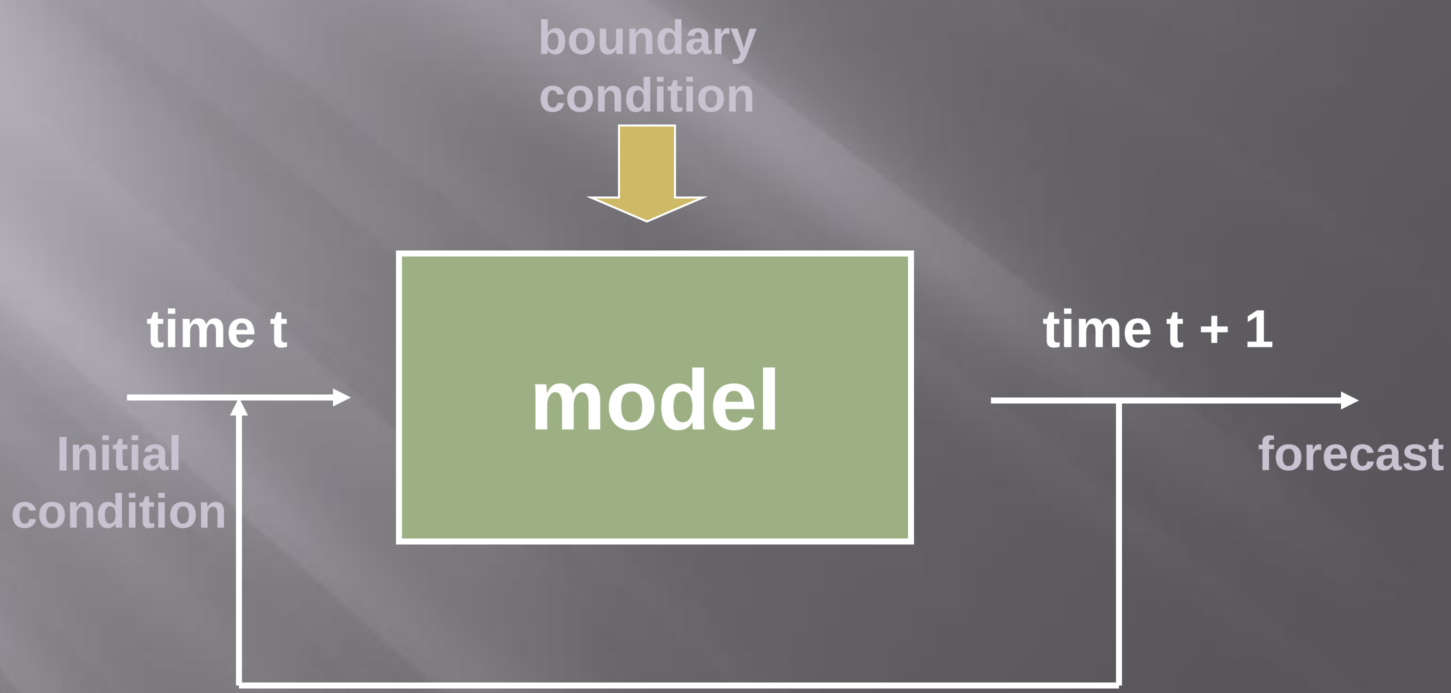
ORIGINS OF FORECAST SKILL OF WEATHER AND CLIMATE EVENTS ON VERIFIABLE TIME SCALES

Willem A. Landman
Stephanie Landman

Operational Organization

A typical run:

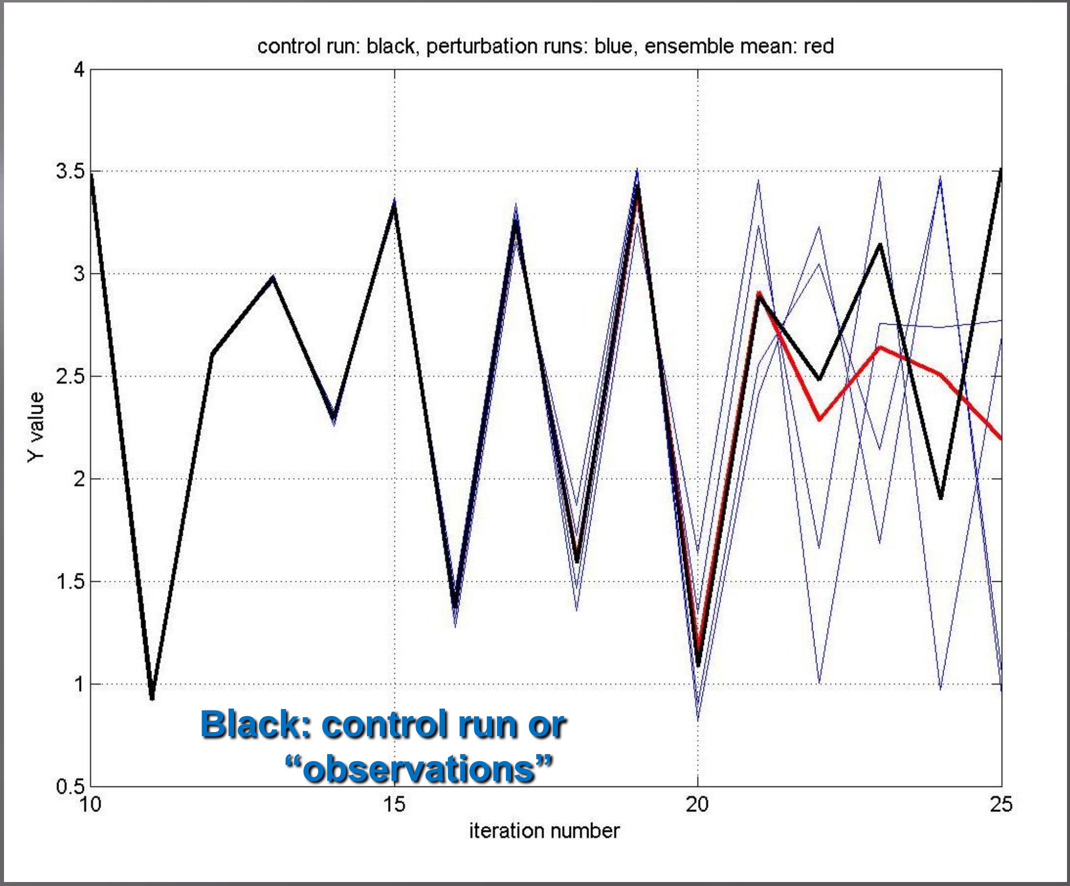


First-order quadratic difference equation

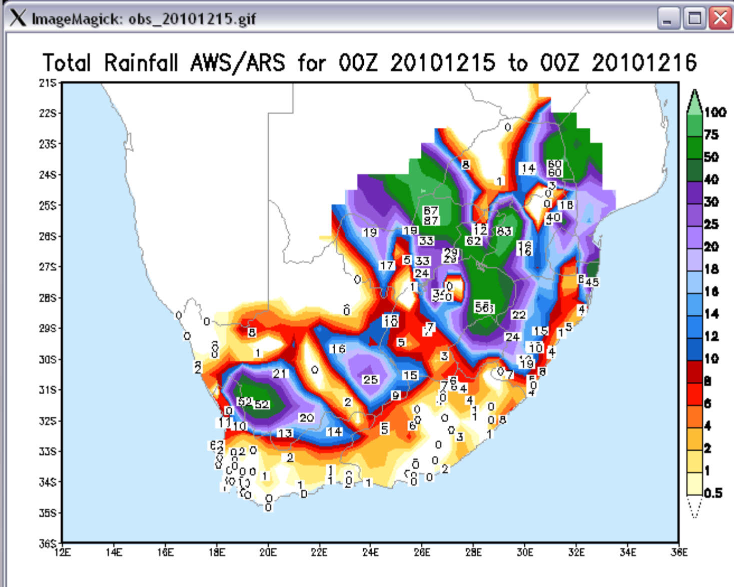
- ***Lorenz*** illustrated the general problem of predictability by considering the first-order quadratic difference equations:

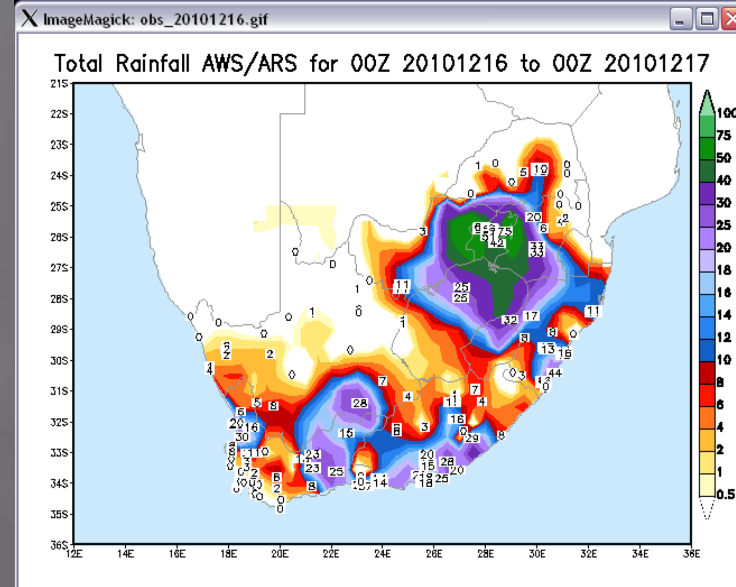
$$Y_{s+1} = aY_s - Y_s^2$$

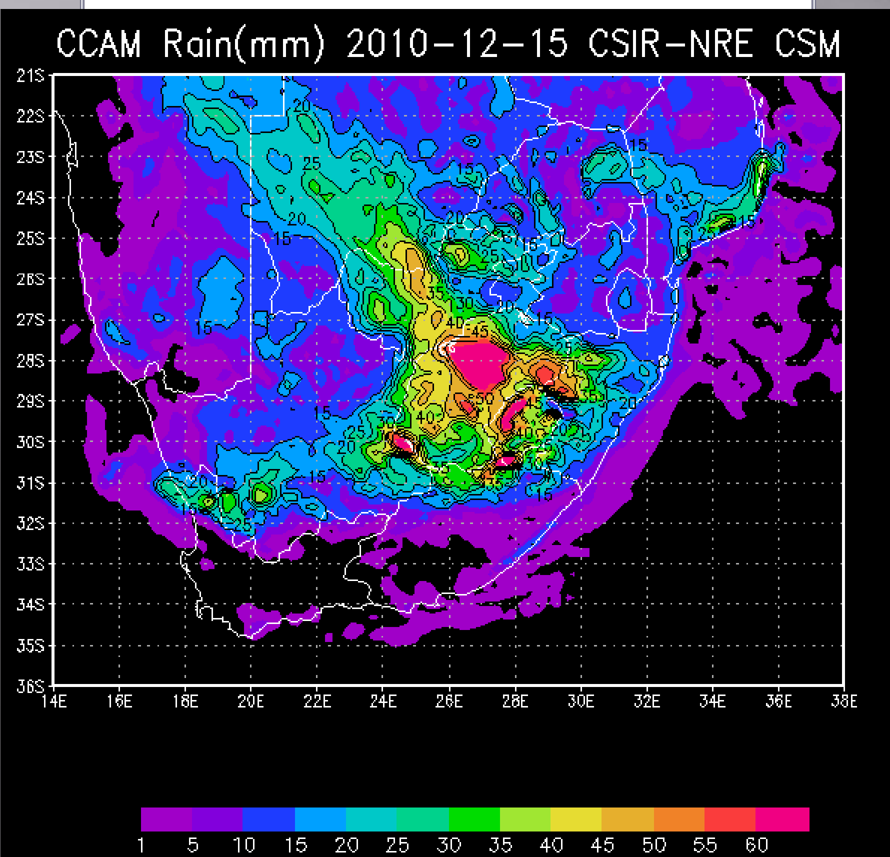
- Figure is for Y(0) = 1.5; a = 3.75

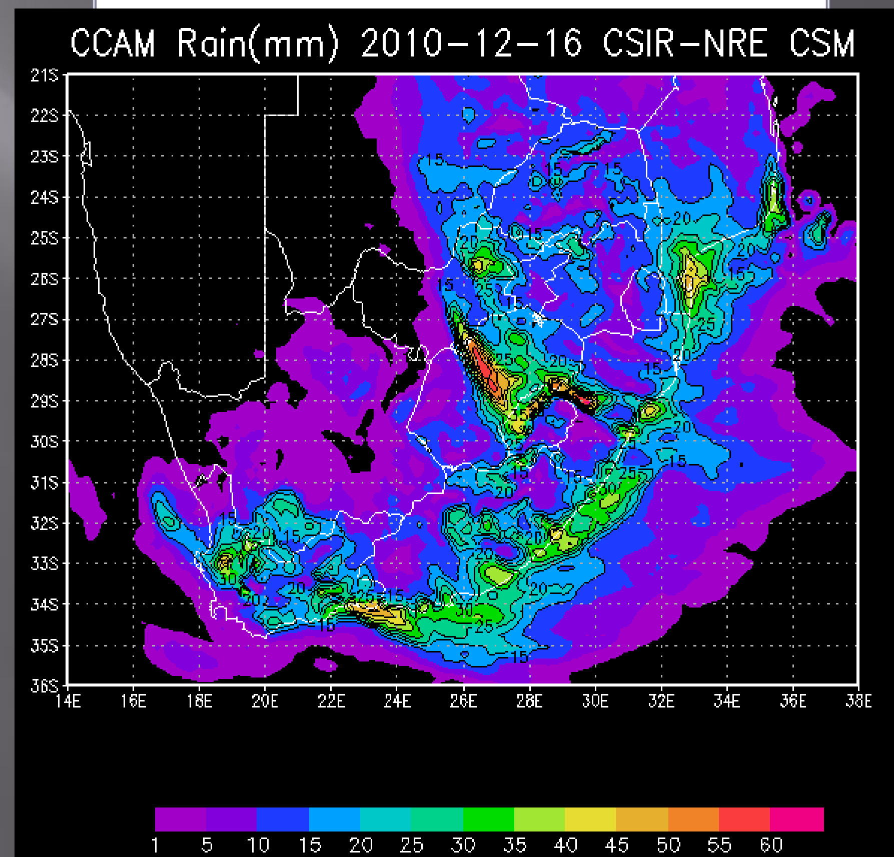


Initial value perturbed by 0.001


ImageMagick: obs_20101215.gif
Total Rainfall AWS/ARS for 00Z 20101215 to 00Z 20101216

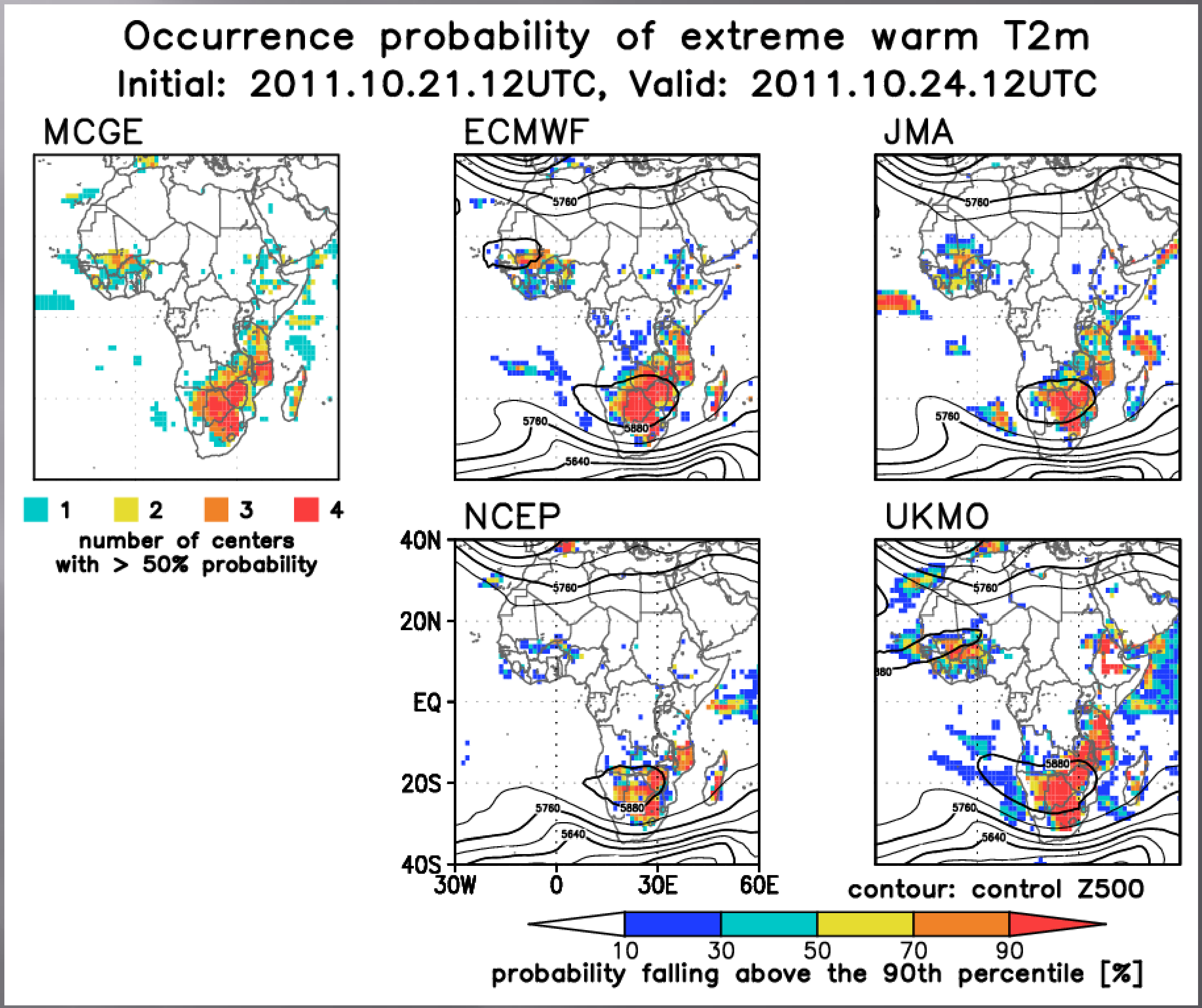

ImageMagick: obs_20101216.gif
Total Rainfall AWS/ARS for 00Z 20101216 to 00Z 20101217

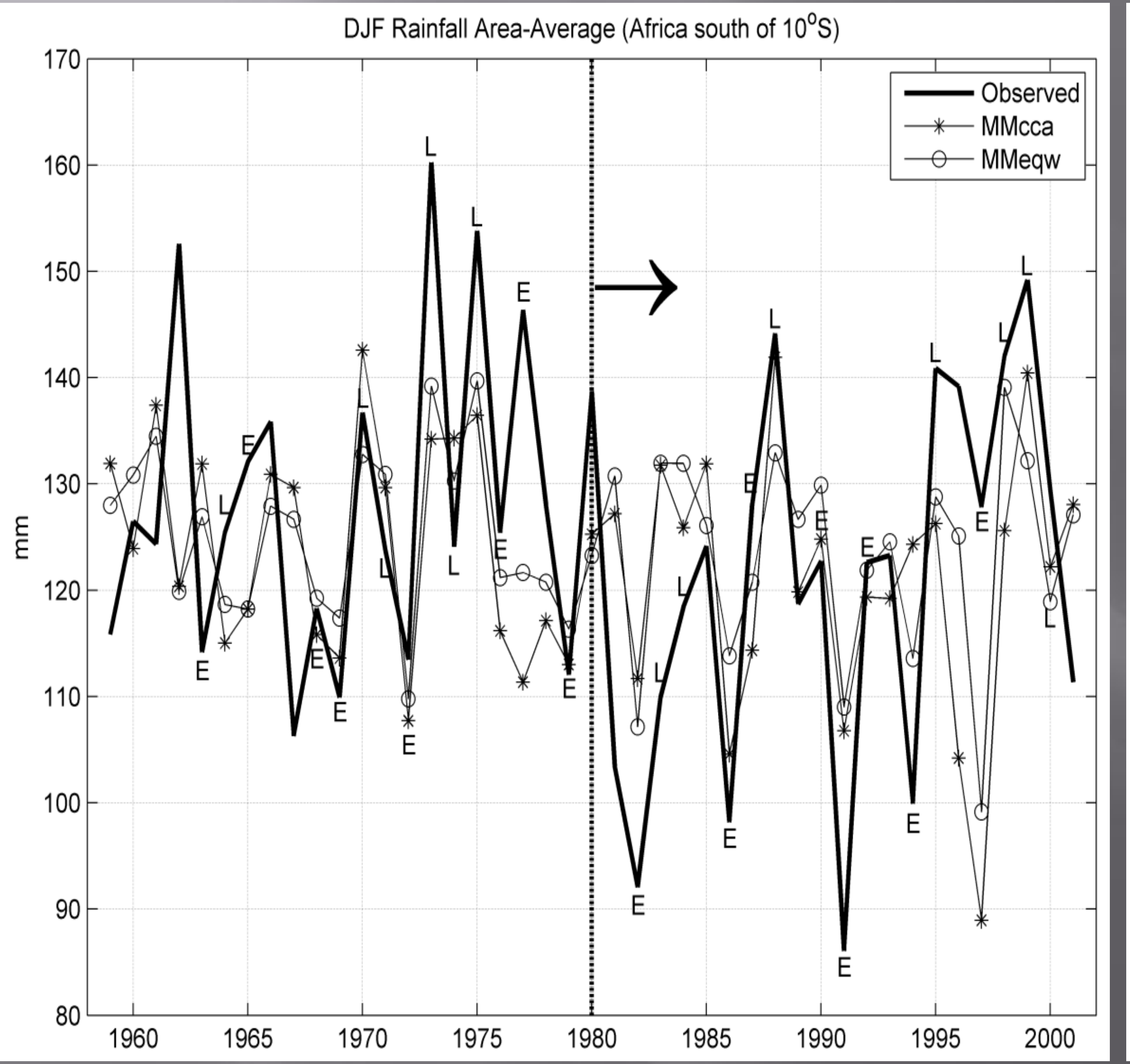

CCAM Rain(mm) 2010-12-15 CSIR-NRE CSM

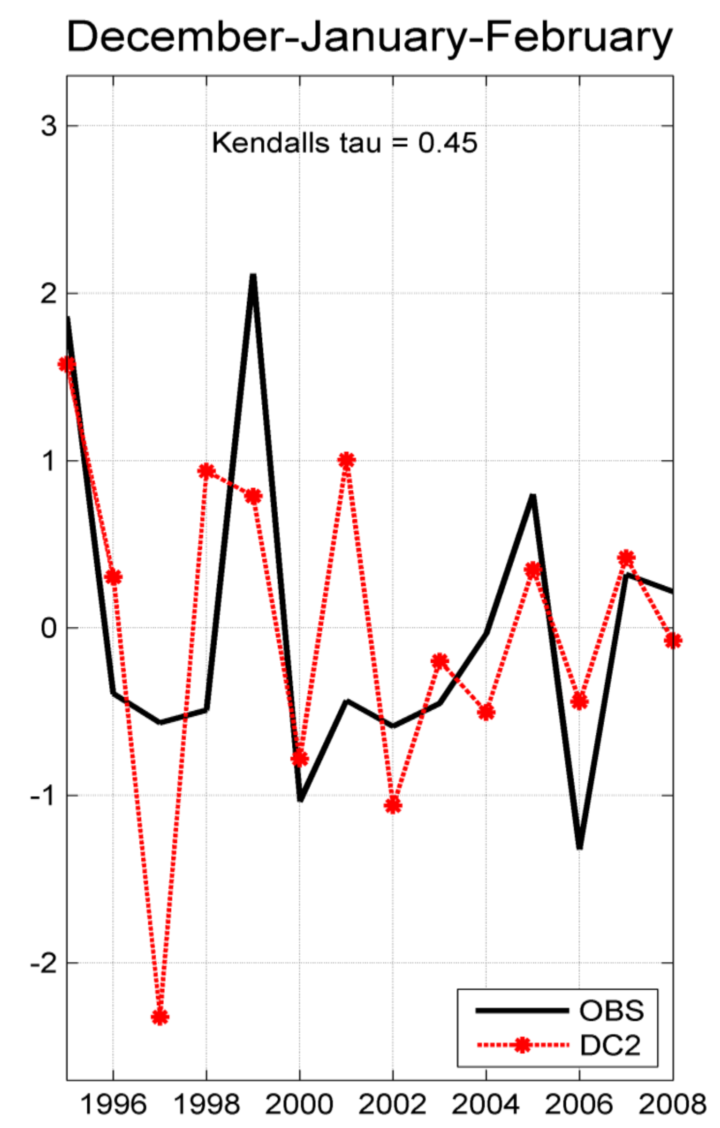

CCAM Rain(mm) 2010-12-16 CSIR-NRE CSM

A recent heat wave

Multi-Centre Grand Ensemble




DJF Rainfall Area-Average (Africa south of 10°S)
Observed
MMcca
MMeqw
mm
170
160
150
140
130
120
110
100
90
80
1960
1965
1970
1975
1980
1985
1990
1995
2000


December-January-February
Kendalls tau = 0.45
OBS
DC2
3
2
1
0
-1
-2
1996
1998
2000
2002
2004
2006
2008

CCA…

- Identifies new variables that maximize the interrelationship between **two** data sets
- This is in contrast to the patterns describing the internal variability within a **single** data set identified in PCA
- It is in this sense that CCA is referred to as "double-barreled" PCA

CCA…

- In multiple regression, the predictand is a scalar
- CCA can also be viewed as an extension of multiple regression to the case of a vector-valued predictand variable
- The predictor: vector of SSTs, SLP, etc.
- The predictand: vector of rainfall stations, etc.
- Widely applied to geophysical data in the form of fields

CCA…

$$\mathbf{V} = \mathbf{A}^T\mathbf{X}$$

$$\mathbf{W} = \mathbf{B}^T\mathbf{Y}$$

each is a linear combination of elements of the respective data vectors **X** and **Y**

A: corresponding vectors of weights of **X**,

B: corresponding vectors of weights of **Y**, (called canonical vectors)

X and **Y**: centered data

Properties of CCA…

- $\text{corr}[V_1,W_1] \geq \text{corr}[V_2,W_2] \geq \ldots \geq \text{corr}[V_M,W_M]$

- $\text{corr}[V_k,W_m] = r_c, \quad k = m,$
  $= 0, \quad k \neq m$

- $\text{var}[V_m] = \text{var}[W_m] = 1, m=1,\ldots,M$

Each of the M successive pairs of canonical variates exhibits a weaker correlation than the previous pair

Canonical correlations, r_c, are correlations between the pairs of canonical variates

Algebraic problem to solve for A and B…

$\psi = \mathbf{A}^T\mathbf{S}_{XY}\mathbf{B} - \frac{1}{2}\lambda(\mathbf{A}^T\mathbf{S}_{XX}\mathbf{A} - 1) - \frac{1}{2}\mu(\mathbf{B}^T\mathbf{S}_{YY}\mathbf{B} - 1)$,

(λ and μ are Langrangian multipliers)

$$\partial\psi/\partial\mathbf{A} = \mathbf{S}_{XY}\mathbf{B} - \lambda\mathbf{S}_{XX}\mathbf{A} = \mathbf{0}$$

$$\partial\lambda/\partial\mathbf{B} = \mathbf{S}^T{}_{XY}\mathbf{A} - \mu\mathbf{S}_{YY}\mathbf{B} = \mathbf{0}$$

…and after some incredible algebra…

The Mathematics of CCA

The CCA eigenvalue problem:

$$(\mathbf{S_{xx}}^{-1}\mathbf{S_{xy}}\mathbf{S_{yy}}^{-1}\mathbf{S_{yx}} - \lambda^2\mathbf{I})\mathbf{A} = \mathbf{0}$$

$$(\mathbf{S_{yy}}^{-1}\mathbf{S_{yx}}\mathbf{S_{xx}}^{-1}\mathbf{S_{xy}} - \lambda^2\mathbf{I})\mathbf{B} = \mathbf{0}$$

The largest eigenvalue $\lambda^2{}_1$ is associated with the first eigenvector $\mathbf{A}_1$ or $\mathbf{B}_1$

$(\lambda^2 = \mathbf{r_c})$

PCA version of CCA

- In practice: sometimes useful to “prefilter” the two fields (predictor and predictand) of raw data
- The two analyses may be truncated at different numbers of principal components
- BEWARE: important information could be lost when truncating the PCA
- PCA necessary when there is strong spatial correlation within fields
- With small sample size, PCA pre-filtering tends to improve stability – necessary for forecasting independent data

CCA as analysis tool

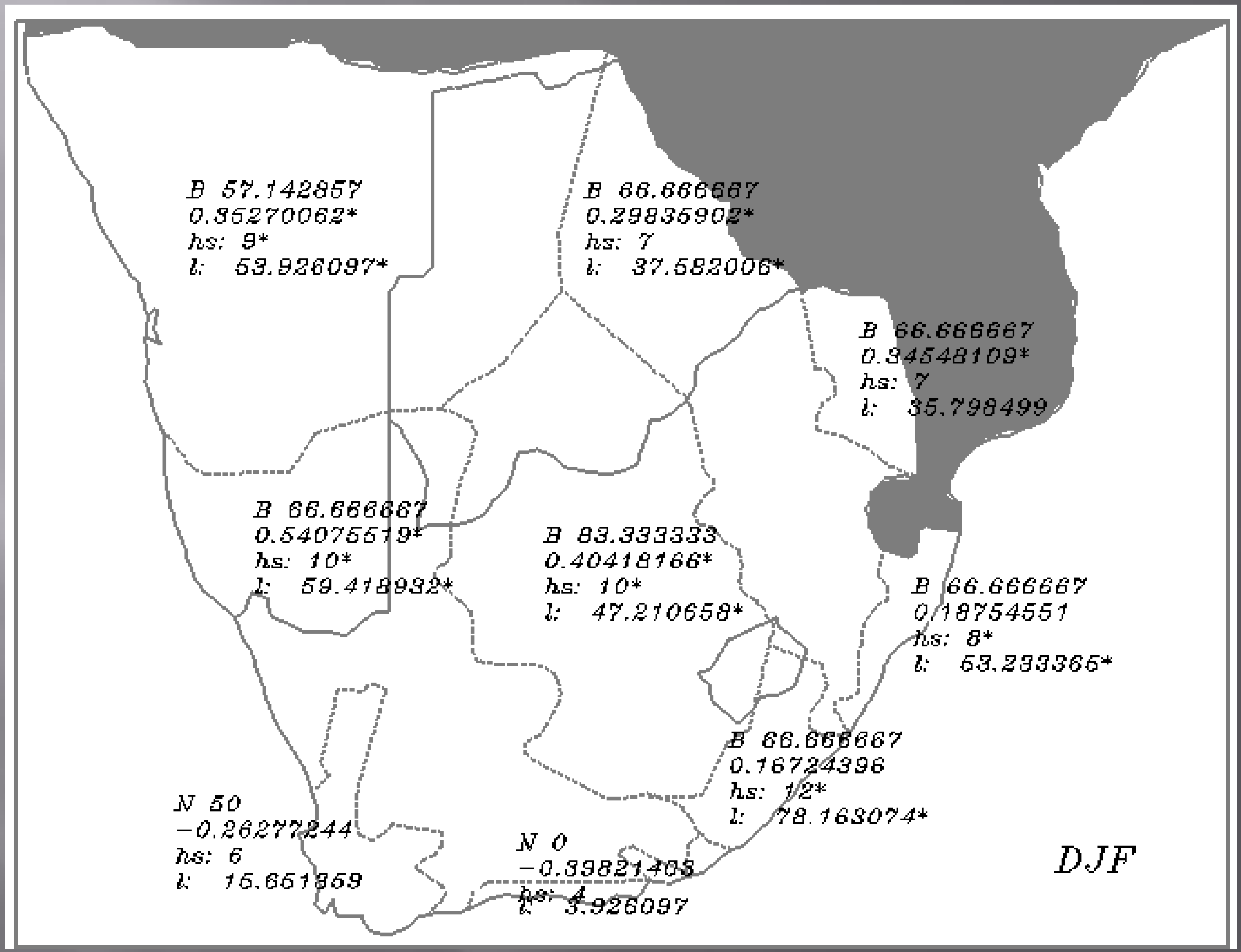
$$\mathbf{\mathit{X}}(x,t) = \Sigma_j \mathbf{r}_j(t)\mathbf{g}_j(x),\ j = 1, 2, \ldots , p$$

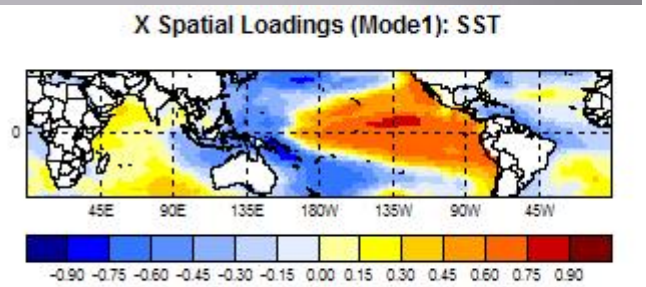
$$\mathbf{\mathit{Y}}(x',t) = \Sigma_k \mathbf{s}_k(t)\mathbf{h}_k(x'),\ k = 1, 2, \ldots , q$$

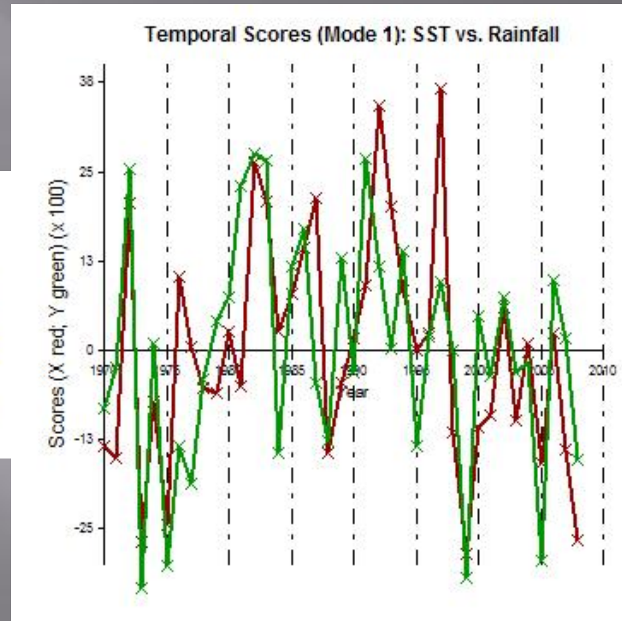
$\mathbf{g}_j$ and $\mathbf{h}_k$ are vectors whose components show the correlation at a specific location between the predictor or the predictand and their respective canonical component time series ($\mathbf{r}_j$ and $\mathbf{s}_k$)

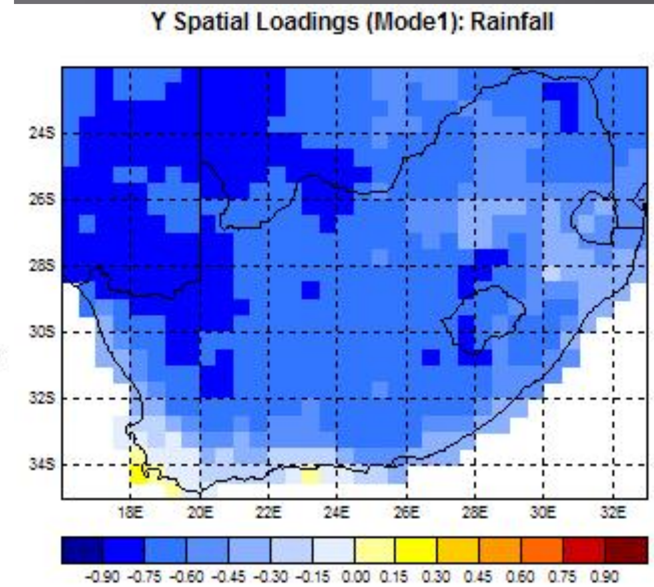
Barnett, T. P., and Preisendorfer, R. W. 1987: Origins and levels of monthly and seasonal forecast skill for United States air temperature determined by canonical correlation analysis, *Monthly Weather Review*, **115**, 1825-1850.

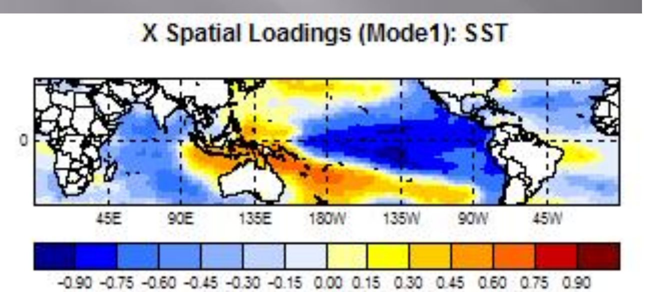
Deterministic statistical model (SST as predictor)

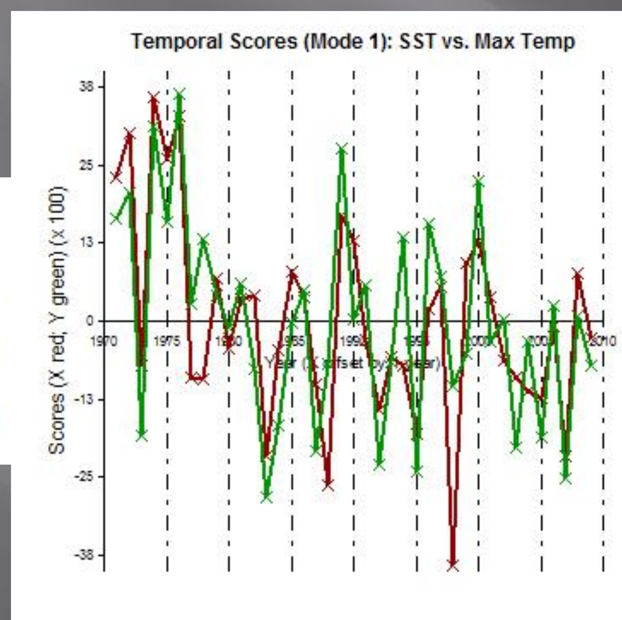


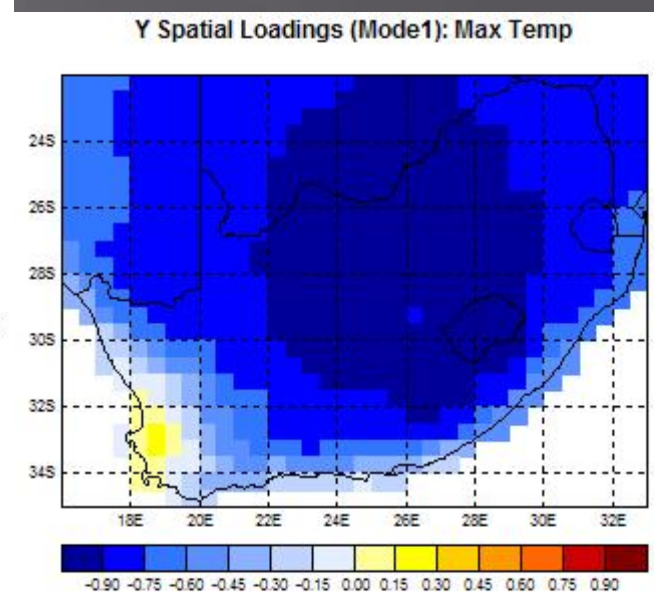

X Spatial Loadings (Mode1): SST
0
45E 90E 135E 180W 135W 90W 45W
-0.90 -0.75 -0.60 -0.45 -0.30 -0.15 0.00 0.15 0.30 0.45 0.60 0.75 0.90


Temporal Scores (Mode 1): SST vs. Rainfall
Scores (X red; Y green) (x 100)
38 25 13 0 -13 -25
2010
Year

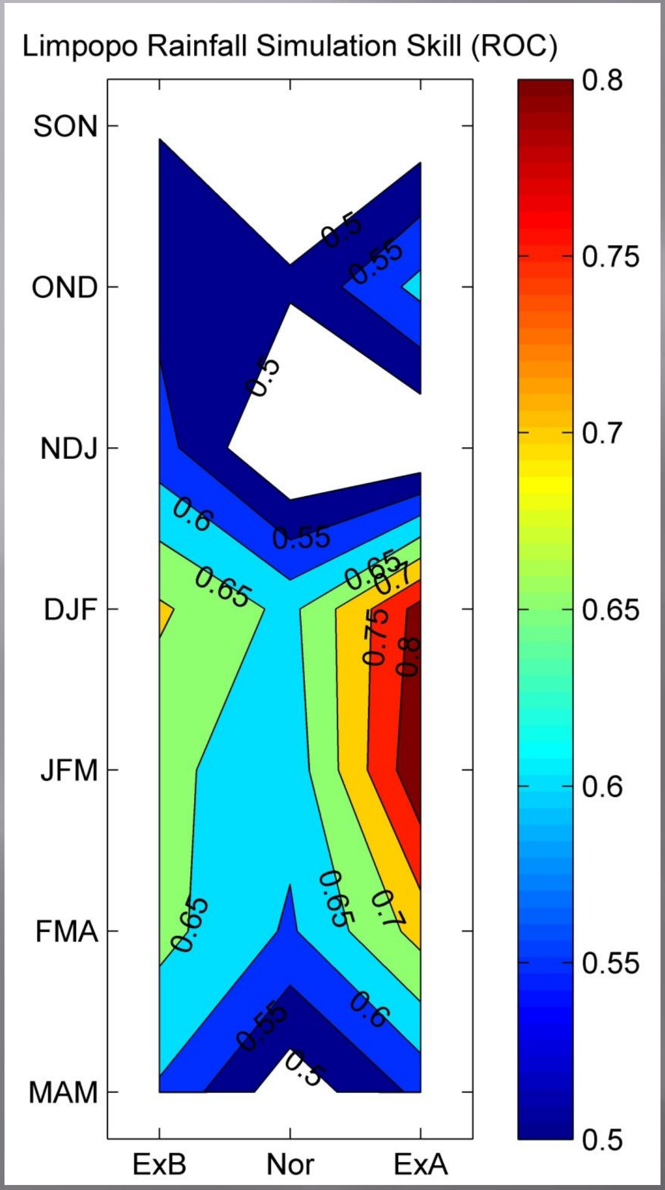

Y Spatial Loadings (Mode1): Rainfall
24S 26S 28S 30S 32S 34S
18E 20E 22E 24E 26E 28E 30E 32E
-0.90 -0.75 -0.60 -0.45 -0.30 -0.15 0.00 0.15 0.30 0.45 0.60 0.75 0.90

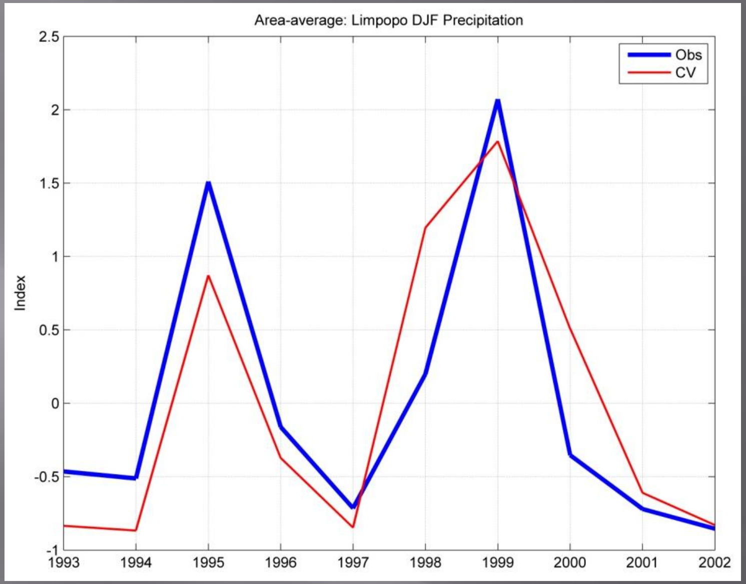

X Spatial Loadings (Mode1): SST
0
45E 90E 135E 180W 135W 90W 45W
-0.90 -0.75 -0.60 -0.45 -0.30 -0.15 0.00 0.15 0.30 0.45 0.60 0.75 0.90

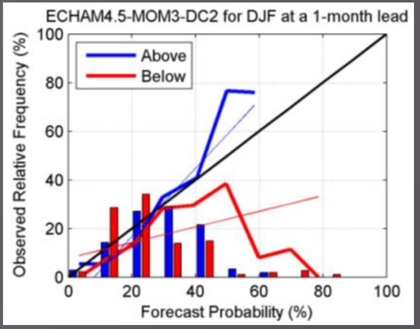

Temporal Scores (Mode 1): SST vs. Max Temp
Scores (X red; Y green) (x 100)
38 25 13 0 -13 -25 -38
1970 1975 2010

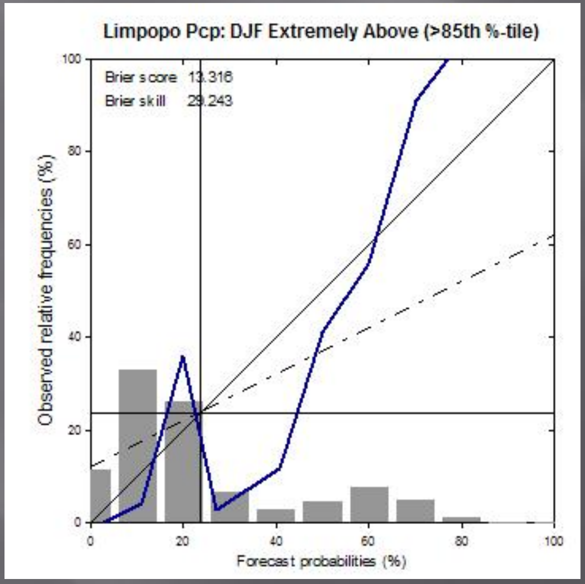

Y Spatial Loadings (Mode1): Max Temp
24S 26S 28S 30S 32S 34S
18E 20E 22E 24E 26E 28E 30E 32E
-0.90 -0.75 -0.60 -0.45 -0.30 -0.15 0.00 0.15 0.30 0.45 0.60 0.75 0.90

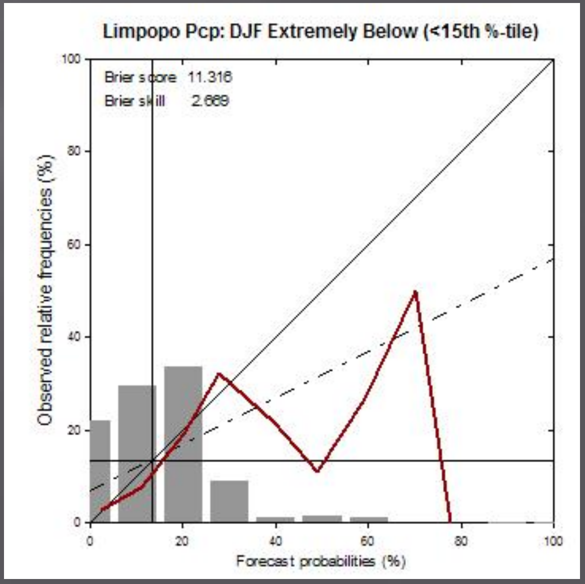
Verification: *Limpopo*

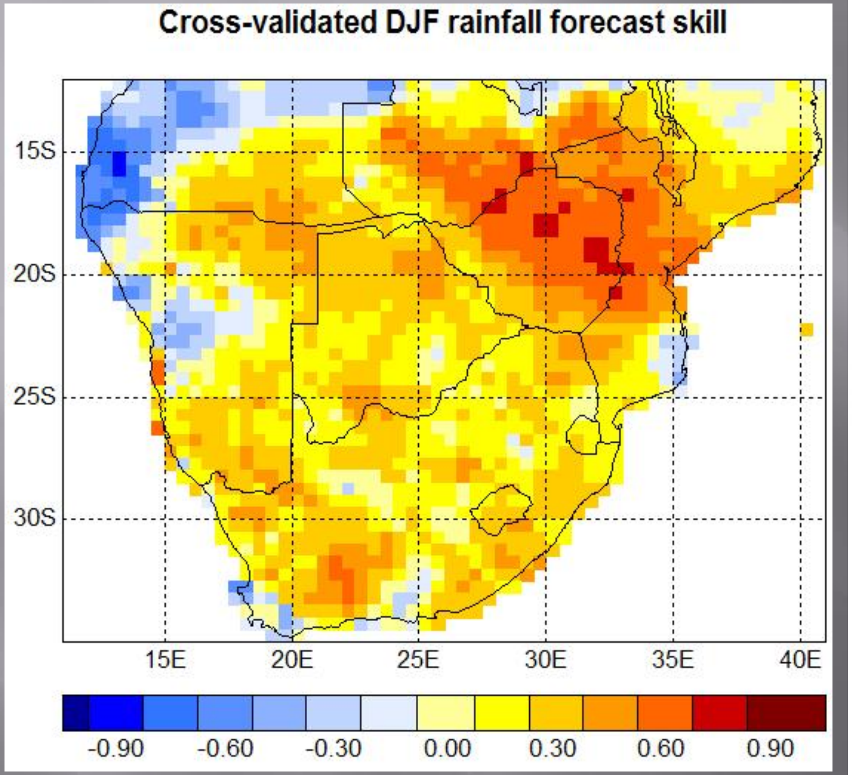




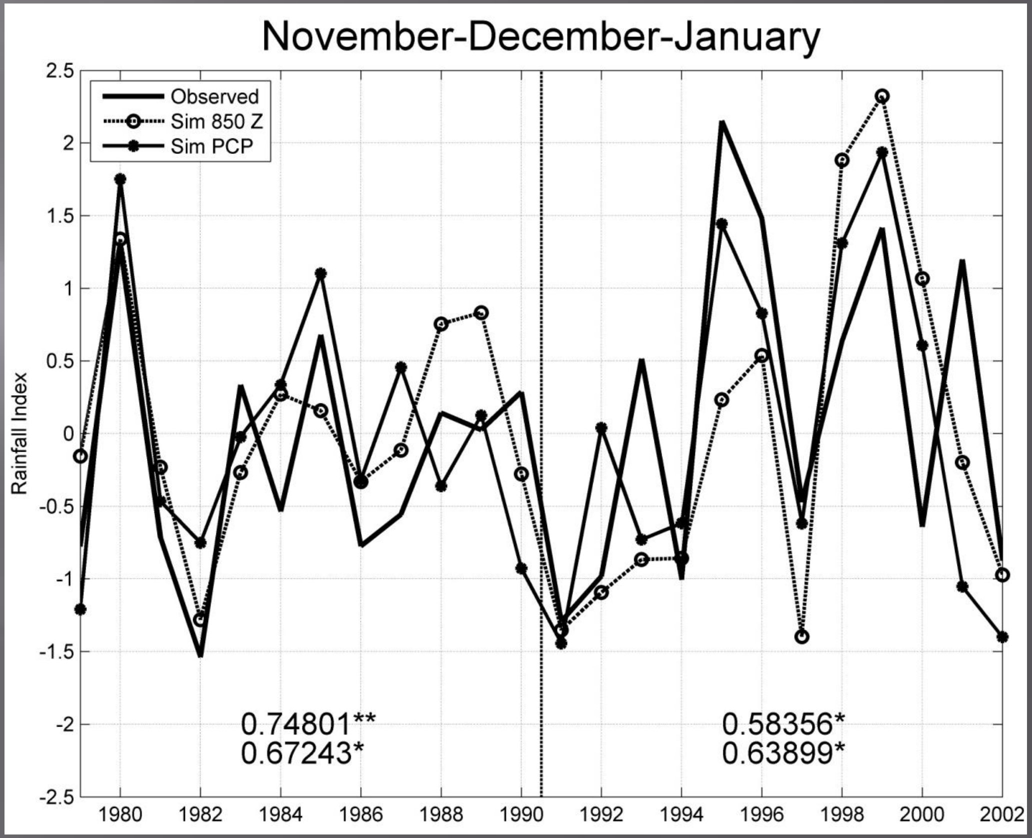




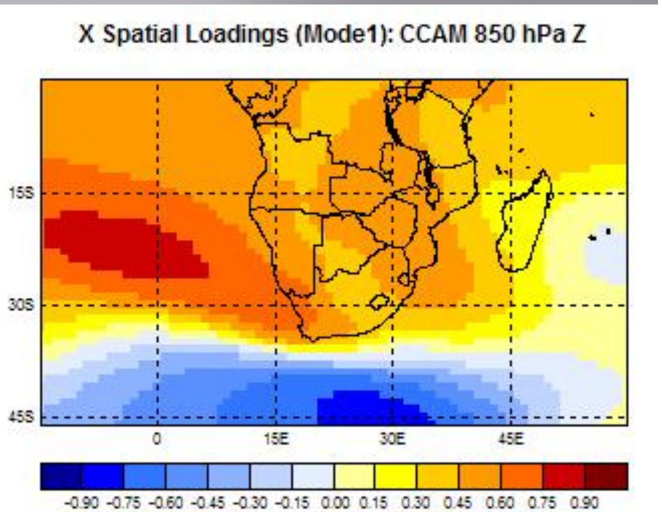


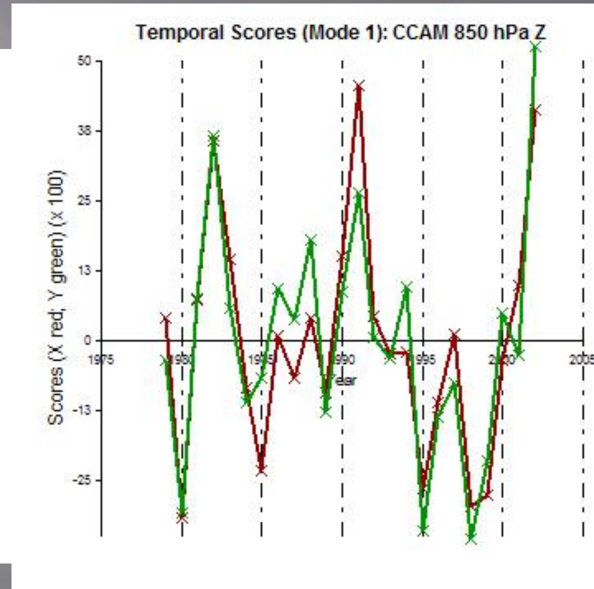


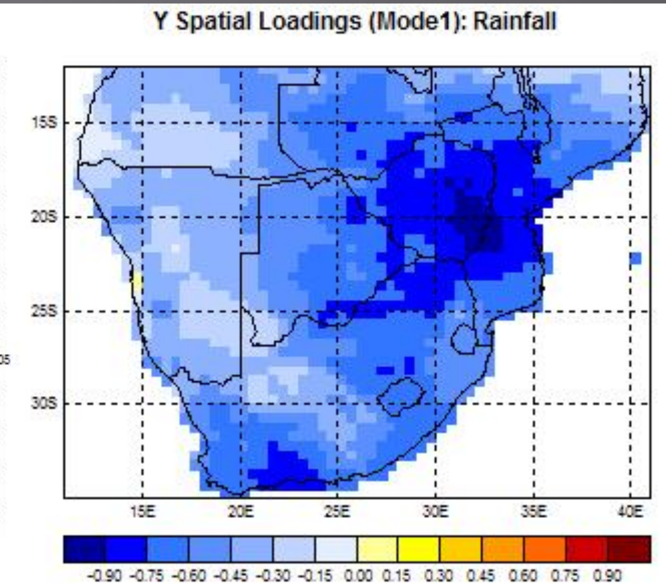
Cross-validated DJF rainfall forecast skill
15S
20S
25S
30S
15E
20E
25E
30E
35E
40E
-0.90
-0.60
-0.30
0.00
0.30
0.60
0.90

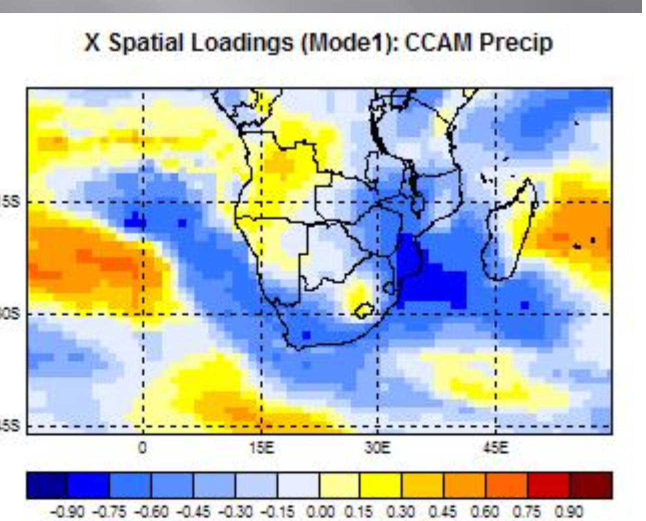


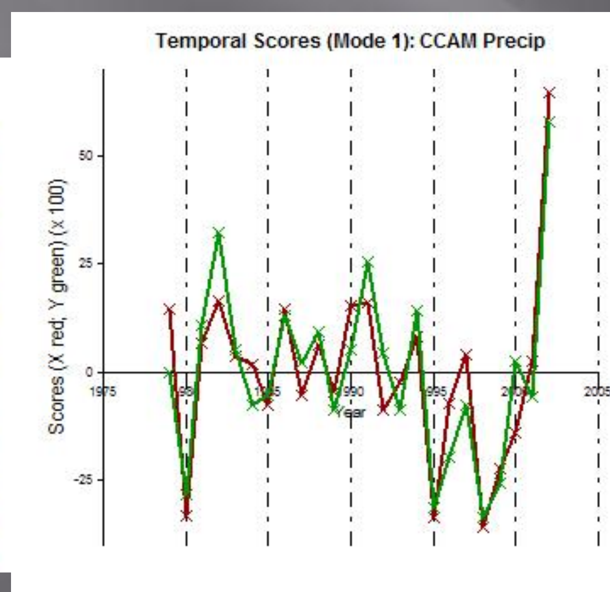
November-December-January
Observed
Sim 850 Z
Sim PCP
Rainfall Index
0.74801**
0.67243*
0.58356*
0.63899*
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002

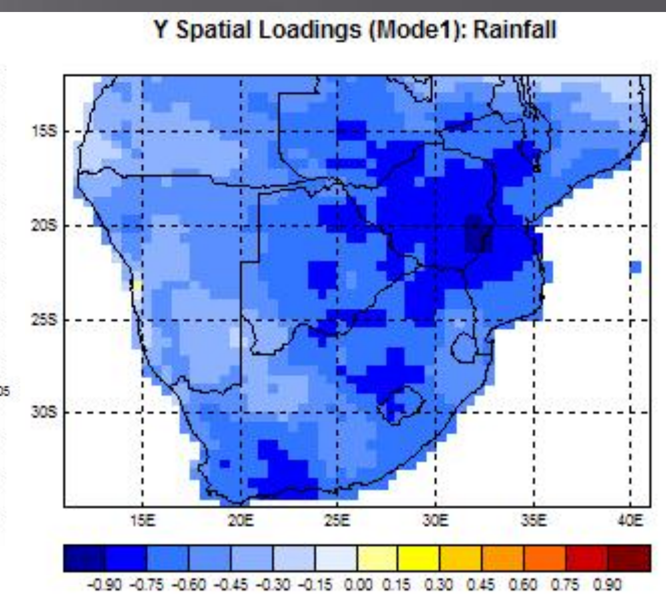

X Spatial Loadings (Mode1): CCAM 850 hPa Z
15S
30S
45S
0
15E
30E
45E
-0.90 -0.75 -0.60 -0.45 -0.30 -0.15 0.00 0.15 0.30 0.45 0.60 0.75 0.90


Temporal Scores (Mode 1): CCAM 850 hPa Z
Scores (X red; Y green) (x 100)
Year
50
38
25
13
0
-13
-25
1975
2005


Y Spatial Loadings (Mode1): Rainfall
15S
20S
25S
30S
15E
20E
25E
30E
35E
40E
-0.90 -0.75 -0.60 -0.45 -0.30 -0.15 0.00 0.15 0.30 0.45 0.60 0.75 0.90


X Spatial Loadings (Mode1): CCAM Precip
0
15E
30E
45E
-0.90 -0.75 -0.60 -0.45 -0.30 -0.15 0.00 0.15 0.30 0.45 0.60 0.75 0.90


Temporal Scores (Mode 1): CCAM Precip
Scores (X red; Y green) (x 100)
Year
50
25
0
-25
1975
2005


Y Spatial Loadings (Mode1): Rainfall
15S
20S
25S
30S
15E
20E
25E
30E
35E
40E
-0.90 -0.75 -0.60 -0.45 -0.30 -0.15 0.00 0.15 0.30 0.45 0.60 0.75 0.90

Final points…

- Models can reliably predict weather and climate variations and extremes
- CCA is a linear technique which can provide some insight into the dynamics of the Earth System
- But…
  - CCA diagnostics are notoriously difficult to interpret physically
  - The weights are defined to maximize the correlation, not maximize the interpretability