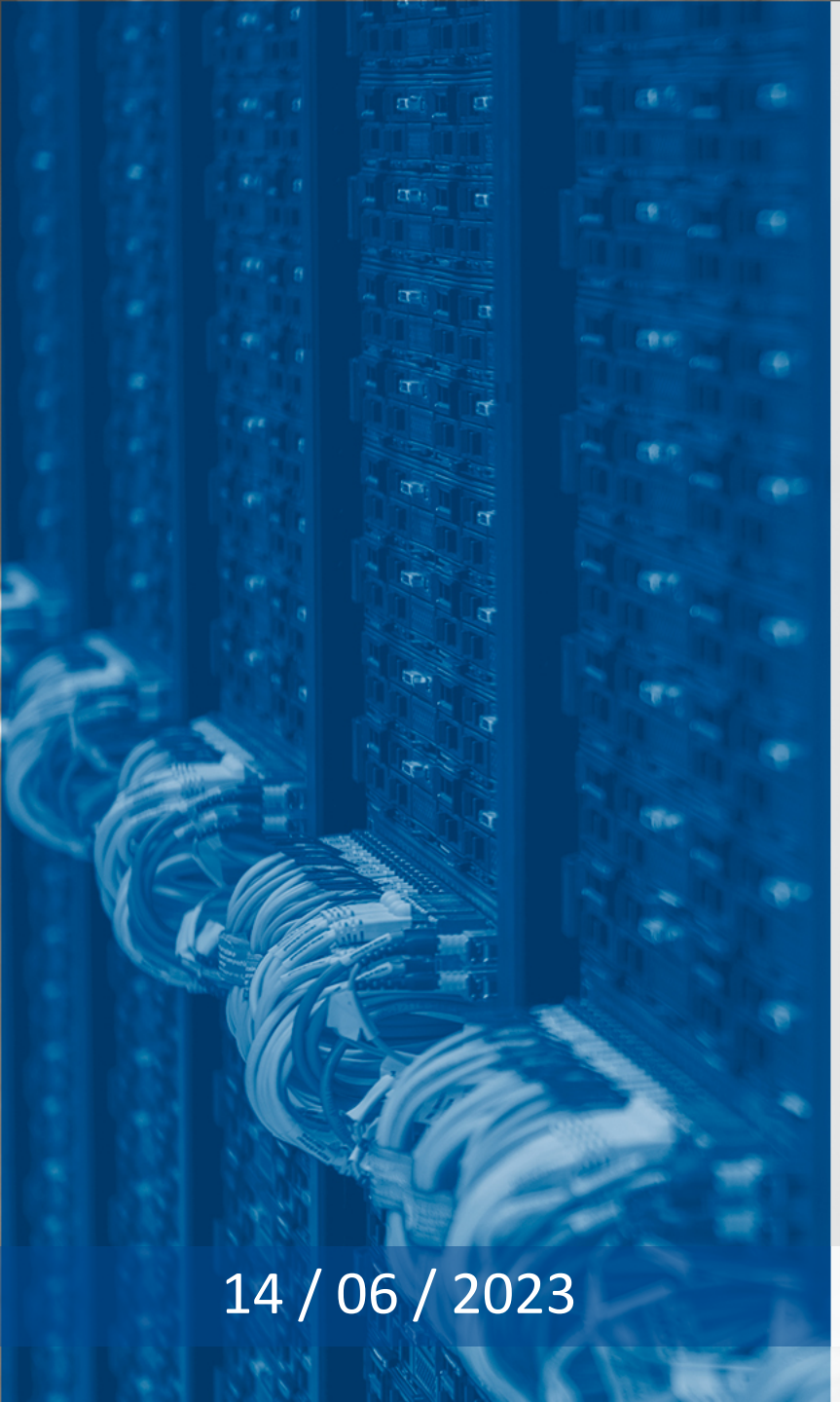


Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Seasonal climate prediction in big cities: from Brazil to Barcelona

Javier Corvillo Guerra

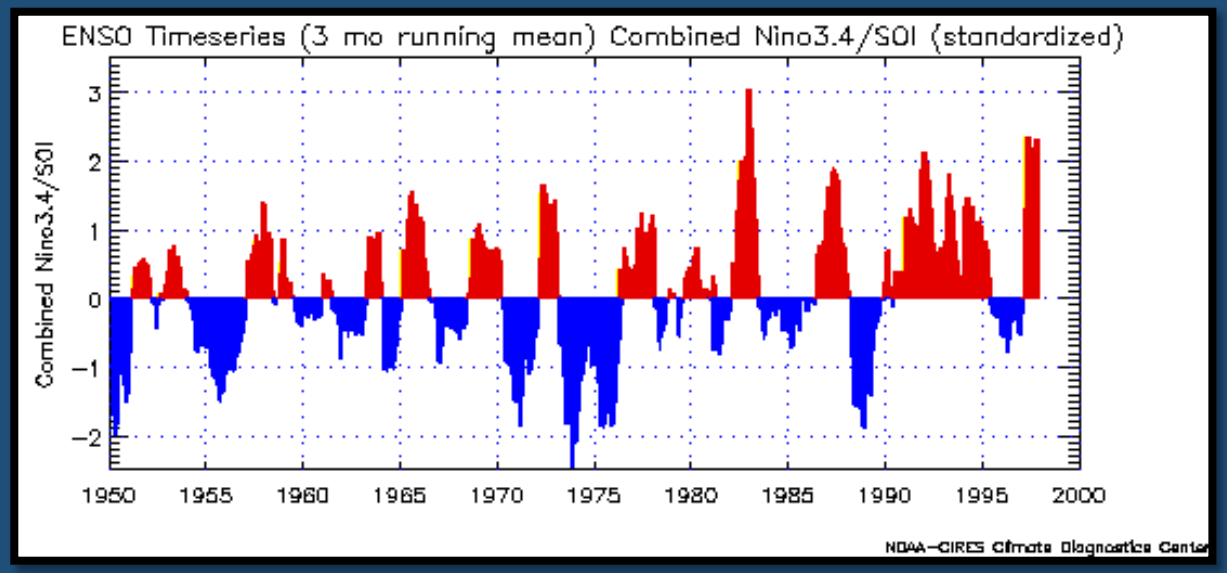
14 / 06 / 2023

Master's Degree in Meteorology and Geophysics

The research leading to this Master's thesis has been carried out at the Earth Sciences department of the **Barcelona Supercomputing Center - Centro Nacional de Supercomputación (BSC - CNS).**

GCMs in seasonal prediction

- GCMs (Global Climate Models) can be applied in seasonal timescales (1 – 12 months), predicting the evolution of the components of the Earth system.
- Internal mechanisms of the Earth System allow for model predictability (skill) over seasonal timescales, increasing forecast quality.



Seasonal predictability sources include ENSO phases, the MJO, convective and orographic effects…

However, GCMs aren’t perfect…

- GCMs outputs are computationally expensive, often constrained to coarse-resolution grids.
- Some coarse-grid GCMs may not capture sub-grid phenomena correctly (e.g, convective effects, thermal heat island) leading to lower model skill.

We shouldn’t use this model resolution for a fine grid analysis! (e.g. a farm or a city)

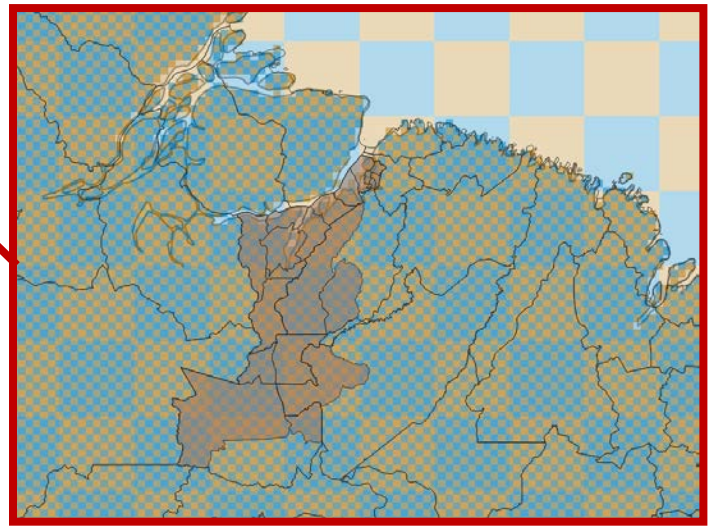
How to solve this? → Statistical Downscaling

From coarse to fine grid: Statistical Downscaling

- SD techniques transform coarse-grid data into fine-grid outputs WHILE correcting model bias (leads to higher skill).

- SD methods are computationally fair and widely used in climate services.

With downscaled outputs, we can thoroughly study sub-grid phenomena with higher precision



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

HARMONIZE

An application for SD methods: the bridge between health and climate services

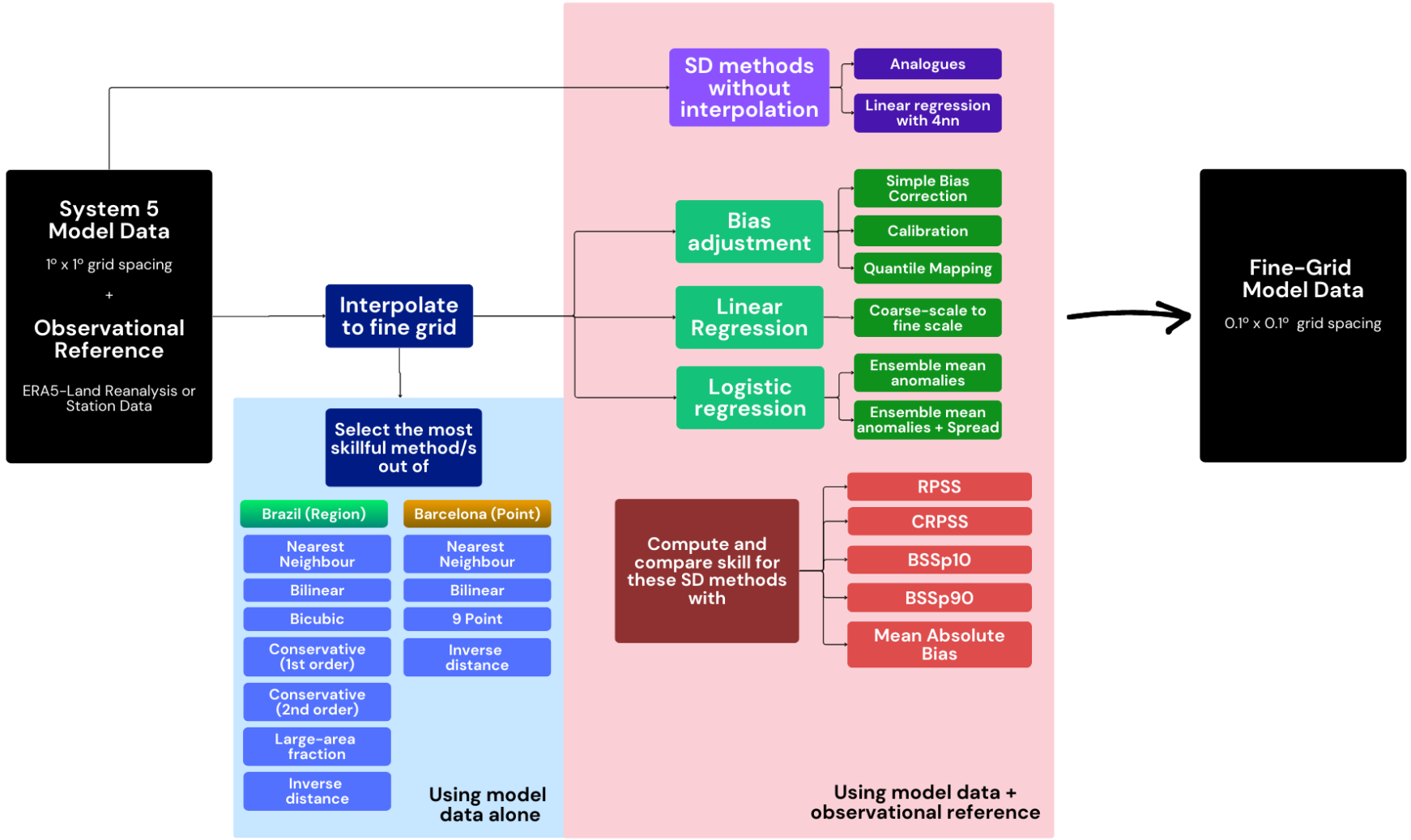
Question: which climatological factors condition the transmissibility of vector-borne diseases such as malaria and dengue? (HARMONIZE Project)

A fine grid output is needed! SD is necessary for a precise first-step analysis

HARMONIZE

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

But there are too many SD methods available…



…which one do we use?

Main goal: Which is the best SD method?

- Throughout this work, **we'll compare the skill of 24 SD methods** used to downscale model data to a region (South America) and to a point (Barcelona) **so as to outline the best one**.

- Outlining the best SD method(s) allows for precise factors conditioning malaria and dengue to be plugged into epidemiological models, allowing health experts to issue early warnings to the public if need be.

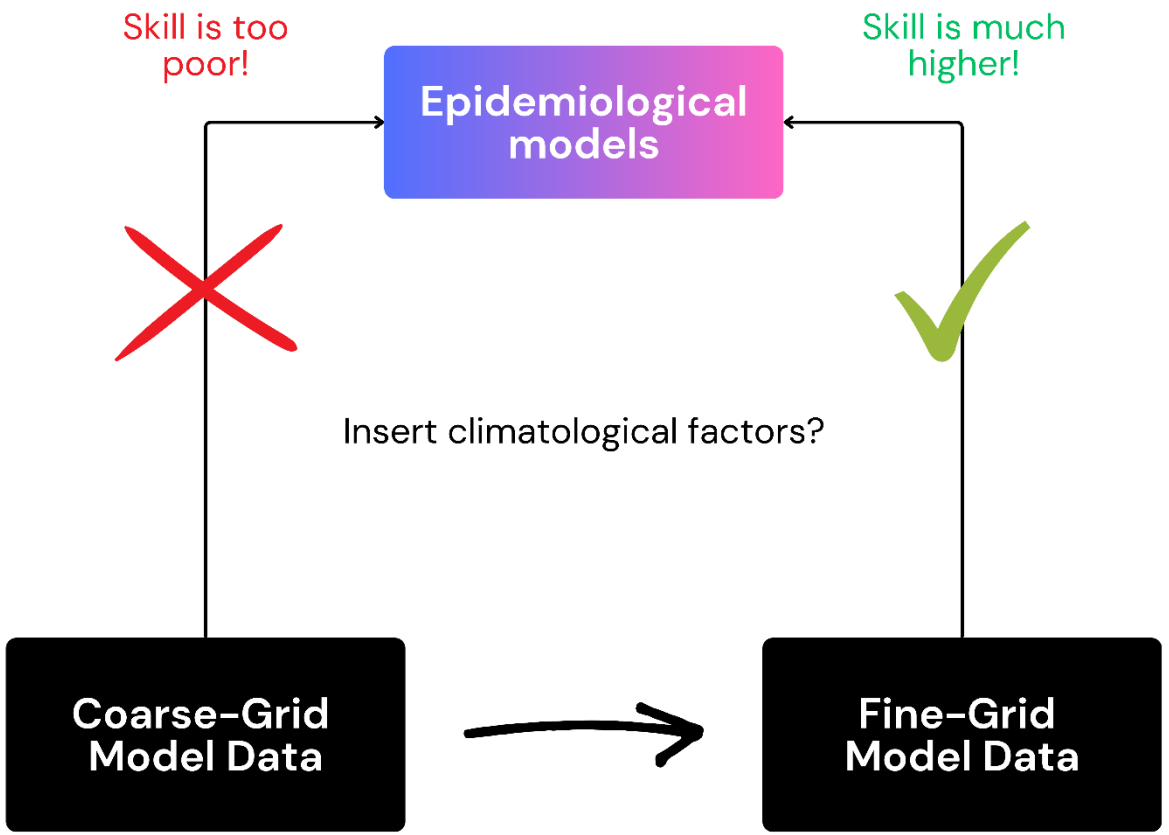
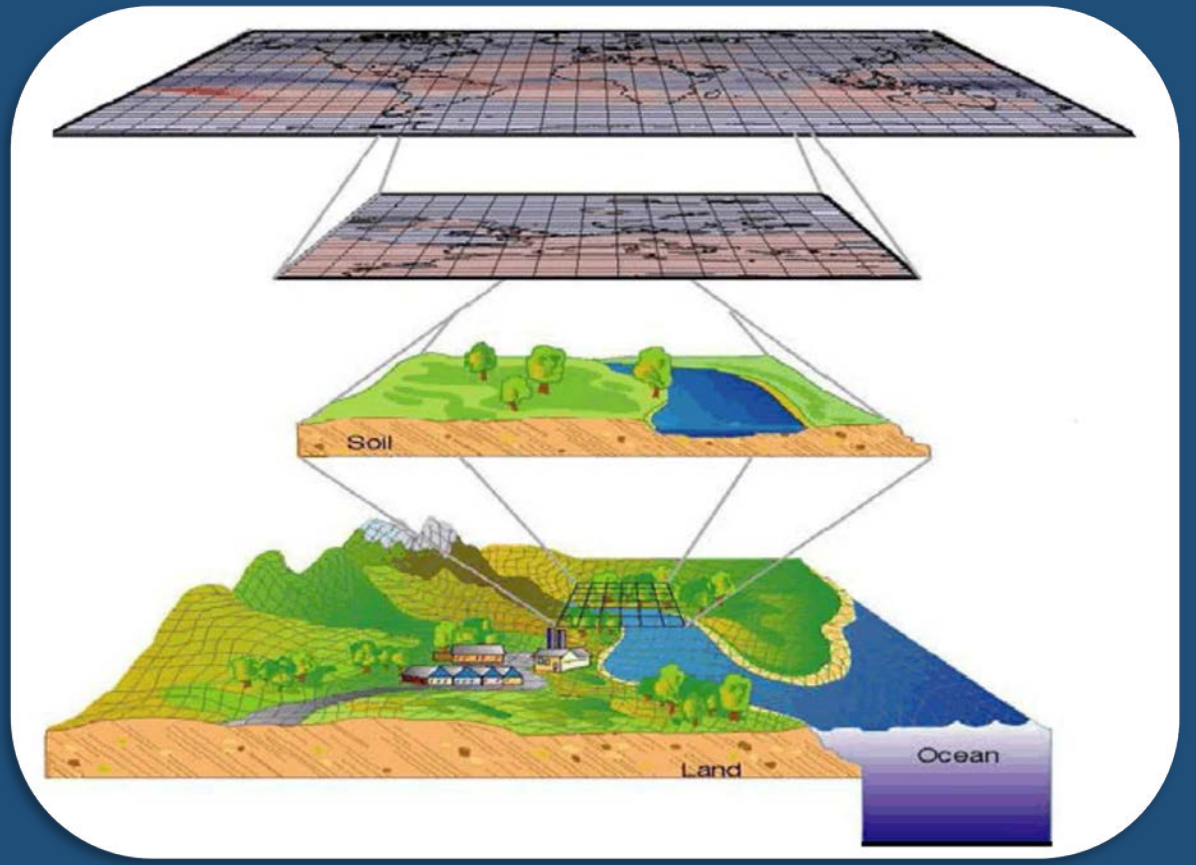


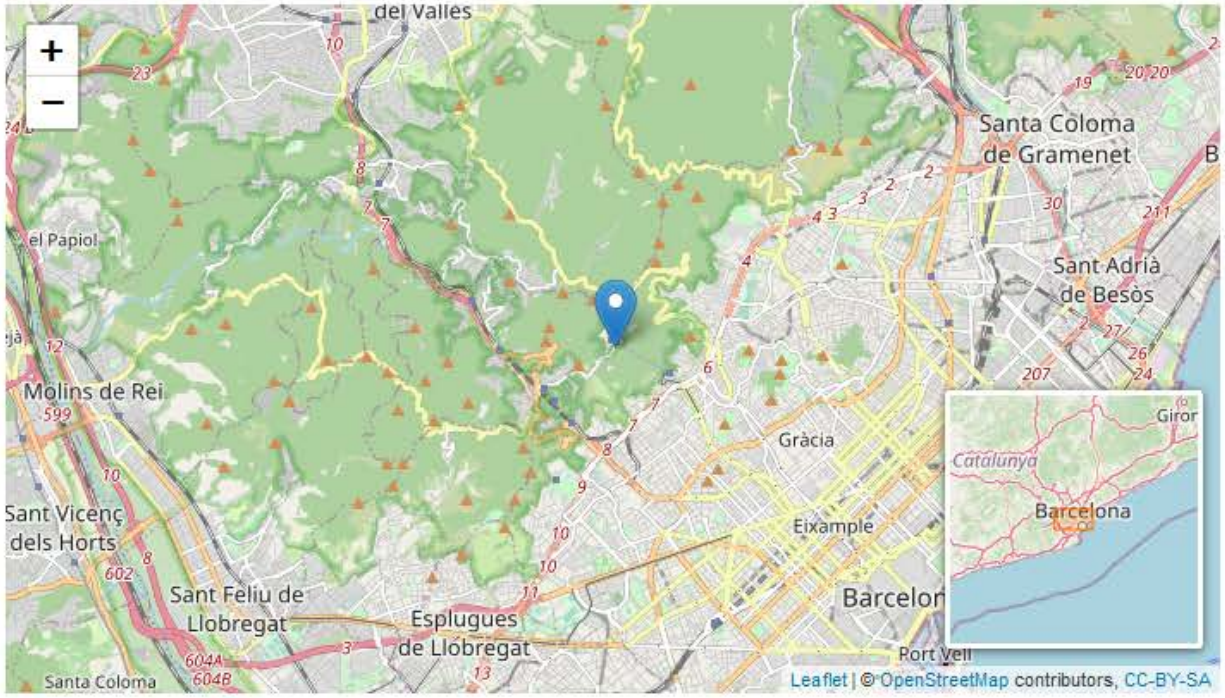
Table of Contents



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Methodology – Target regions

Downscaling to a region (South America)



Downscaling over a point (Barcelona, Observatori Fabra)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

HARMONIZE

Methodology – Datasets

Seasonal Prediction Outputs (SEAS5 Model)

Coupling model: IFS + NEMO (both models with native horizontal resolution 0.25° x 0.25°)

Horizontal grid spacing: Outputs regridded to 1° x 1° (~100 km)

Hindcast period: 1993 - 2018

Native time resolution: 6h

Ensemble members: 25

ERA5-Land Reanalysis Data (South America)

Horizontal grid spacing: 0.1° x 0.1° (~10 km)

Available period: 1993 - 2018

Native time resolution: Hourly

Station Data (Observatori Fabra)

Available period: 1993 - 2018

Native time resolution: Daily

Main Variables

2m air temperature (tas)

Daily minimum temperature (tasmin)

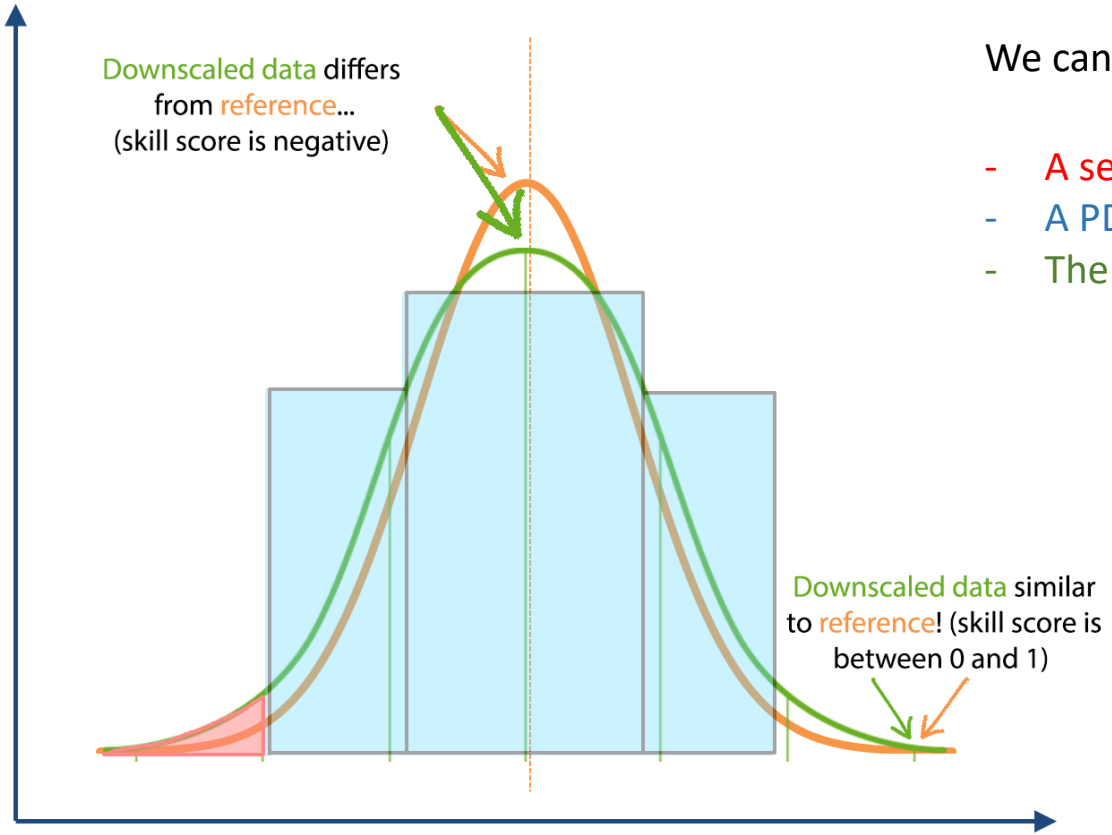
Daily maximum temperature (tasmax)

24h-accumulated precipitation (prlr)

Methodology – Forecast verification metrics

- Downscaled data is obtained by training model data with observational reference (using leave-one-out cross-validation). **How do we validate the skill of downscaled data to outline the best SD method?**

- **Forecast verification metrics (skill scores) compare the PDF of the SD outputs to that of the observational reference.**



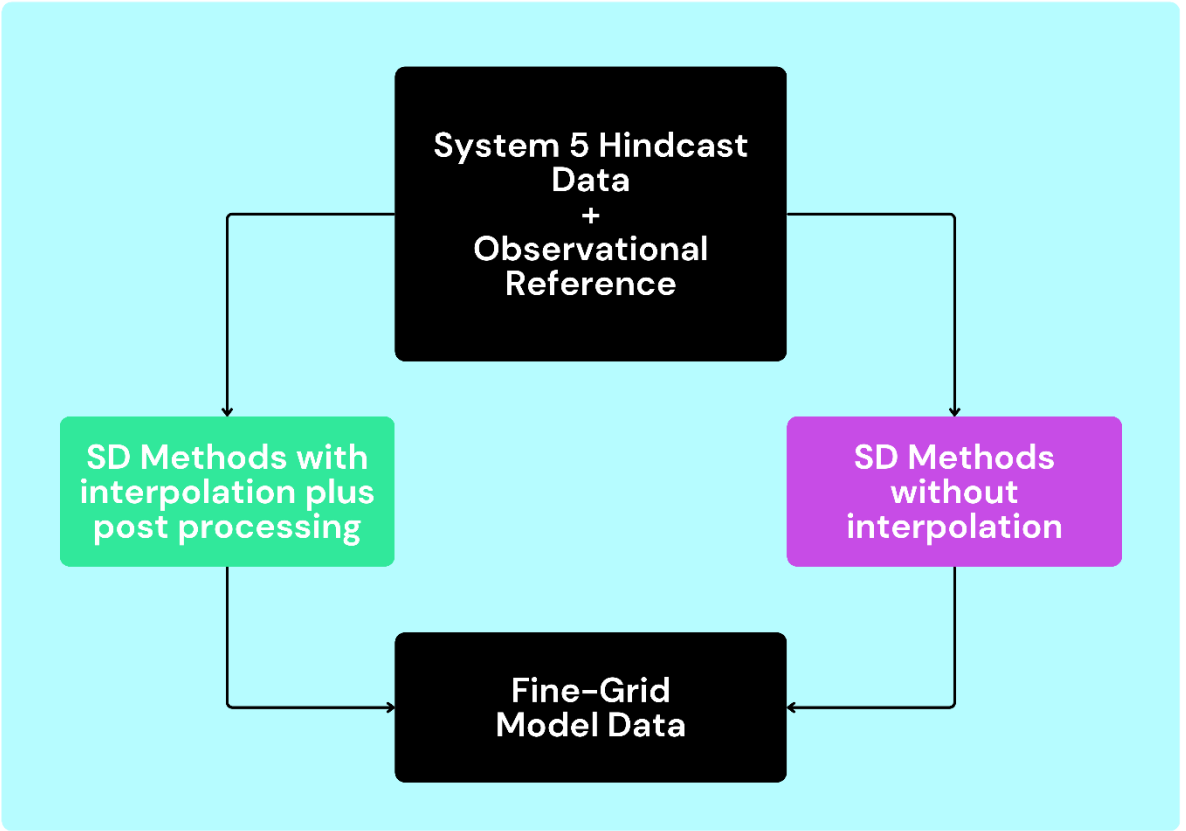
We can prioritize…

- A set percentile (Brier Skill Score)
- A PDF discretized by terciles (RPSS)
- The PDF as a whole (CRPSS)

Methodology – SD Methods

The 24(!) tested methods can be grouped in two big categories:

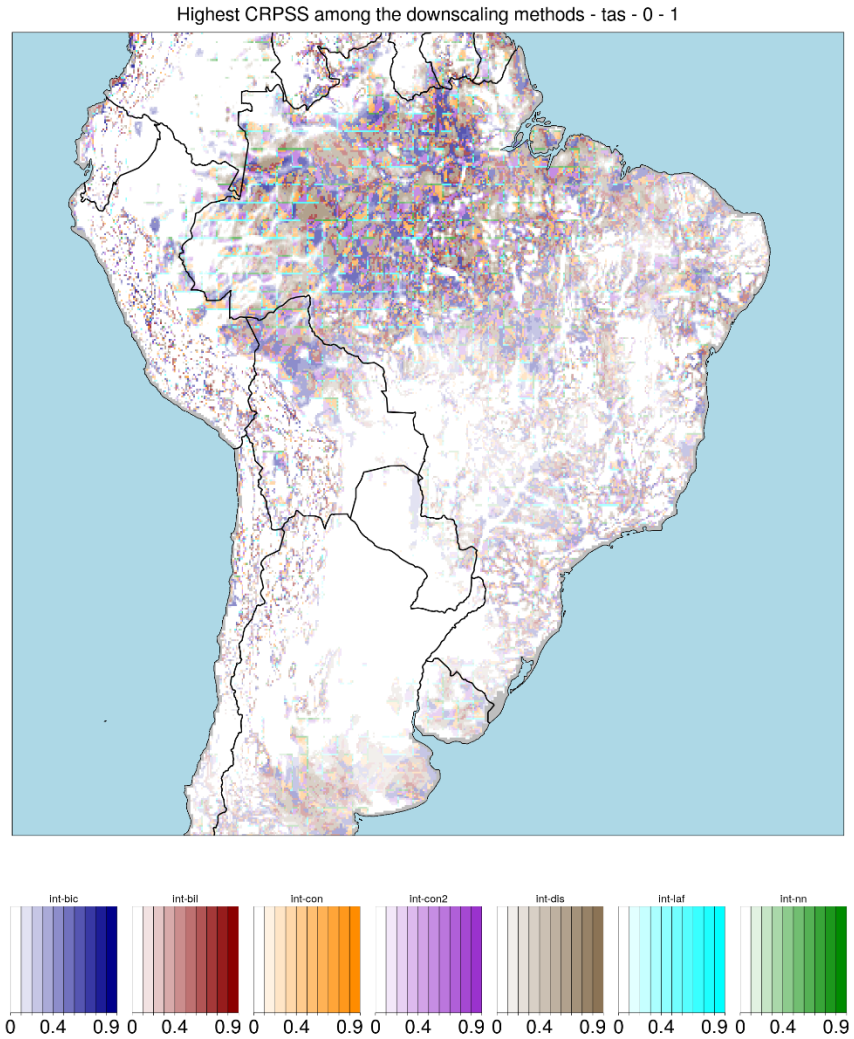
- Interpolation plus post processing techniques (bias correction, linear regression...)**. We need to outline the best interpolation method first**
- Non-interpolation methods (analogues and linear regression with 4nn)

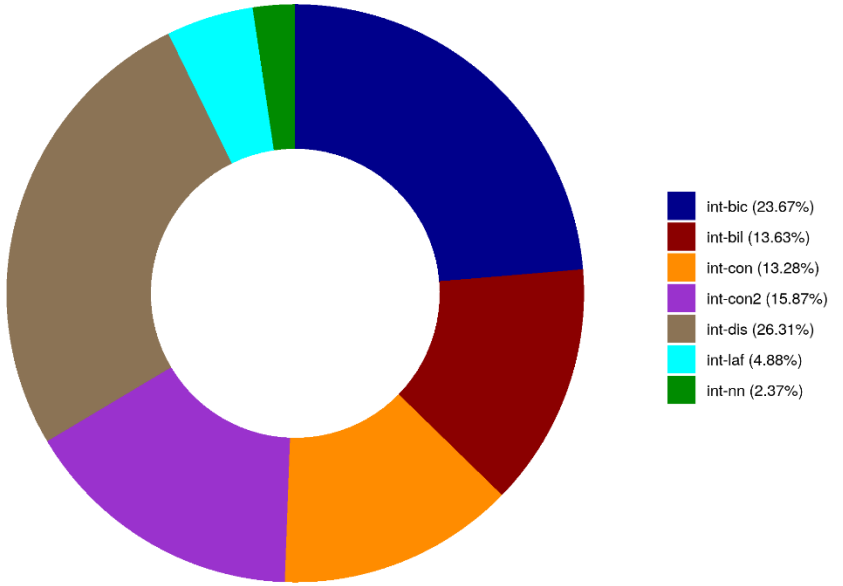


Downscaling to a region (South America)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Which interpolation methods are better?





Interpolation methods

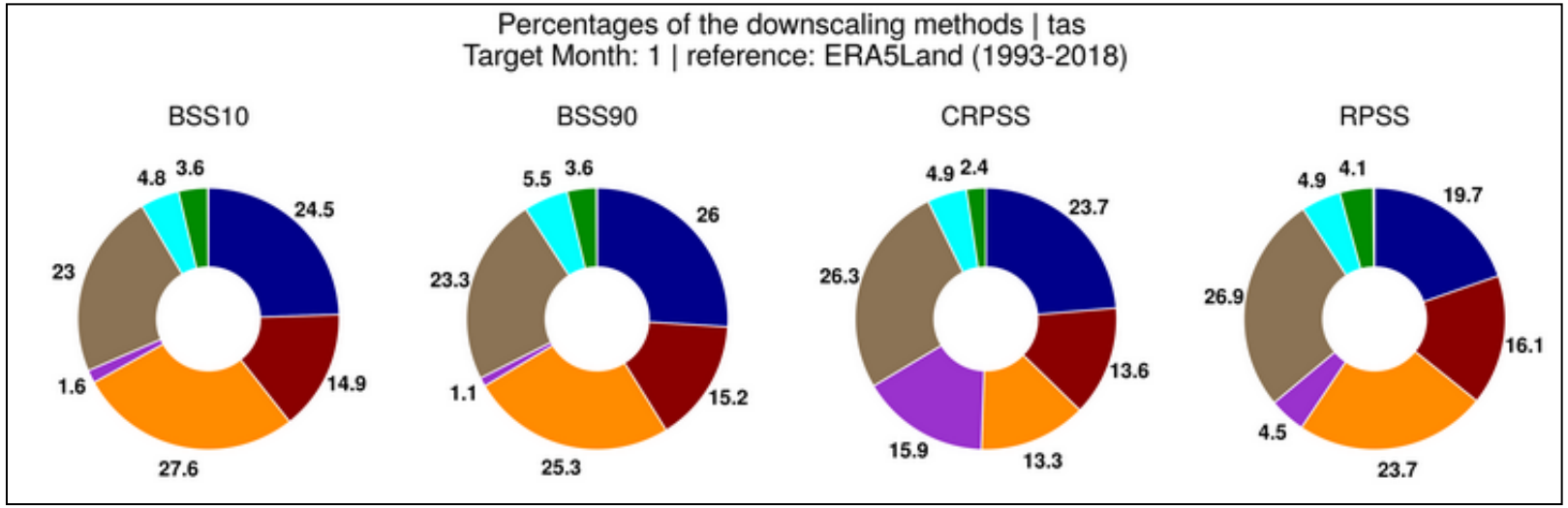
- Nearest Neighbour
- Bilinear
- Bicubic
- Conservative (1st order)
- Conservative (2nd order)
- Large-area fraction
- Inverse distance

Using model data alone

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

HARMONIZE

Which interpolation methods are better?



Three interpolation methods hold the same weight across the analysis…but **precipitation is constrained to the conservative method** (conservation of mass!)

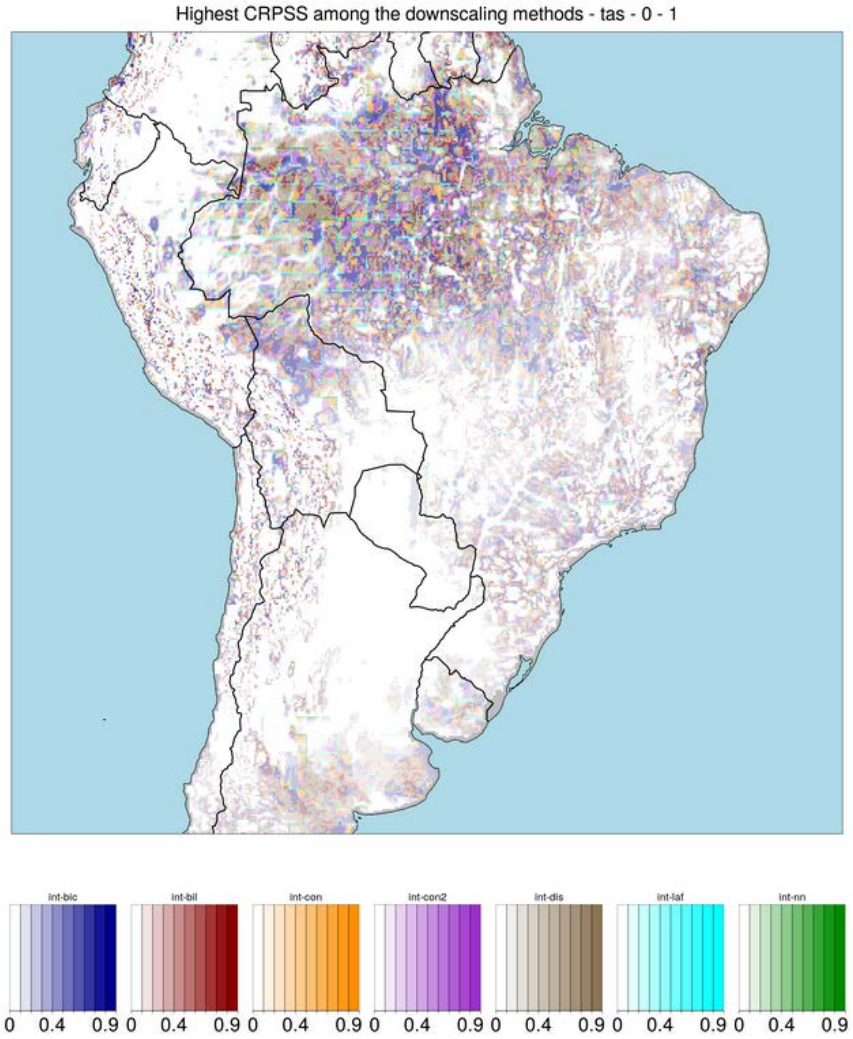
These best methods will be accounted for in following post processing techniques

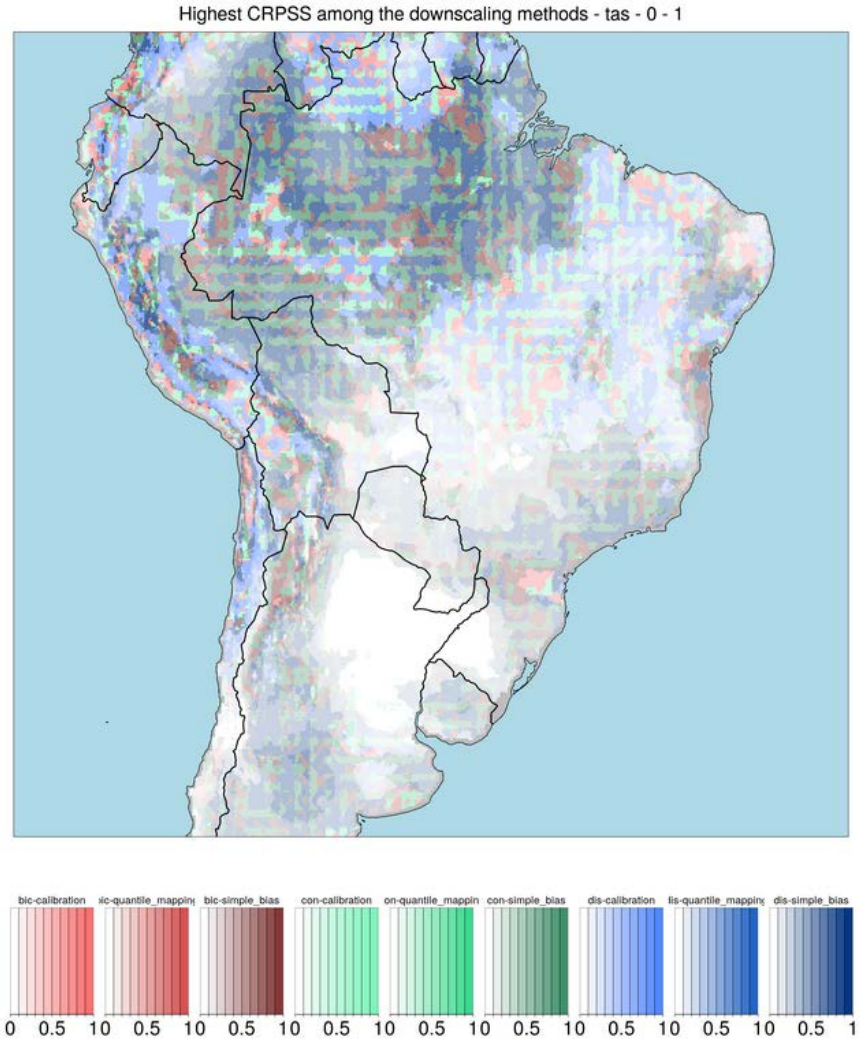
Interpolation methods

- Nearest Neighbour
- Bilinear
- Bicubic
- Conservative (1st order)
- Conservative (2nd order)
- Large-area fraction
- Inverse distance

Using model data alone

From model data alone…

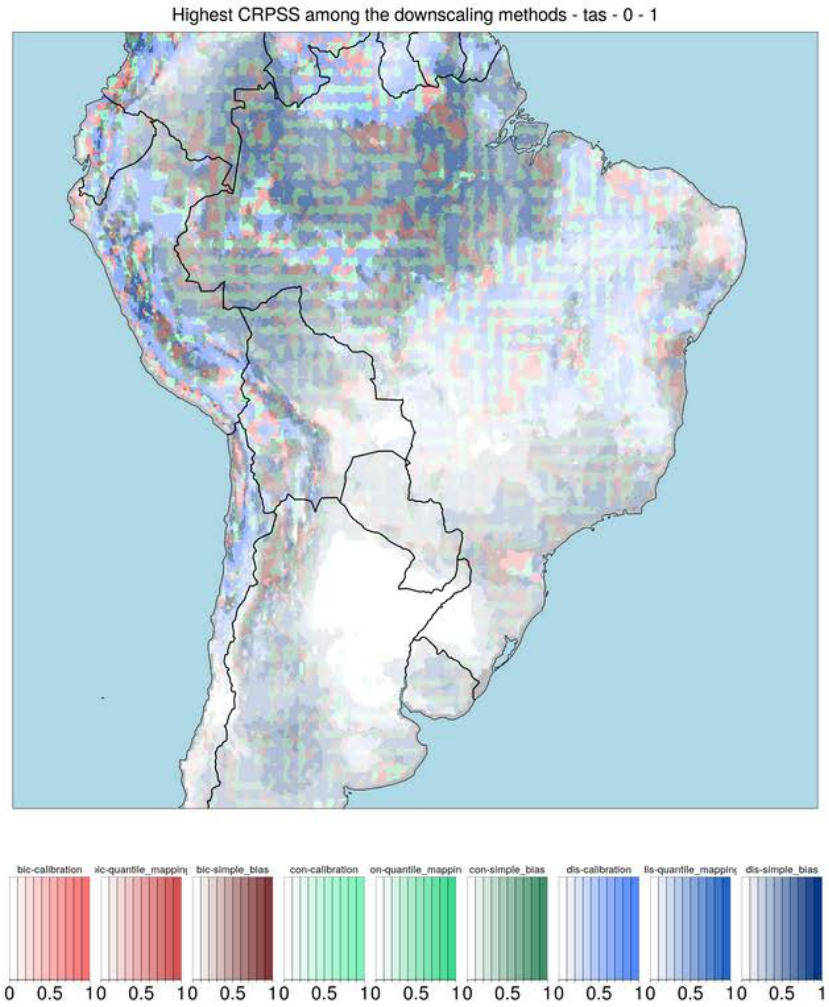


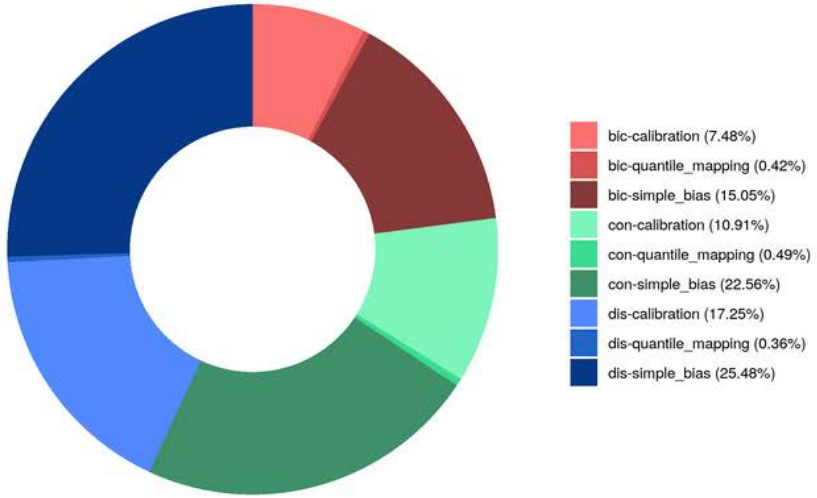


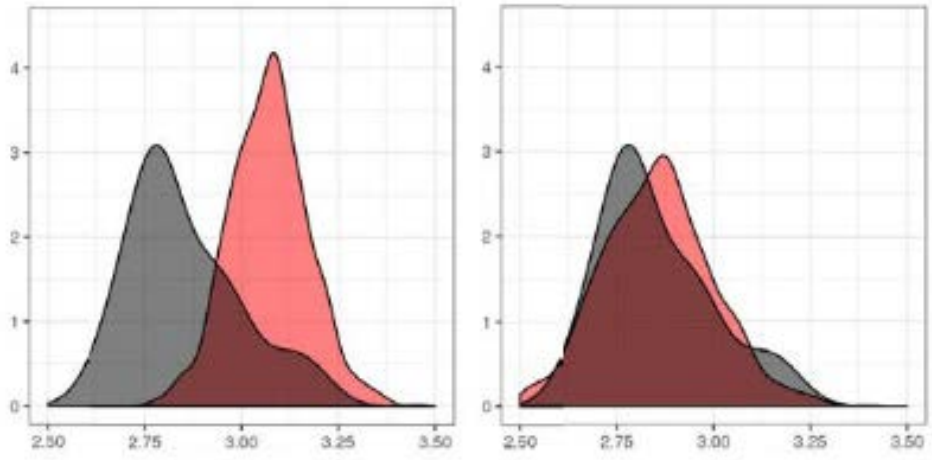
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

…to a bias corrected output

Interpolation plus bias correction







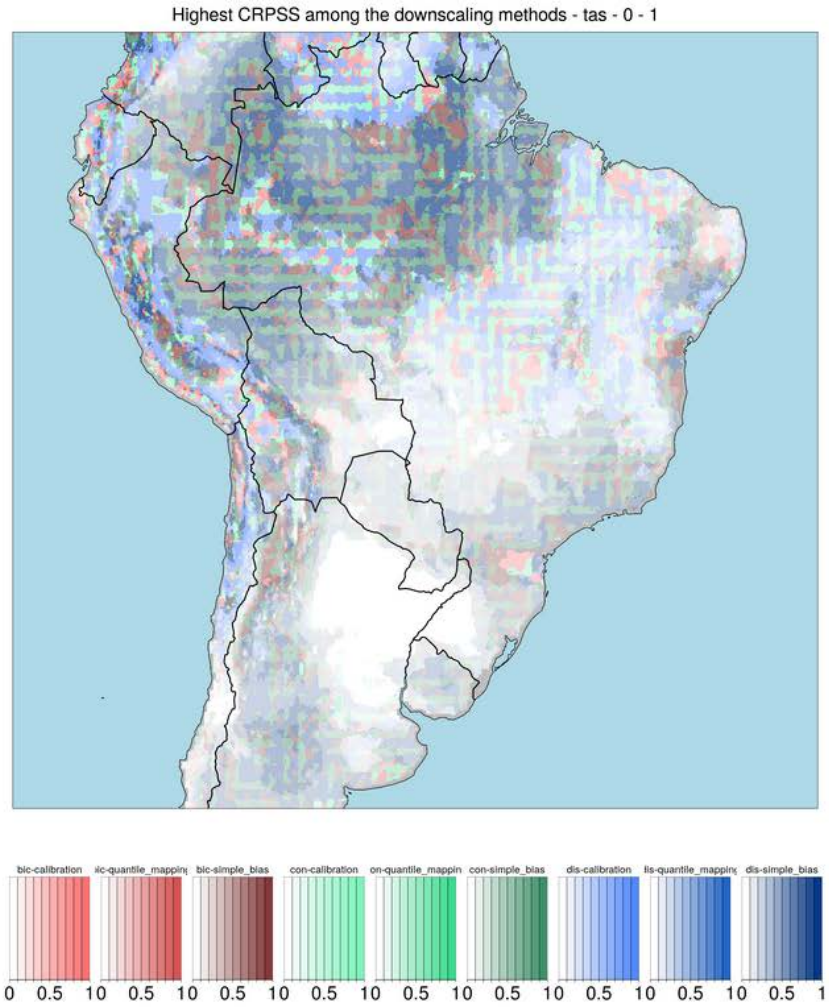
Interpolation plus bias correction methods

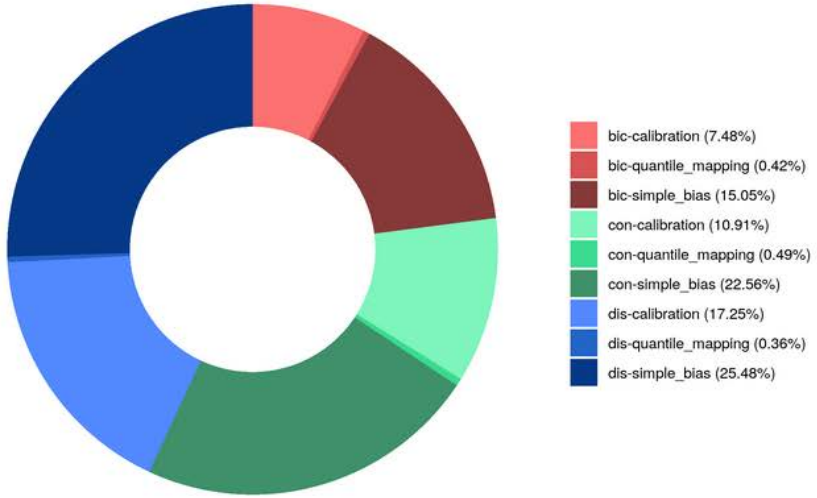
- Calibration
- Simple Bias
- Quantile Mapping

Using model data + observational reference

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Interpolation plus bias correction





Interpolation plus bias correction methods

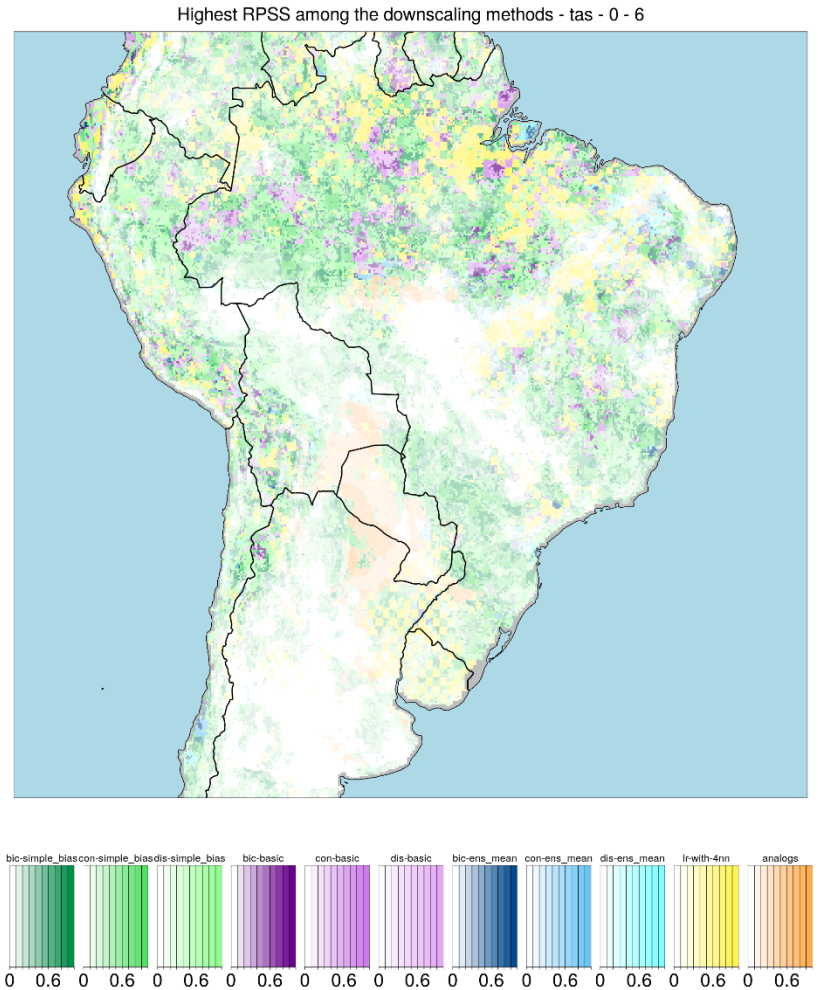
- Calibration
- Simple Bias
- Quantile Mapping

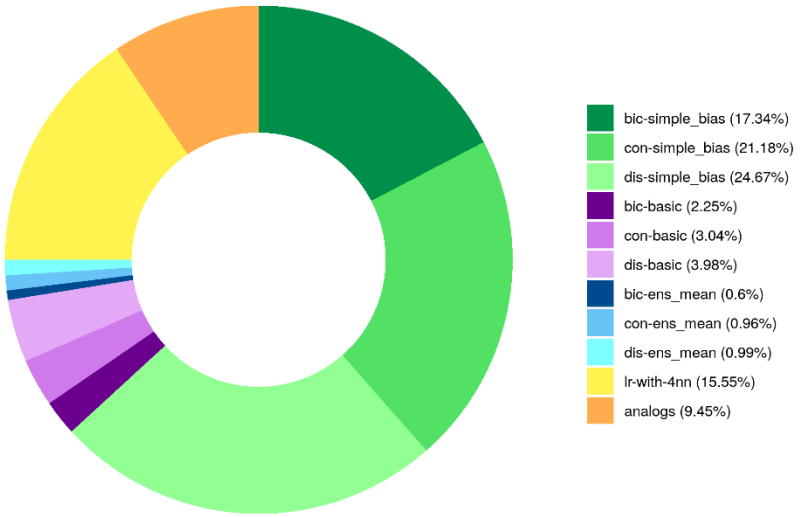
Using model data + observational reference

Other interpolation plus post processing techniques have the same positive skill regions, but **only bias adjustment techniques are exceptionally good in correcting orographic effects**

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Full comparison (best SD methods)





Best SD methods

Interpolation plus post processing	Without interpolation
Int + Simple Bias Correction	Linear Regression w/ 4nn
Int + Linear Regression	Analogues
Int + Logistic Regression w/ ensemble mean	

Using model data + observational reference

Interpolation plus simple bias is the best overall SD method

Linear regression w/ 4nn is the 2nd best SD method (sometimes even challenging simple bias in some months)

Analogues method is fairly variable (heavily relies on hindcast length and ensemble size)

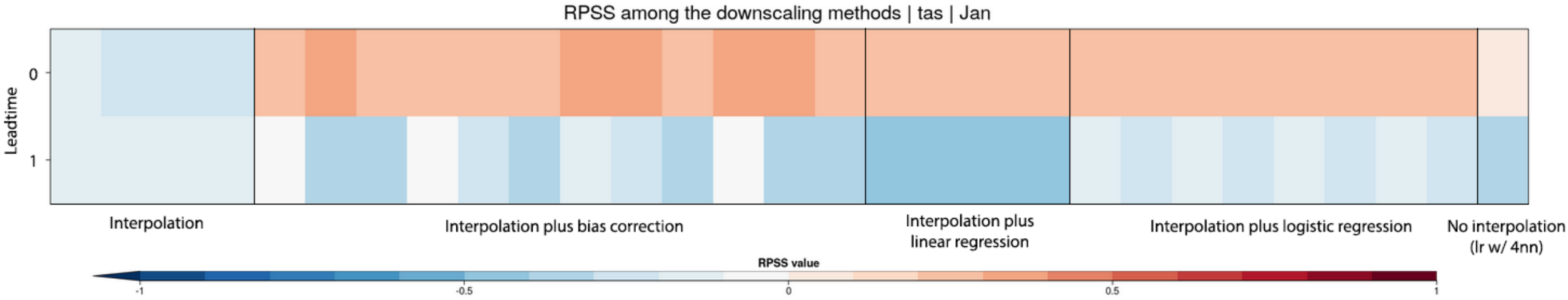
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

HARMONIZE

Downscaling to a point (Observatori Fabra, Barcelona)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Full comparison (all SD methods)



Interpolation plus bias correction and interpolation logistic regression are the best SD methods…but there are no evident patterns to discern which of the two is better

Representative errors may lead to unexpected skill values (skill usually decreases with leadtime…)

Conclusions and Future Work

Region-wide skill comparison
(South America)

- **Initial SEAS5 outputs present positive skill over the Amazon basin over seasonal timescales**.

- **Interpolation + bias correction techniques are particularly useful in orographic regions**. In southern regions, the analogues method is rather more advisable.

- **Interpolation + simple bias adjustment is the best SD method overall**. Linear regression w/ 4nn follows as the second best SD method, while the analogues' method can presents a variable result depending on the target month to be observed.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Conclusions and Future Work

Point-location skill comparison (Barcelona, Observatori Fabra)

- **Int + bias correction and int + logistic regression are the best SD methods over a target point in Barcelona**, though there are no evident patterns to prioritise one over the other.
- **Representativeness errors are responsible for the low skill values** noted for the downscaled predictions to Fabra Observatory, as the station observational reference captures point variability much better than SEAS5 data.

Future Work

- **Some SD methods (e.g. analogues method) should be further tested with a higher hindcast range and ensemble size**, to ensure whether the downscaled results obtained here are of use to any potential user.
- For a point analysis, assessing SEAS5 data against a model-driven observational reference (e.g. reanalysis) could lead to better skill results.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación


Thanks for listening!
BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

Want to see more results?

Let's take this somewhere else…

Barcelona Supercomputing Center
Centro Nacional de Supercomputación