

Lecture 4. The Basics of Downscaling Methodology

Yun-Young Lee
7 Sep 2016



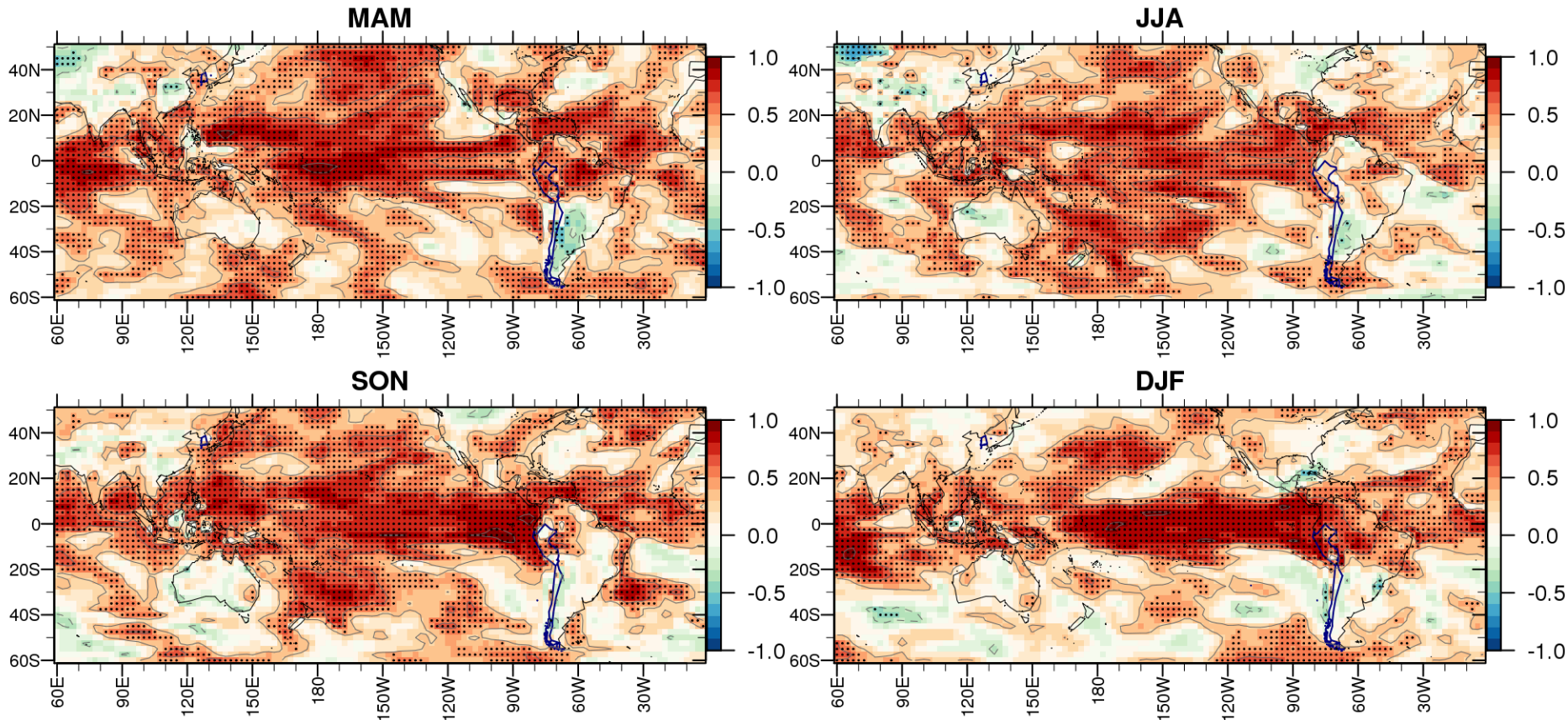
Why downscaling is required?



1. Low skill of dynamical seasonal forecast...

- near surface temperature...

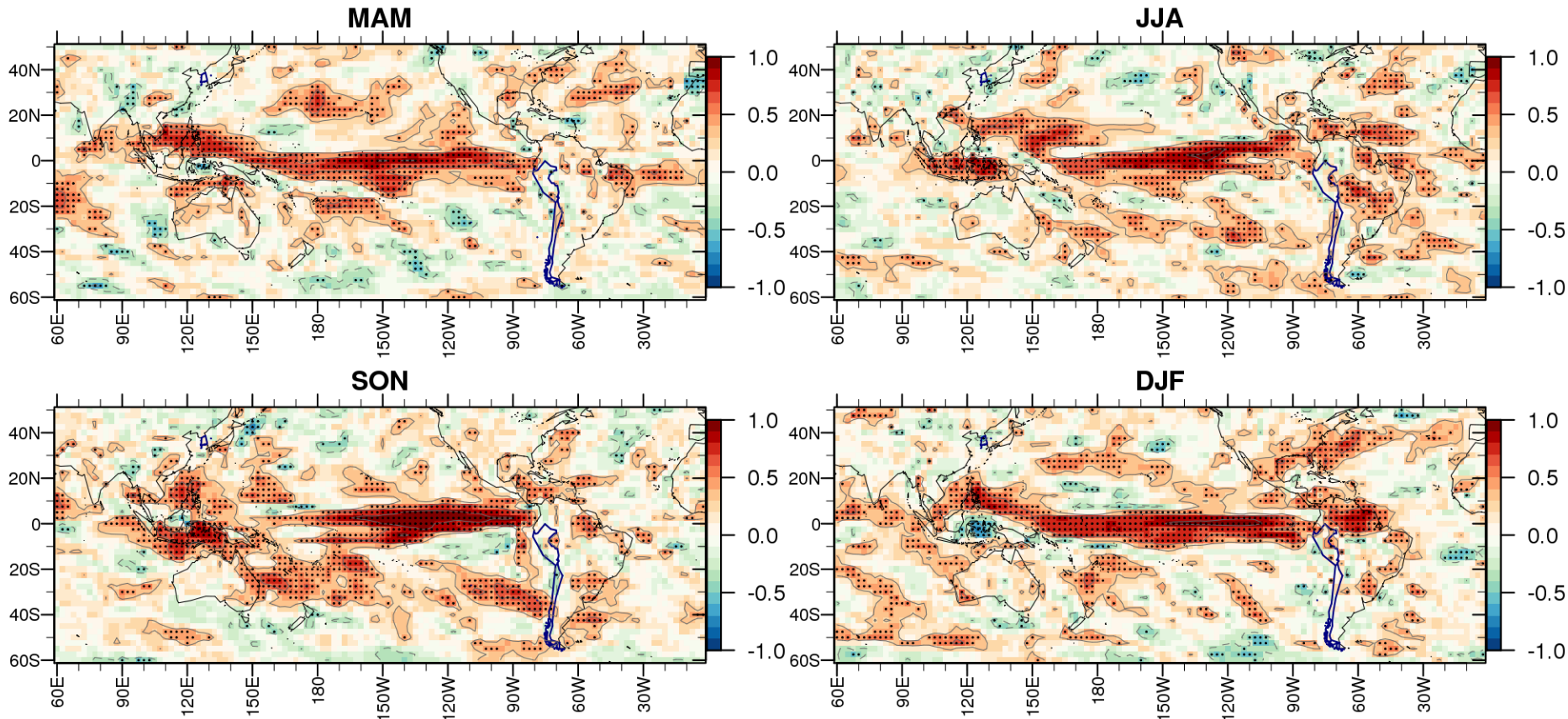
TCC (NNR2 vs SCM), t2m



1. Low skill of dynamical seasonal forecast...

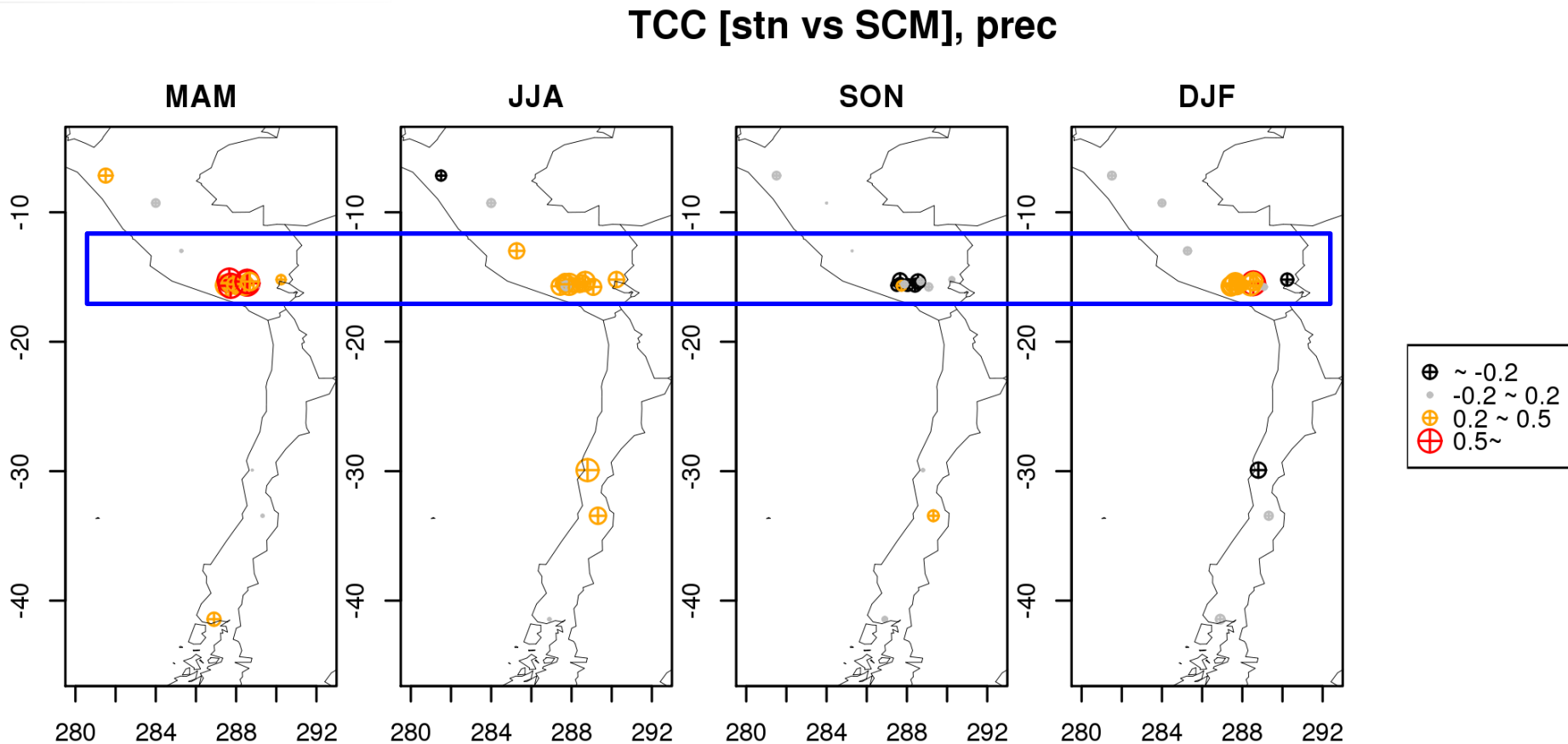
- **Rainfall ?**
- tropical region only?
- South America?

TCC (NRR2 vs SCM), prec



1. Low skill of dynamical seasonal forecast...

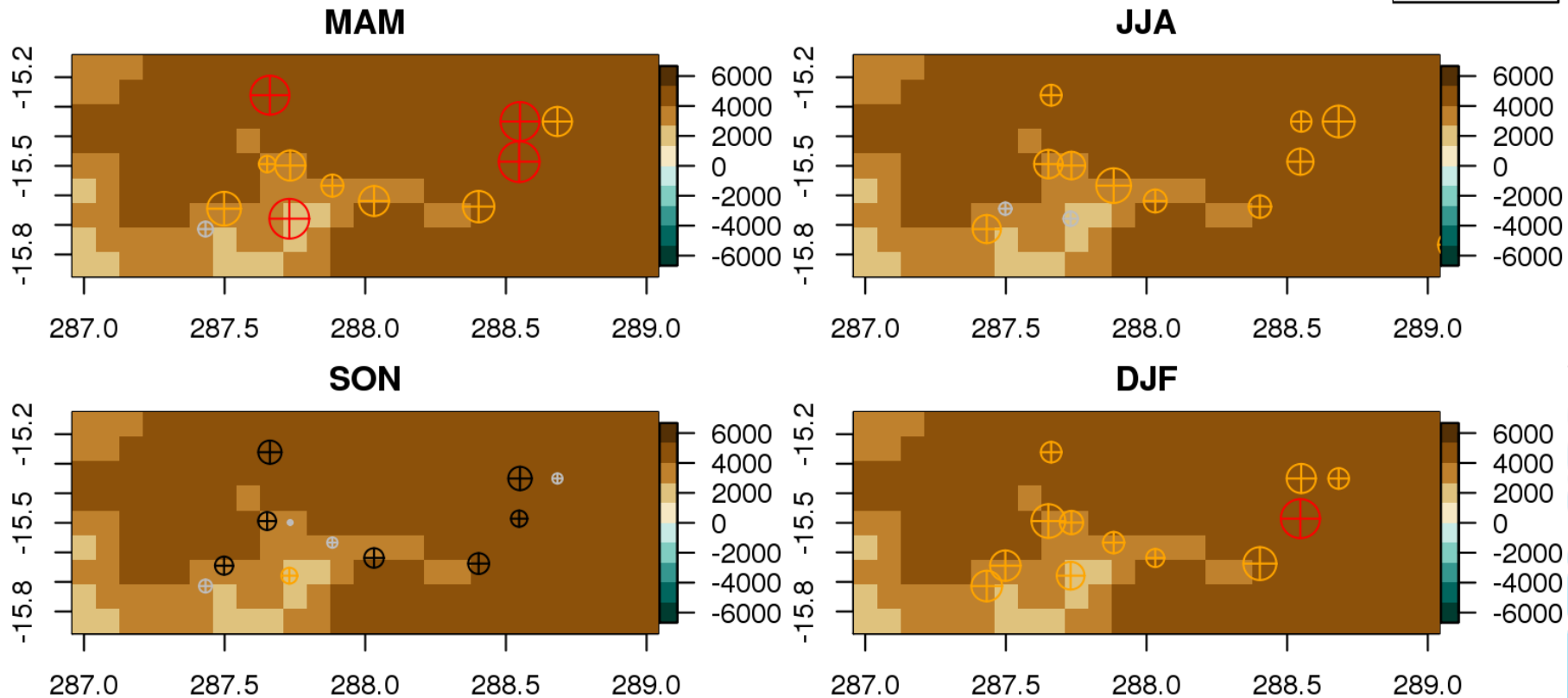
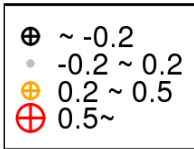
- In particular stations...



1. Low skill of dynamical seasonal forecast...

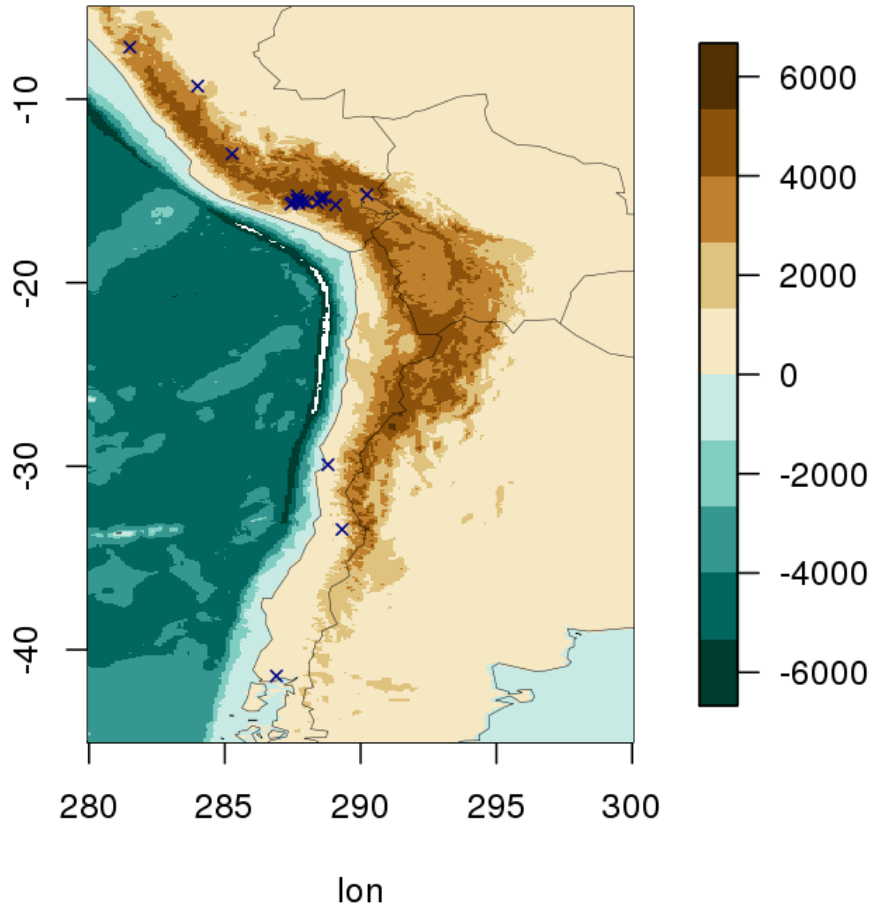
- In particular stations...

TCC [stn vs SCM], prec



Peru & Chile

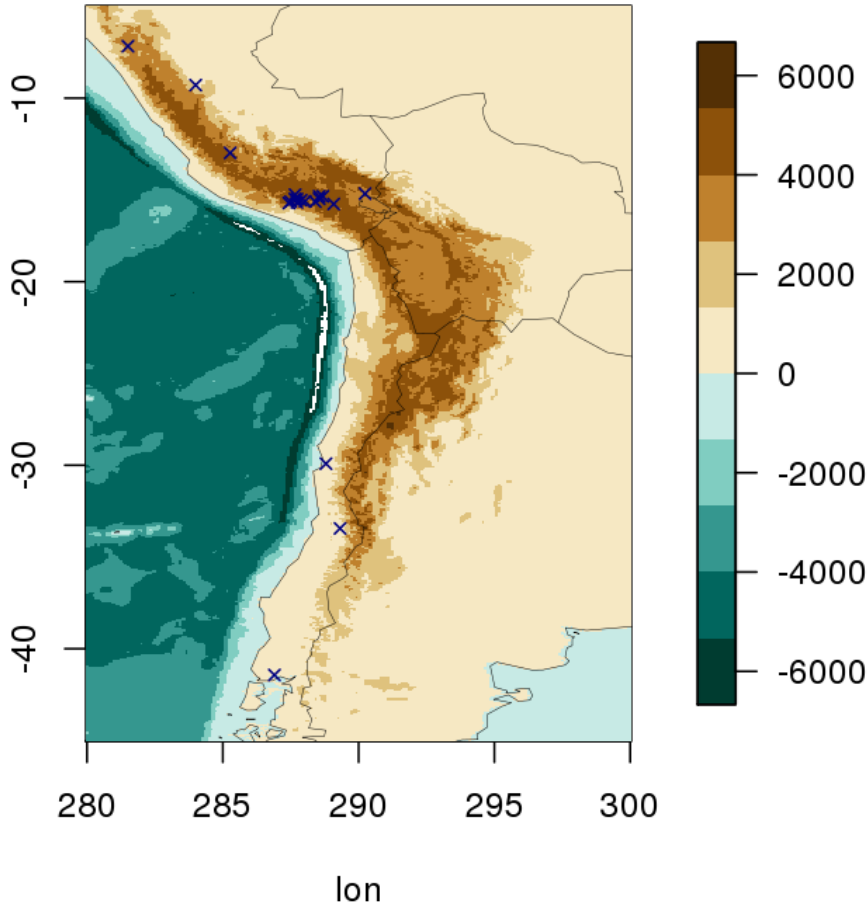
20 stations (Peru + Chile)



- **Pacific Ocean** to the west and the **Andes mountains** to the east running parallel to the Pacific Ocean
- Three regions: the costa (coast), the sierra (highlands), and the selva (jungle).
- Ocean climate vs Mountain climate?

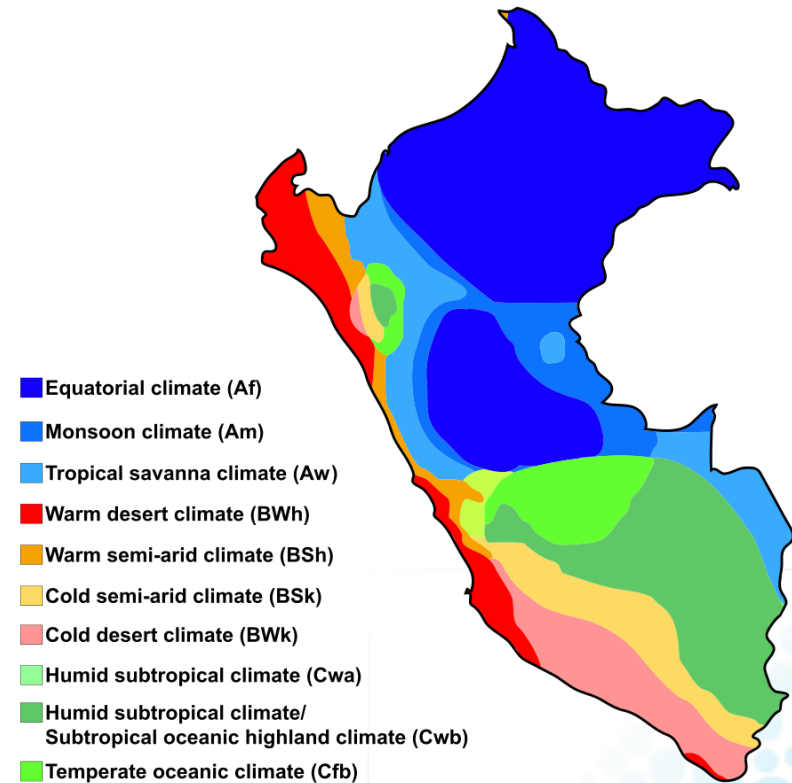
2. Climate Locality

20 stations (Peru + Chile)



- Complicate climate of Peru

Peru map of Köppen climate classification

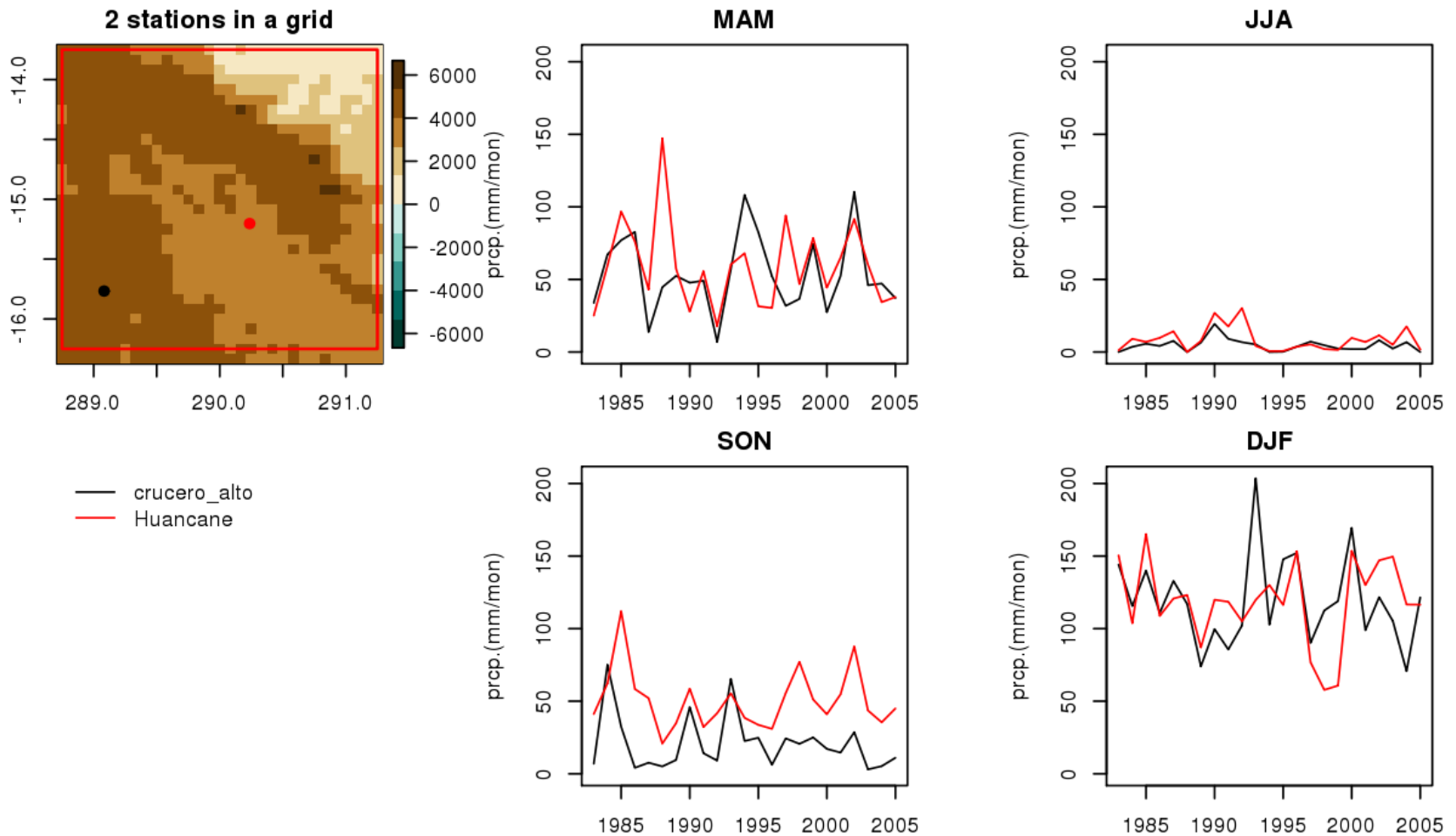


✓ Can 2.5 by 2.5 grid pixel simulate *steep terrain* effect therefore *locality of station climate*? Not really.

2. Climate Locality

How different is the climate between 2 adjacent stations **in the east of Arequipa?**
But, they are in one grid! OMG!!!

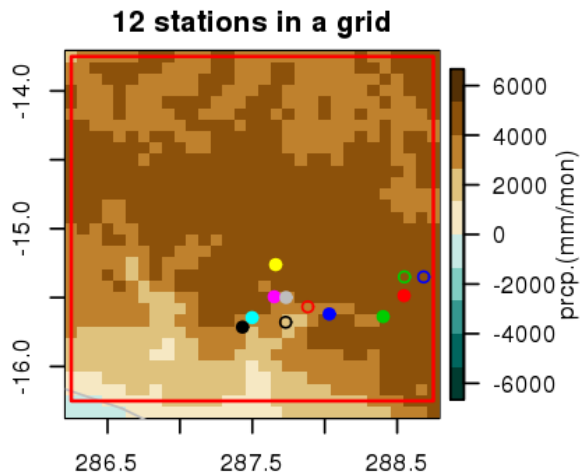
East of Arequipa [290, -15]



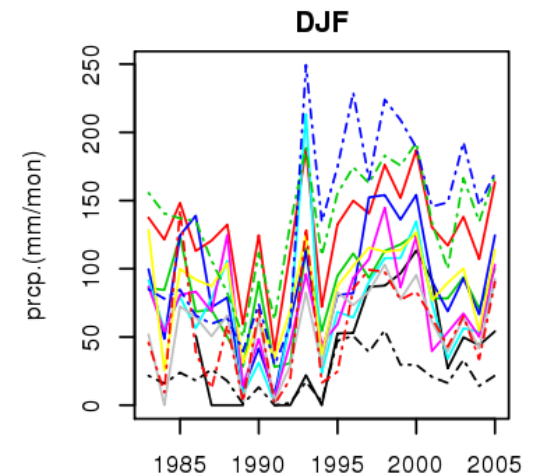
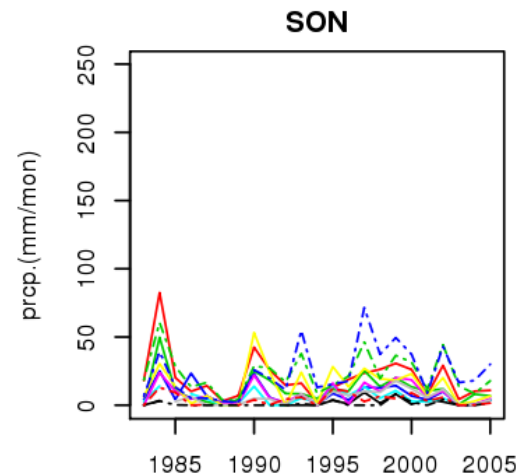
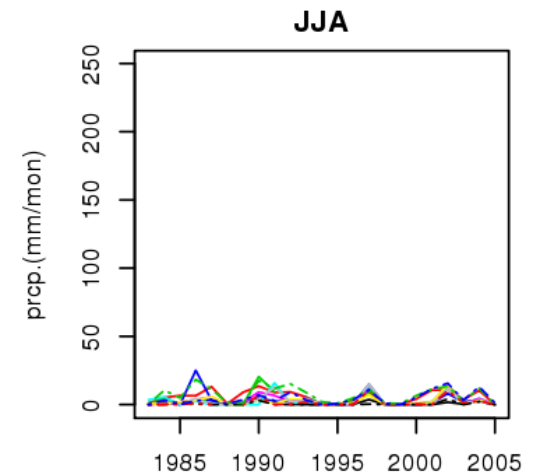
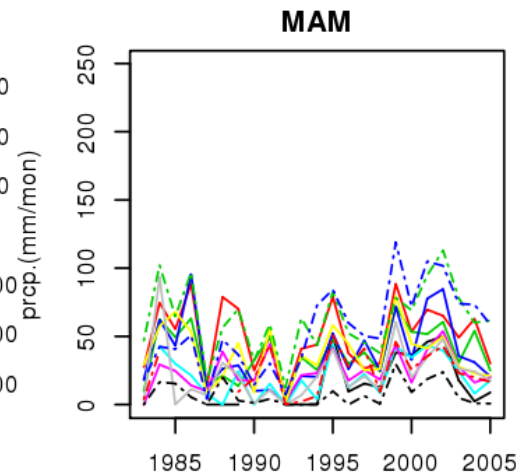
2. Climate Locality

How different is the climate between 12 adjacent stations near Arequipa?
But, they are in one grid! OMG!!!

Arequipa [287.5, -15]

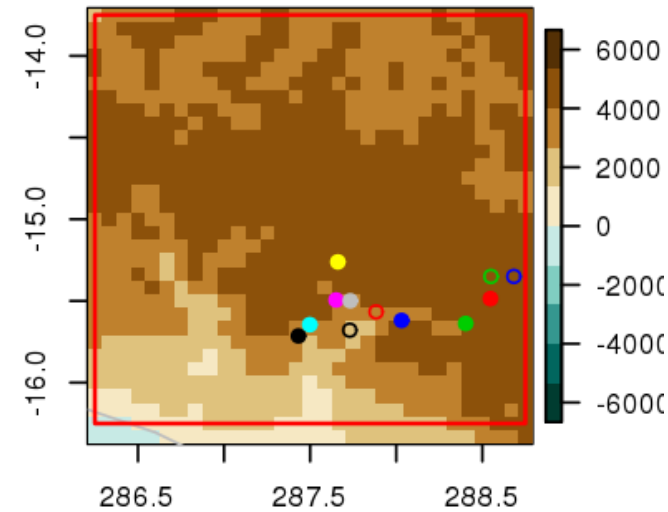


- pampacolca
- sibayo
- chival
- cabanaconde
- machahuay
- andahua
- orcopampa
- chachas
- - - ayo
- - - choco
- - - tisco
- - - porpera



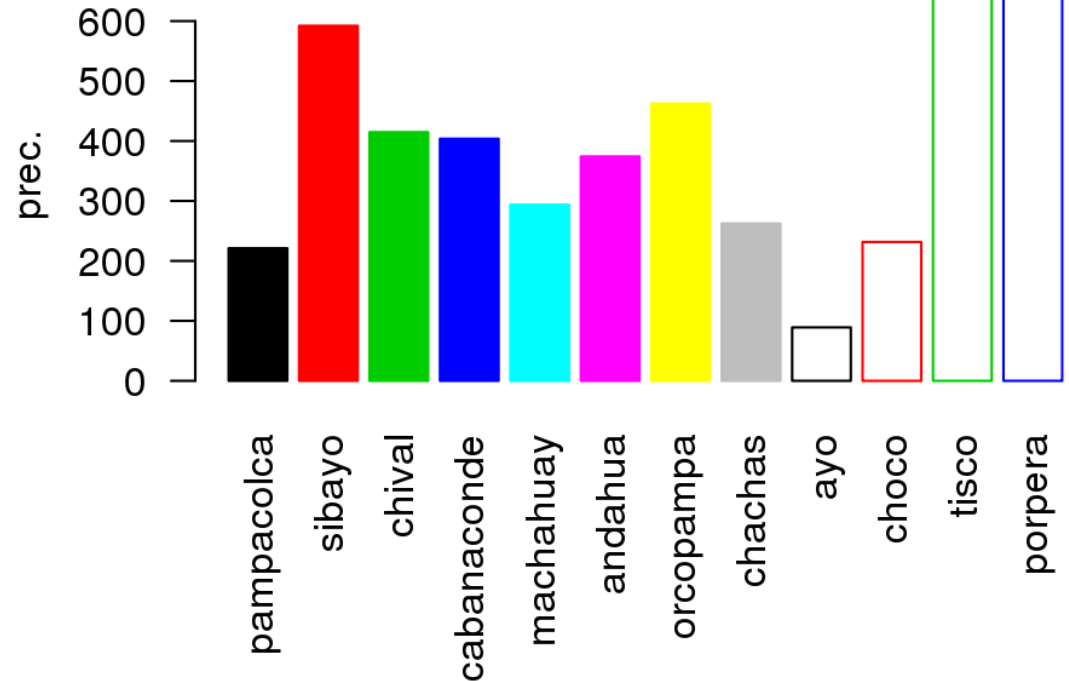
2. Climate Locality: annual rainfall

12 stations in a grid



- pampacolca
- sibayo
- chival
- cabanaconde
- machahuay
- andachahua
- orcopampa
- chachas
- - - ayo
- - - choco
- - - tisco
- - - porpera

Annual rainfall

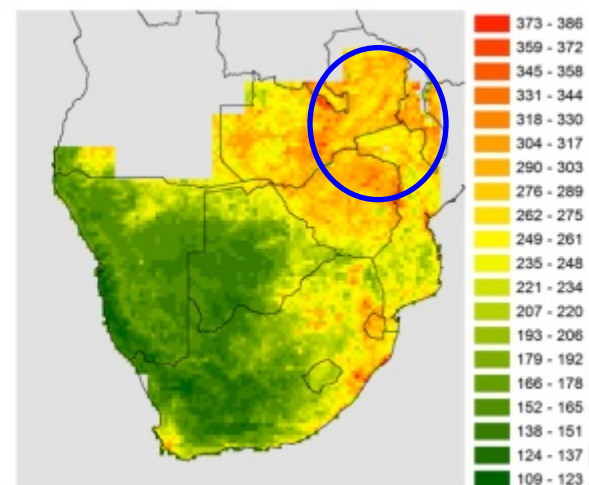
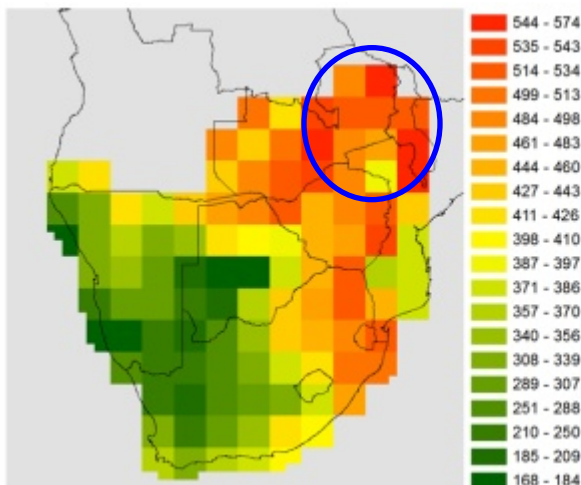




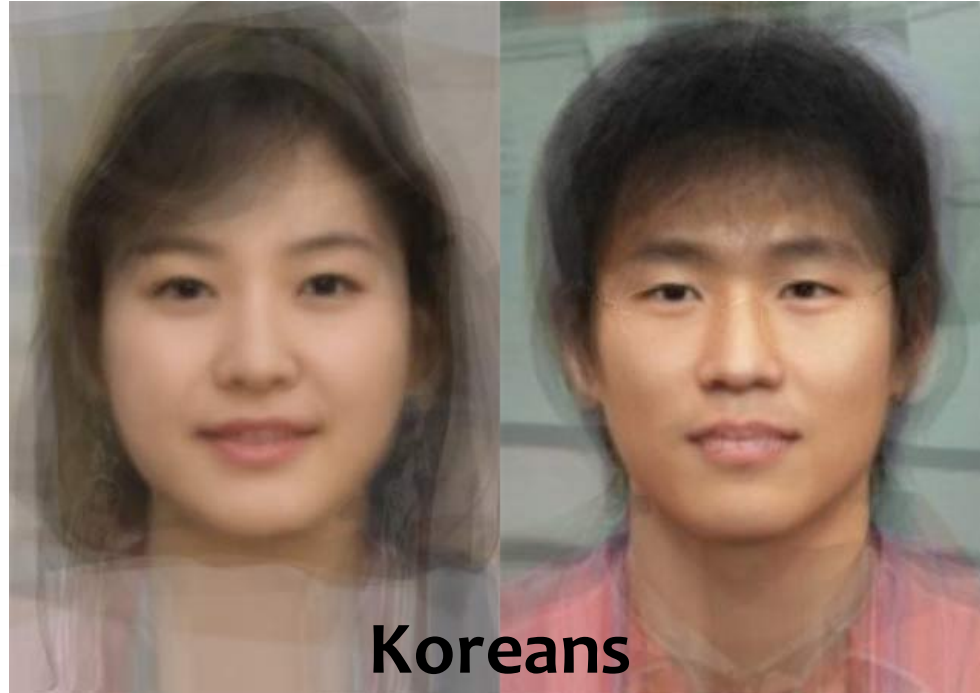
What is downscaling?



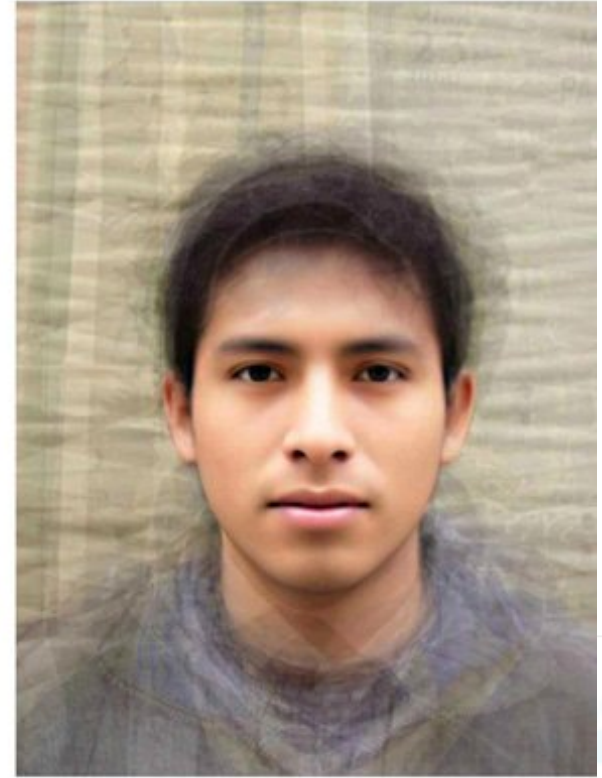
Downscaling



Average face (1)



Average face (2)



the Face of Tomorrow

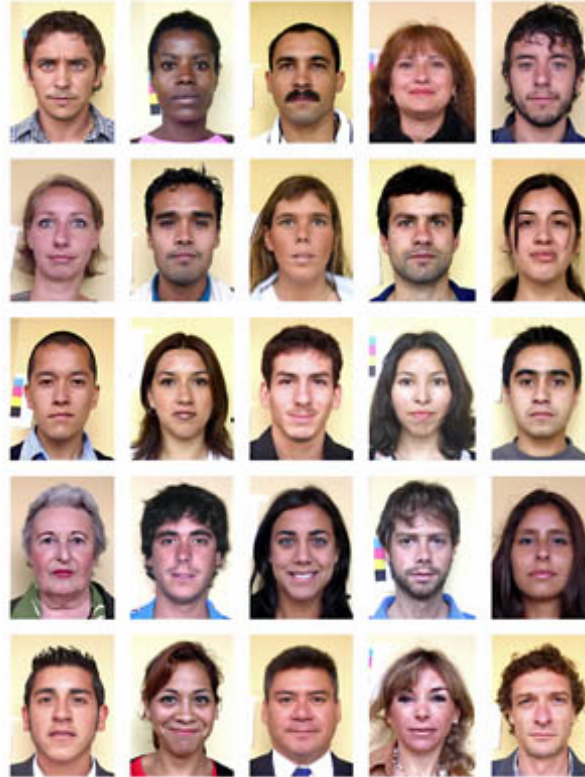
Mike Mike

faceoftomorrow.com

Lima
Jiron de la Union

2004.07.13

Average face (3)



the Face of Tomorrow

Mike Mike

faceoftomorrow.com

Santiago de Chile

Los Leones

2004.02.25

Downscaling is like... isolating each of your faces from “Average Face” of Peruvian and Chilean.

Possible? Not really.

If we know **information of your ancestors looking** such as *how they were similar/different to the average face*, then?

Maybe possible...

In seasonal prediction,
how to draw the **information of a particular “point” from COARSE grid model data** ?

We need *additional information* to downscale!

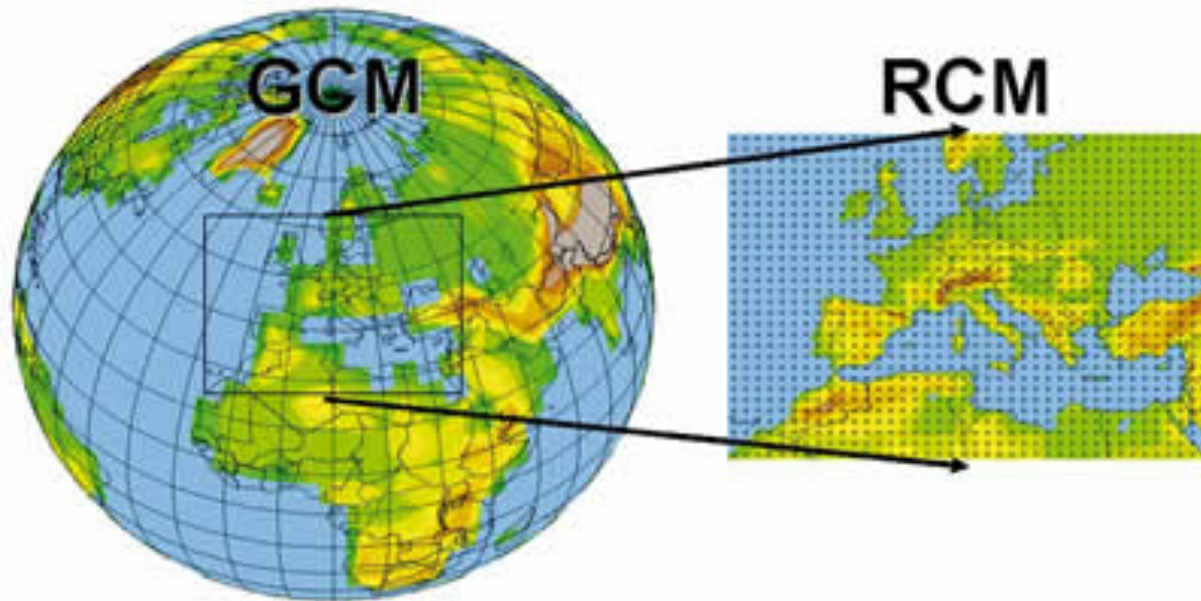
How to overcome?

- Resolve the (dynamical/physical) process determining point value
- Find a ***relationship between the point and large scale field***

What is dynamical downscaling?

- Simply, it is running RCM.

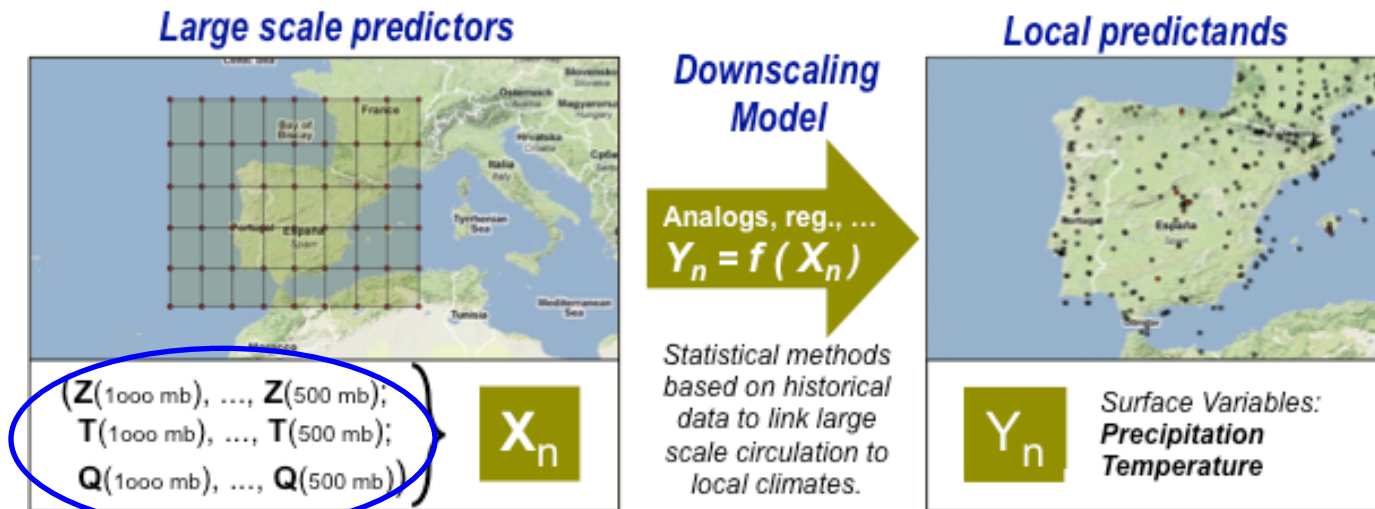
BC from GCM, IC \rightarrow solving dynamic equations!



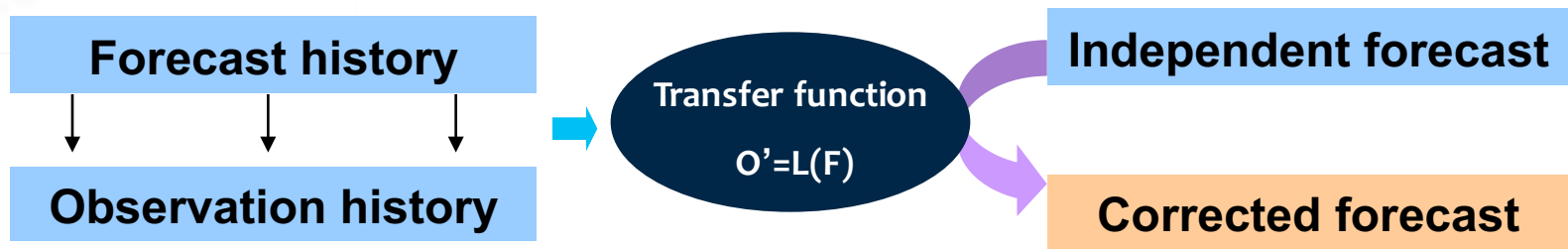
Statistical downscaling

- Based on statistical relationship between **precipitation/temperature** at particular stations and **adjacent/remote upper level atmospheric fields**

→ Developing simple downscaling model



Statistical Post-process

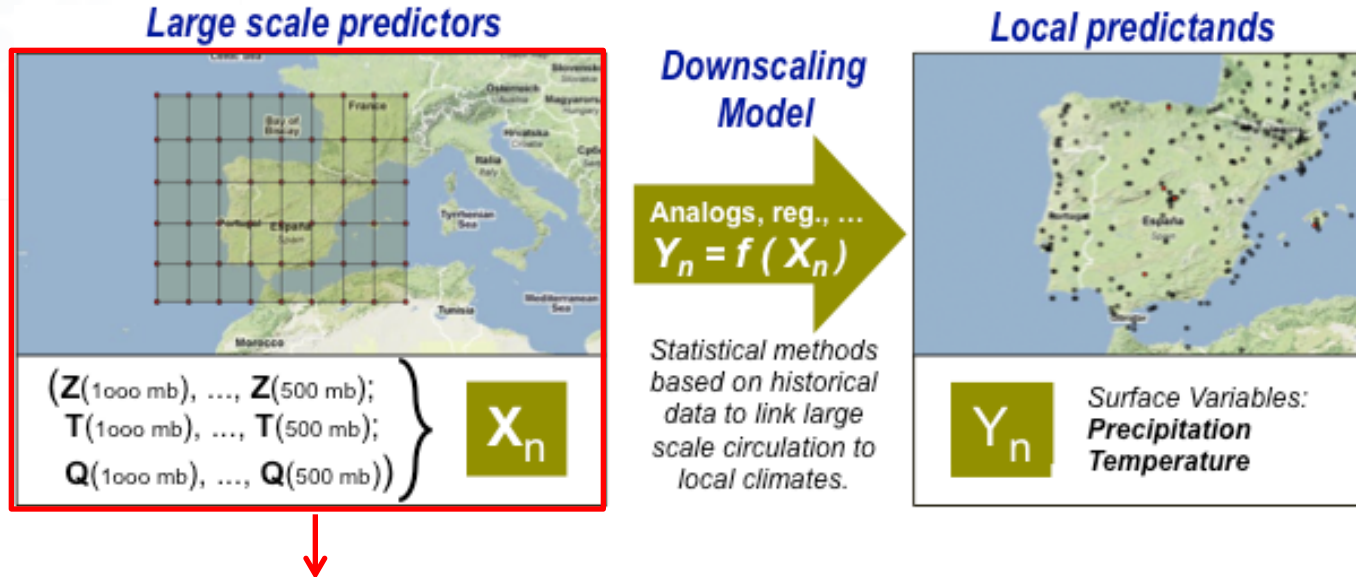


There are many approaches in post-process, All of them share similar assumption.: **Statistics between forecast and observation is stationary**

If statistics is not stationary, post-process will not work in independent forecast

Thus, statistical stability is a rule of thumb in the statistical post-process (avoiding overfitting)

Accuracy of Large Scale Information?



Simulated Large scale information is correct? **NOPE!!!**

Downscaling = Statistical Post-process

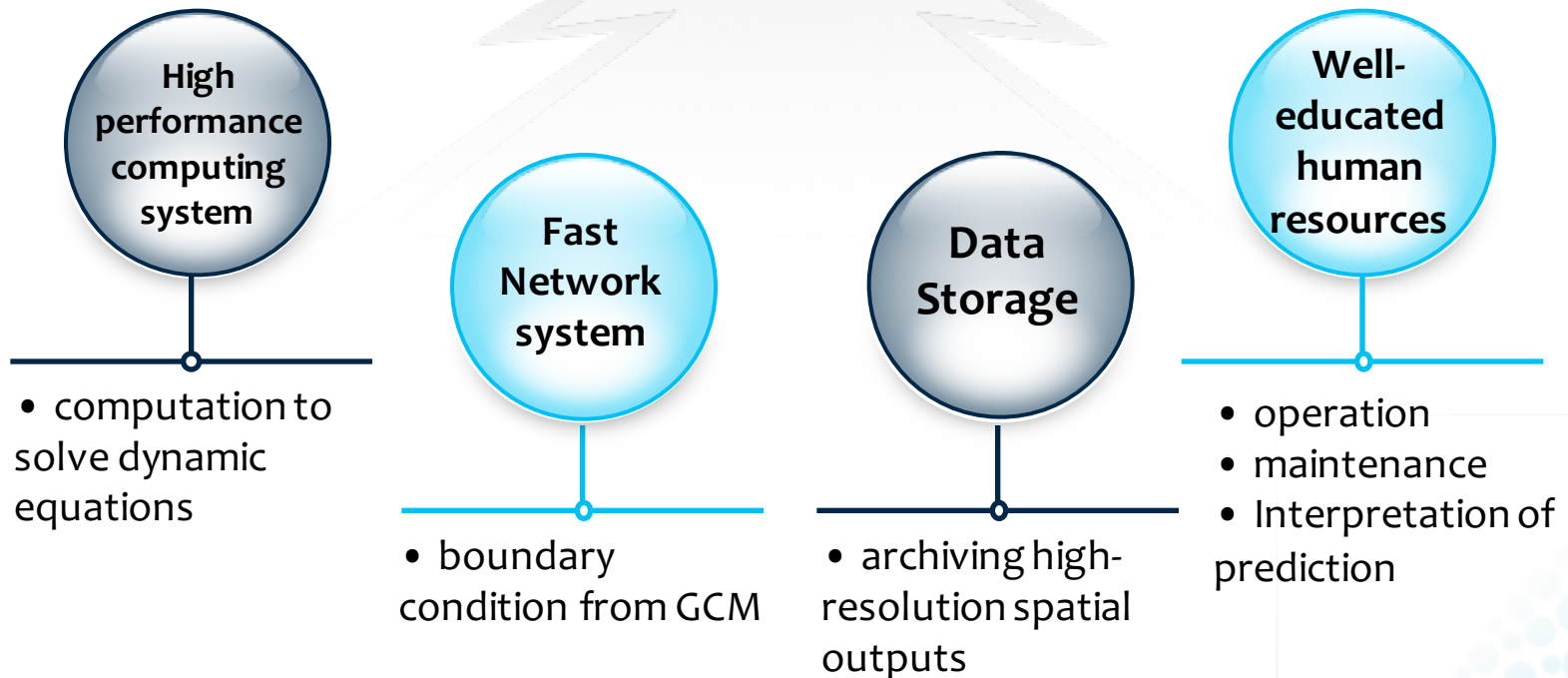


Why statistical downscaling rather than dynamical downscaling?



Dynamic downscaling is costing... - _ -

Dynamical downscaling



Statistical downscaling forecast



- ✓ No HPC needed
- ✓ No huge data storage needed
- ✓ Less dependent on internet speed

- ✓ Easy to interpret (simple but based on dynamical relationship)

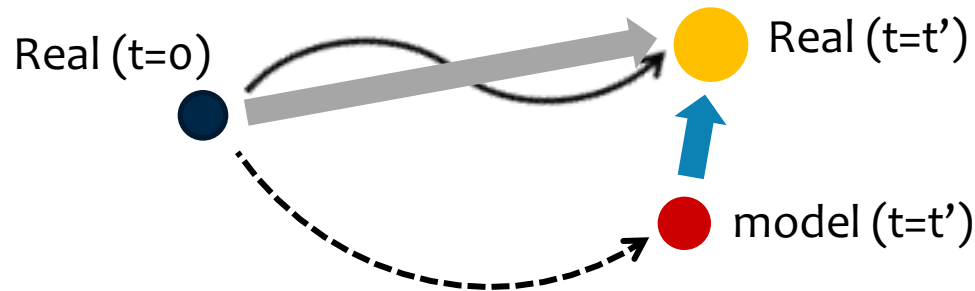
Approach

- Statistical downscaling forecast based on *past forecast*

$y(s, t)$: observation

$x(s, t)$: forecast

$$y'(t) = f(x(t), \alpha), \alpha = g(x(1 : t - 1), y(1 : t - 1))$$



- The most common way : Regression

$$\sum_j b_j y_j = \sum_i a_i x_i + \epsilon$$

If i & $j = 1$: Linear regression
 $i > 1, j = 1$: Multiple regression
 i & $j > 1$: CCA, SVD, etc

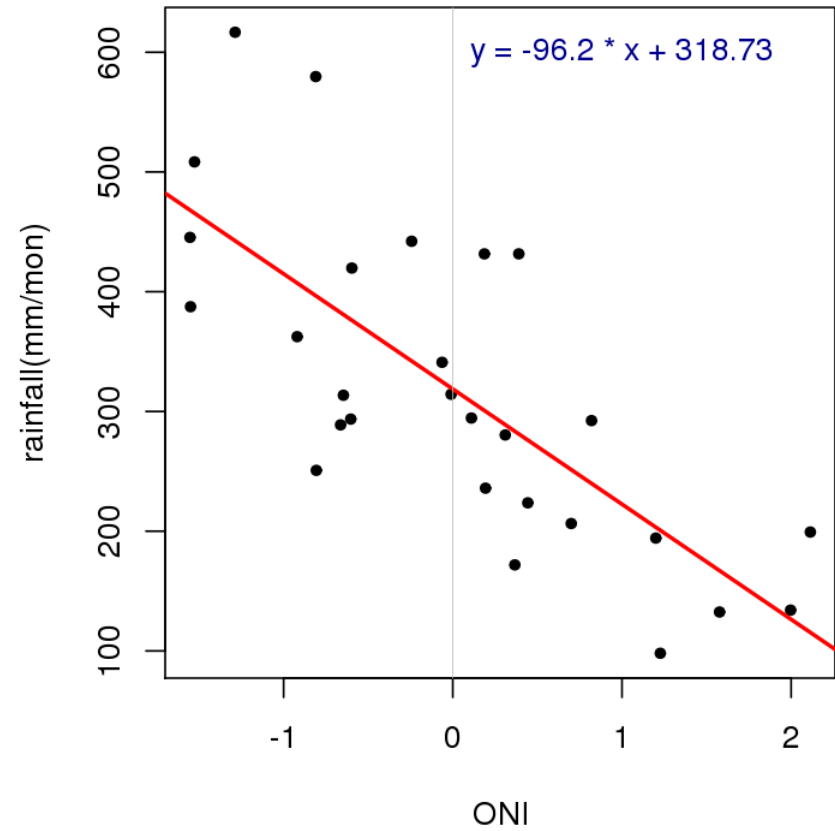
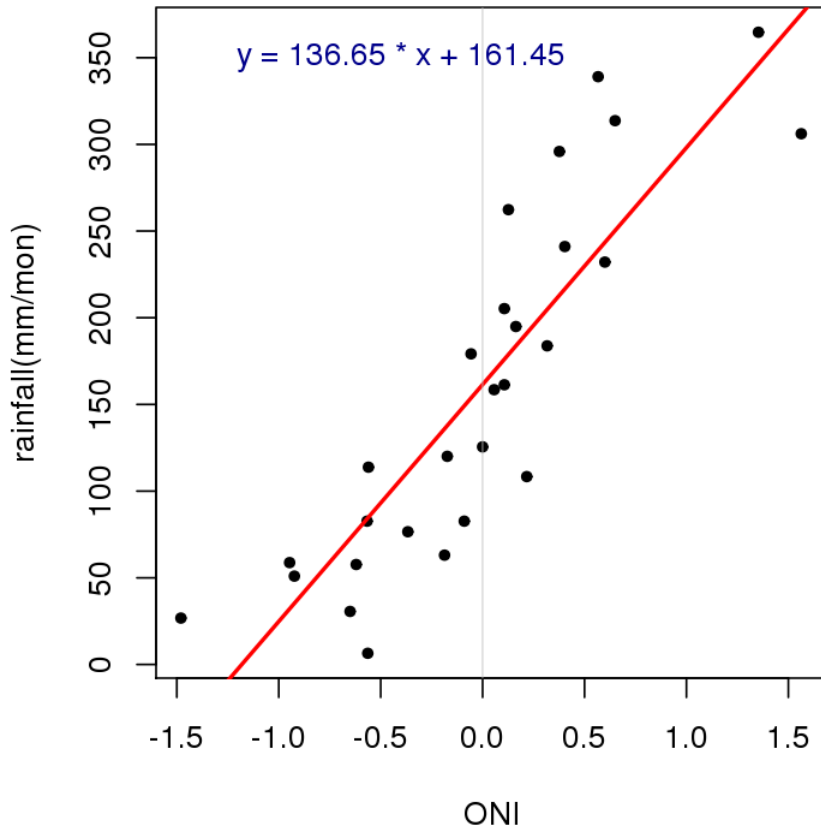
Regression concept!

Scatter plot & least square fitting!

$$Y = a * x + b$$

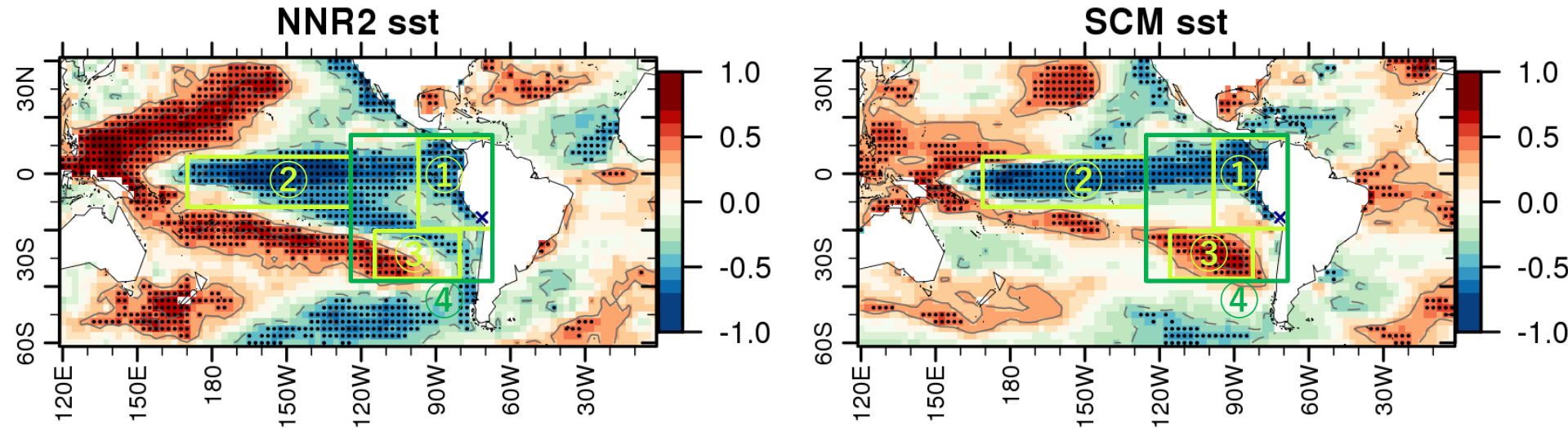
x: predictor (ONI)

y: predictand (rainfall)



Station to LSMP relationship

sibayo [FMA]



Domain candidates: ①②③: monopole
④: dipole

+ area: ESP, part of SPCZ (domain ③)

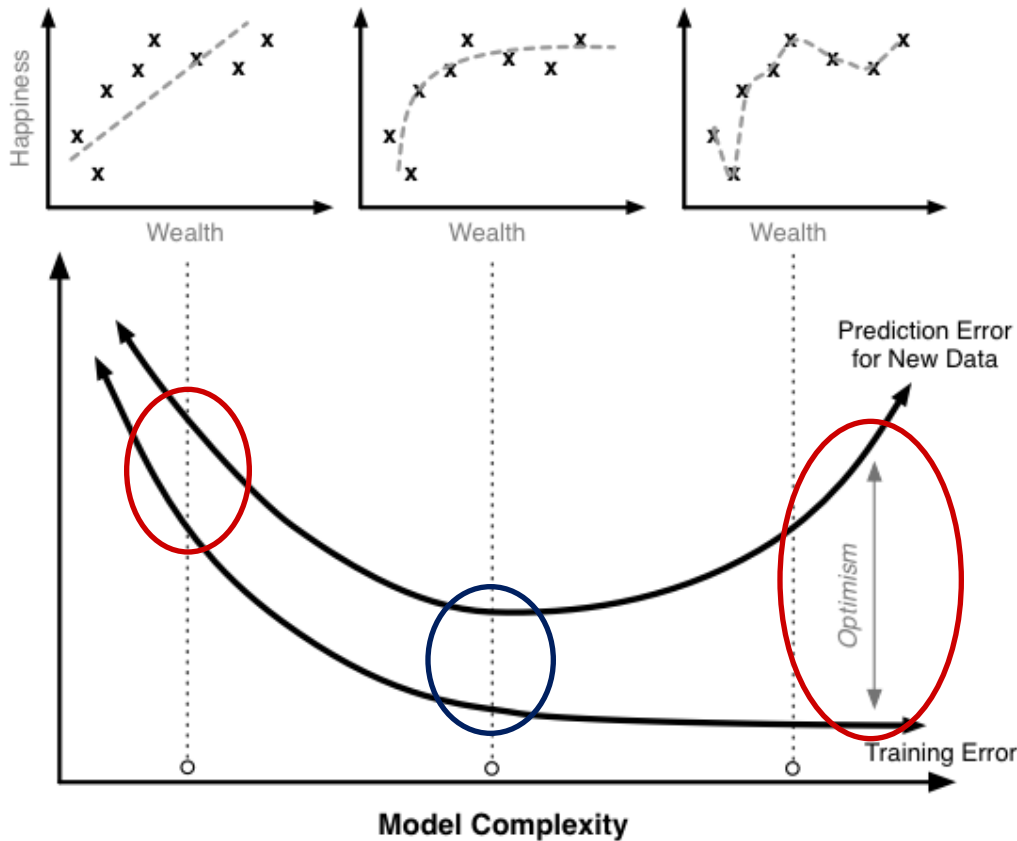
SST↑ → PREC. ↑

- area: equatorial central to eastern Pacific (domain ①②)

SST↓ → PREC. ↑

Weakness : **overfitting**

■ Consider potential predictability



If model output is fitted to the unpredictable noise : Overfitting.
What if we remove “noise” in the observation?

CLIK downscaling

➡ A way to **localize** existing coarse climate information

CLIK downscaling is mainly based on station to Large Scale Meteorological Field (LSMF) relationship. ($Y = a \cdot X + b$) By utilizing the simulated LSMF (X, predictor), CLIK estimates seasonal mean precipitation/temperature (Y, predictand) at specific station.

Dynamical fcst.

Simulated LSMF from Individual models

- Hypothesis: The station to LSMF relationship is well replicated in individual models

Statistical dwncsc.

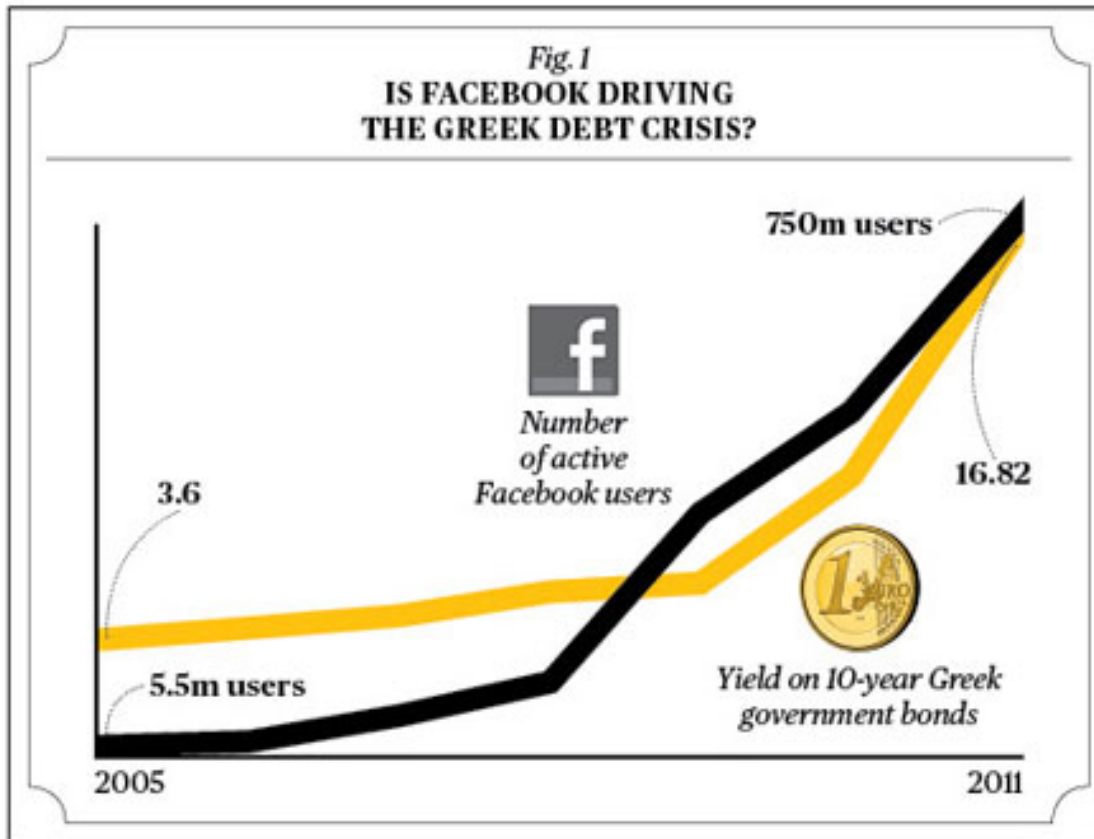
Station to LSMF regression relationship

- Predictor: LSMF
- Predictand: Prec. at specific station

A kind of hybrid system
for point-wise seasonal forecast

Predict yield of Greek bonds with number of Facebook users

Is it appropriate?



If yes, why?

If not, why?

From *business week*

Predict global average temperature with Carbon dioxide concentration

Is it appropriate?

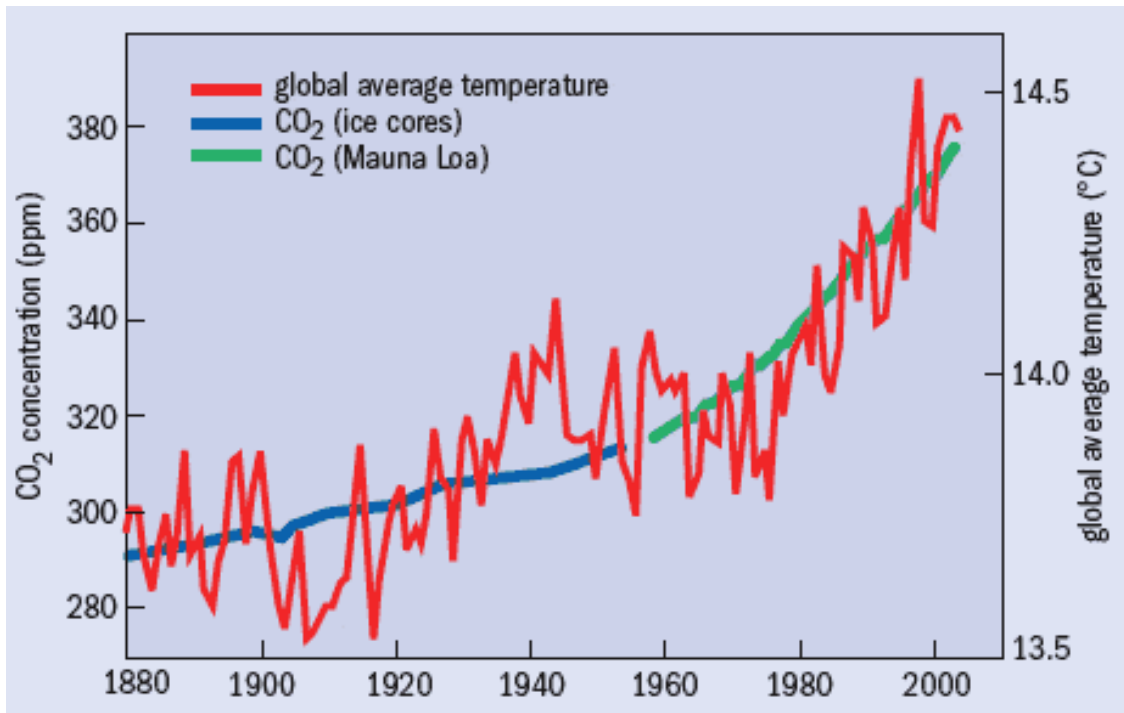


Image Source: images.iop.org

If yes, why?

If not, why?

The most important things...

1. Physical understanding of,
 - What weather event/system consists of your seasonal climate (LOCAL, predictand)
 - What external (slow varying factor) controls the weather system (GLOBAL, predictor)

→ Finding **predictors** (large scale meteorological patterns (circulations) associated with local prcp/temp of your station)
2. whether GCM (MME) is able to reproduce those patterns/relationship?

→ **Applicability** of downscaling



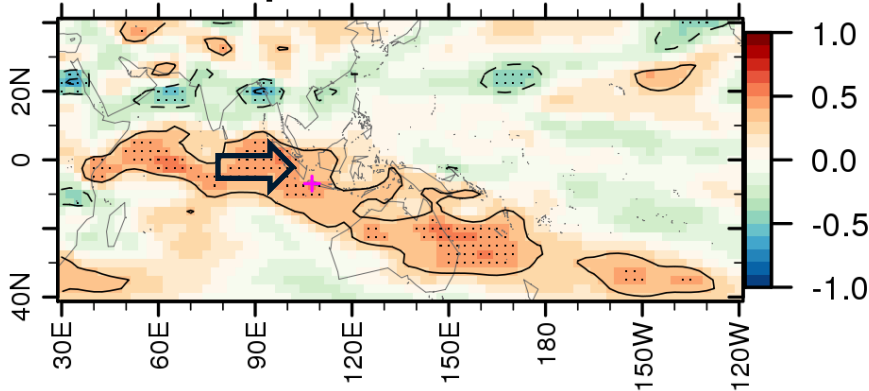
Example of Domain Selection



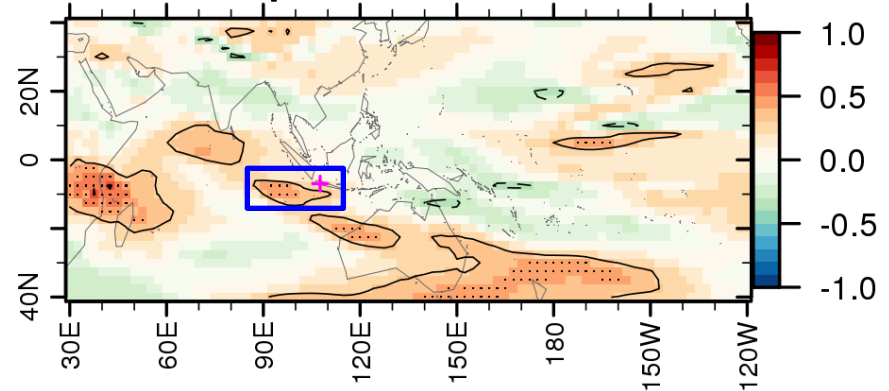
Maritime Continent: Indonesia (1)

JJA (Jakarta)

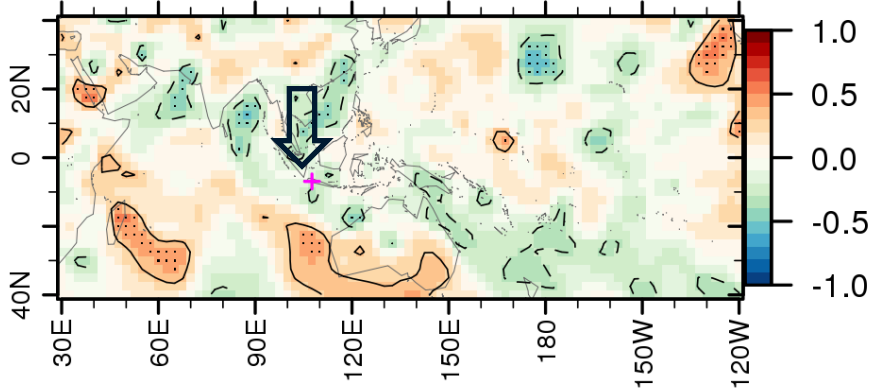
stn prec. VS obs u850



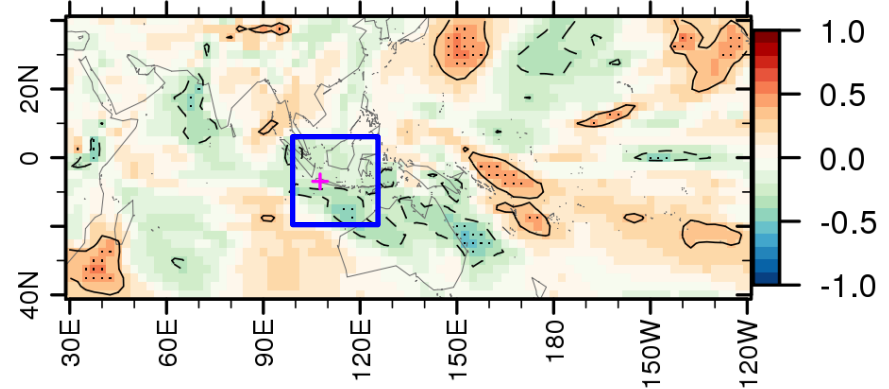
stn prec. VS scm u850



stn prec. VS obs v850



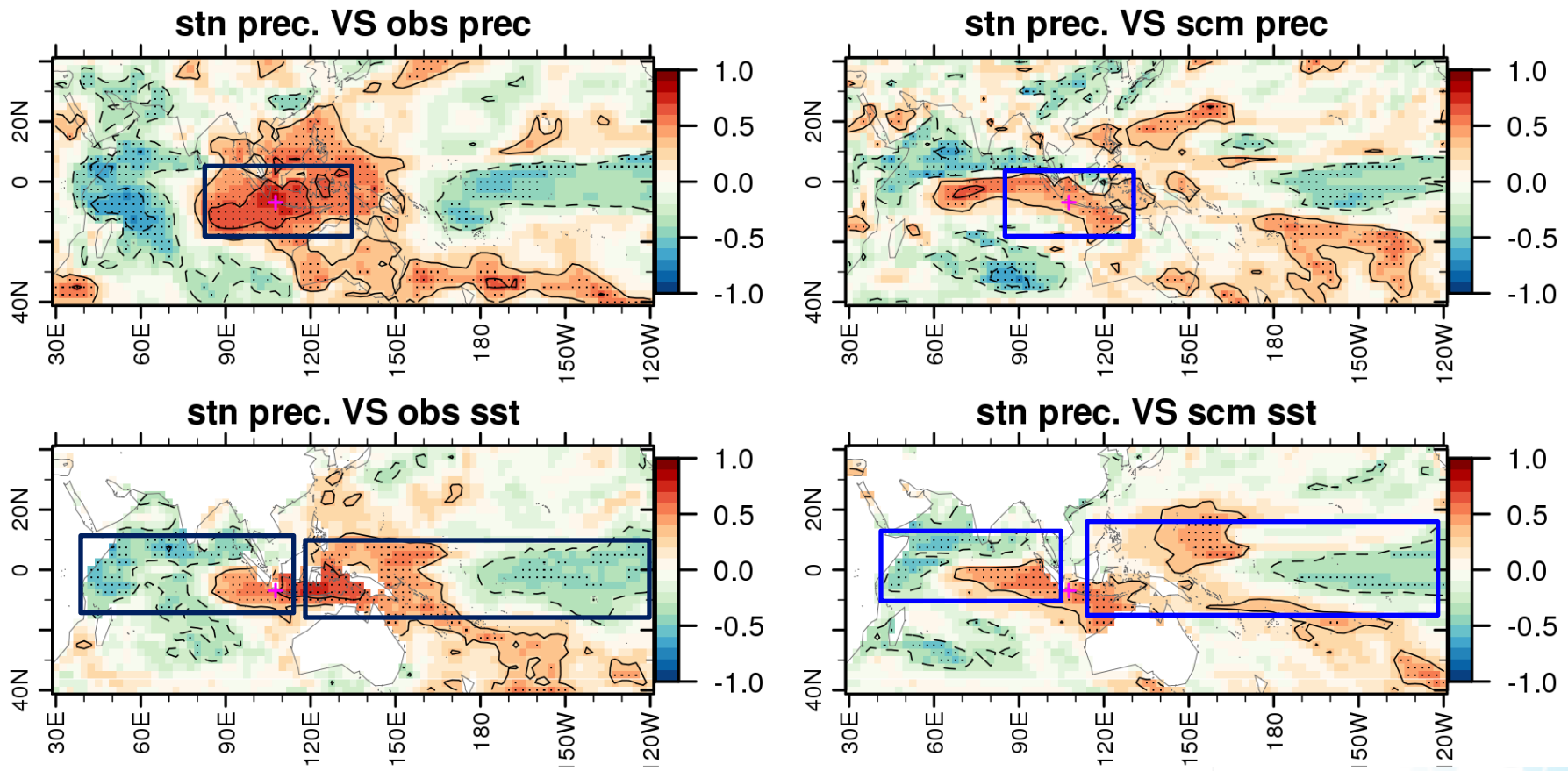
stn prec. VS scm v850



Weakening of Mean monsoon flow (southwesterly)
-> Increase SST + Rainfall

Maritime Continent: Indonesia (2)

SON [Jakarta]

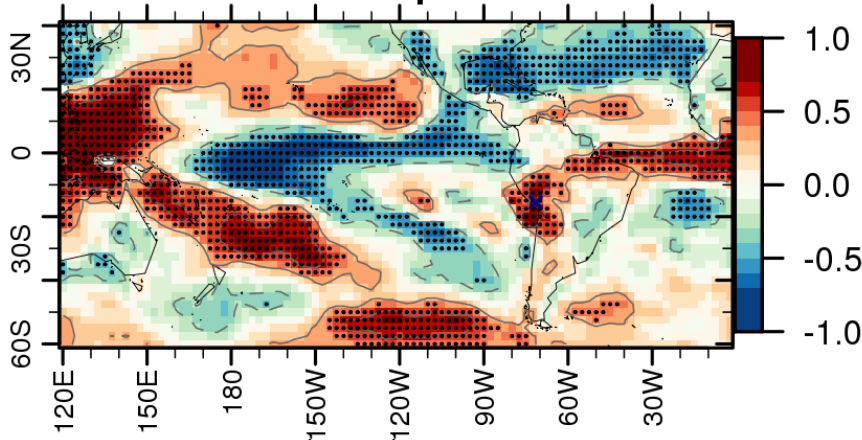


La Nina (and negative IOD) signature

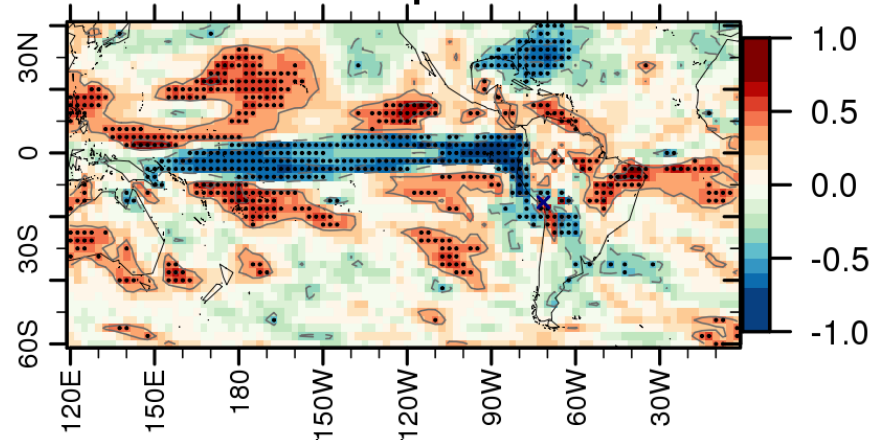
West South America: Sibayo

sibayo [FMA]

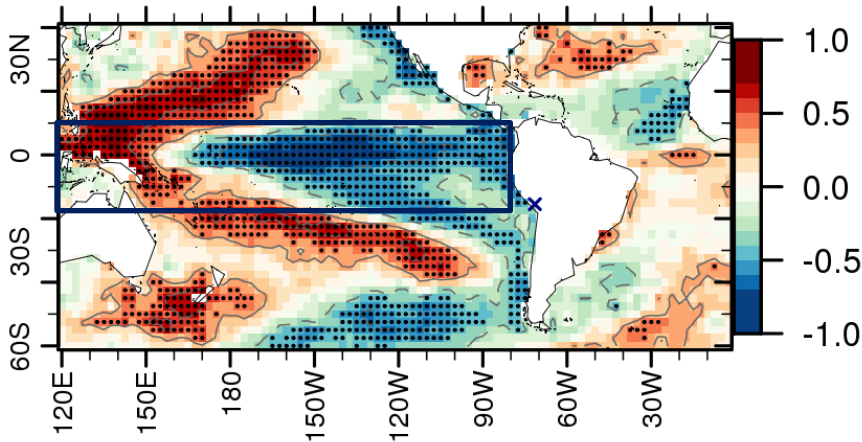
NNR2 prec



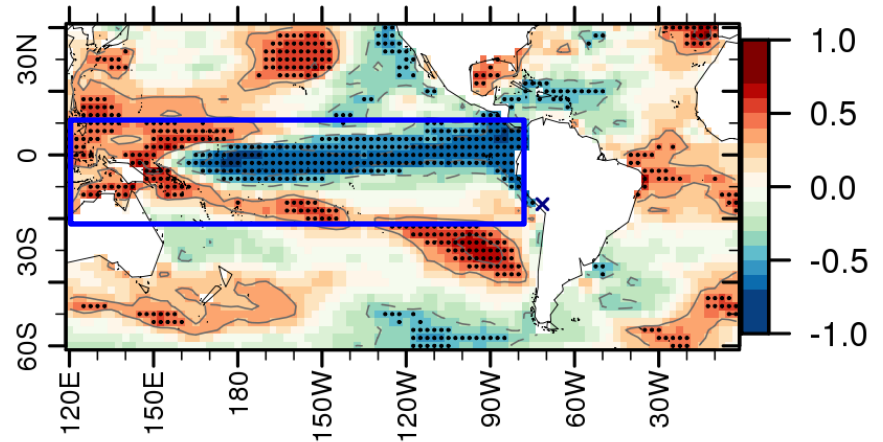
SCM prec



NNR2 sst

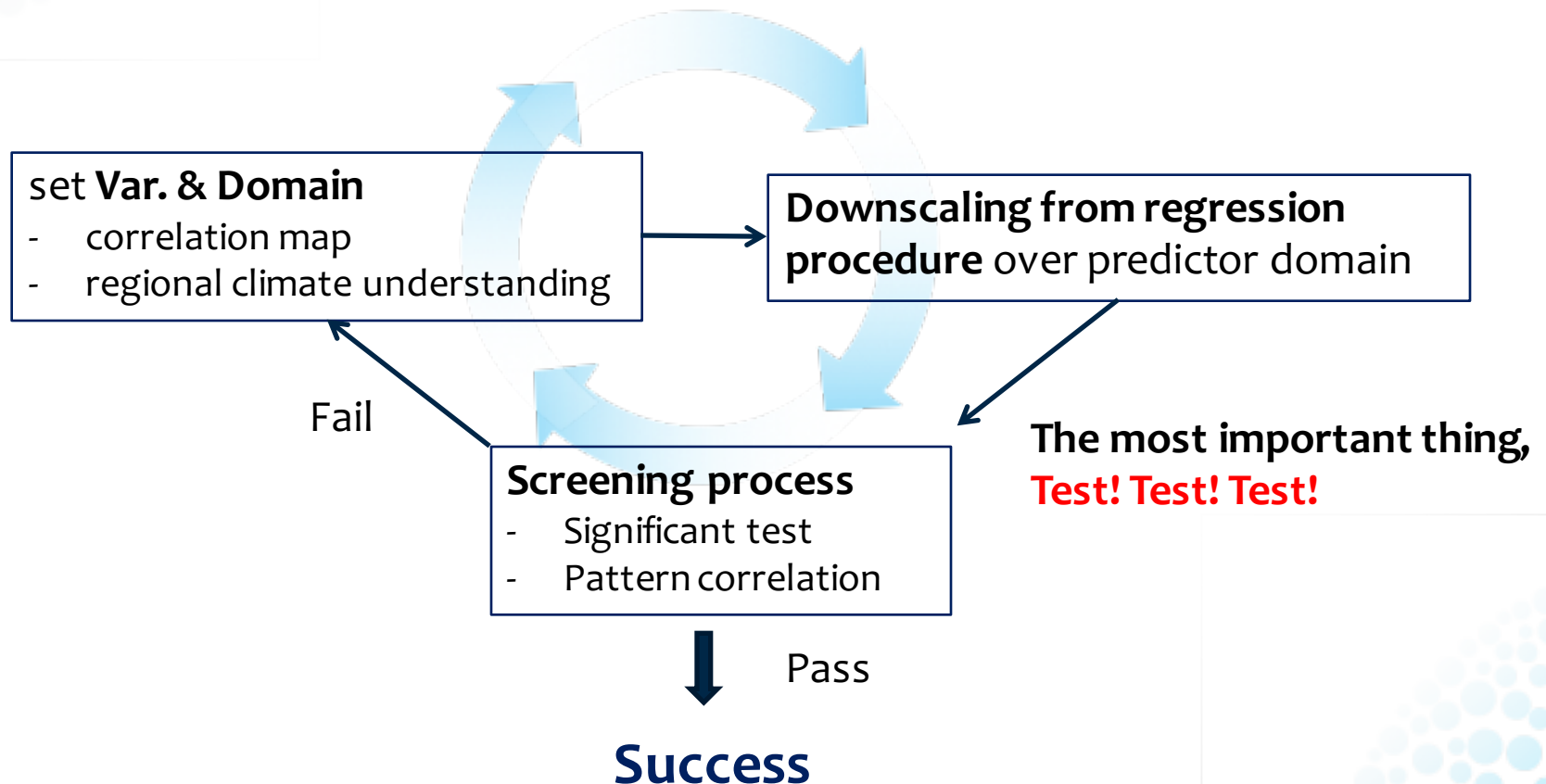


SCM sst



La Nina signature

CLIK downscaling procedure

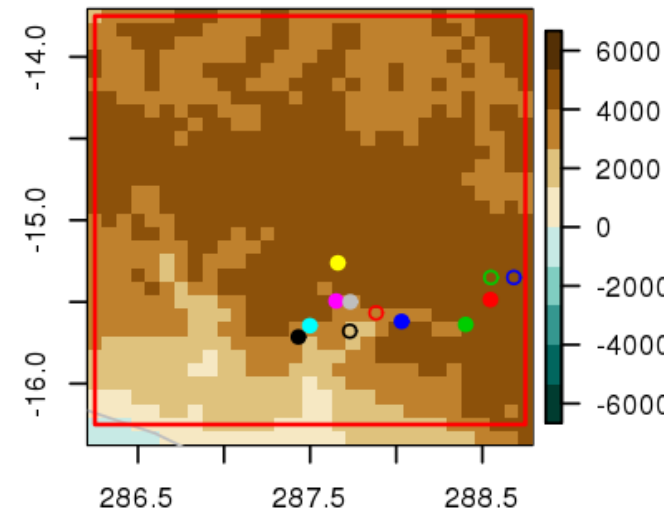




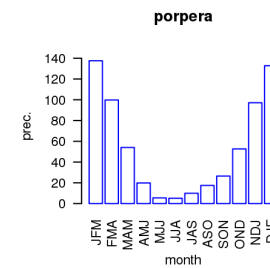
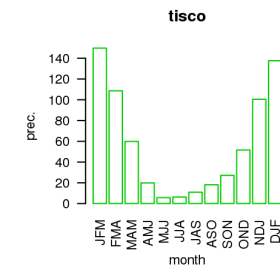
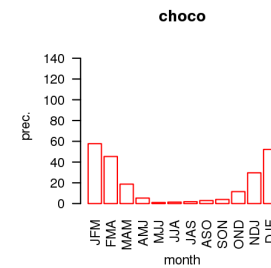
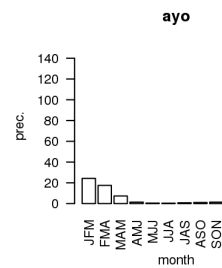
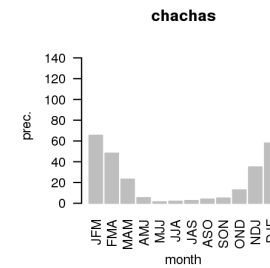
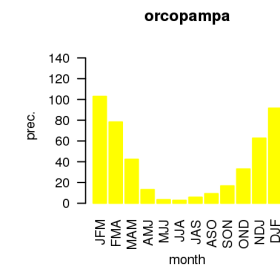
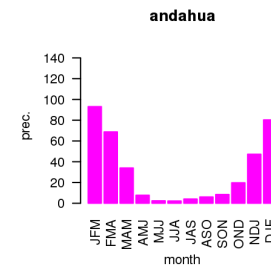
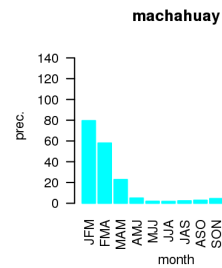
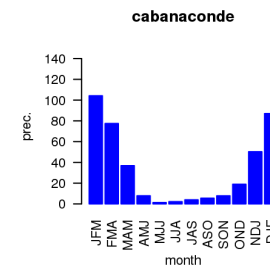
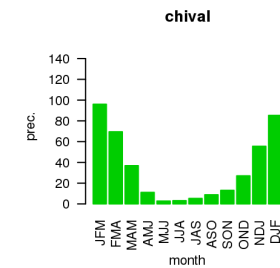
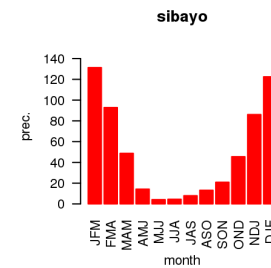
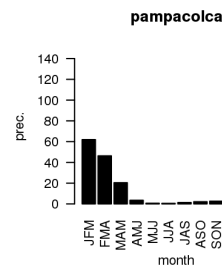
Thank you

2. Climate Locality: seasonal cycle

12 stations in a grid



- pampacolca
- sibayo
- chival
- cabanaconde
- machahuay
- andahua
- orcopampa
- chachas
- - - ayo
- - - choco
- - - tisco
- - - porpera



Our final goal

: Predicting next season climate at particular station point (your city, town, village)

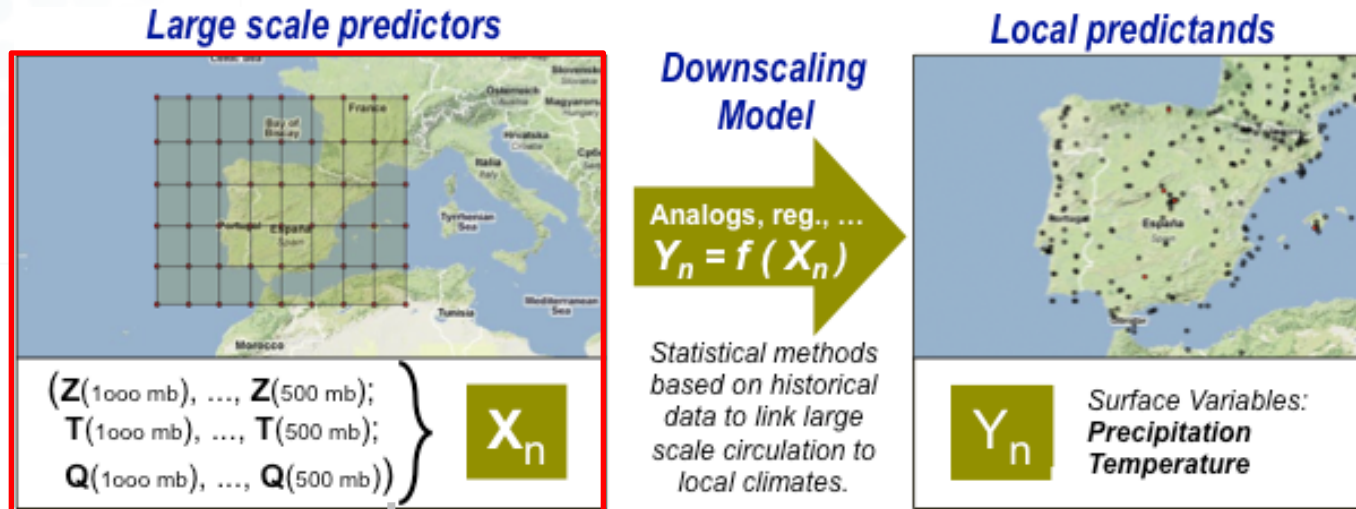
✓ *What we have*

: Prcp/temp station data (past) & coarse GCM data (past & future)

✓ *What we know*

: past relationship (station point –large scale field)

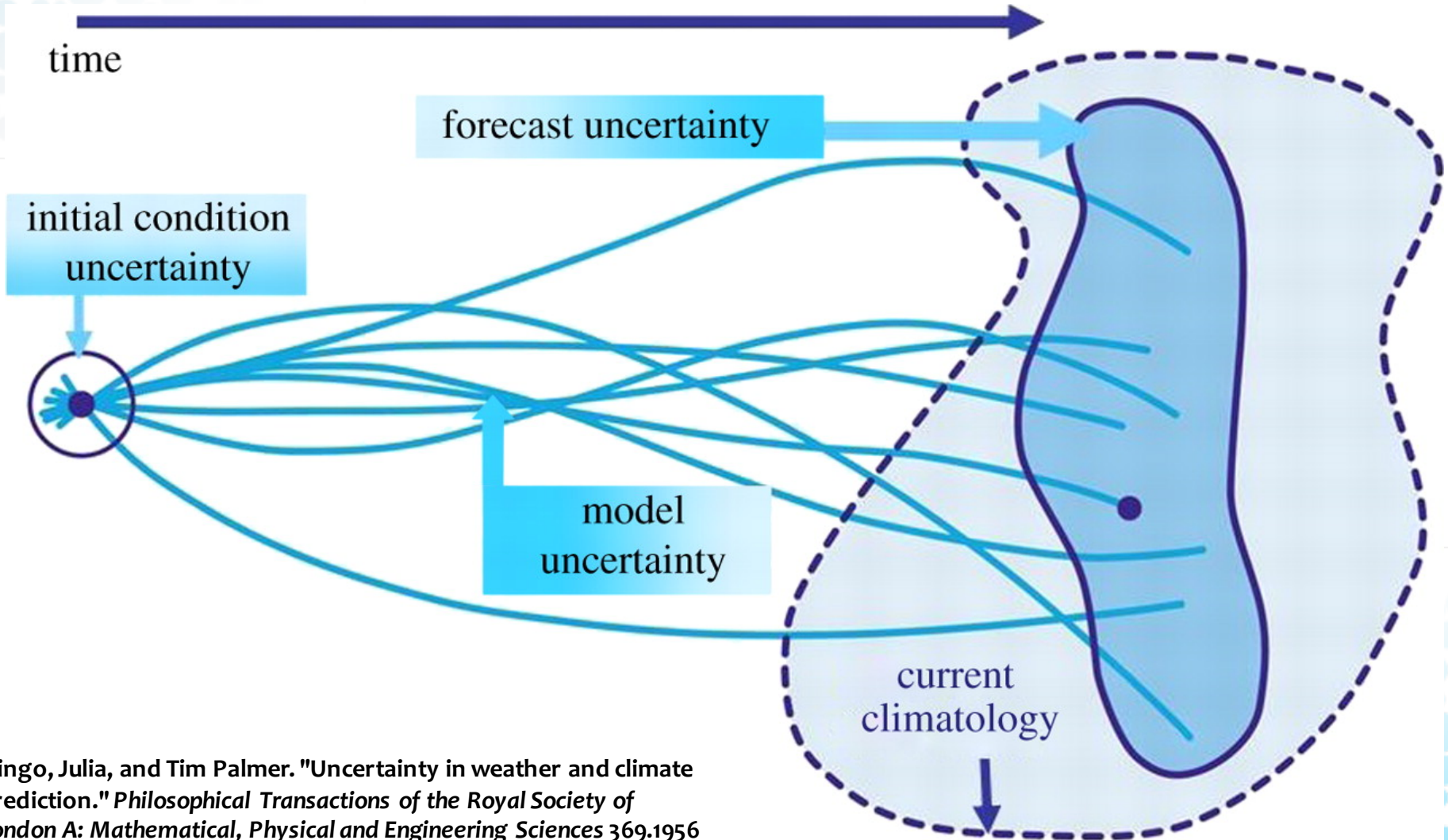
Accuracy of Large Scale Information?



Simulated Large scale information is correct? **NOPE!!!**
Under this circumstance,

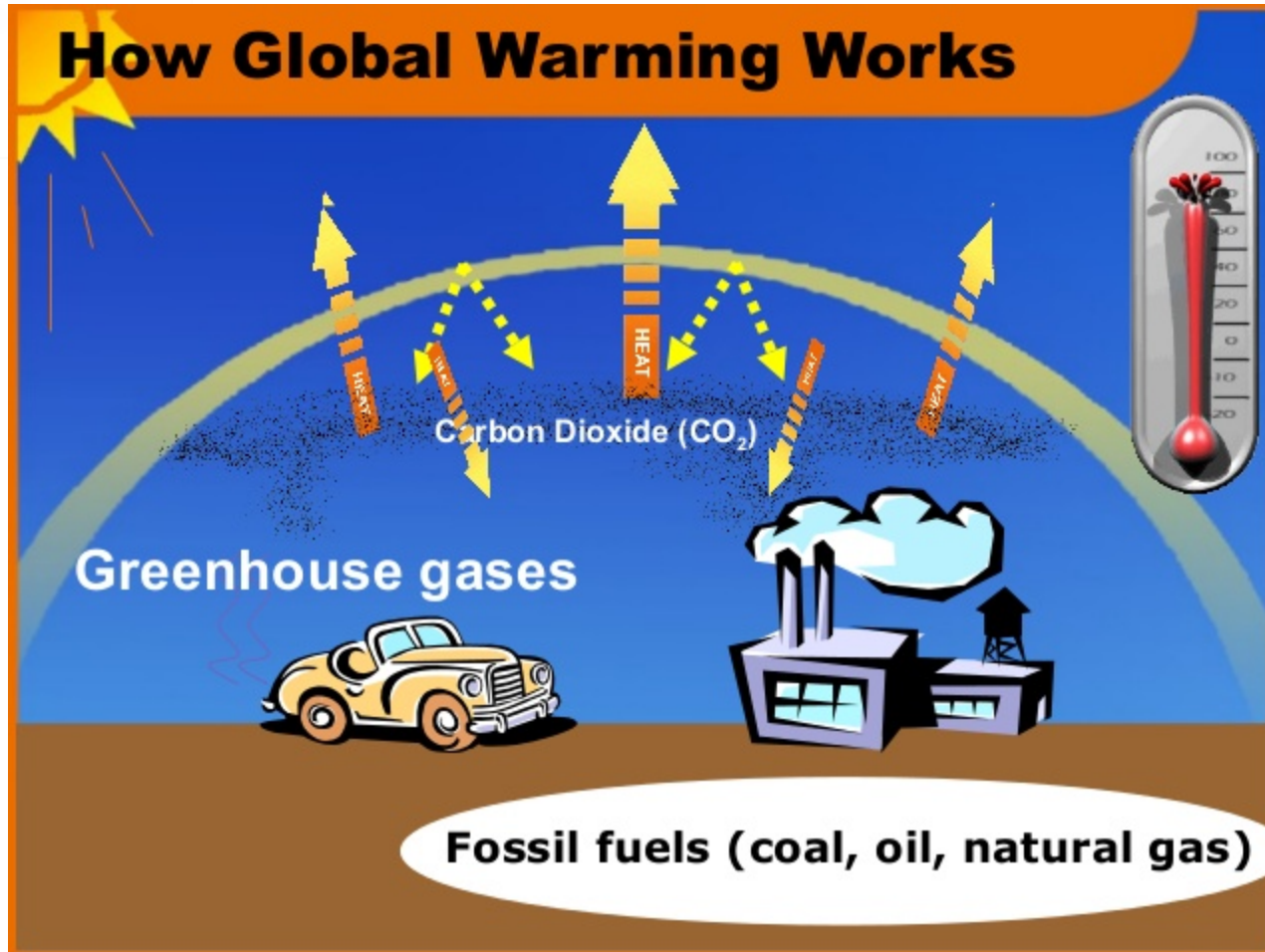
Statistical downscaling can be understood as *a statistical "forecast" based on "forecasted" large scale information.*

Uncertainty in climate prediction



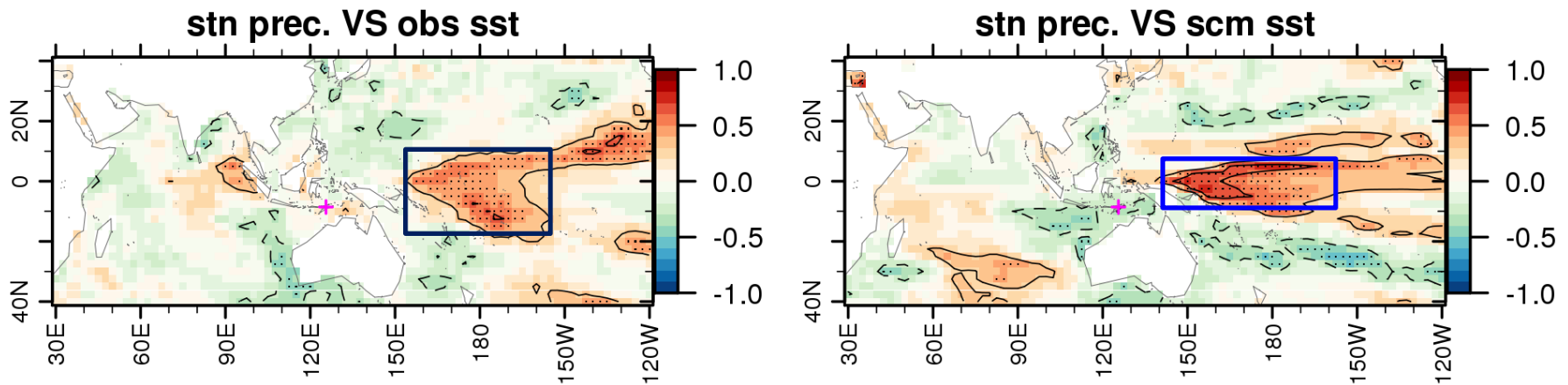
Slingo, Julia, and Tim Palmer. "Uncertainty in weather and climate prediction." *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 369.1956 (2011): 4751-4767.

Increase Greenhouse effect due to CO₂ increase



Maritime Continent: East Timor

DJF [Dili]

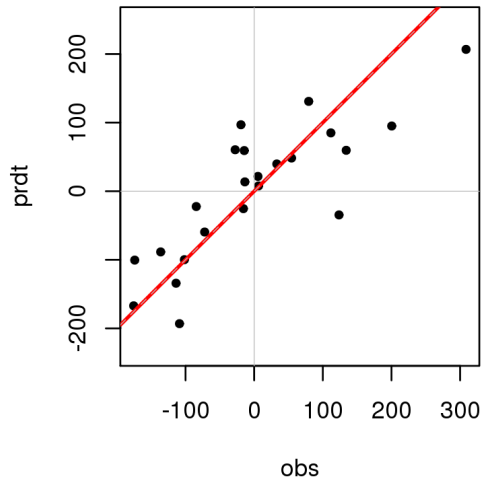


**Difficult but Some interesting signal at equatorial
Central Pacific (El Nino Modoki?)**

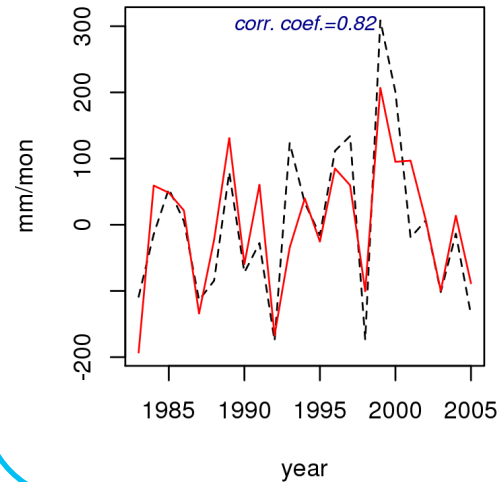
Central Pacific: Fiji

Fiji, 91680, Nadi Airport [JFM]

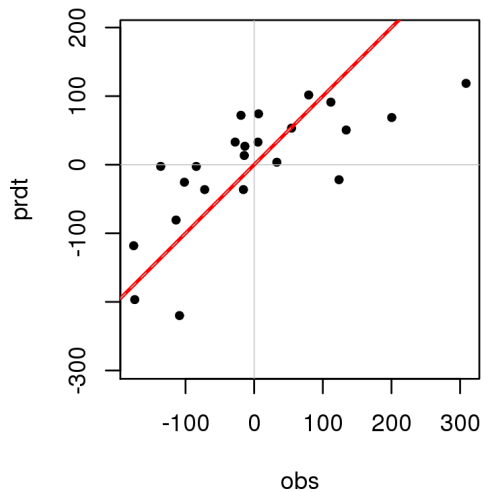
Prec. Anom. down. from v850



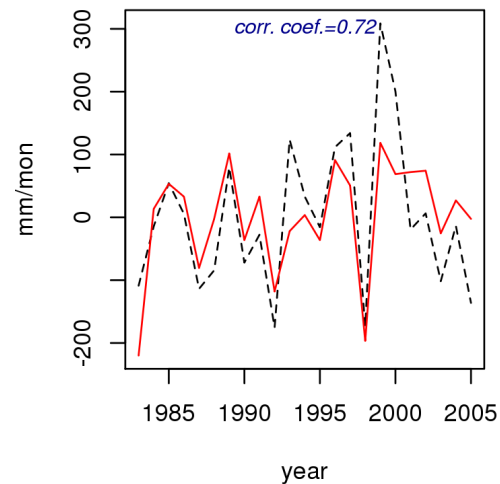
Prec. Anom. down. from v850



Prec. Anom. down. from slp



Prec. Anom. down. from slp



Should remind!

1. Based on *past* relationship. It doesn't guarantee that it works in present/future.
2. Can be *subjective* (dependent on *human*)
 - Which variable is the best?
 - Which domain is the base for predictor?

***Stationary* statistics & *physical* understanding** should be based.