Introduction to the Climate Predictability Tool

Simon J. Mason

simon@iri.columbia.edu

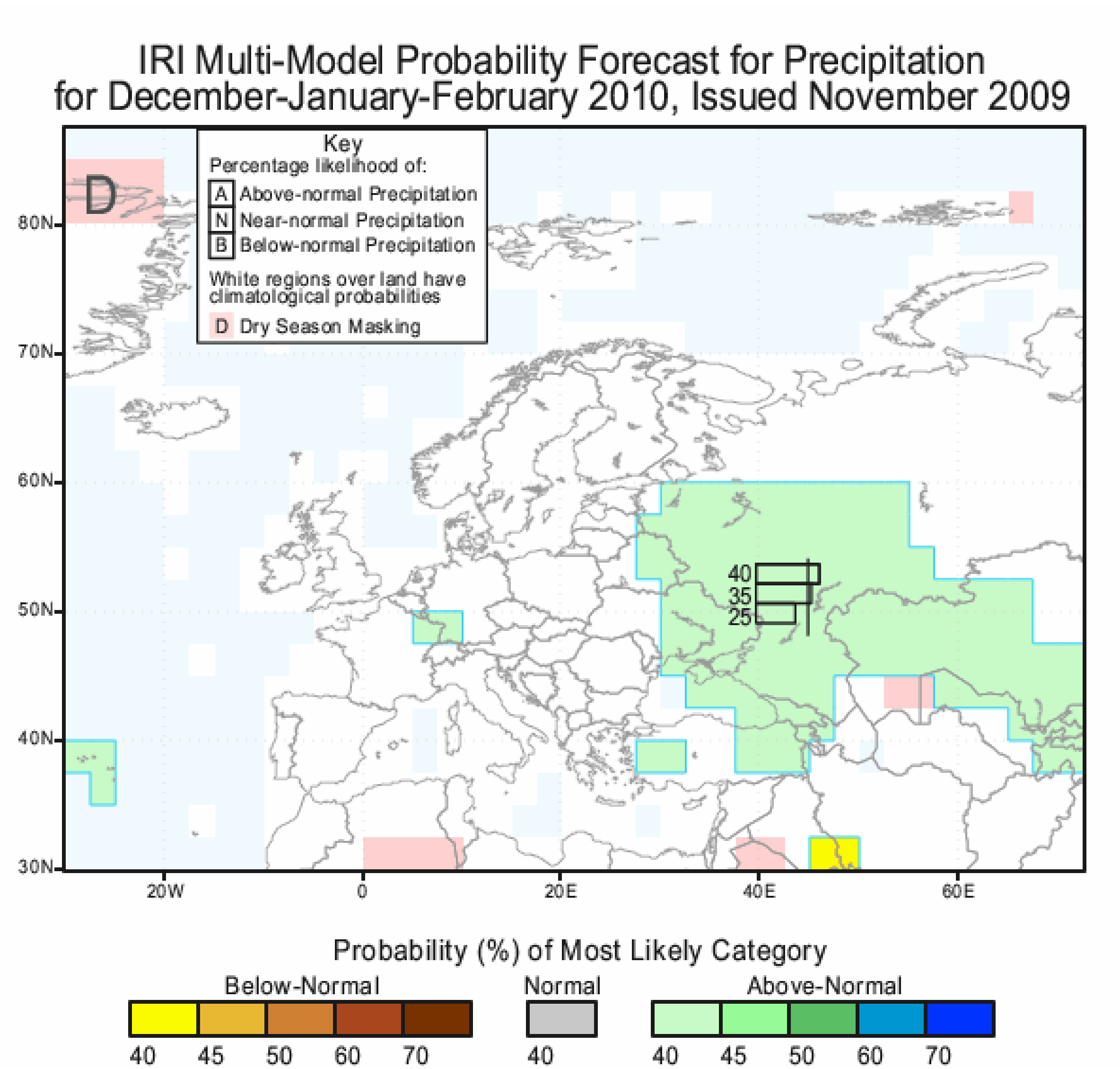
International Research Institute for Climate and Society

The Earth Institute of Columbia University

The Fourth Session of Southeastern Europe Climate Outlook Forum (SEECOF-4)

Belgrade, Serbia, 22 – 26 November 2010

Seasonal Forecast Formats



Why the Limited Use of Forecasts?

Gaps in Data

- Insufficient observations, and poor quality
- Downscaling or “right-scaling”

Gaps in Climate Services

- Poor skill, and insufficient lead-time
- Poor communication

Gaps in Practice

- Relevance of forecast
- Unclear what action to take

Gaps in Policy

- Institutional constraints to taking action

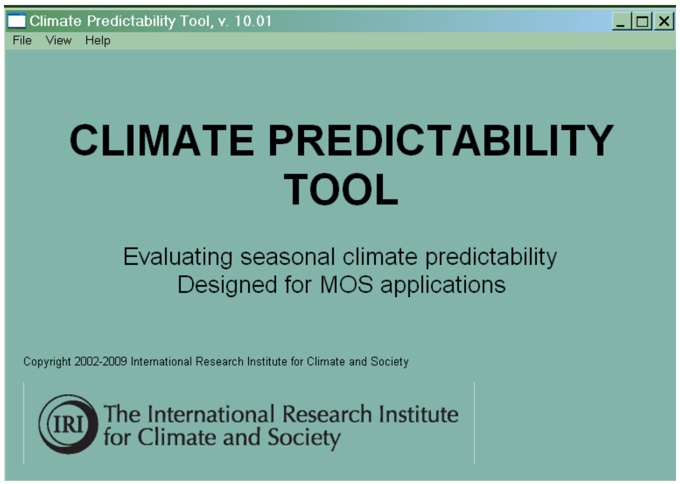
CPT - Goal

Widespread creation and communication of quality-controlled seasonal climate forecasts that address specific needs of different user groups.

To produce forecasts that people can and want to use!

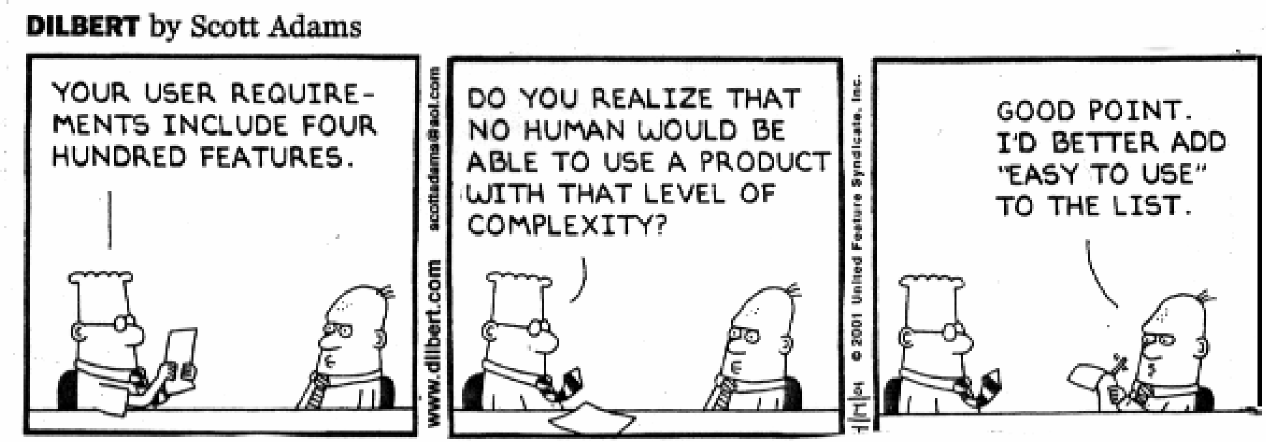
What is CPT?

Climate Predictability Tool (CPT) is an easy-to-use software package for making tailored seasonal climate forecasts.



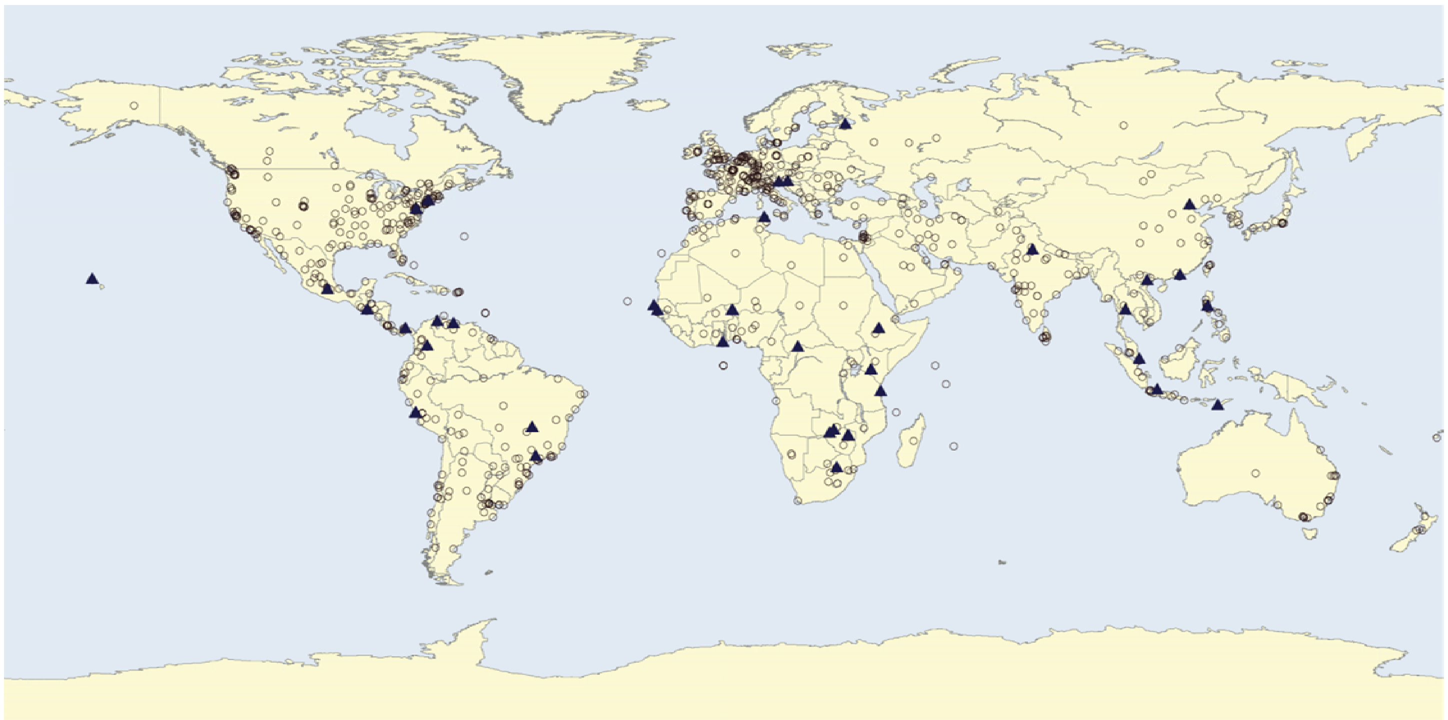
Versions:

- Windows 95+ 
- Batch
- Generic GUI version (under development)



CPT Use

CPT downloads (circles) and known CPT courses (triangles) from 2003 to 2009



History of CPT

CPT 0 : MATLAB code for performing canonical correlation analysis (CCA).

CPT1 : (Dec 2002) translated into 650 Lines of Fortran 77 interfacing to LAPACK SVD routines

CPT2 : (Aug 2003) converted to Fortran 95; GUI; validation; new forecasts

CPT3 : (Feb 2004) mapping of station data; handling of missing values

CPT4 : (Feb 2005) improved graphics

CPT5 : (Aug 2005) forecast uncertainty; WMO SVSLRF verification; tailoring

CPT 6 : (Nov 2005) multiple users; exceedance probabilities

CPT 7 : (Aug 2006) data transformation; improved retroactive procedure

CPT 8 : (May 2007) retroactive forecast probabilities and verification; multiple regression

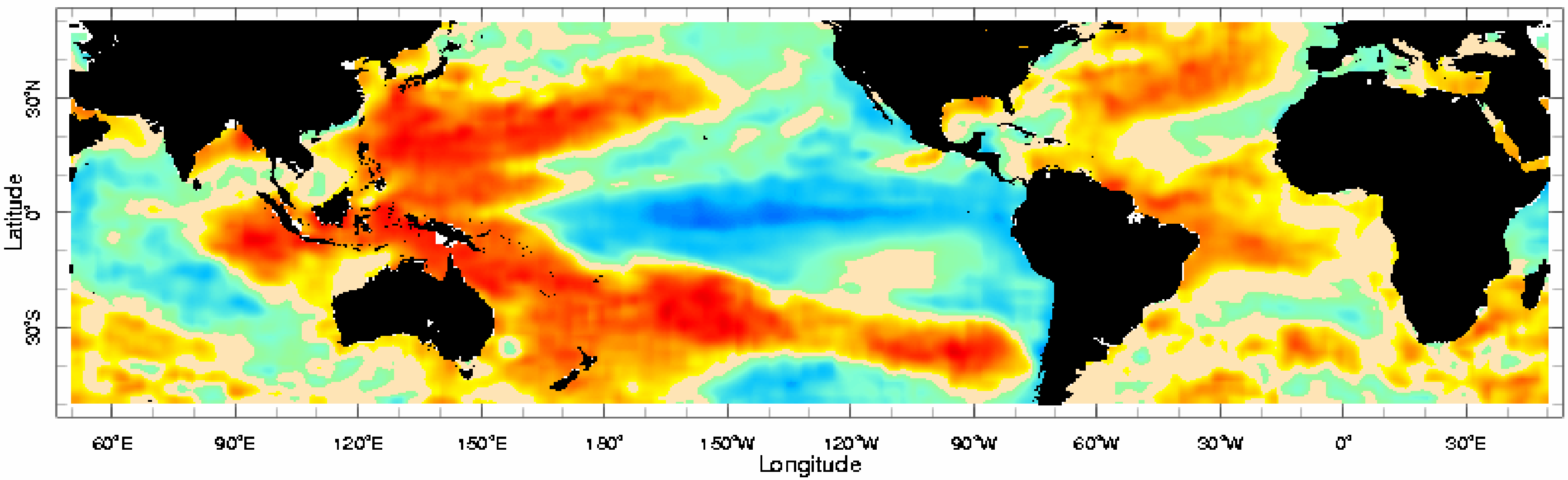
CPT 9 : (Mar 2008) DLLs; major internal restructuring

CPT10 : (Oct 2009) multiple fields; extended EOFs; new interface; ensemble forecasts; new verification procedures; new input data formats

CPT11 : (Nov 2010) multilingual interface; simplified input data formats; probabilistic verification scores

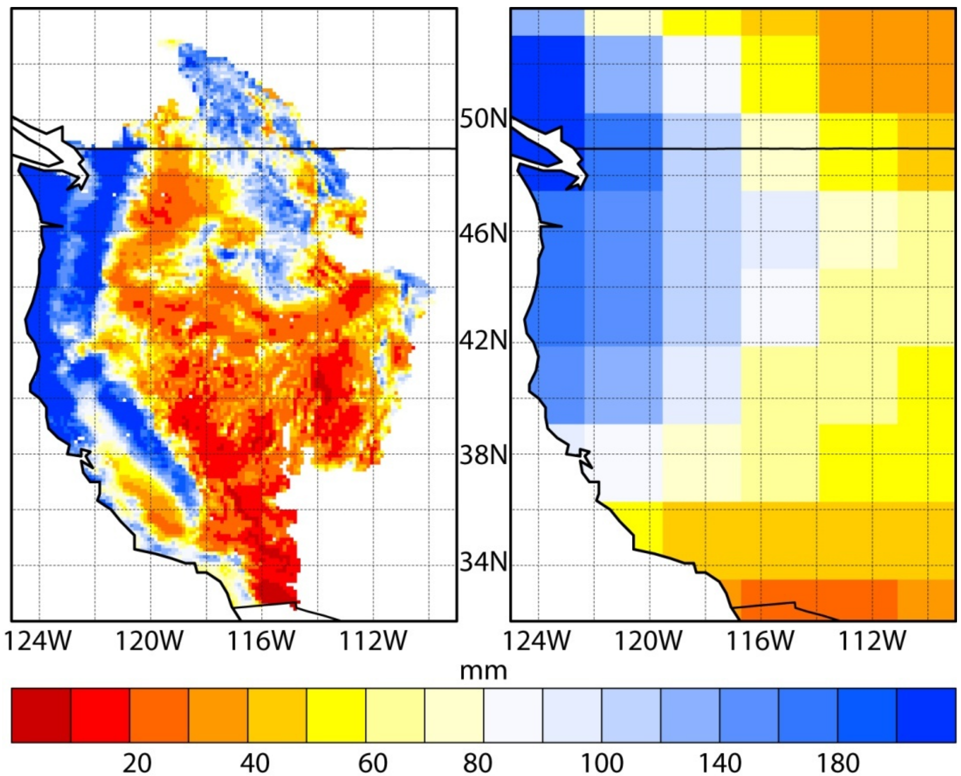
CPT – Statistical Forecasting Tool

Correlation between Philippines rainfall and global sea-surface temperatures



Downscaling

- GCM gridpoint forecasts are area-averages, and may not produce information at a scale that is useful for practical applications.
- A typical GCM grid represents about a 60,000 km^2 area.

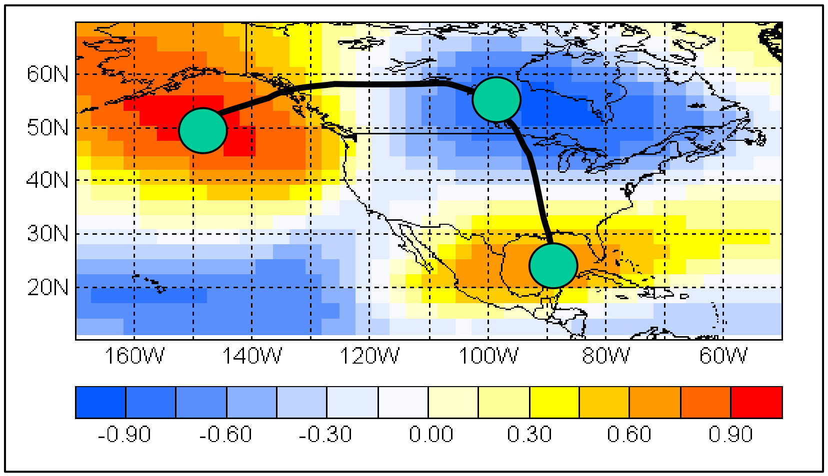


Annual mean observed and simulated precipitation over western North America.

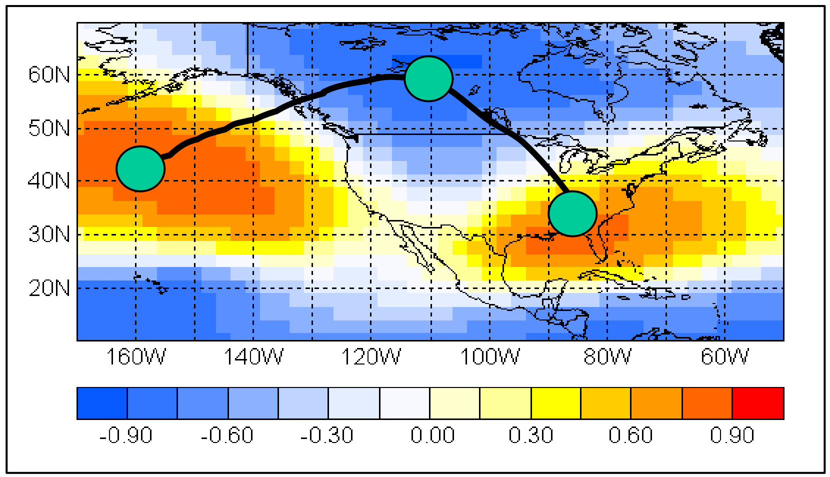
CPT – Downscaling Tool

Important climate features may be displaced in GCMs relative to observations

ECHAM 4.5 “PNA” Pattern



NCEP Reanalysis PNA



Why CPT?

CPT was developed to address some problems in producing seasonal climate forecasts at a number of the RCOFs:

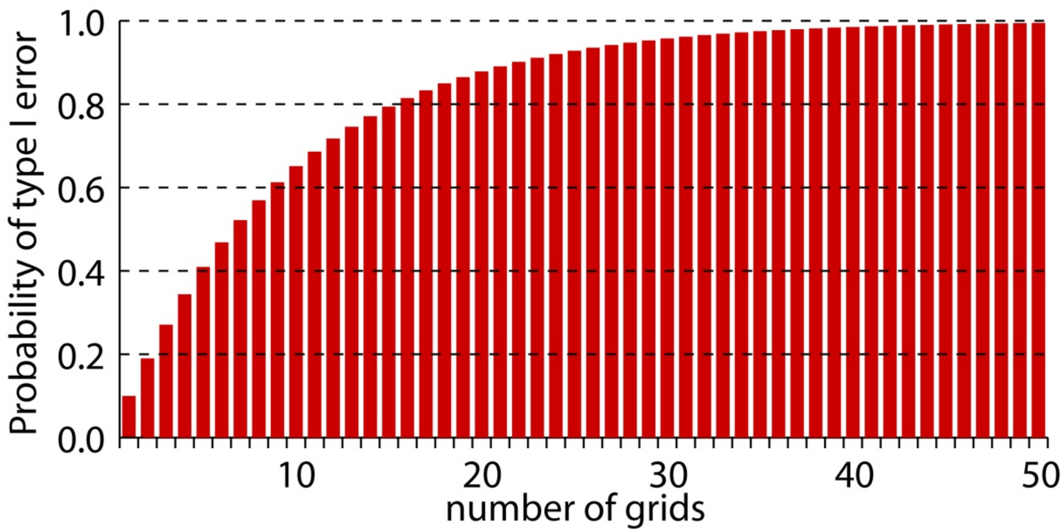
Slow production time - expensive pre-forum workshops expensive, and limited availability of monthly updates;

Artificial skill, and lack of vigorous performance evaluation;

Minimal consideration of global products.

Why Not Regression?

- **Multiplicity** – too many predictors



Why Not Regression?

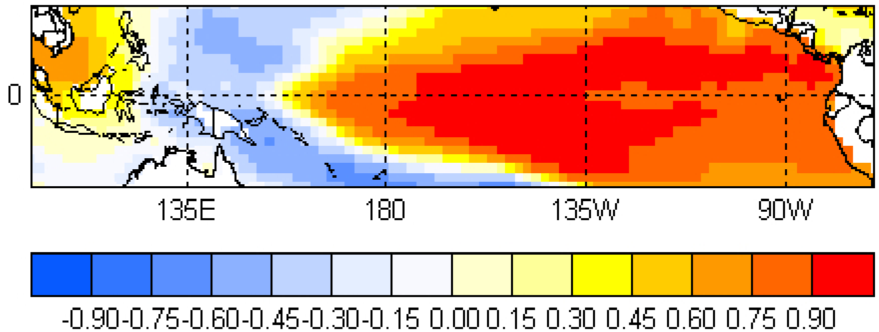
- **Multicoliearity** – similar predictors

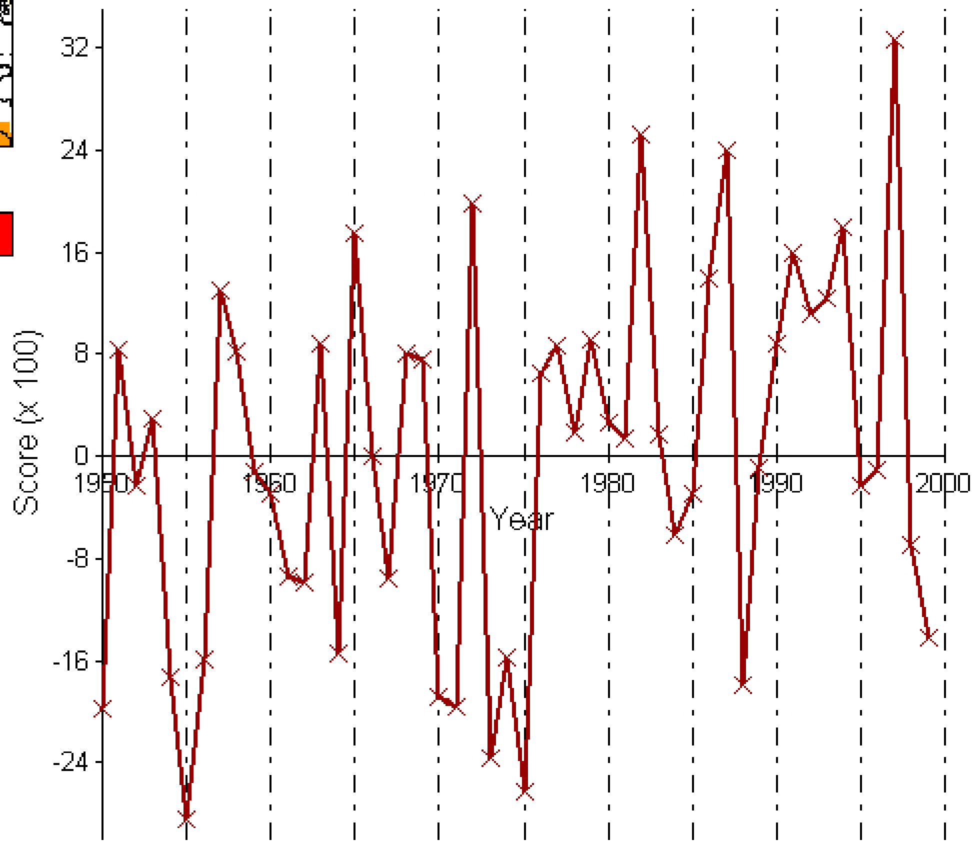
$$\text{Nino3.4}_{\text{Mar}} = \beta_0 + 0.761 \times \text{Nino3.4}_{\text{Feb}} + \varepsilon$$

$$\text{Nino3.4}_{\text{Mar}} = \beta_0 + 0.628 \times \text{Nino3.4}_{\text{Jan}} + \varepsilon$$

$$\text{Nino3.4}_{\text{Mar}} = \beta_0 + 1.216 \times \text{Nino3.4}_{\text{Feb}} - 0.395 \times \text{Nino3.4}_{\text{Jan}} + \varepsilon$$

What is PCA?





First principal component of October – December 1950 -1999 sea-surface temperatures.

Why PCA?

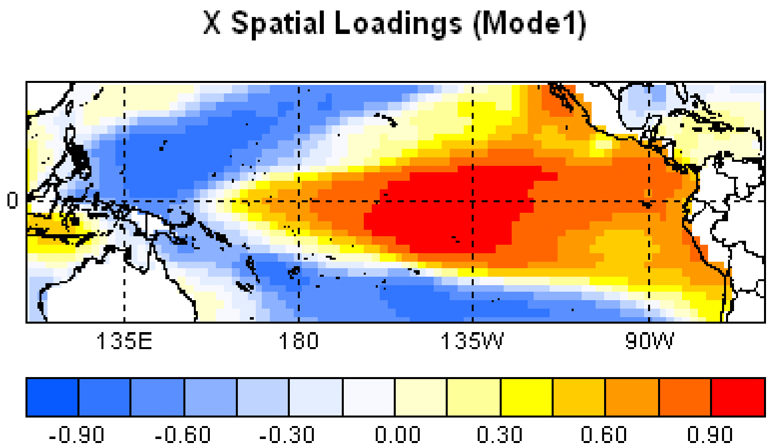
When using principal components of sea-surface temperatures the components have desirable features:

1. They explain maximum amounts of variance, and therefore are representative of sea temperature variability over large areas;

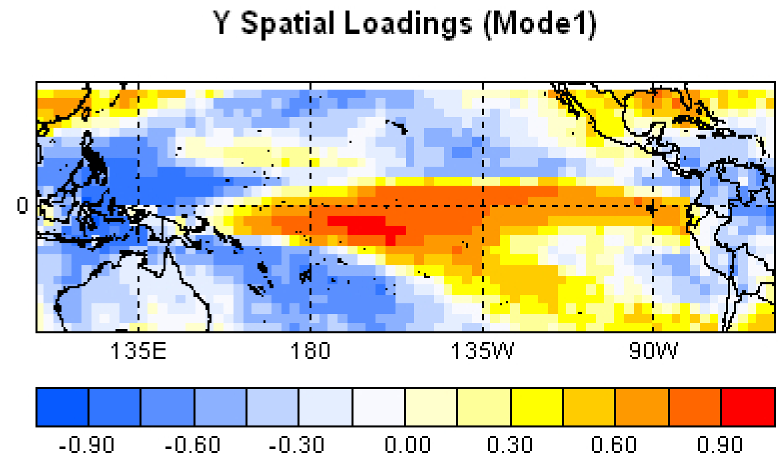
2. They are uncorrelated, and so errors in estimating the regression parameters are minimized.

3. Only a few need be retained and so the dangers of fishing are minimized.

CCA

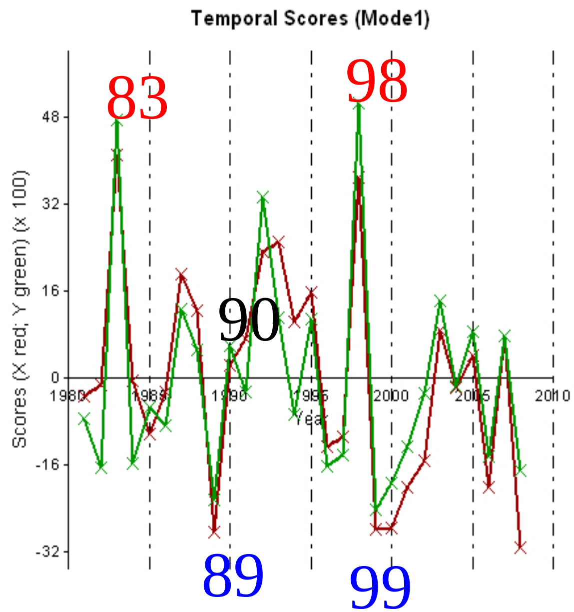


SST during DJF



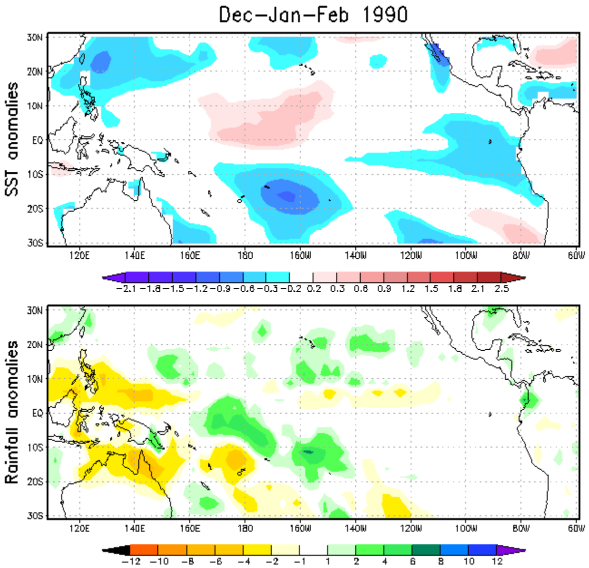
Rainfall during DJF

<u>Note :</u>
(1983;1998) and (1989;1999) were cases of strong associations but opposite.
1990 was year of weak or no association.

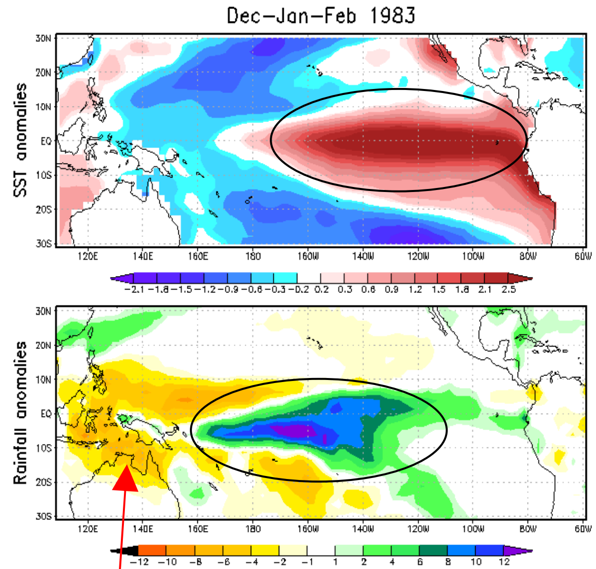


Degree of association between those two patterns from 1981 to 2008.

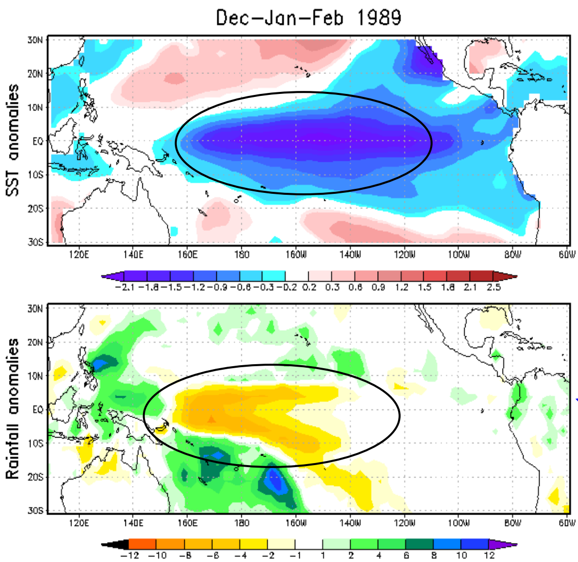
CCA



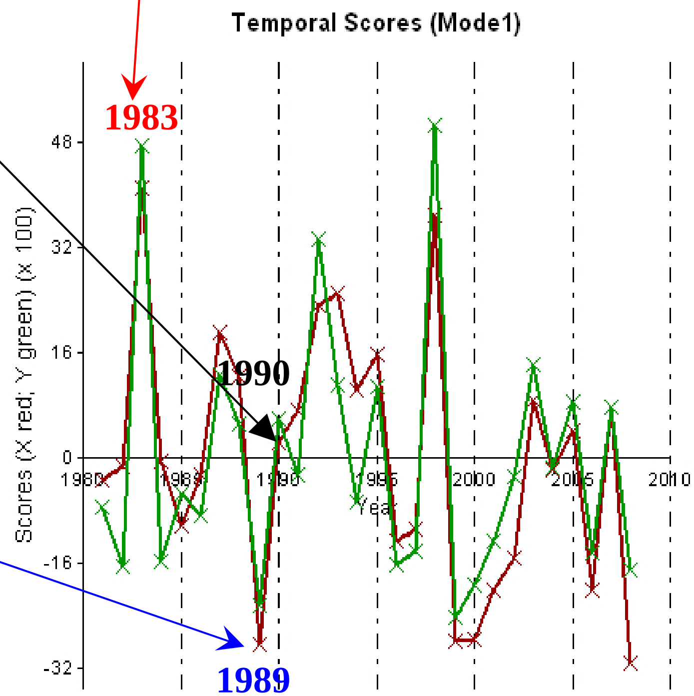
Neutral:
Call it normal



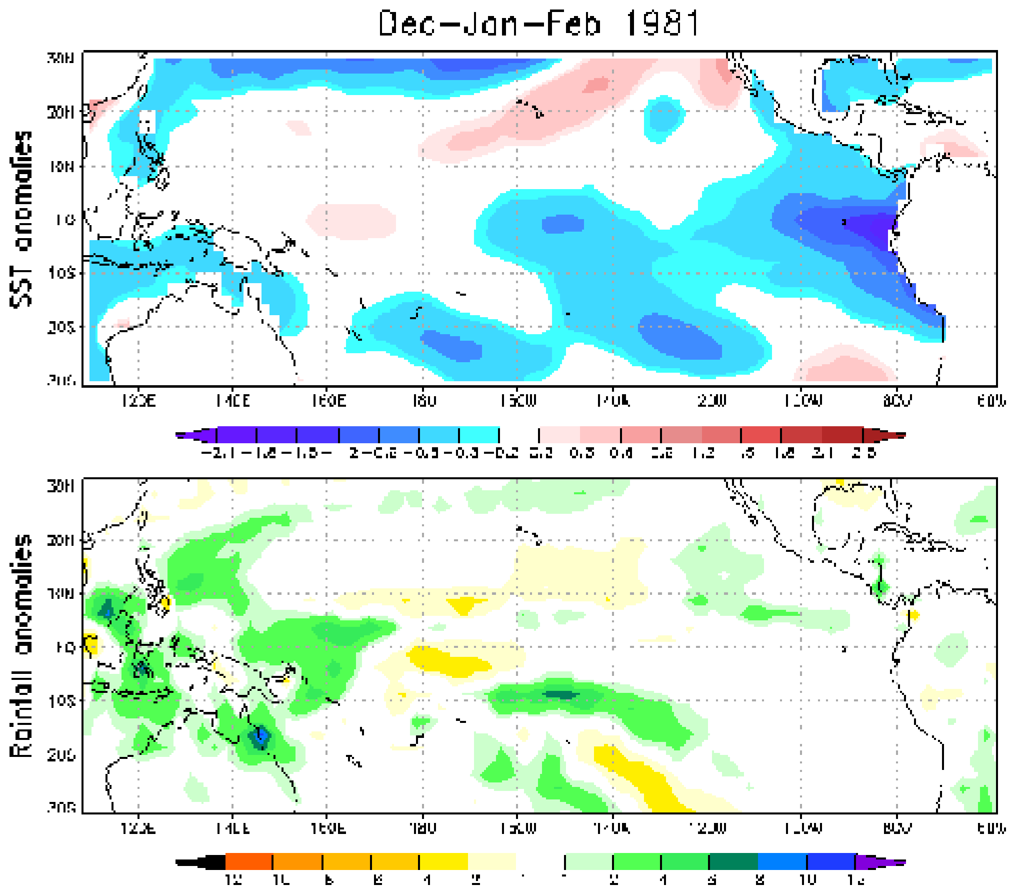
El Nino
teleconnection

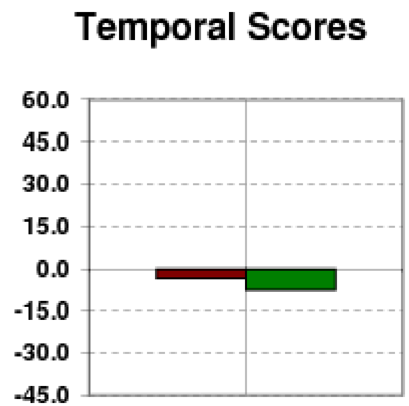


La Nina teleconnection

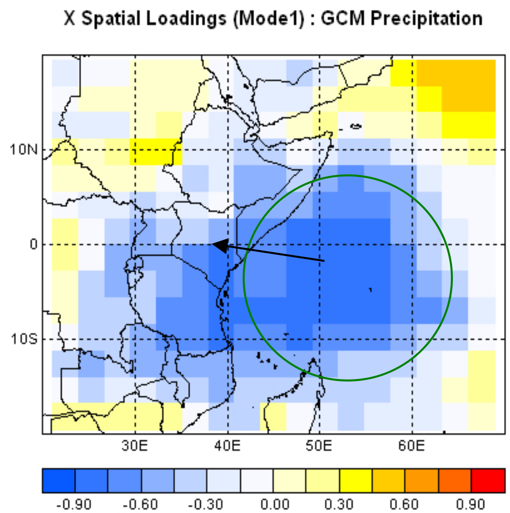


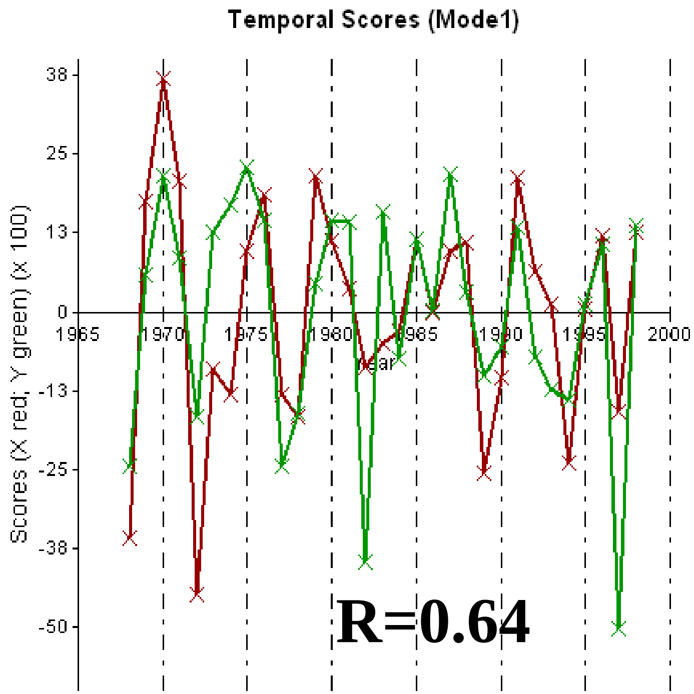
how the two patterns are actually associated

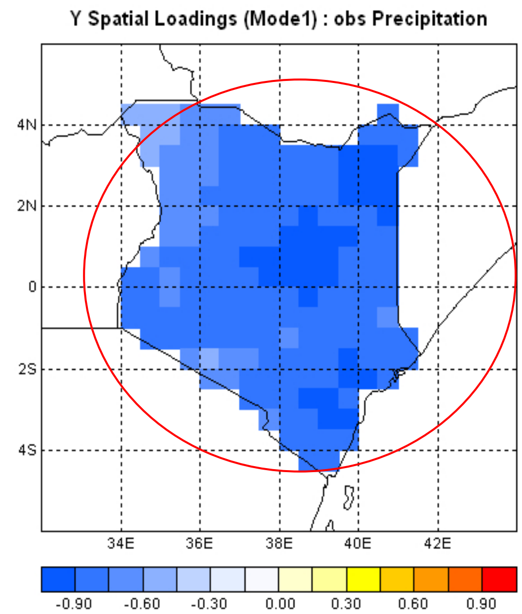




CCA : bias correction



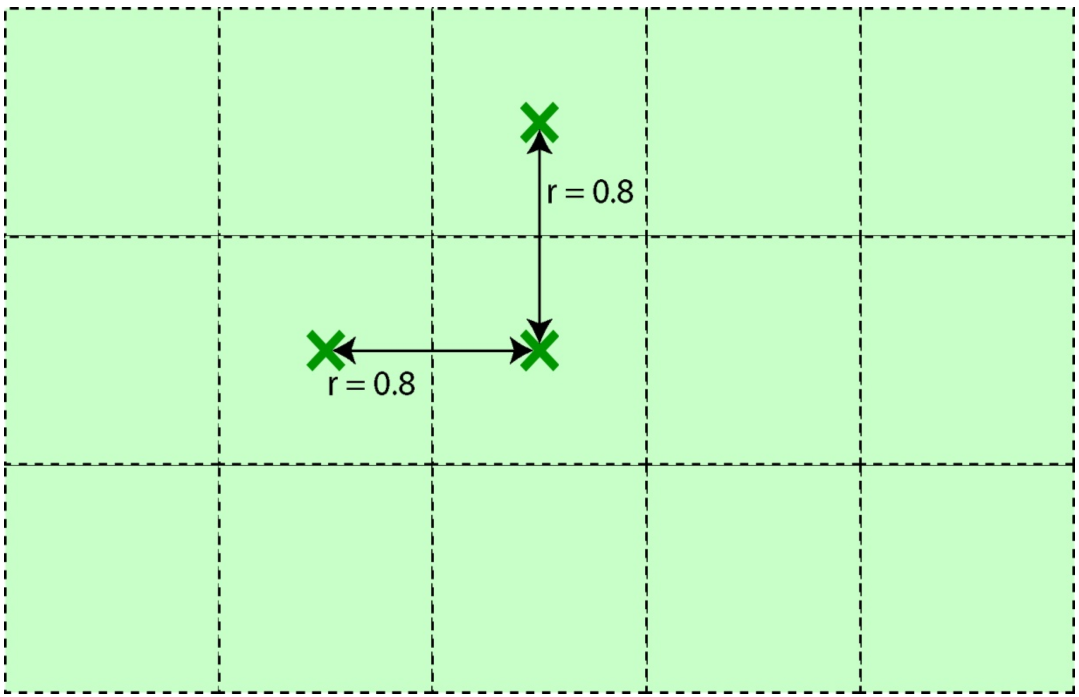




The CCA shows a strong association between GCM rainfall (X) and observed rainfall (Y) with r=0.64 showing the GCM capability to capture inter-annual rainfall variability, *but* it misses the location of the rainfall in Kenya (spatial bias).
If we looked at GCM rainfall in Kenya (targeted region) we would miss the signal.

Buell patterns

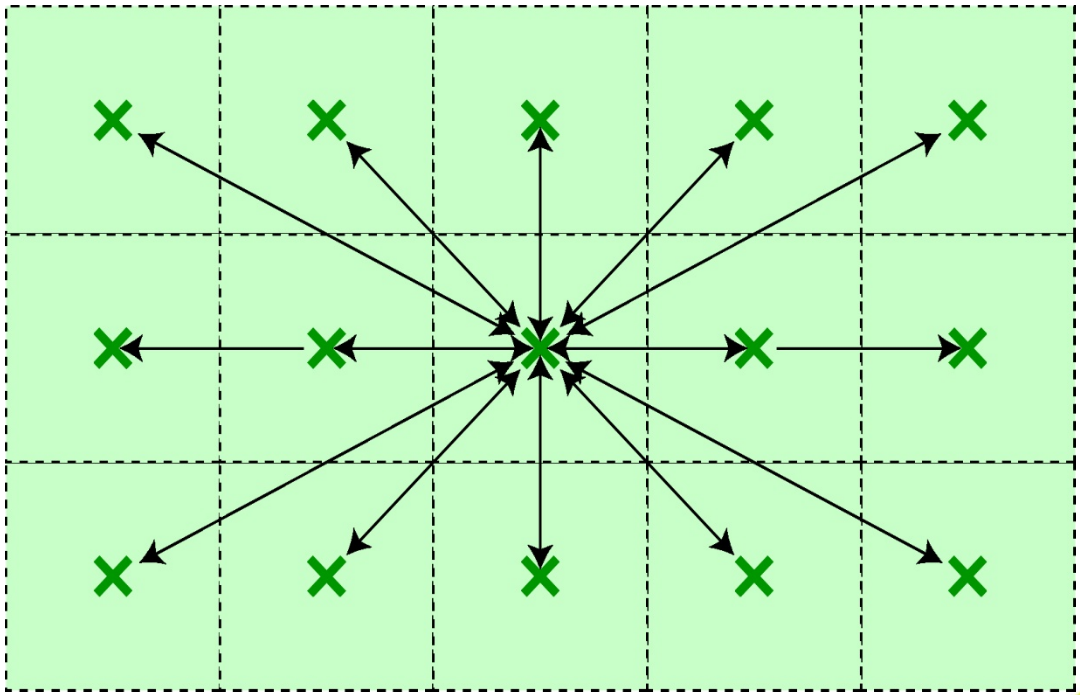
Imagine a rectangular domain in which all the points are strongly correlated with their neighbours.



Buell patterns

The points in the middle of the domain will have the strongest average correlations with all other points, simply because their average distance to all other grids is a minimum.

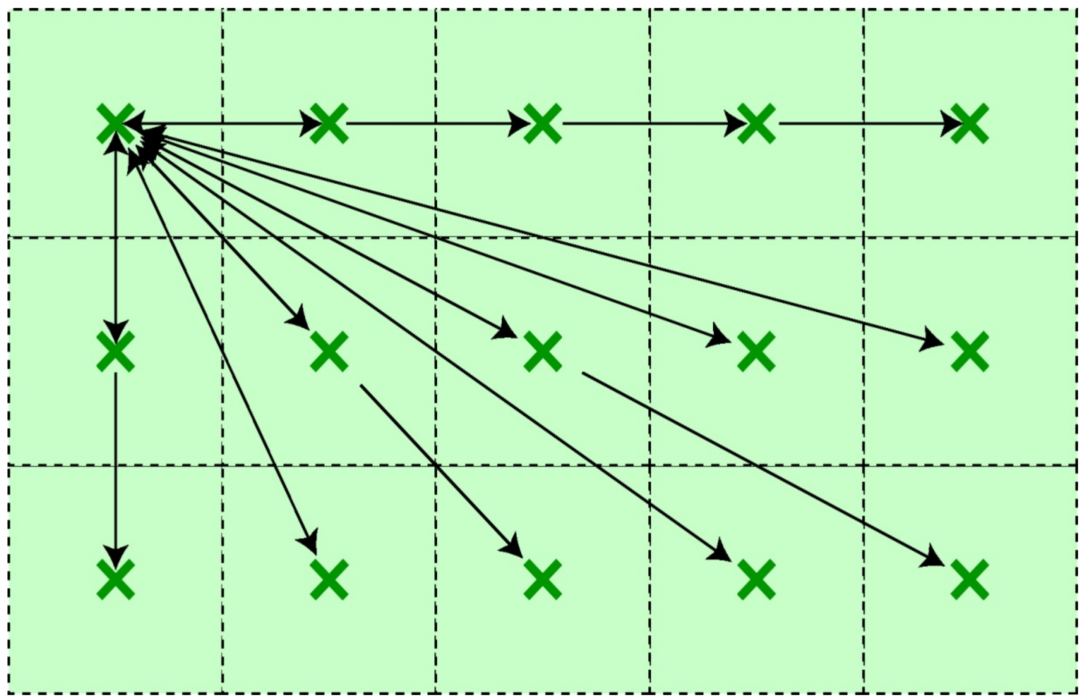
The strong correlations between neighbouring grids will be represented by PC 1, with the central grids dominating.



Buell patterns

The points in the corners of the domain will have the weakest average correlations with all other points, simply because their average distance to all other grids is a maximum.

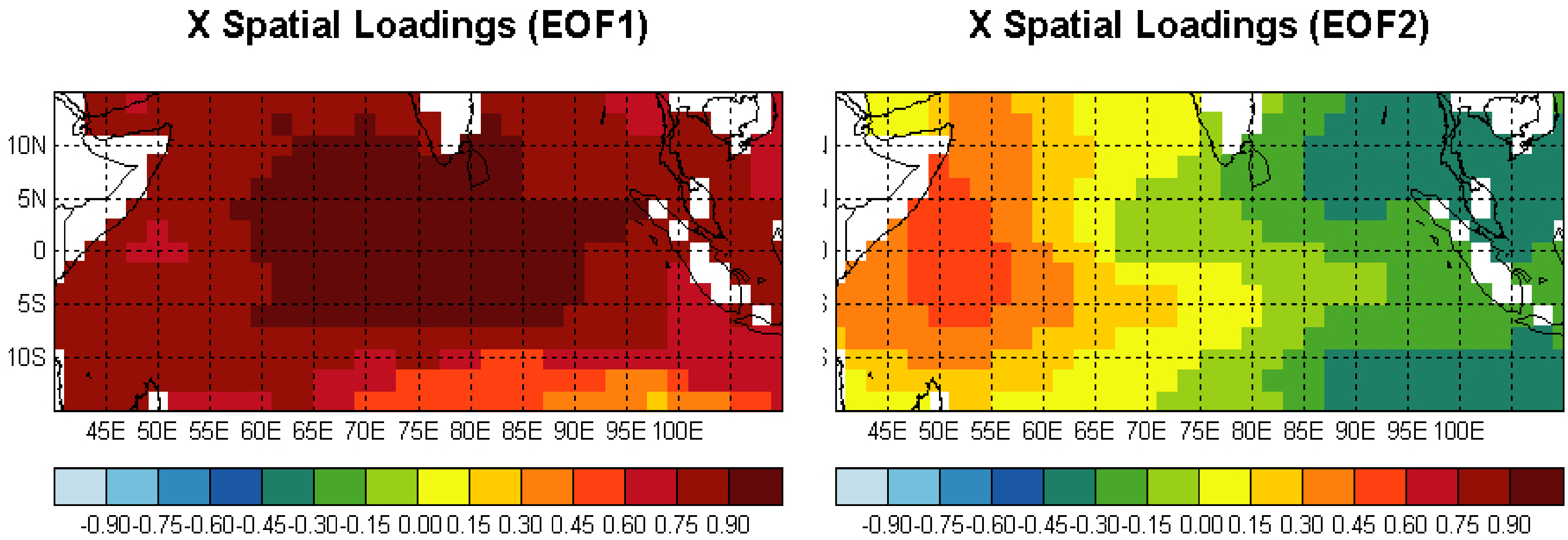
The weak correlations between distant grids will be represented by PC 2. The direction of the dipole reflects the domain shape.



Buell patterns?

PC1: positive loadings virtually everywhere, with strongest loadings in the centre

PC2: positive loadings on one side and negative on the other

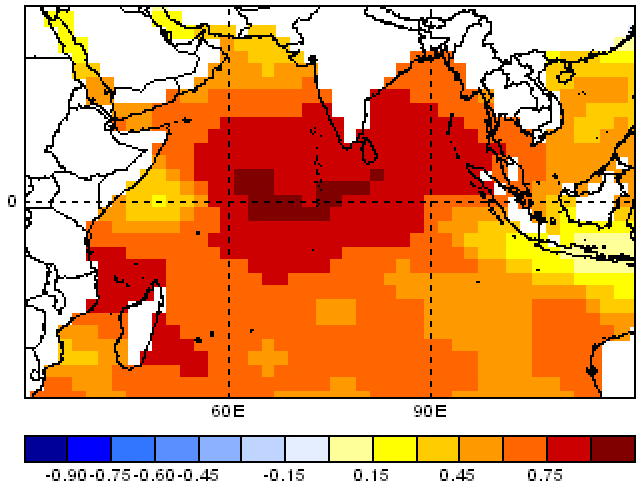


Are these real, or are they a function of the domain shape?

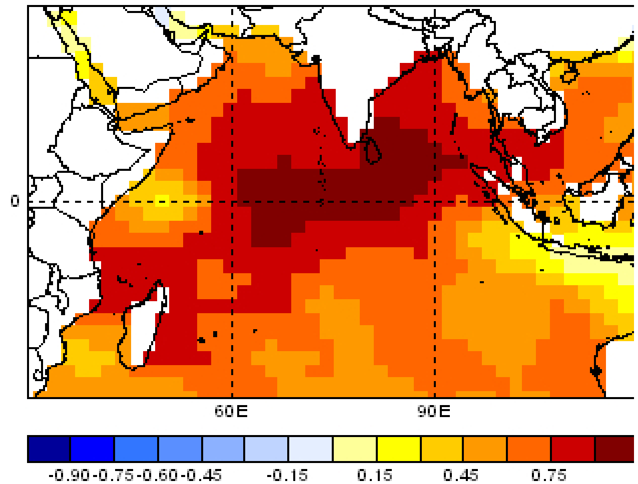
Buell Patterns in Extended EOFs

PC1: positive loadings virtually everywhere, with strongest loadings in the centre and in the central month

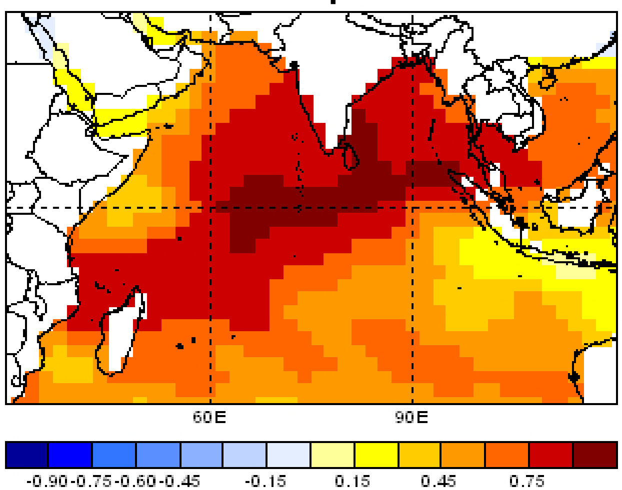
Jul



Aug



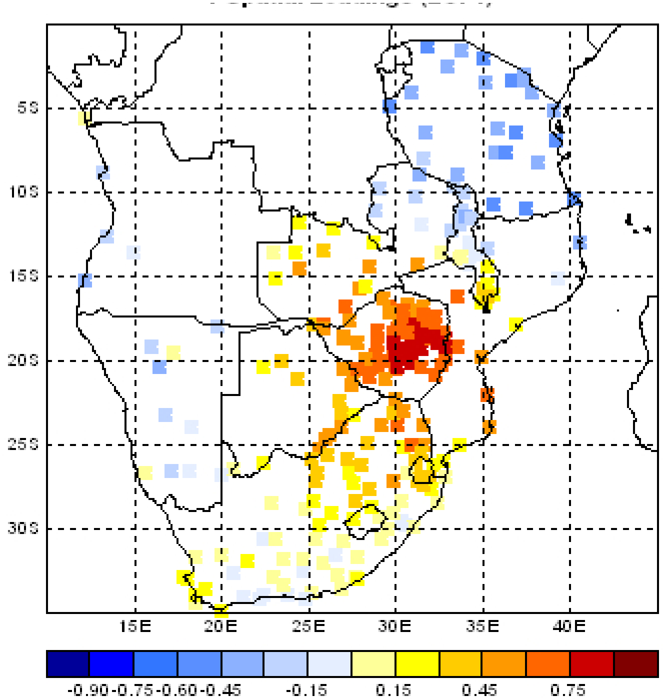
Sep



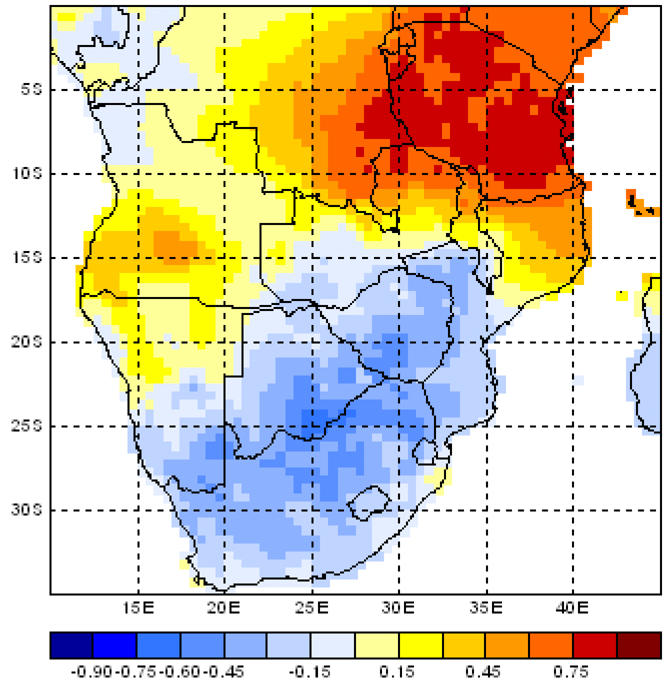
Scores and loadings for first principal component of July, August, September 1951 – 2000 sea-surface temperatures.

Buell Patterns in Station Data

With station data the problem can be even worse because of station clustering



Loadings for first principal component of Oct – Dec rainfall for 1961 – 2000 using station data.



Loadings for first principal component of Oct – Dec rainfall for 1961 – 2000 using gridded data.

Cross-Validation

Cross validation (without a hyphen) means angry validation.

Cross-validation (with a hyphen) is a method for estimating how well a model's predictions will be.

Cross-validation

Leave-one-out cross-validation

1951	Predict 1951	Training period				
1952	Training period	Predict 1952	Training Period			
1953	Training period		Predict 1953	Training period		
1954	Training period			Predict 1954	Training Period	
1955	Training period				Predict 1955	Training period

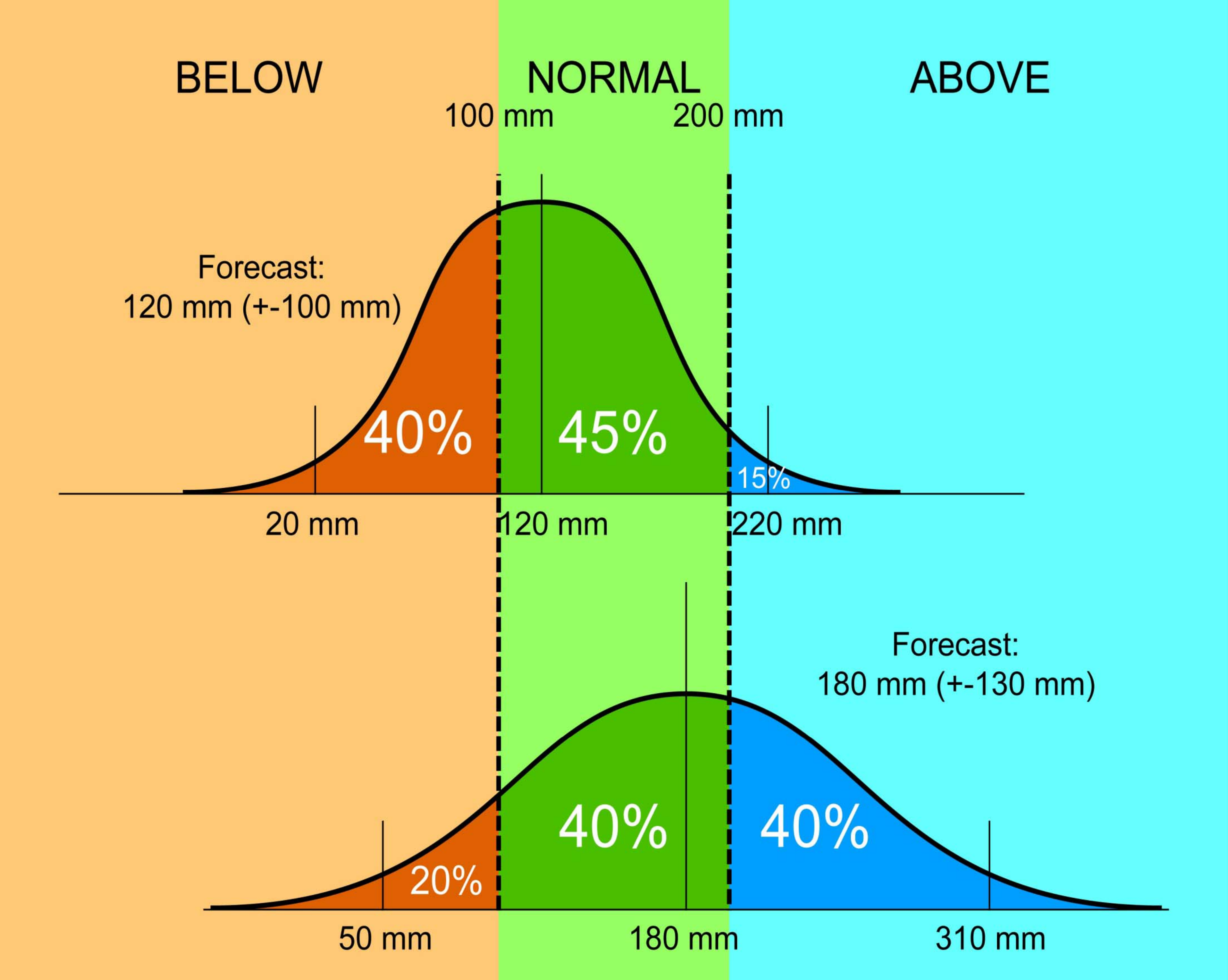
… then correlate 1951–2000.

Leave-*k*-out cross-validation

1951	Predict 1951	Omit 1952	Omit 1953	Training period			
1952	Omit 1951	Predict 1952	Omit 1953	Omit 1954	Training period		
1953	Omit 1951	Omit 1952	Predict 1953	Omit 1954	Omit 1955	Training period	
1954	Training period	Omit 1952	Omit 1953	Predict 1954	Omit 1955	Omit 1956	Training period
1955	Training period		Omit 1953	Omit 1954	Predict 1955	Omit 1956	Omit 1957

Retrospective forecasting

<table>
<tr><td>1981</td><td>Training period
(1951-1980)</td><td>Predict
1981</td><td colspan="4">Omit
1982+</td></tr>
<tr><td>1982</td><td colspan="2">Training period
(1951-1981)</td><td>Predict
1982</td><td colspan="3">Omit
1983+</td></tr>
<tr><td>1983</td><td colspan="3">Training period
(1951-1982)</td><td>Predict
1983</td><td colspan="2">Omit
1984+</td></tr>
<tr><td>1984</td><td colspan="4">Training period
(1951-1983)</td><td>Predict
1984</td><td>Omit
1985+</td></tr>
<tr><td>1985</td><td colspan="5">Training period
(1951-1984)</td><td>Predict
1985</td></tr>
</table>



BELOW
NORMAL
ABOVE
100 mm
200 mm
Forecast:
120 mm (+-100 mm)
40%
45%
15%
20 mm
120 mm
220 mm
Forecast:
180 mm (+-130 mm)
20%
40%
40%
50 mm
180 mm
310 mm