million parameter
✤ ~ 3h to train on GPU (depends on the target domain size)
✤ ~ 1 min to predict

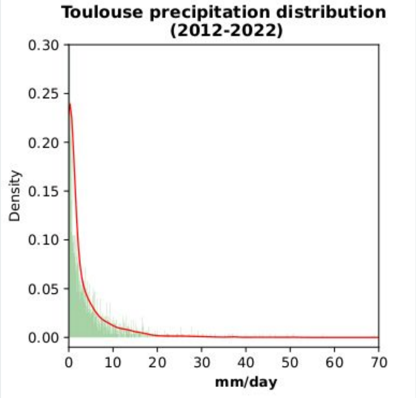
Emulator conception

RCM: ALADIN63 (12km, driven by CMIP5 runs)

- **Target variables :** Daily Temperature & Precipitations
- Inputs : **Daily description of the atmospheric conditions**
  - ➢ Geopotential, temperature, wind components, humidity at 3 vertical levels + external forcing (aerosols, Greenhouse gases)
  - ➢ 2 steps standardization : Temporal and spatial information given separately
- **Neural network** architecture : **UNet** based
  - ➢ Efficient management of multidimensional data
  - ➢ **Fully convolutional** : helps the network to better capture the spatial relationship (Gonzalez-Abbad et al. 2023)
- Loss function designed for precipitation
  Penalizes stronger an underestimated heavy precipitation
  The parameter is the quantile value for the precipitation at a given day/point, following a Gamma distribution fitted at the grid point.



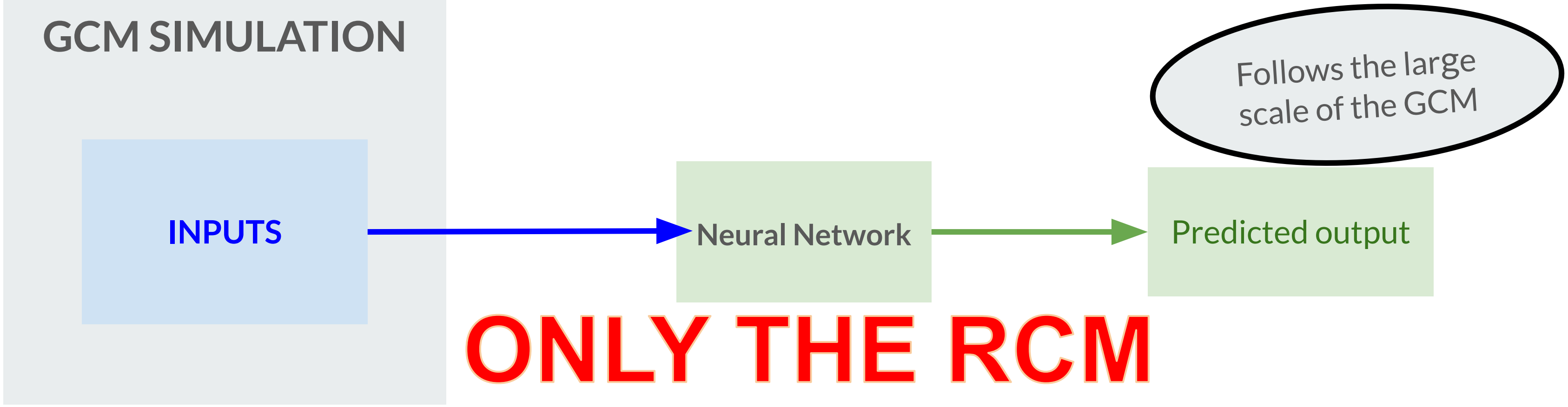
Normally distributed
⇒ MSE is well adapted



Right skewed
⇒ Asymmetric loss function, to specifically focuses on extremes

$$L(y,\hat{y}) = |y-\hat{y}| + (\beta * max(0, y-\hat{y}))$$

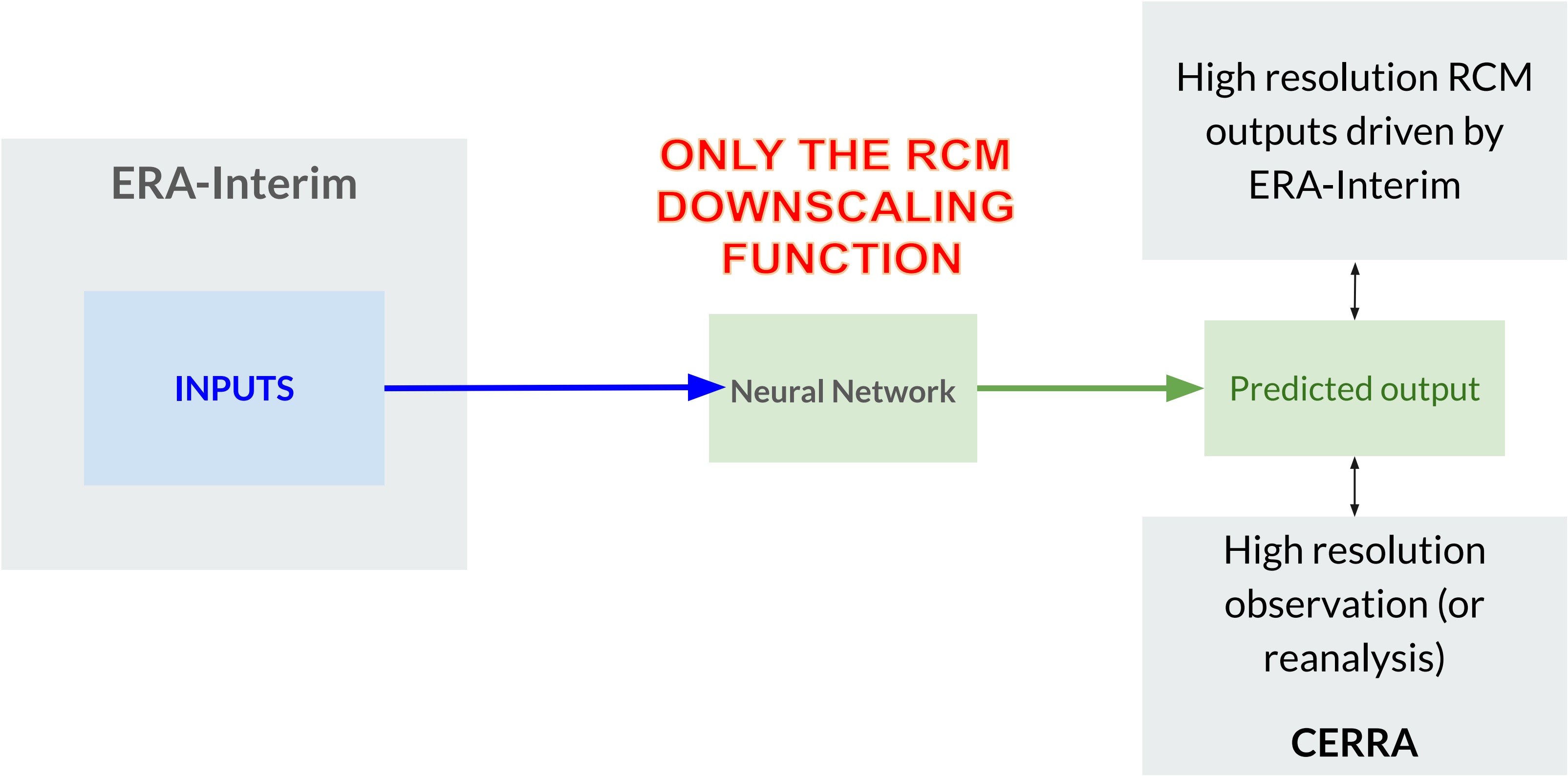
How to apply the emulator?



ONLY THE RCM DOWNSCALING FUNCTION

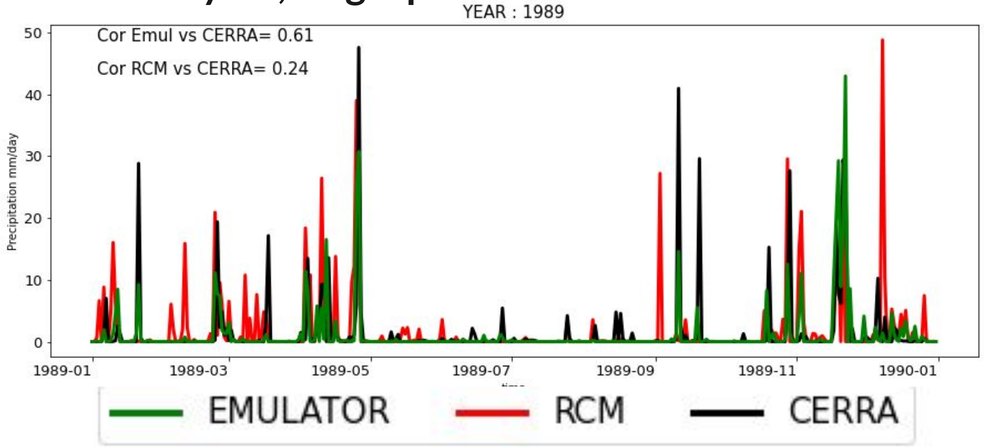
Short Evaluation

Evaluation: Application to low-res reanalysis

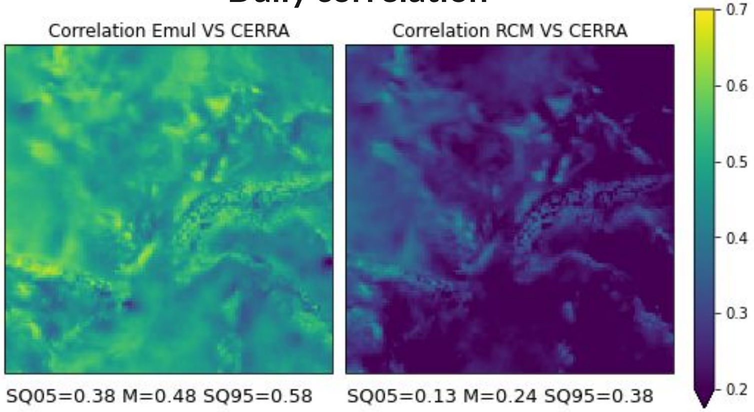


The case of precipitations

1 year, single point South France

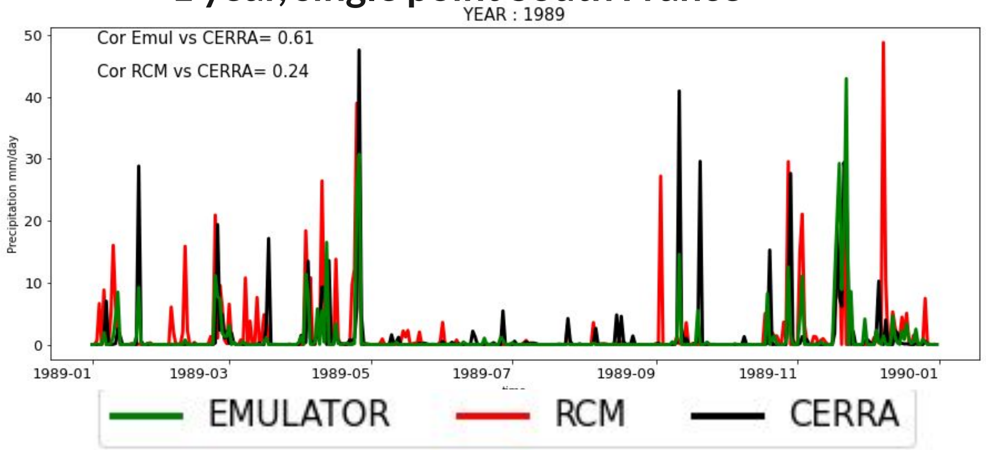


Daily correlation

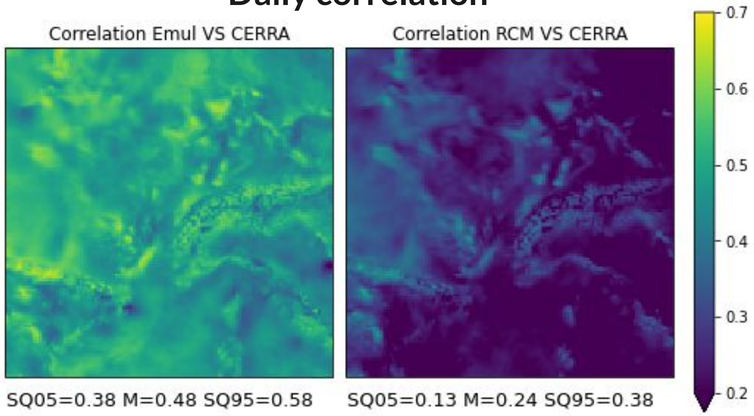


The case of precipitations

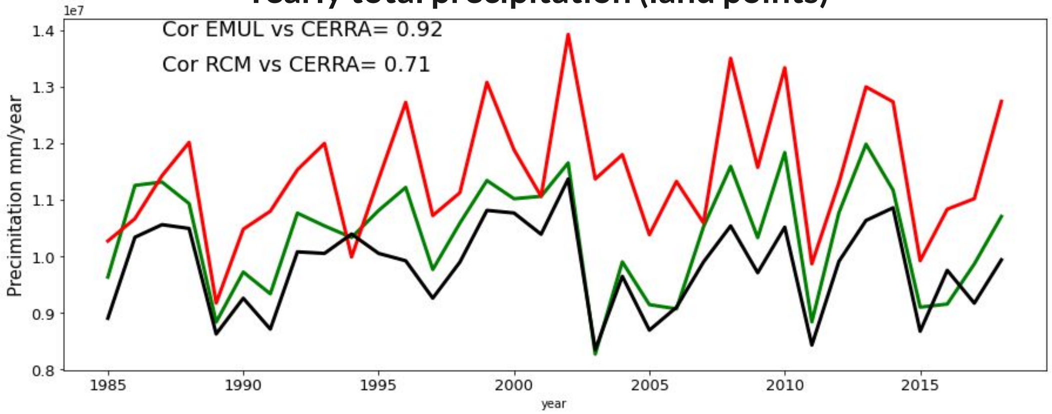
1 year, single point South France



Daily correlation



Yearly total precipitation (land points)



⇒ **The emulator follows better the observational (CERRA) time series** at the daily and grid point scale but also for the interannual variability, than the RCM.

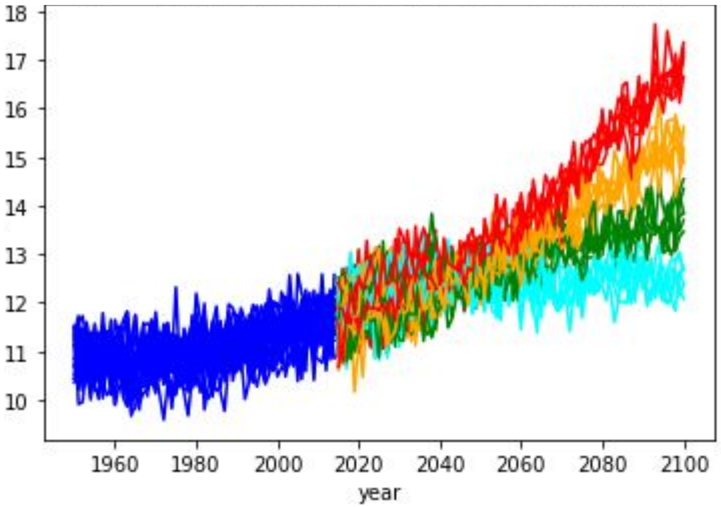
Application:
Let’s downscale somthing big

Application: let's downscale something big.

CNRM-CM6 , 150 km.

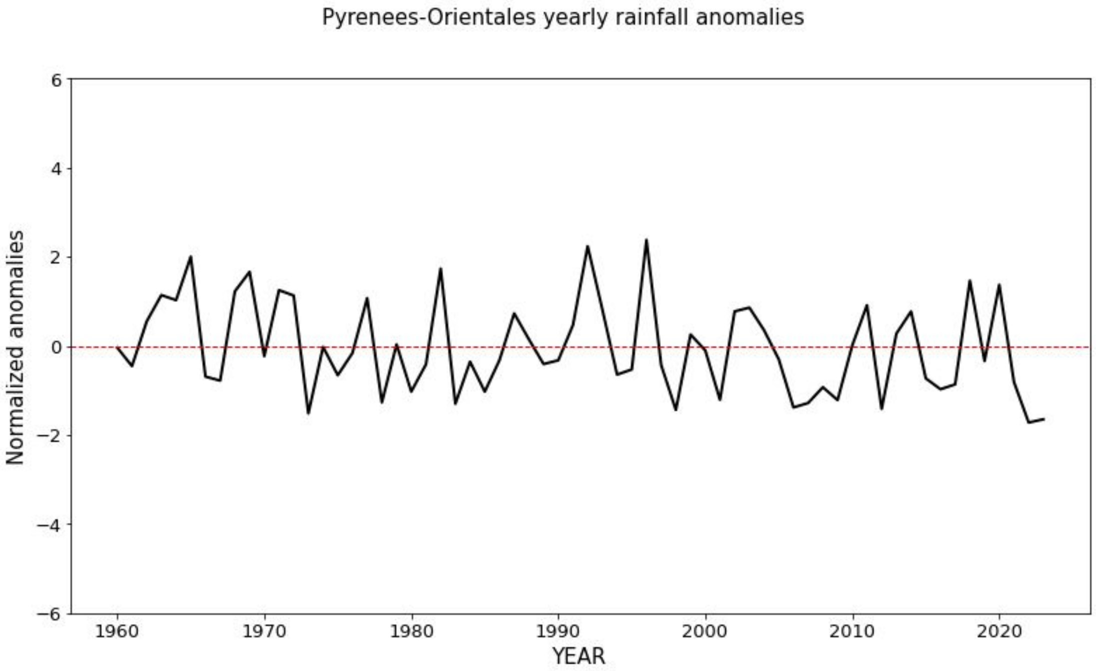
- historical : 1950-2014, 22 members
- Projections : 2015 -2100
  - 4 scenarios SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5
  - 6 members each
- Projections sort-term: 2015-2039
  - 4 scenarios SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5
  - 24 members each

Yearly mean temperature, European domain



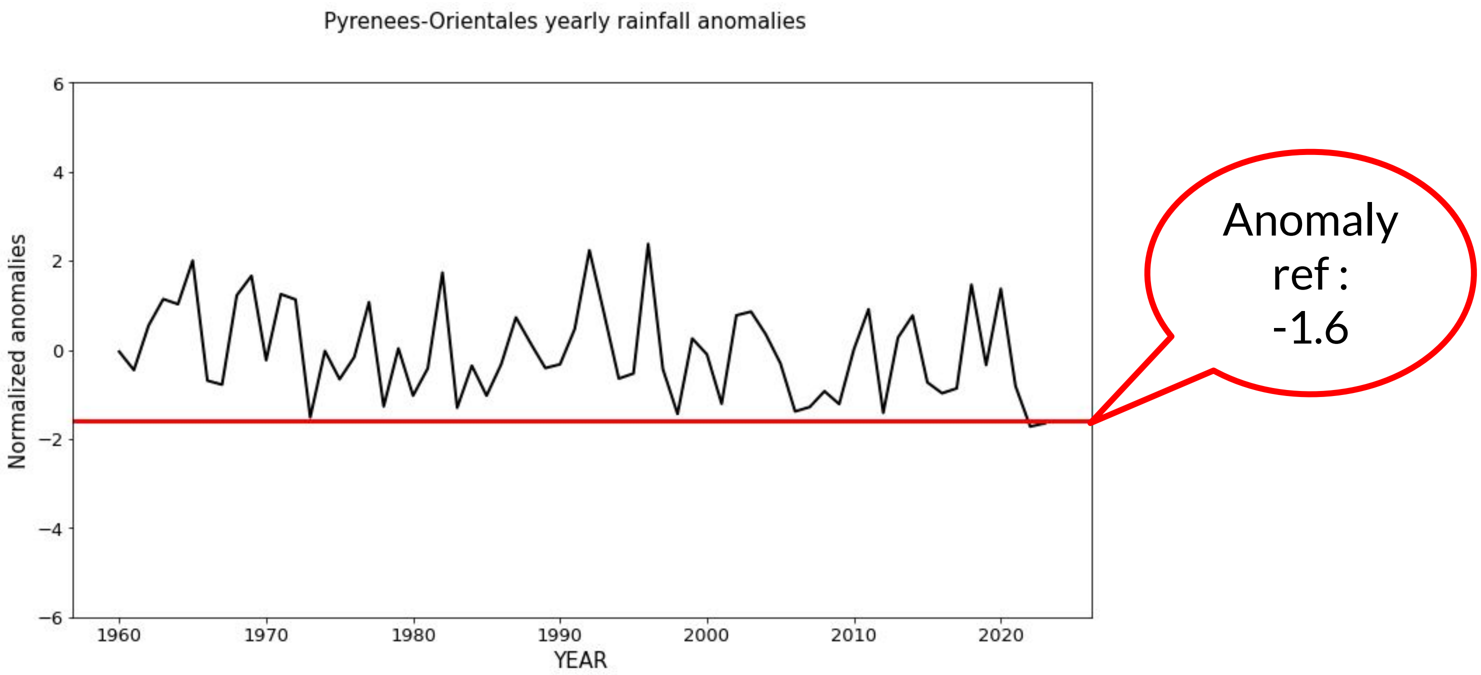
Case study : Pyrenees-Orientales drought 2022-23

⇒ Yearly cumulated precipitation over the P-O territory,
normalised (normally distributed so wrt mean and standard deviation)



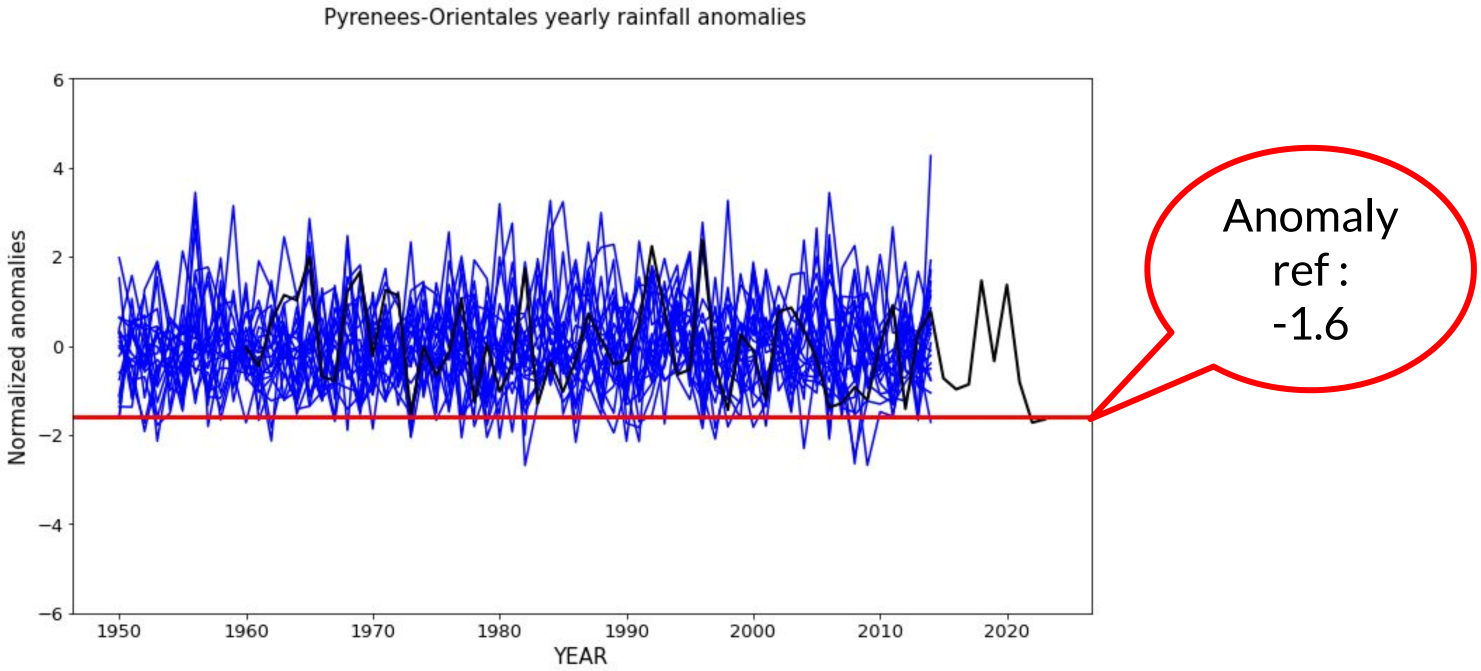
Case study : Pyrenees-Orientales drought 2022-23

⇒ Yearly cumulated precipitation over the P-O territory,
normalised (normally distributed so wrt mean and standard deviation)



Case study : Pyrenees-Orientales drought 2022-23

⇒ Yearly cumulated precipitation over the P-O territory,
normalised (normally distributed so wrt mean and standard deviation)



Case study : Pyrenees-Orientales drought 2022-23

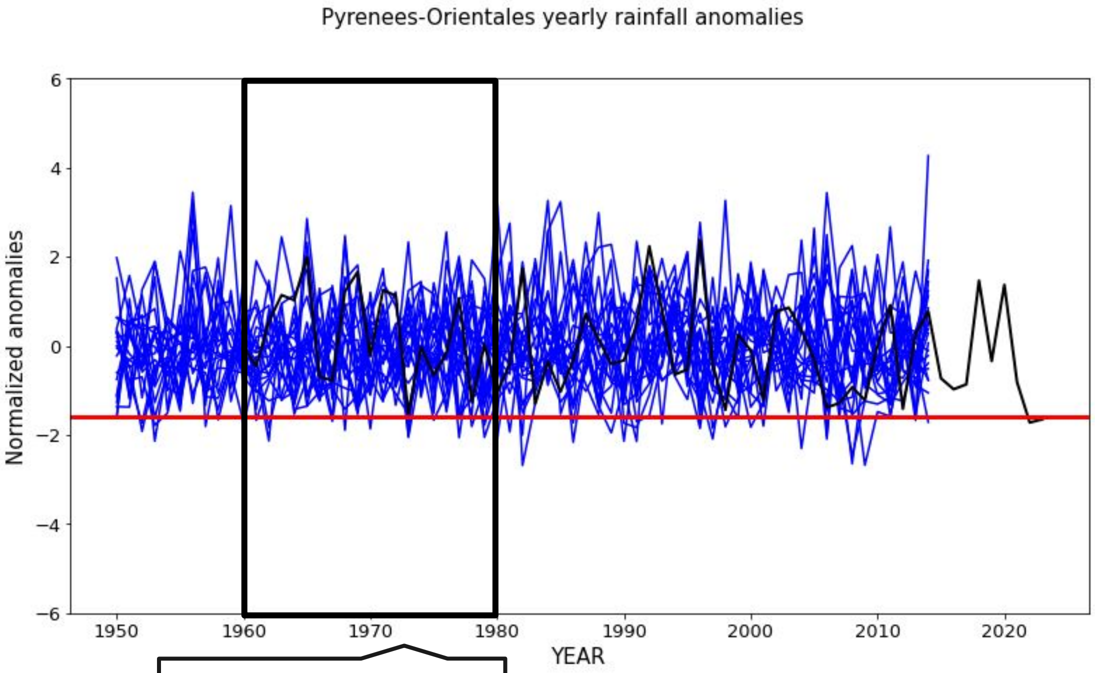
⇒ Yearly cumulated precipitation over the P-O territory,
normalised (normally distributed so wrt mean and standard deviation)

Proportion of anomalies
1 yr < ref (anomaly 1.6):

- historical (1960-80):4%

2 yr consecutive < ref (anomaly 1.6):

- historical (1960-80): 0



22 members * 20 ans
→ 462 ans

Case study : Pyrenees-Orientales drought 2022-23

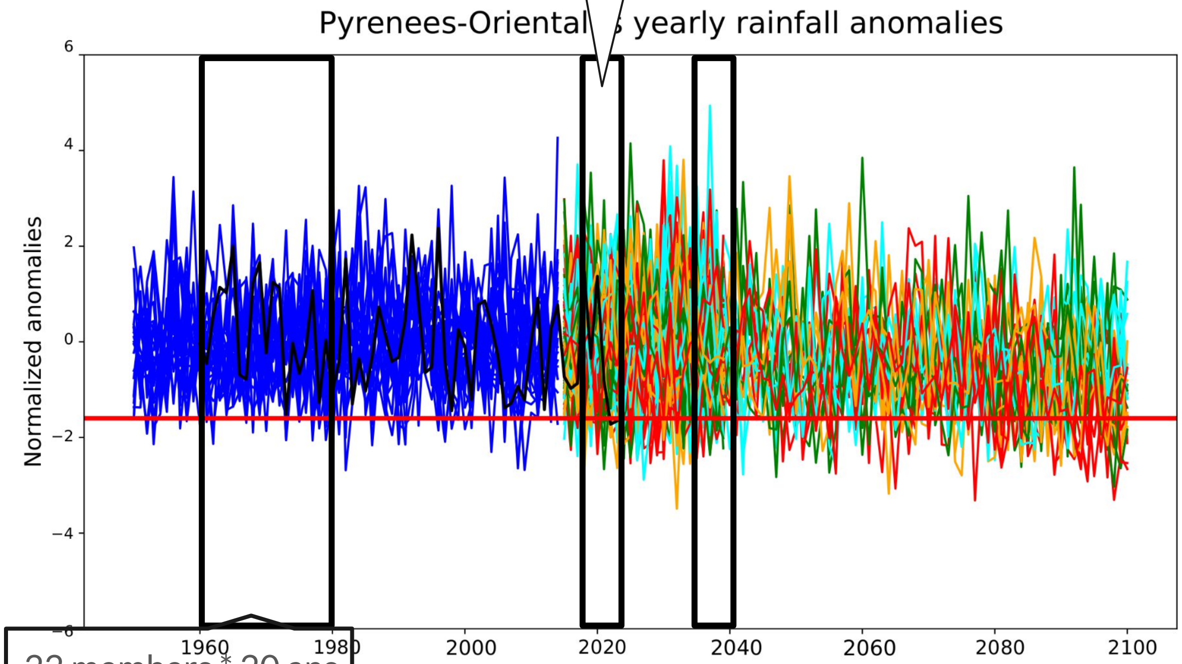
620 yrs

Proportion of anomalies
1 yr < ref (anomaly 1.6):

- historical (1960-80):4%
- 2020-2024 : 4%
- 2035-2039 : 6%
-

2 yr consecutive < ref (anomaly 1.6):

- historical (1960-80): 0
- 2020-2024 : 1% (11)
- 2035-2039 : 2%
-



22 members * 20 ans
→ 462 ans

Case study : Pyrenees-Orientales drought 2022-23

Proportion of anomalies
1 yr < ref (anomaly 1.6):

- historical (1960-80):4%
- 2020-2024 : 4%
- 2035-2039 : 6%
- 2095-2099 : 28%

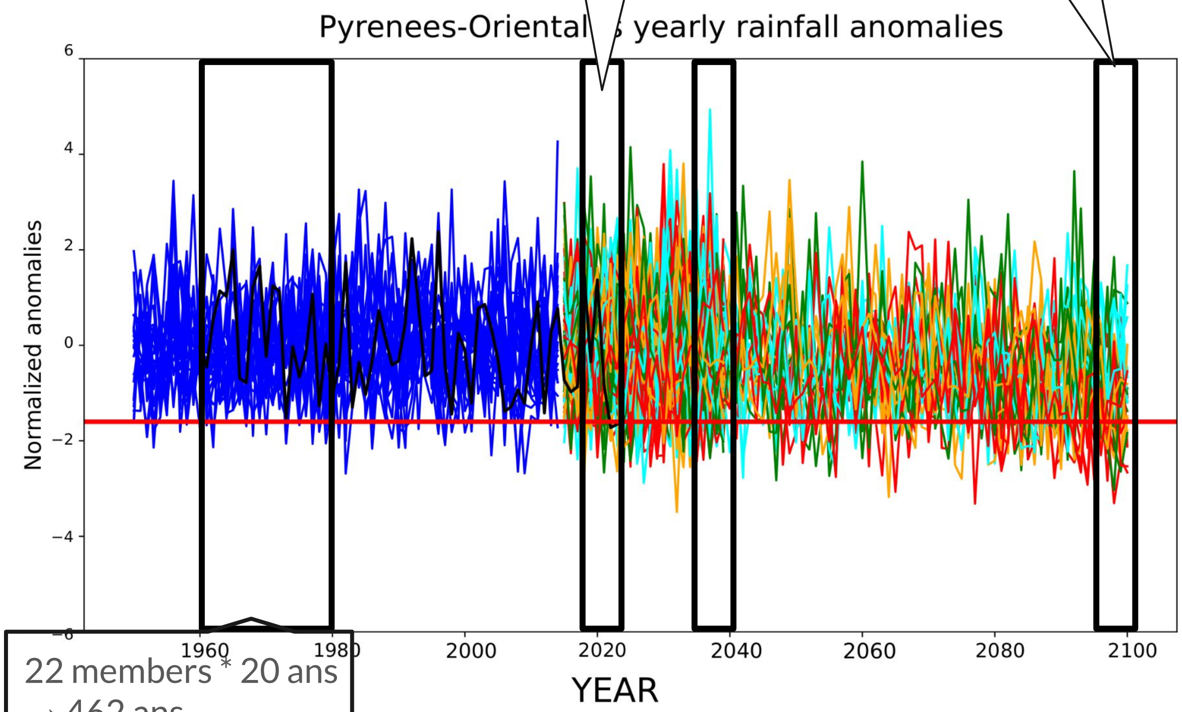
2 yr consecutive < ref (anomaly 1.6):

- historical (1960-80): 0
- 2020-2024 : 1% (11)
- 2035-2039 : 2%
- 2095-2099 : 20%

  3yrs⇒ 10%

620 yrs

96 yrs



22 members * 20 ans
→ 462 ans

Conclusion

- We trained an RCM-emulator for precipitation on existing RCM simulations.
- We set a training strategy that forces the emulator to follow the GCM large scale.
- We validated the emulator by downscaling low resolution reanalysis.
- The emulator shows better consistency with the observations chronology.

- We downscaled a big ensemble from CNRM-CM6 (120 members for near future period)
- The very large ensemble allows us to study the local climate change, for example over the Pyrénées-Orientales, and severe drought.
- We find that drought as 2022-2023 seemed not really possible before, rare today, but maybe common in a far future.
- We also see a trend in the future with less precipitation and “very wet years”, and less variance.

But some limitations maybe...