Geophysical Applications of Singular Spectrum Analysis (SSA) and Multivariate Singular Spectrum Analysis (MSSA)

References:

1. ADVANCED SPECTRAL METHODS FOR CLIMATIC TIME SERIES

   M. Ghil,[1] M. R. Allen,[2] M. D. Dettinger,[3] K. Ide,[1]
   D. Kondrashov,[1] M. E. Mann,[4] A. W. Robertson,[1]
   A. Saunders,[1] Y. Tian,[1] F. Varadi,[1] and P. Yiou[5]

   Received 28 August 2000; revised 3 July 2001; accepted 18 September 2001; published XX Month 2001.

2. Dennis Hartmann's notes.
3. http://www.atmos.ucla.edu/tcd/ssa/

You can use many software packages to do SSA or MSSA

MATLAB

Splus

SAS

Fortran

C++

SSA

SSA is desinged to extract information from a noisy timeseries. We can extract trends, oscillatory patterns and we can do singal/noise enhancement.

SSA	Fourier Analysis
Basis function: T-EOFs (time varying EOFs)	Basis function: Sines and Cosines
Data Adaptive	Fixed Sinusoidal
T-PCs are functions of time	Time independent

Let’s assume you have a timeseries T of size n.

$$\hat{T} = \hat{T}_1, \hat{T}_2, \hat{T}_3, \ldots., \hat{T}_{n-1}, T_n$$

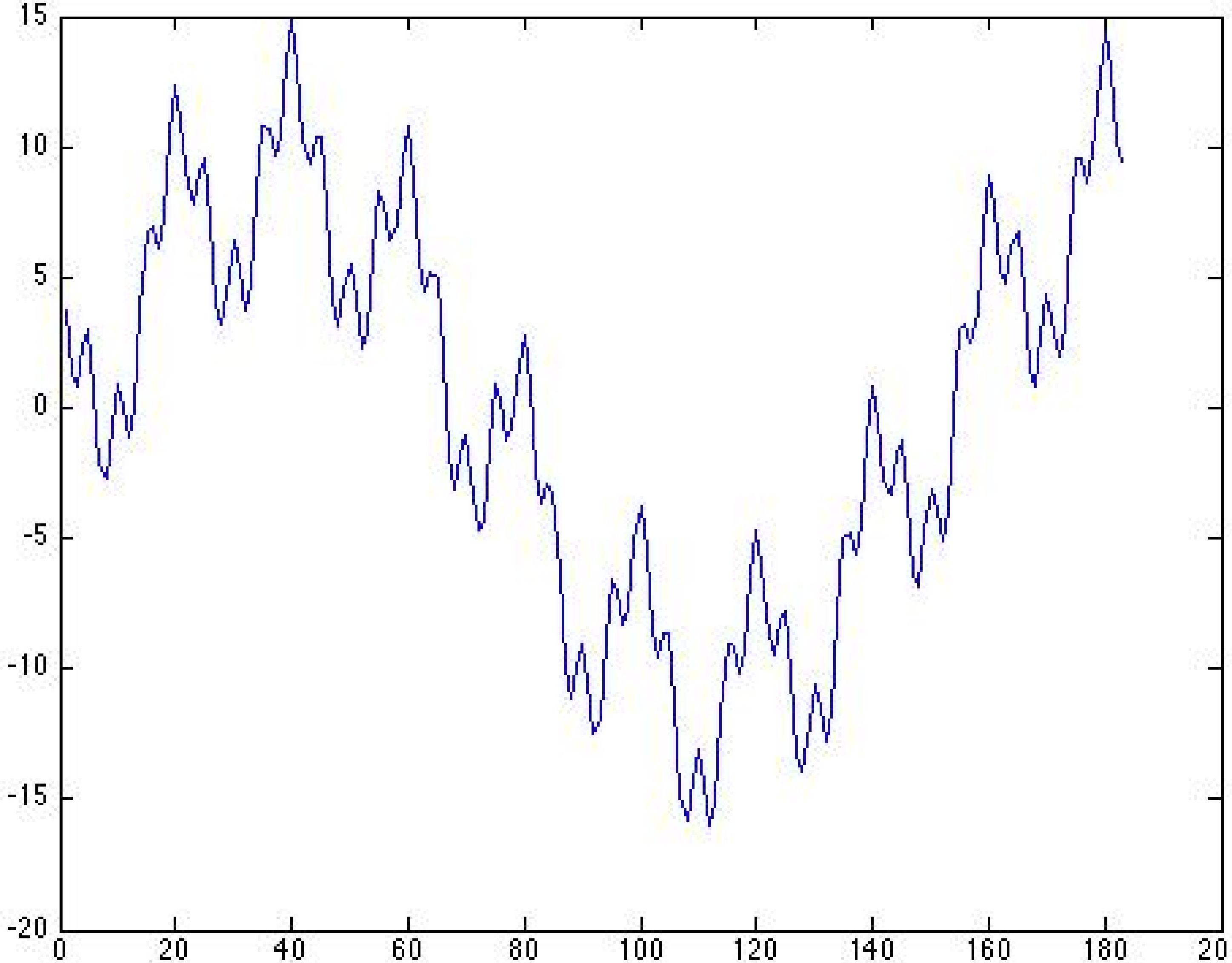
First we subtract the mean of the timeseries (or normalize)

$$Ti = \hat{T}i - \bar{T}$$

$$T = T_1, T_2, T_3, \ldots., T_{n-1}, T_n$$

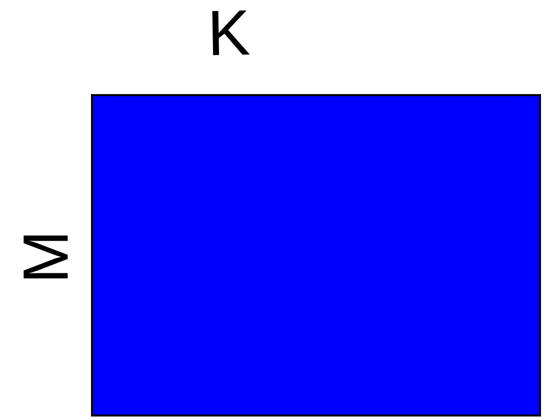
We select a window of size M. The physical processes that we are interested must occur within this window. The window length should not exceed 1/3 timeseries length.

$$T = T_1, T_2, T_3, \ldots., T_{n-2}, T_{n-1}, T_n$$



15
10
5
0
-5
-10
-15
-20
0
20
40
60
80
100
120
140
160
180
200

We first construct a “trayectory matrix”, by passing a delay window of length M on the timeseries.

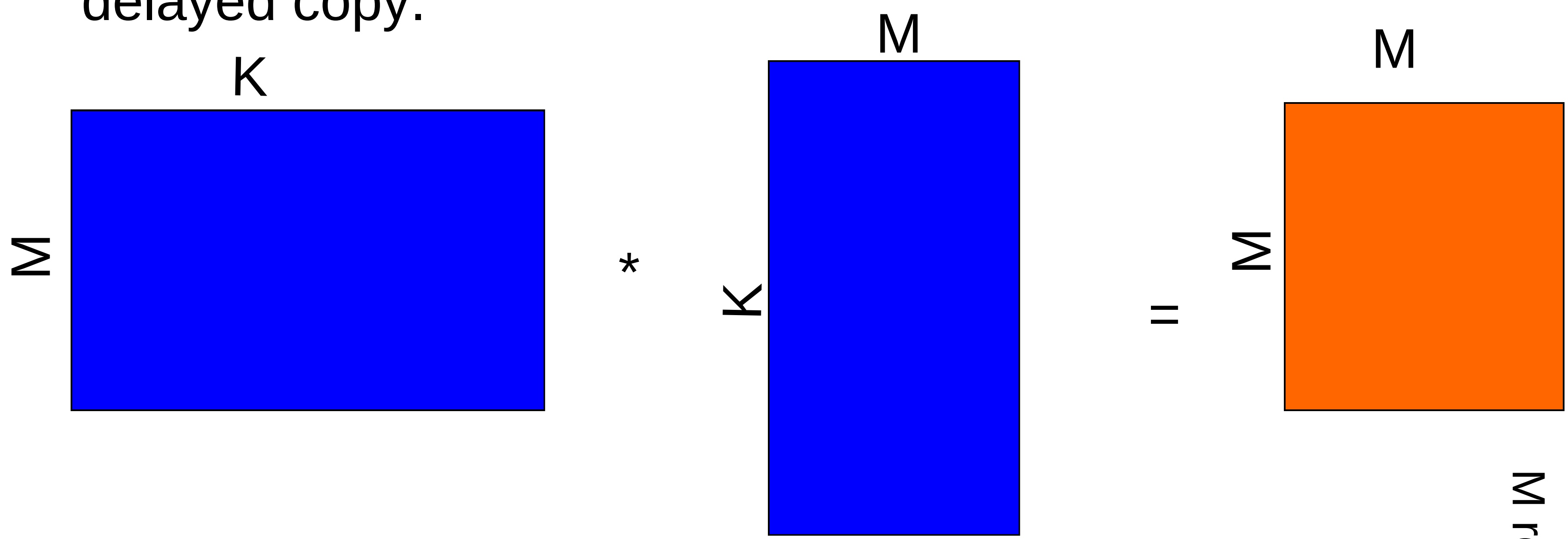


K columns (K=N-M+1)

$$X = \begin{pmatrix} T_1 & T_2 \; \dots & T_K \\ T_2 & T_3 \; \dots & T_{K+1} \\ & & \\ T_M & T_{M+1} \; \dots & T_N \end{pmatrix}$$

M rows (window length)

Then we calculate the covariance (correlation) matrix, so we are looking at how the timeseries is correlated with its delayed copy.

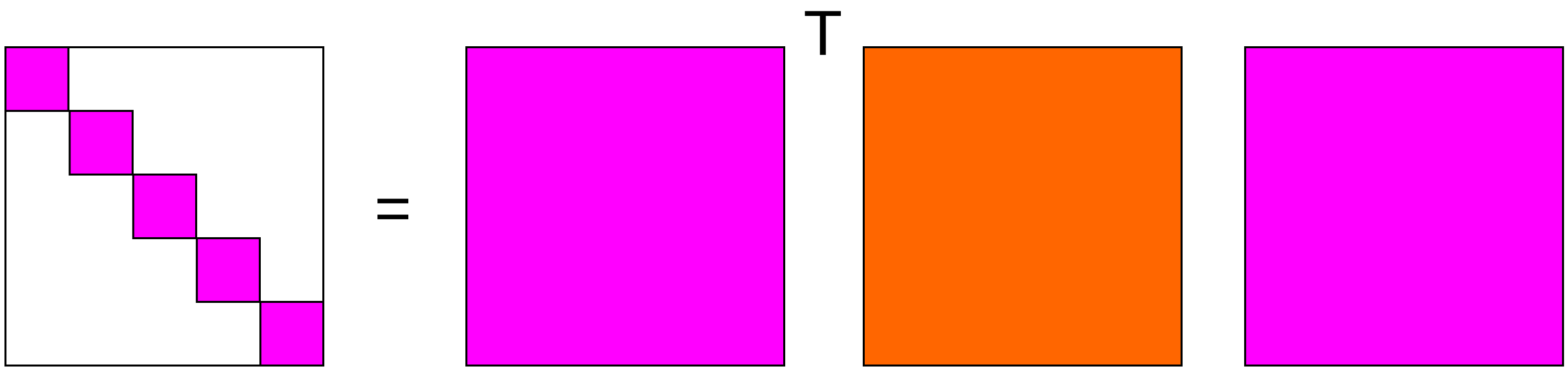


$$C = 1/(K-1)\,[X\,X^T] \quad = \quad \begin{pmatrix} C_{11} & C_{12} \dots & C_{1M} \\ C_{21} & C_{22} \dots & C_{2M} \\ C_{M1} & C_{M2} \dots & C_{MM} \end{pmatrix}$$

M rows (window length)

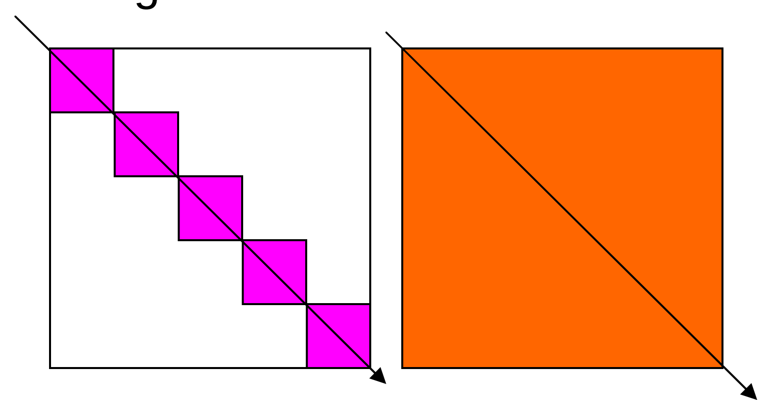
M columns (window length)

We perform eigenanalysis on the covariance matrix.

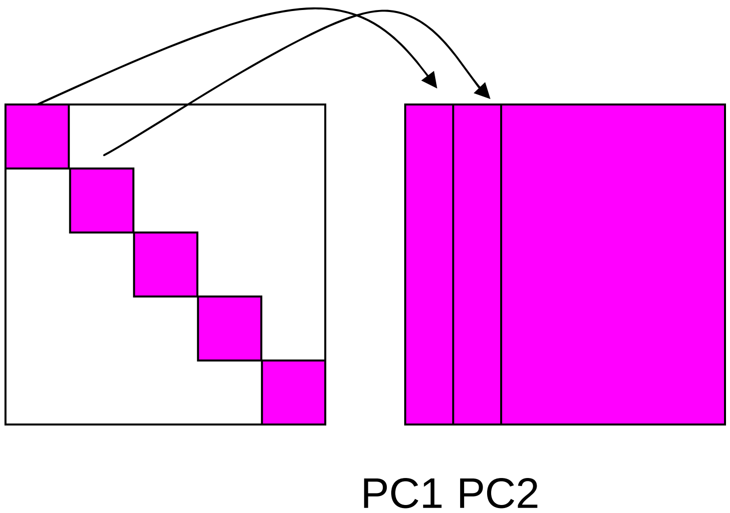


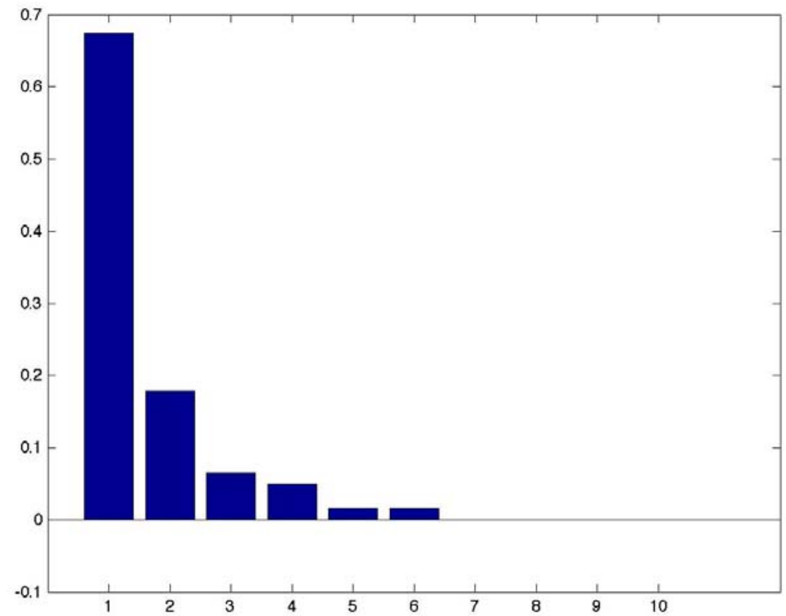
$$\Lambda = E^T C E = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_M \end{pmatrix}$$

The set of eigenvectors e_i and eigenvalues λ_i represent a coordinate transformation where C becomes a diagonal.



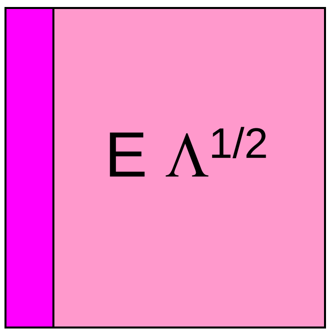
The eigenvector matrix will be of size M x M

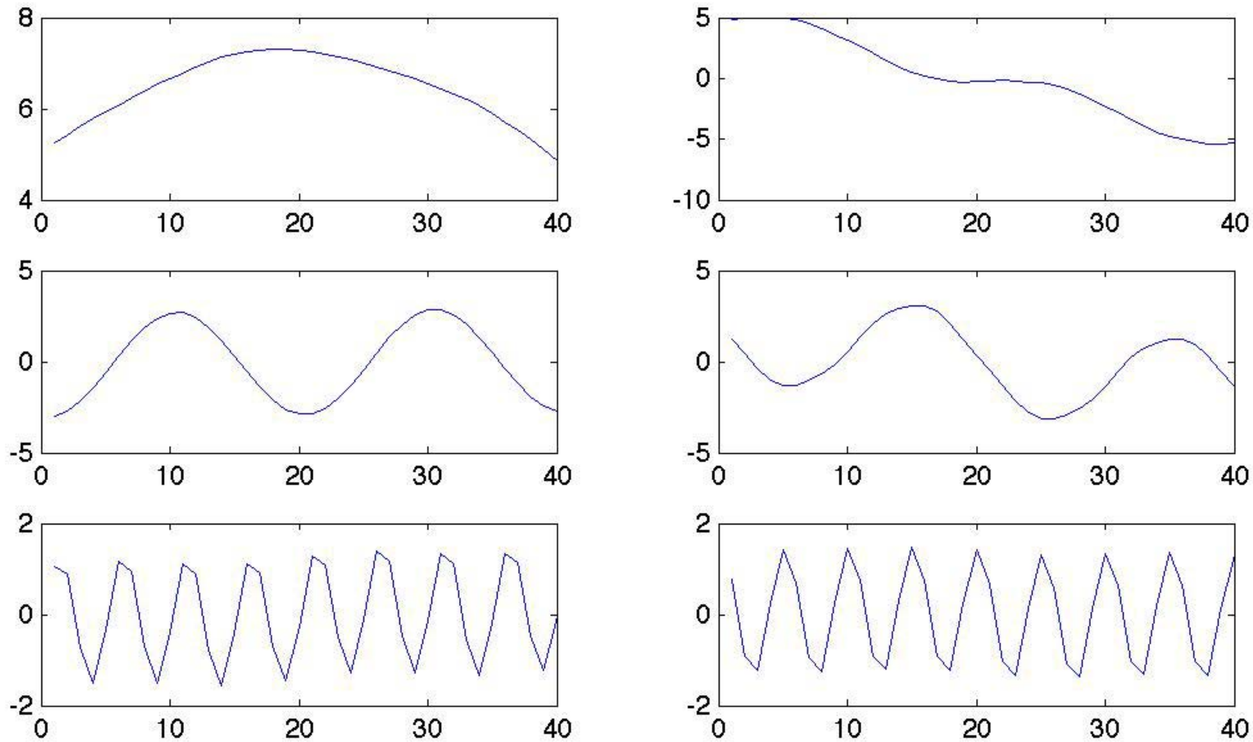




Just as in PCA, the first PC explains the maximum amount of variance, the second, the maximum amount of the remaining variance …

To visualize the data, we scale the eigenvector by the amplitude that it represents.

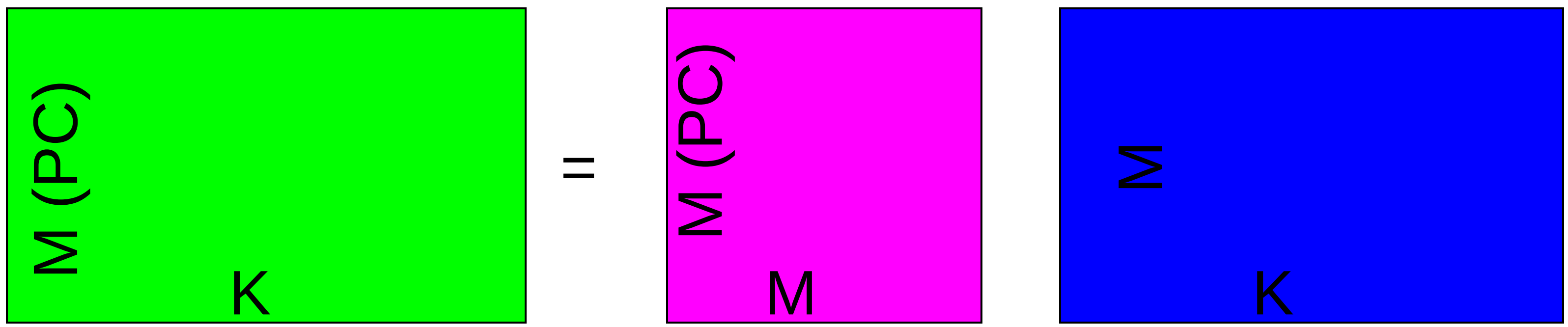




Notice that the eigenvectors that are clearly oscillatory come in pairs that represent the same frequency.

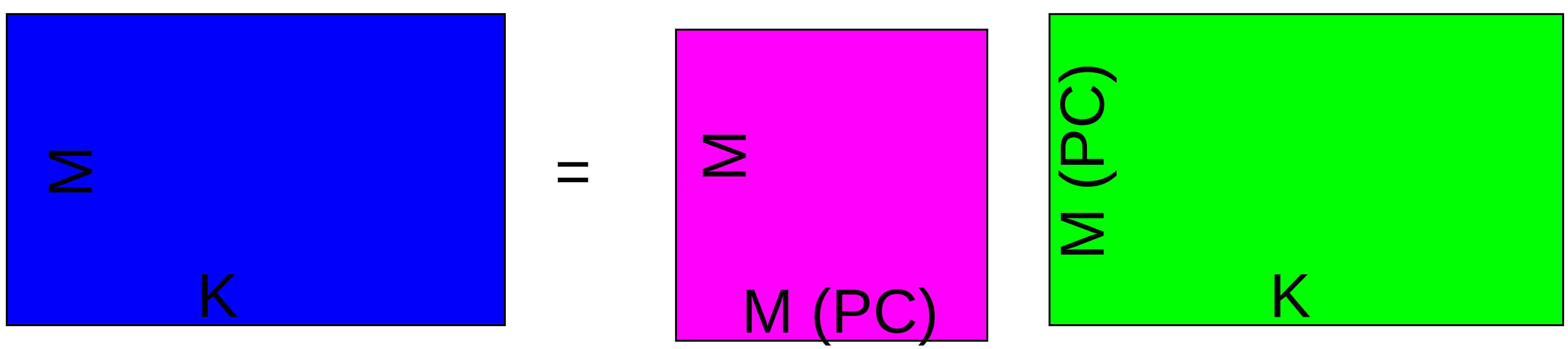
We define the PCs, just as in PCA:

$$PC = E^T X$$



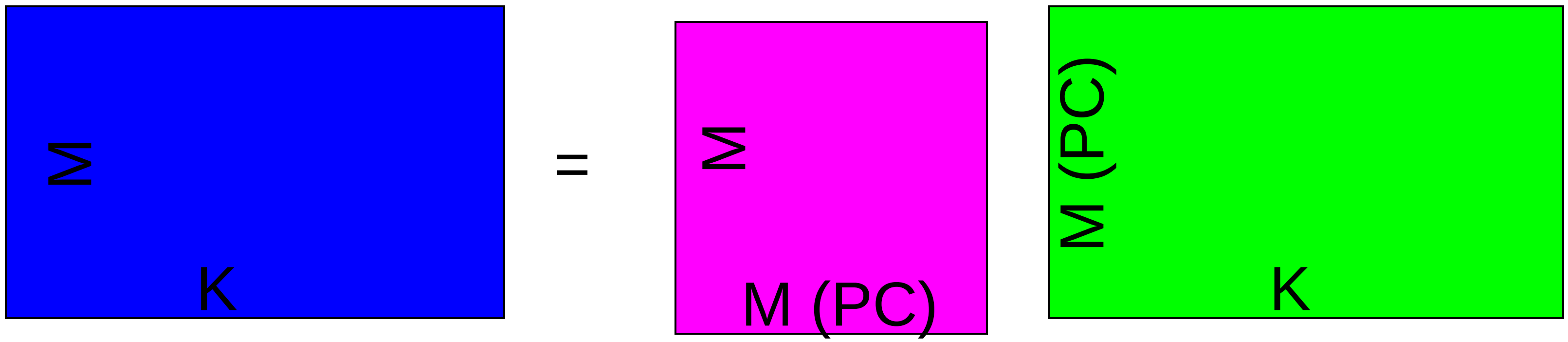
“Y contains the principal component scores, the amplitudes by which you multiply the eigenvectors to get the original data back.”

$$X = E\ PC$$



“Y contains the principal component scores, the amplitudes by which you multiply the eigenvectors to get the original data back.”

X = E PC



But this step is a bit tricky because remember that you want to reconstruct your original timeseries, but the one in X is truncated.

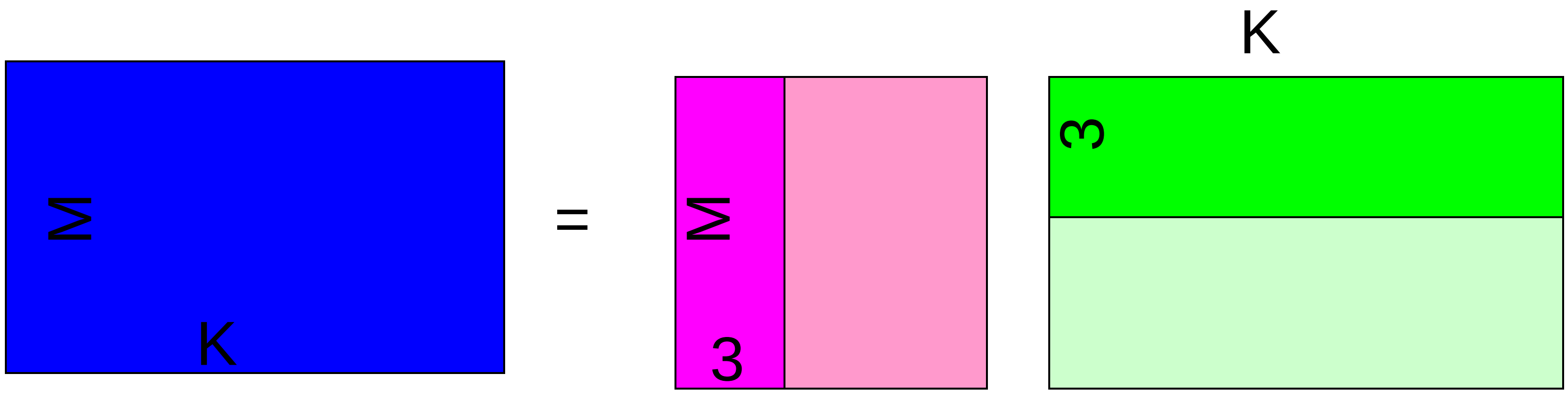
$$
\begin{aligned}
R_l^k(i) &= \frac{1}{i}\sum_{j=1}^{i} PC^k(i-j+1)E_l^k(j) \quad \text{for} \quad [1 \le i \le M-1] \\
&= \frac{1}{M}\sum_{j=1}^{M} PC^k(i-j+1)E_l^k(j) \quad \text{for} \quad [M \le i \le N-M+1] \qquad (\\
&= \frac{1}{N-i+1}\sum_{j=i-N+M}^{M} PC^k(i-j+1)E_l^k(j) \quad \text{for} \quad [N-M+2 \le i \le N]
\end{aligned}
$$

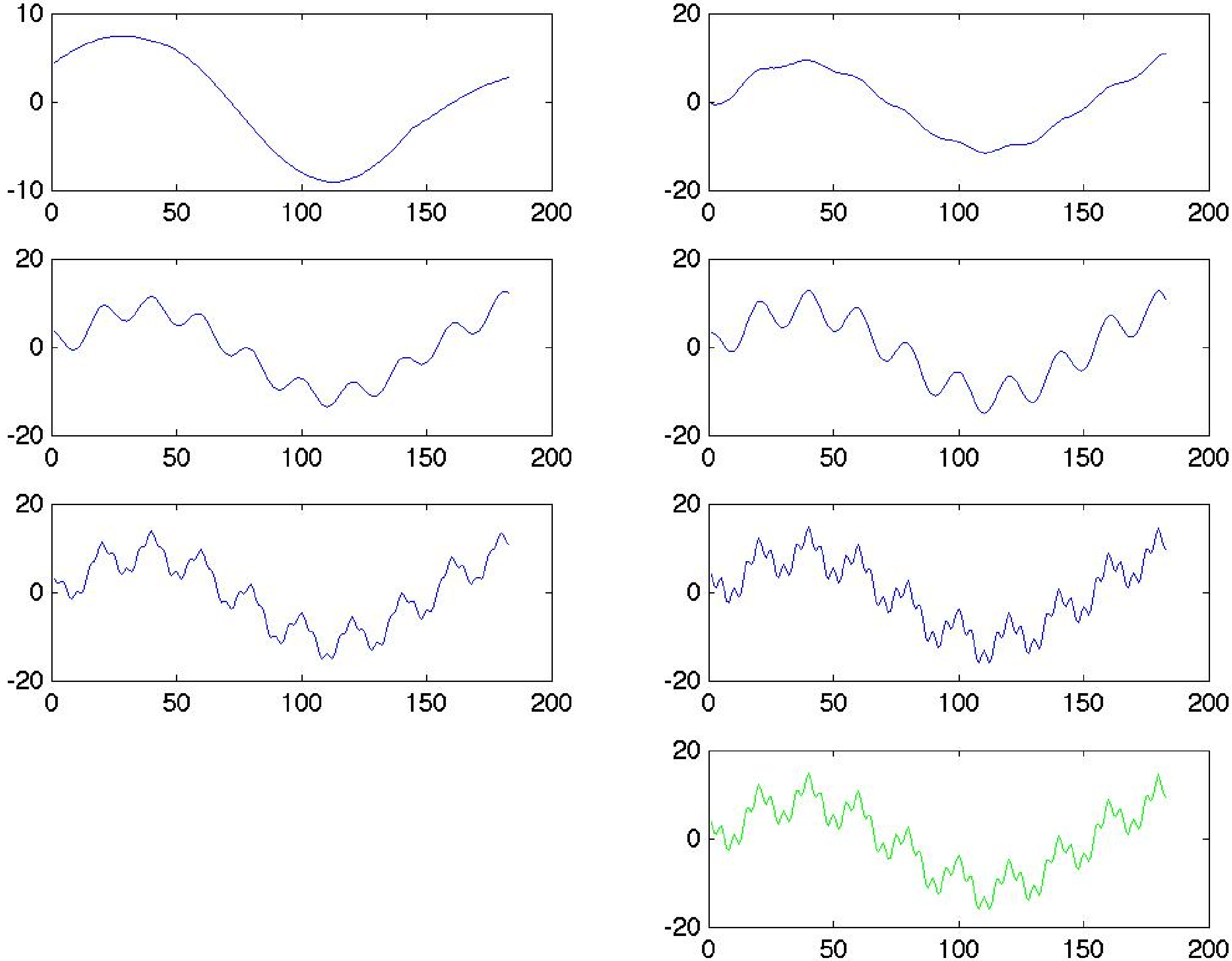
$$R_{\mathcal{K}}(t) = \frac{1}{M_t} \sum_{k \in \mathcal{K}} \sum_{j=L_t}^{U_t} A_k(t - j + 1)\rho_k(j); \qquad (11)$$

here $\mathcal{K}$ is the set of EOFs on which the reconstruction is based. The values of the normalization factor M_t, as well as of the lower and upper bound of summation L_t and U_t, differ between the central part of the time series and its end points [*Ghil and Vautard*, 1991; *Vautard et al.*, 1992]:

$$(M_t, L_t, U_t) = \begin{cases} \left(\frac{1}{t}, 1, t\right), & 1 \leq t \leq M - 1, \\ \left(\frac{1}{M}, 1, M\right), & M \leq t \leq N', \\ \left(\frac{1}{N - t + 1}, t - N + M, M\right), & N' + 1 \leq t \leq N. \end{cases} \qquad (12)$$

You can also partially reconstruct the timeseries using only the dominant PCs (let's say 3), this is VERY useful.




10
0
-10
0 50 100 150 200
20
0
-20
0 50 100 150 200
20
0
-20
0 50 100 150 200
20
0
-20
0 50 100 150 200
20
0
-20
0 50 100 150 200
20
0
-20
0 50 100 150 200
20
0
-20
0 50 100 150 200

```
for i=1:183
   U(i) = 10*sin(2*pi()*i/150)+2*cos(2*pi()*i/5)+ 4*cos(2*pi()*i/20);
end


% First, let's remove the mean
T = U - mean(U);


% The maximum lag will be 40 days
% Construct trayectory matrix
M = 40;
N = 183;
K = N-M+1;

for l=1:40
   X(l,:) = T(l:N-M+l);
end

% Calculate the covariance matrix
C = (1/(K-1)) * X * X';

% Perform Eigenanalysis
[VEC,VAL] = eig(C);

% Check variances and calculate the trace
VarTOT = trace(VAL);
check = trace(C);

% Look at the variances for first 10 PCs
Var = diag(real(VAL));
RelVar = real(Var)/VarTOT;
figure(1)
bar(RelVar(1:10))
saveas(gcf,'RelativeVar.jpg')

% Look at the eigenvectors
figure(2)
for i=1:6
   vecplot = VEC(:,i)*VAL(i,i)^0.5
subplot(4,2,i), plot(vecplot)
end
saveas(gcf,'eigenvectors.jpg')
```

```
% Calculate the PCs
PC = VEC' * X;


%Calculate reconstructed components
RC = zeros(N,K);
for k = 1:M
   k
   for i = 1:M-1
      xi = 0.;
      for j = 1:i
         xi=xi + PC(k,i-j+1)*VEC(j,k);
      end
      RC(i,k) = xi/i;
   end
   for i = M:K

      xi = 0.;
      for j = 1:M

         xi=xi + PC(k,i-j+1)*VEC(j,k);
      end;
      RC(i,k) = xi/M;
   end
   for i = K+1:N
      xi = 0.;
      for j = i-N+M:M
         xi=xi + PC(k,i-j+1)*VEC(j,k);
      end;
      RC(i,k) = xi/(N-i+1);
   end
end

figure(3)
for i=1:6
   RCtot = sum(RC(:,1:i),2);
   subplot(4,2,i), plot(RCtot)
end
subplot(4,2,8), plot(T,'g')
saveas(gcf,'Reconstruction.jpg')
```

http://www.atmos.ucla.edu/tcd/ssa/

Home News Preprints Presentations Research People Links

SSA-MTM Toolkit for Spectral Analysis

Latest news: SSA-MTM Toolkit goes 64-bit! HOT!

Presentations

What it is / What it does

Availability

Grant support

SSA Wikipedia

Documentation

Computer Lab Exercises from Summer School on Climate Variability and Climate Change

SSA Gap-Filling paper!

Review Paper in Rev. Geophys.!

- **User's Guide**
  - Latest version (version 4.4)

Frequently asked questions

Bibliographic references to SSA and MTM

Who we are

Other info

M-SSA

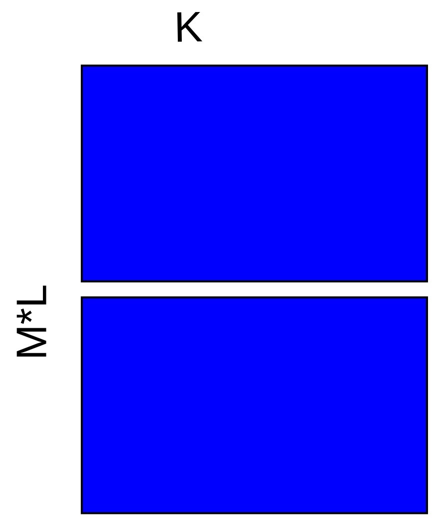
M-SSA is the multivariate extension to SSA. You are essentially extracting common trends and oscillatory patterns from a group of variables.

Mathematically it is very similar to SSA, now you have L variables (in this case L=2).

K columns (K=N-M+1)

$$X = \begin{pmatrix} T_1 & T_2 \ldots & T_K \\ T_2 & T_3 \ldots & T_{K+1} \\ & & \\ T_M & T_{K+1} \ldots & T_N \\ S_1 & S_2 \ldots & S_K \\ S_2 & S_3 \ldots & S_{K+1} \\ & & \\ S_M & S_{K+1} \ldots & S_N \end{pmatrix}$$

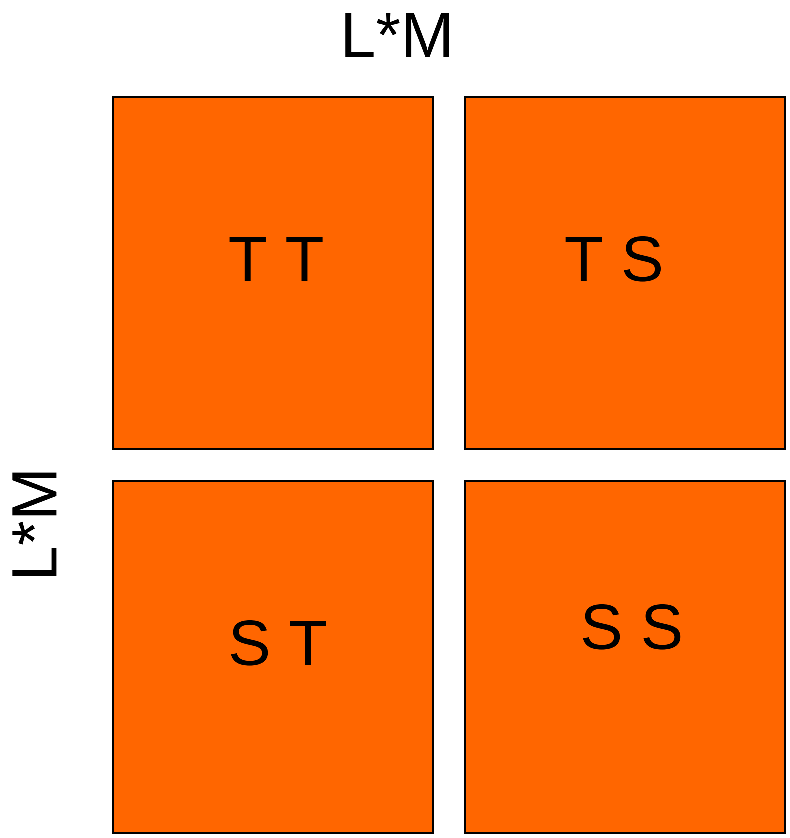
M * L rows (window length)



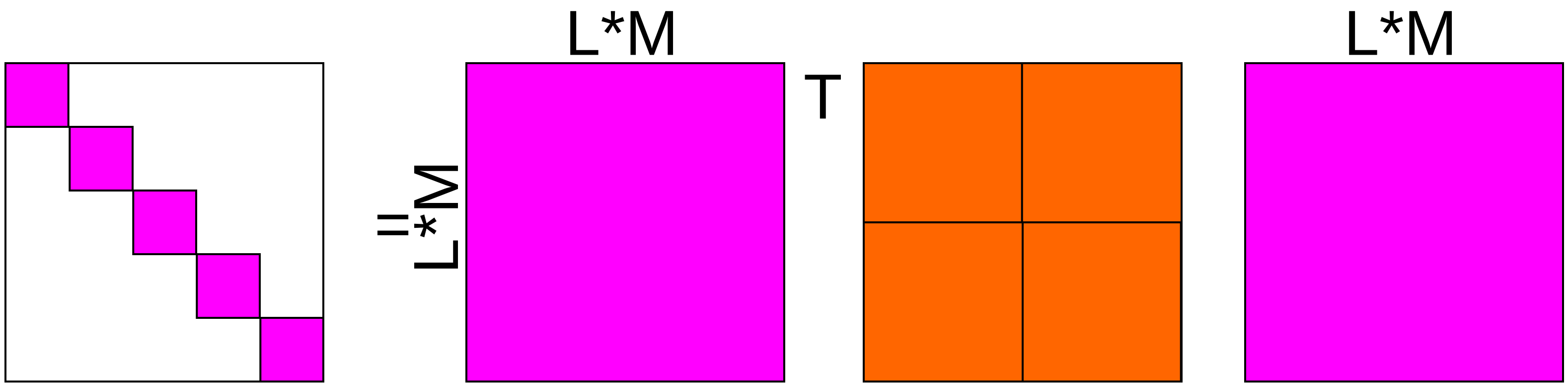
If you are dealing with different variables (say winds and precipitation) you must normalize the data before doing any analysis. That way you are comparing apples to apples.

The covariance matrix now has dimensions L*M x L*M, and it represents the autocovariance and cross-covariance terms

C =

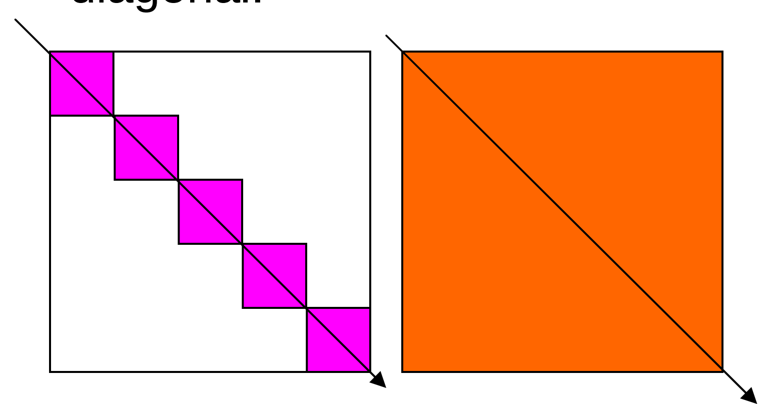


We perform eigenanalysis on the covariance matrix. You now have L*M eigenvectors and eigenvalues



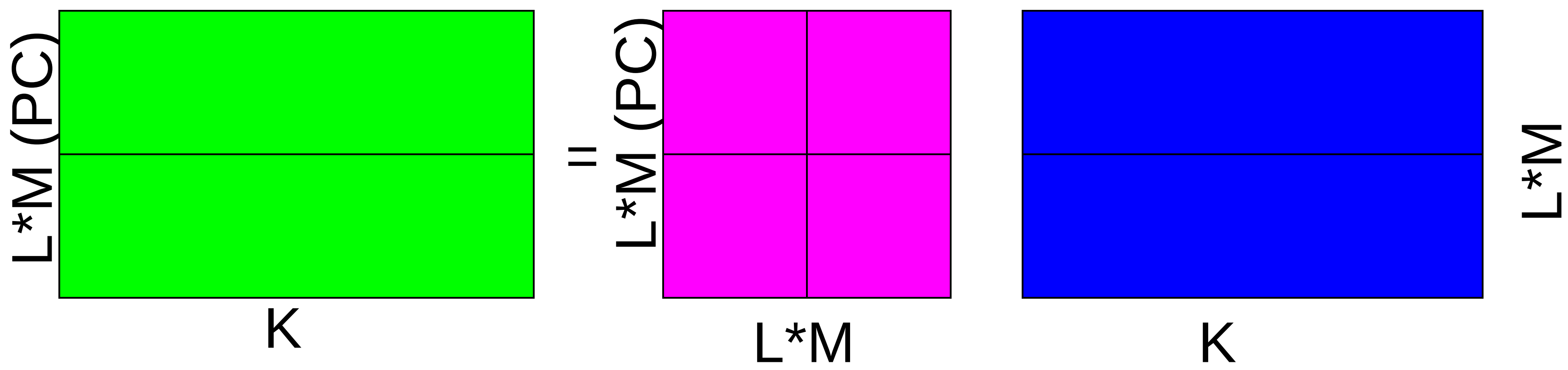
$$\Lambda = E^T C E = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_{L*M} \end{pmatrix}$$

The set of eigenvectors e_i and eigenvalues λ_i represent a coordinate transformation where C becomes a diagonal.



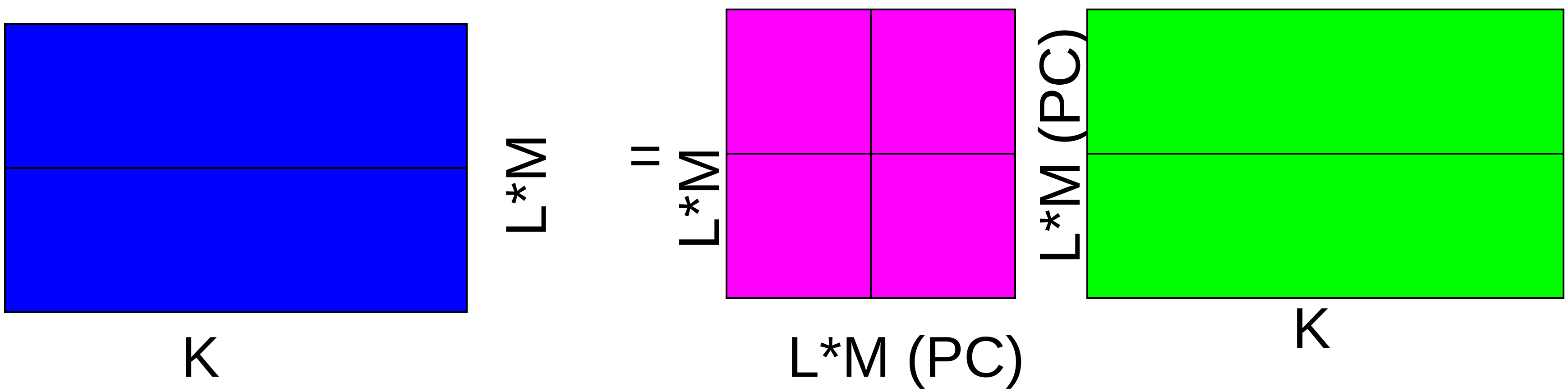
The PCs will have dimensions L*M x K

$$PC = E^T X$$



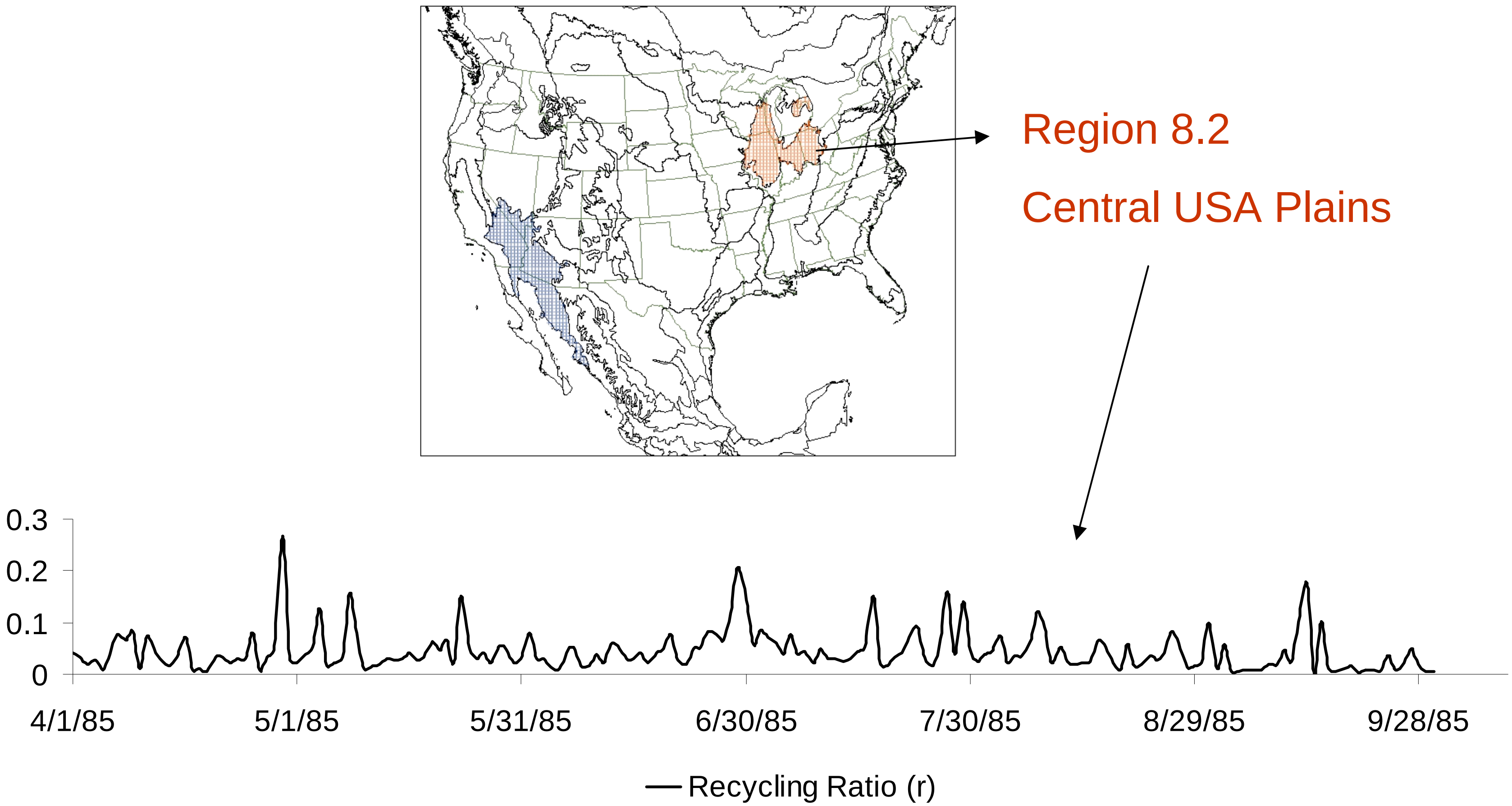
The reconstruction is essentially the same, except that the messy loop will have an outer “if” statement representing the variable (loop not shown).

X = E PC



Application of SSA and M-SSA to land-atmosphere Interactions

Precipitation recycling is the contribution of local evapotranspiration to precipitation events.



The time series of r indicates the daily fraction of precipitation that came from evaporative origin.

Our objective is to understand what causes the spatio-temporal variability of precipitation recycling.

$$R = 1 - \exp\left[-\int_0^{\tau} \frac{\varepsilon}{\omega} d\tau' \right]$$

Inherent to the model

Depends on the size of the region and the velocity of the winds.

Ratio of Evaporation over total precipitable water. E/w

Not explicitly included in model formulation

Precipitation

Temperature

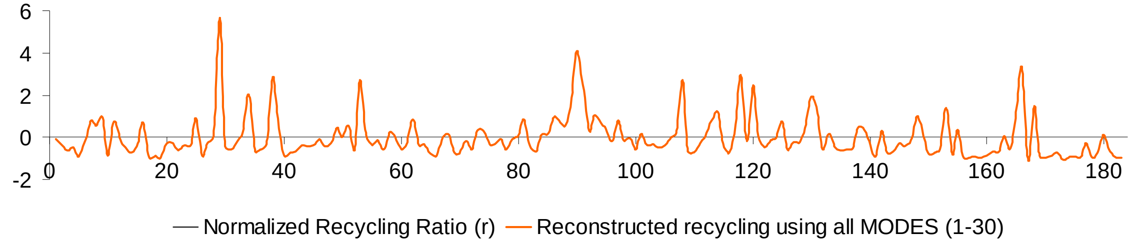
Sensible Heat Flux

Atmospheric Humidity

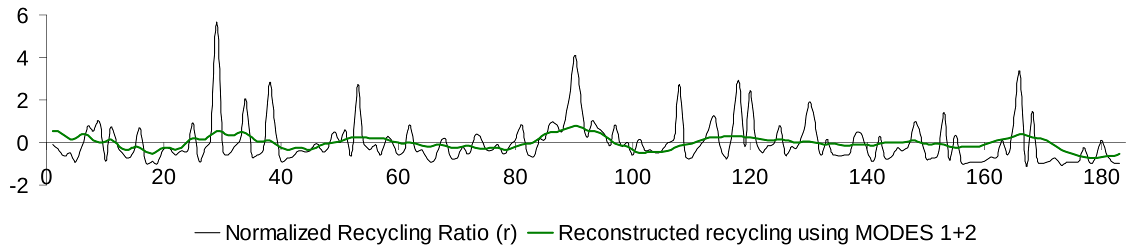
We begin with SSA:

- You can use EOF analysis to look at the structure of a timeseries by doing and eigenanalysis of the lagged covariance matrix. (Ghil et al. 2002)

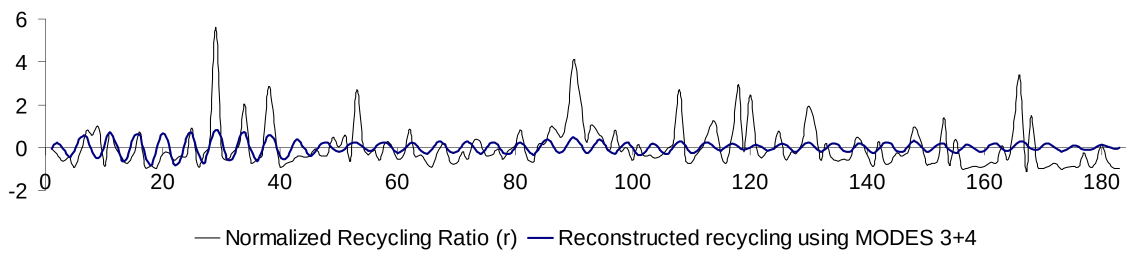
The time series of recycling ratio can provide useful information about the physical process that produced it. SSA separates the time series into components that are statistically independent at zero lag. It will provide information about trends, oscillatory patterns and noise.

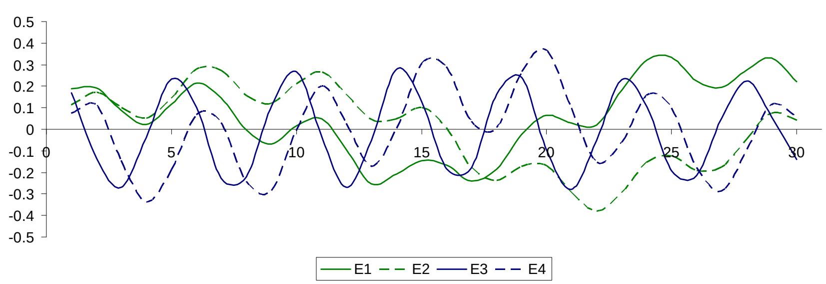


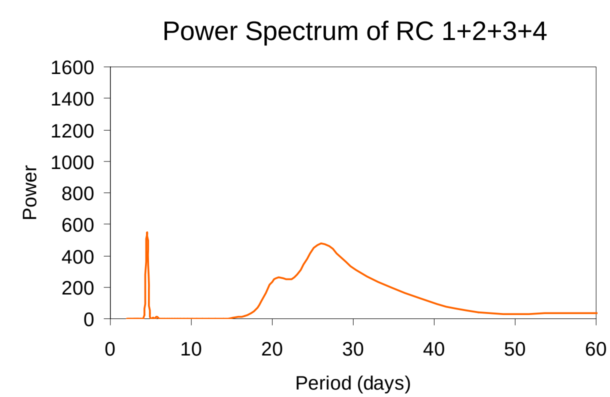
Region 8.2



The SSA of the recycling ratio shows the dominance of two periodicities, one of peak 26 days, and the other of peak 4.6 days. The combination of the first four modes accounts for 25% of the variability in the data.

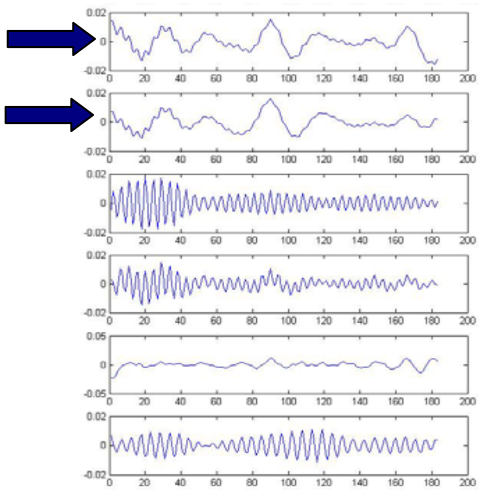




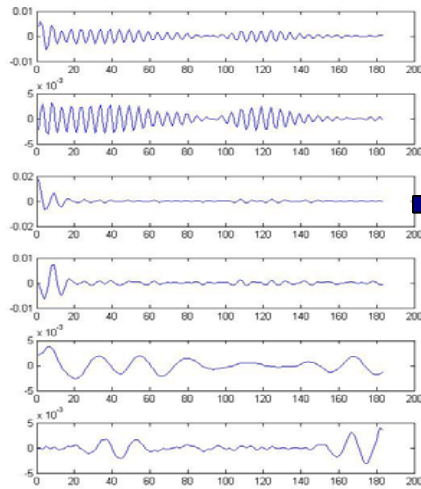


Similarly, we can perform SSA on other variables.

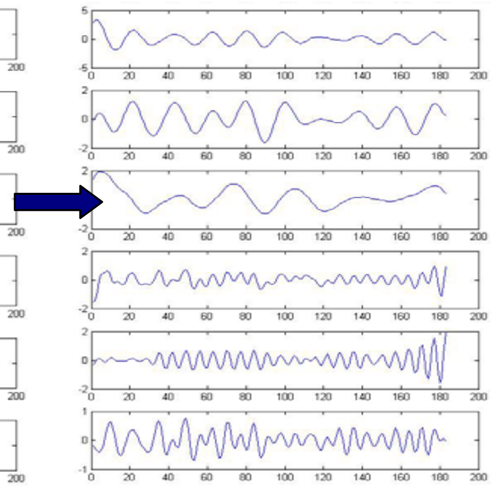
Recycling Ratio



E/w

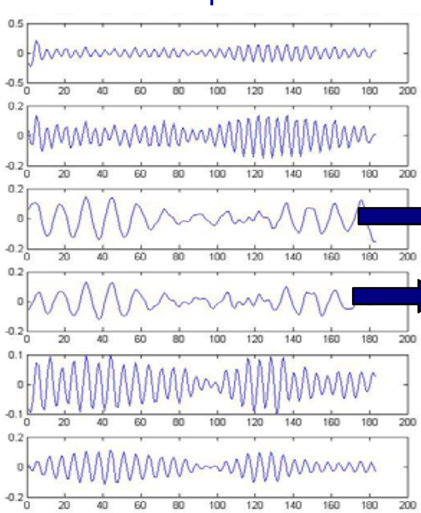


Winds

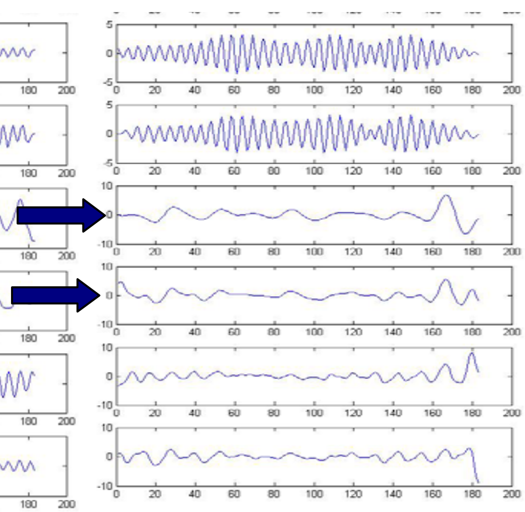


Approximately a 26 day periodicity

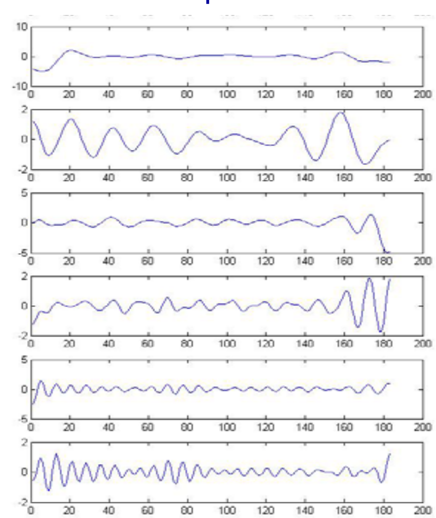
Precipitation



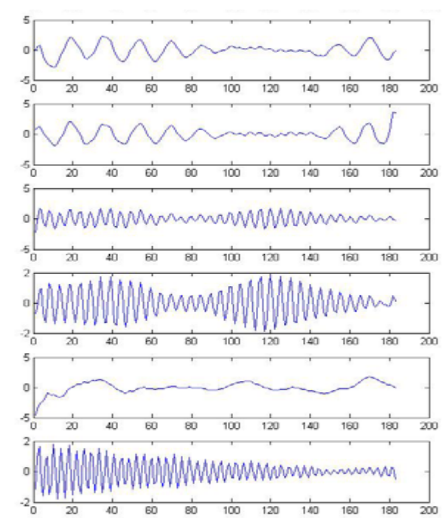
Sensible Heat Flux



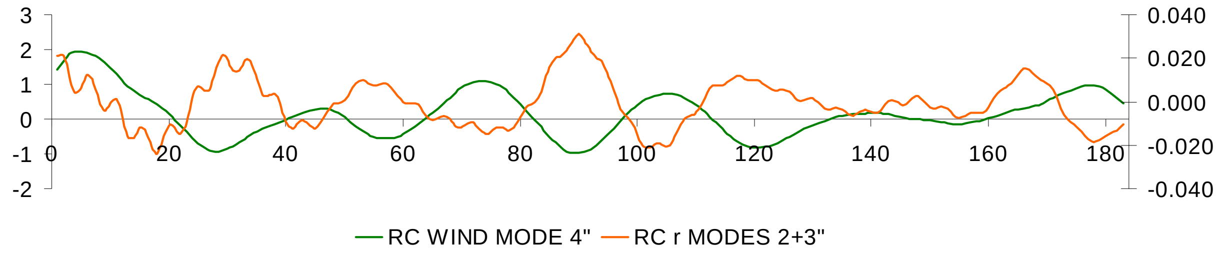
Temperature

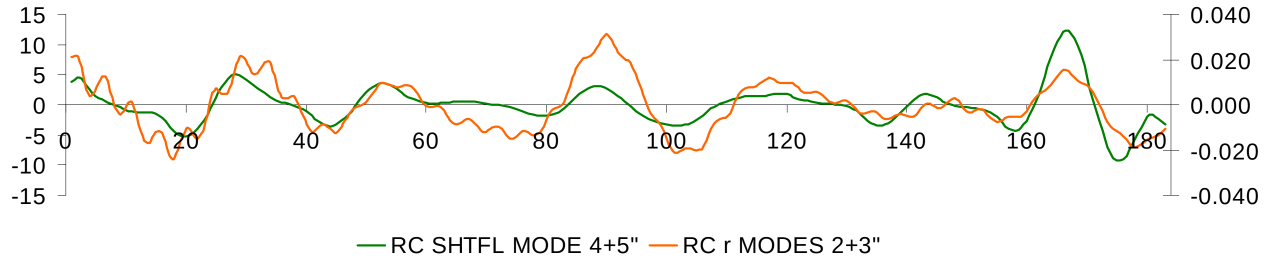


Atmospheric Humidity (Modified Dew Point Depression)

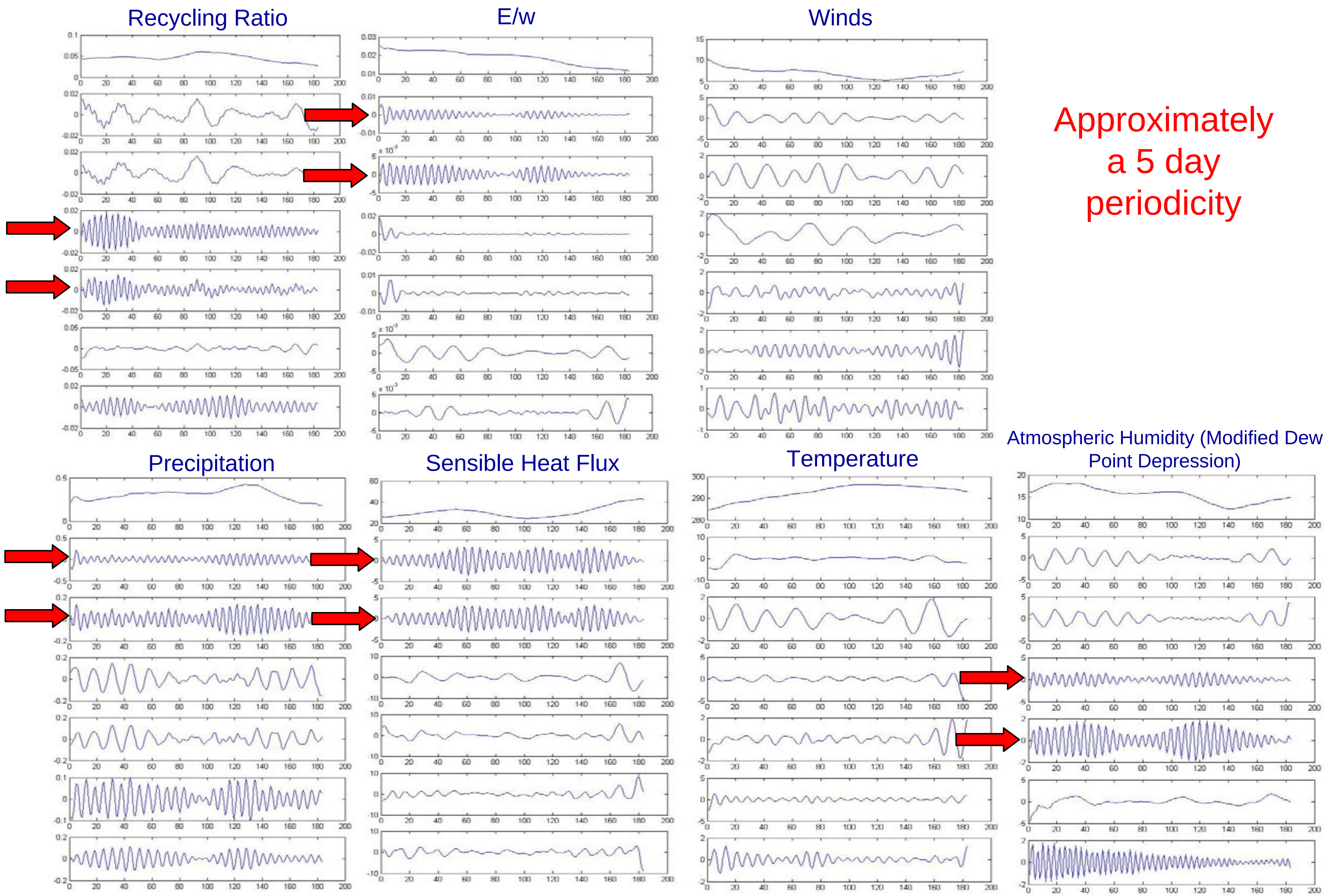


Looking only at the twenty six day periodicity. Recycling is out of phase with the winds, and in phase with the sensible heat flux.

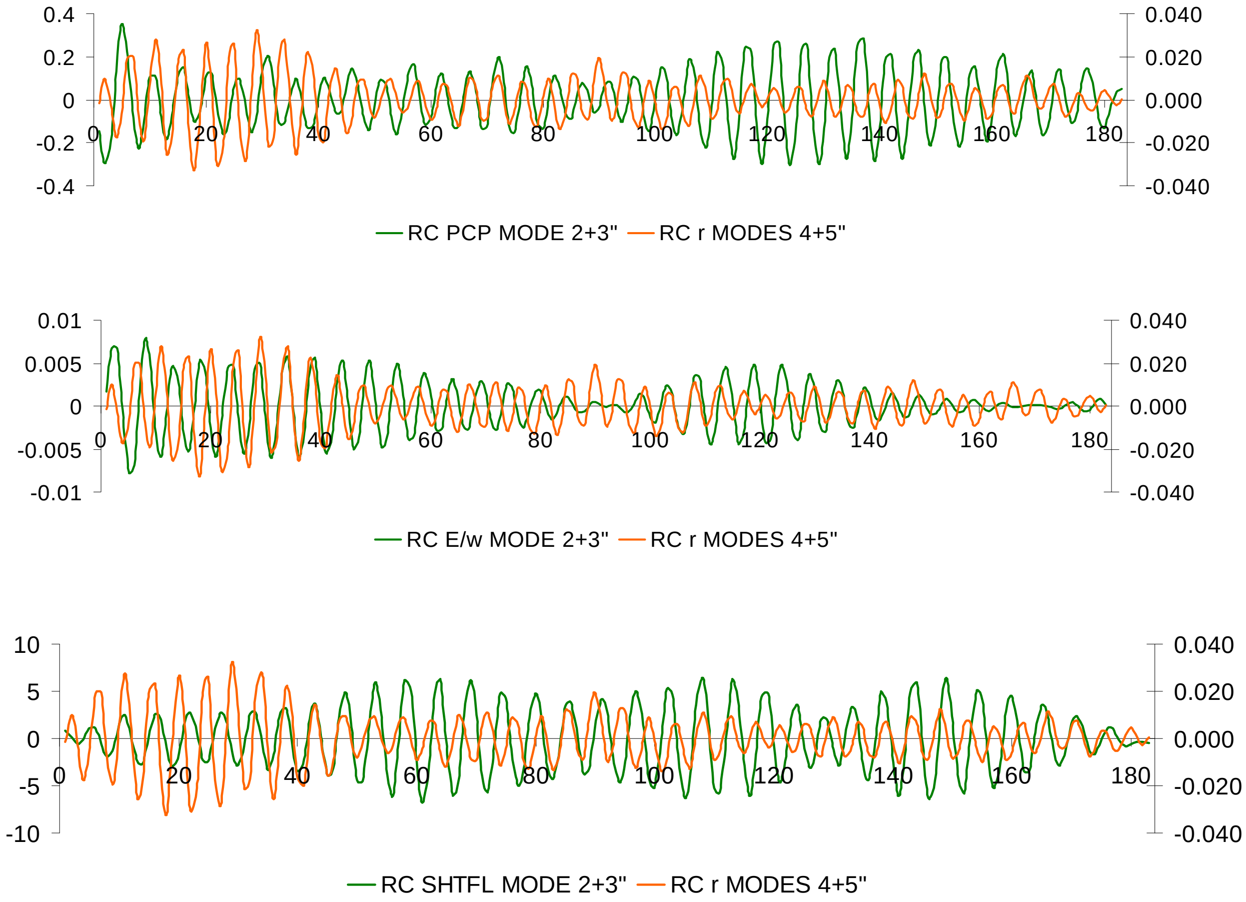




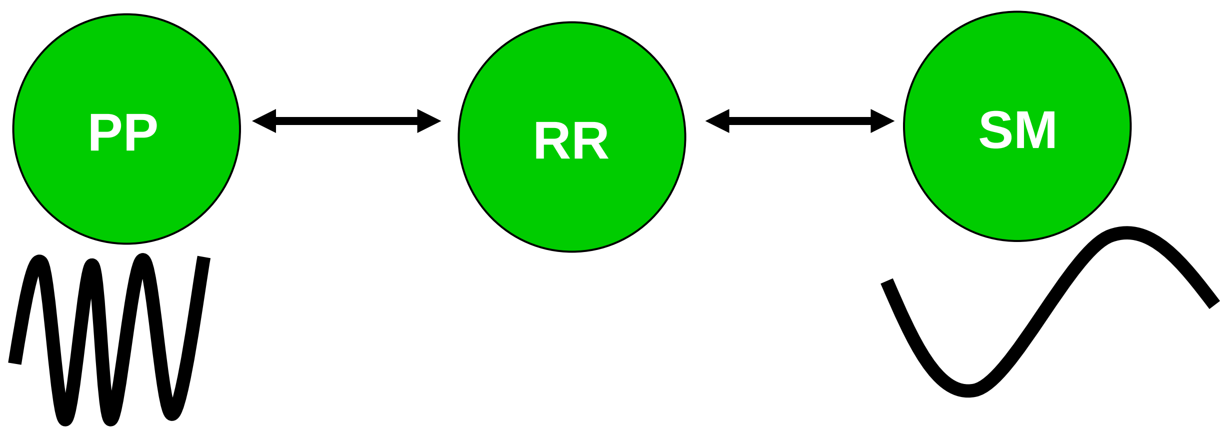
Similarly, we can perform SSA on other variables.



Looking only at the five day periodicity. The in-phase or out of phase relations are difficult to clearly establish.

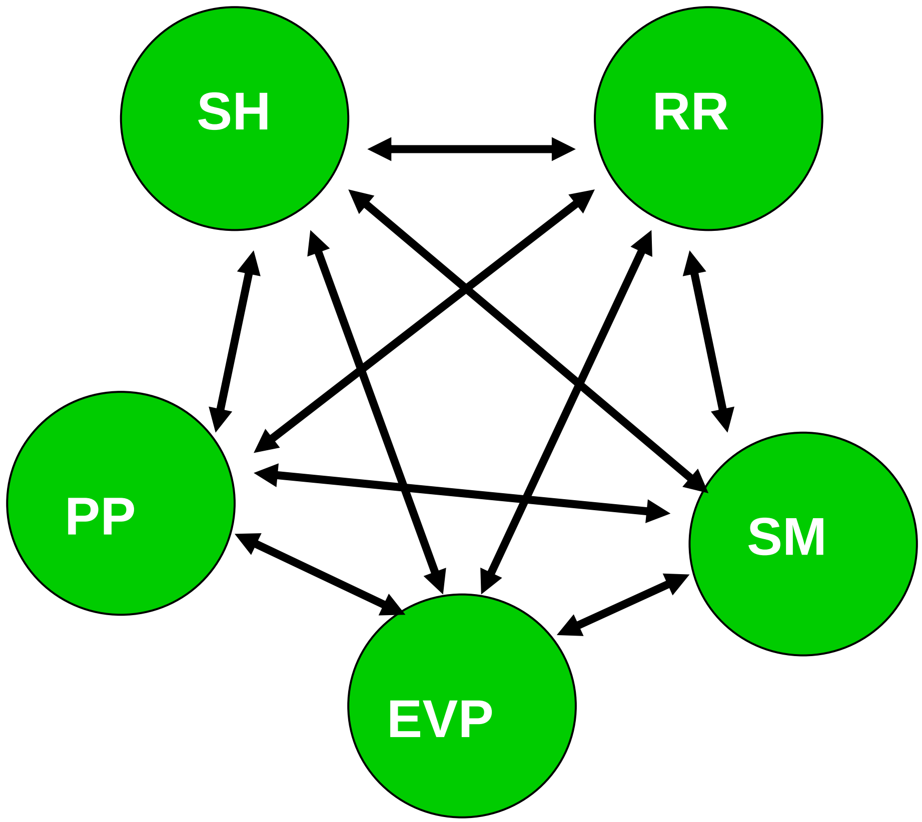


We need a different methodology if we want to find relations among variables. Multivariate singular spectrum analysis (M-SSA) enables us to go the extra step.



We need a methodology to account for the different timescales in land-atmosphere processes.

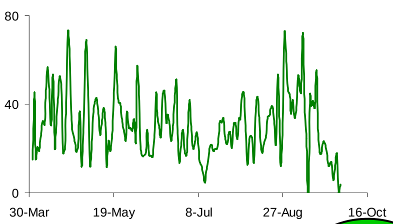
Multivariate Singular Spectrum Analysis allows us to account for the different timescales of the system.

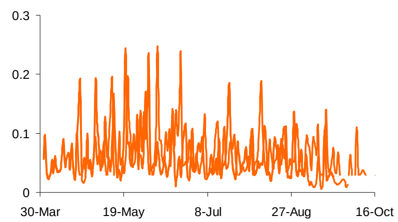


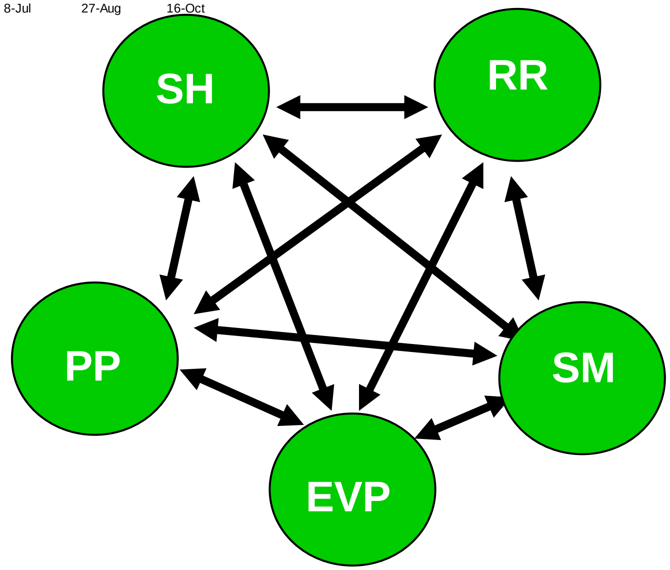
18 Variables

Variable	Notation
Precipitation	PP
Convective Precipitation	PC
Convective Available Potential Energy	CA
Precipitable Water	PW
Cloud Base Height	HC
Zonal Moisture Flux	QU
Meridional Moisture Flux	QV
Moisture Flux Divergence	DV
u	Uw
v	Vw
Wind	WW
Potential Evapotranspiration	PE
Humidity Index	HI
Evapotranspiration	ET
Sensible Heat	SH
Surface Air Temperature	TM
Soil Moisture (0-200 cm)	SM
Recycling Ratio	RR

The purpose of M-SSA is to maximize the joint variance of all the variables.





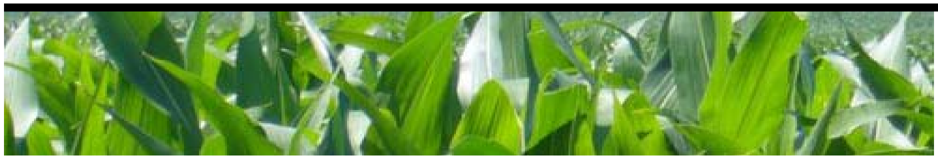


$$X = \begin{pmatrix} RR_1 & RR_2 & & RR_M & SH_1 & SH_2 & & SH_M & \vdots & \vdots & \vdots \\ RR_2 & RR_3 & & RR_{M+1} & SH_2 & SH_3 & & RR_{M+1} & & & \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & & & \\ RR_K & RR_{K+1} & & RR_{K+M} & SH_K & SH_{K+1} & & SH_{K+M} & & & \end{pmatrix}$$

$$C = X^T X$$

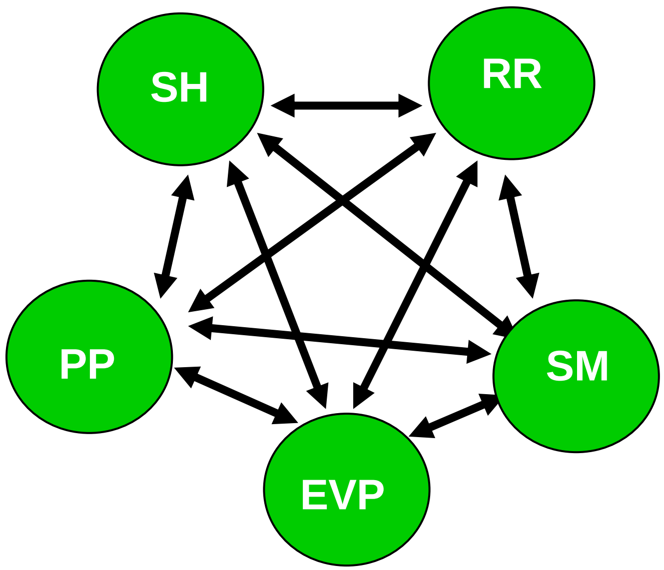
Lagged autocorrelation and cross correlation matrix

In a way that accounts for both their autocorrelation and the correlation between different variables.



Midwestern United States

The purpose of M-SSA is to maximize the joint variance of all the variables.



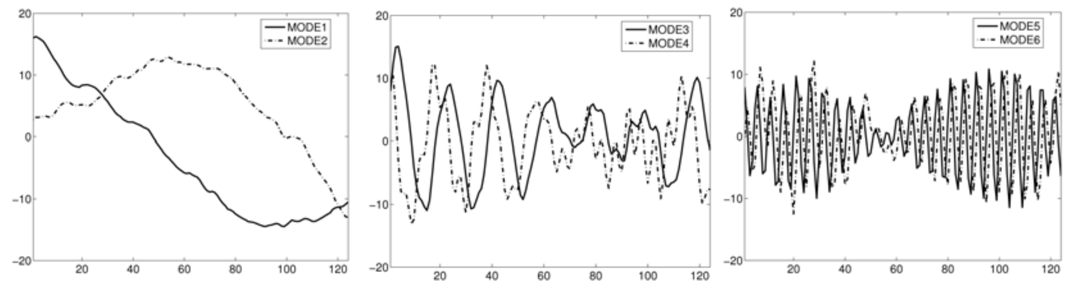
C = (Lagged autocorrelation and cross correlation matrix)

$$\Lambda = E^T C E$$

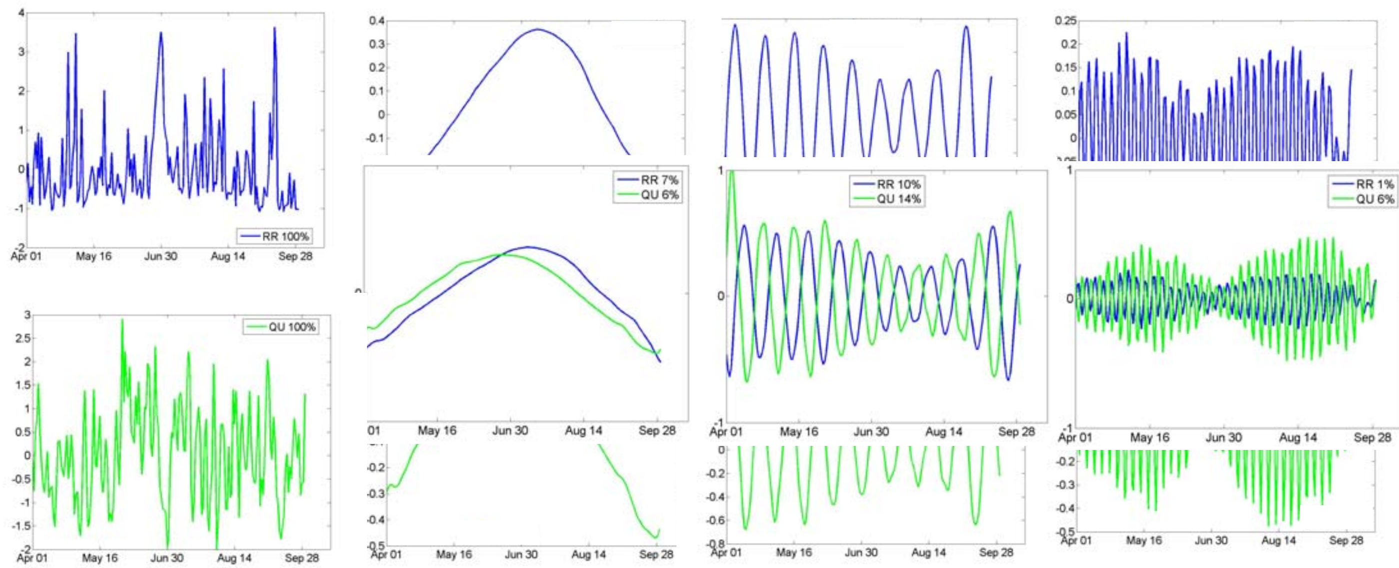
An Eigenanalysis of C, produces temporal structures that explain the maximum possible amount of the temporal autocorrelation and cross correlation

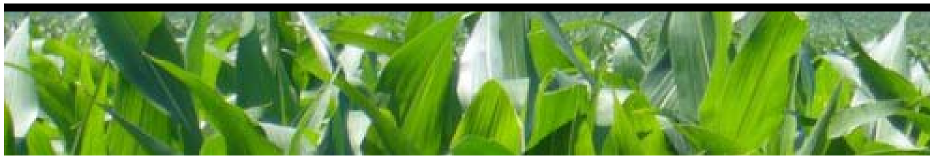
M-SSA can extract trends, oscillations and noise in a group of variables.

Extract the Modes that account for the maximum joint variance.



Extract the reconstructed components for each variable.



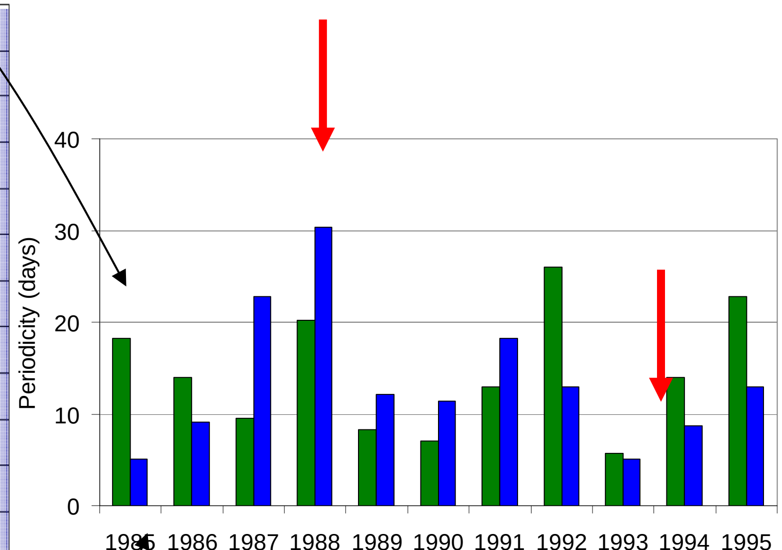


Midwestern United States

For each of the 11 years we extract the two dominant oscillatory components.

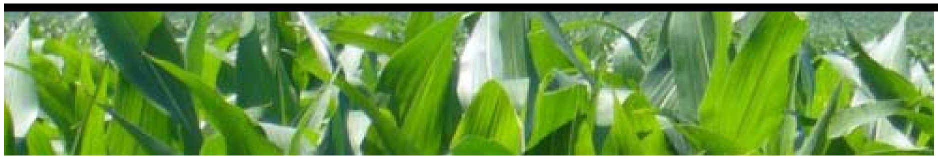
Year	Oscillatory Component 1				Oscillatory Component 2			
	Mode	Mode	Periodicity	% variance	Mode	Mode	Periodicity	% variance
1985	3	4	18.20	7.5	5	6	5.06	7.1
1986	4	5	14.00	6.1	7	8	9.10	5.5
1987	5	6	9.58	6.6	7	8	22.75	4.7
1988	5	6	20.22	5.5	7	8	30.33	4.9
1989	5	6	8.27	5.5	9	10	12.13	4.5
1990	3	4	7.00	5.6	5	6	11.38	5.3
1991	4	5	13.00	7.2	6	7	18.20	6.2
1992	5	6	26.00	6.2	7	8	13.00	5.3
1993	4	5	5.69	5.2	7	8	5.06	4.2
1994	4	5	14.00	6.6	7	8	8.67	5.1
1995	5	6	22.75	5.5	7	8	13.00	4.9

Component 1



Component 2

1988 (drought) has the longest periodicities and 1993 (flood) has the shortest.

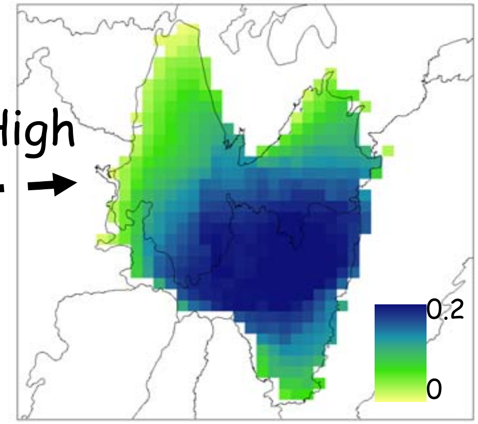


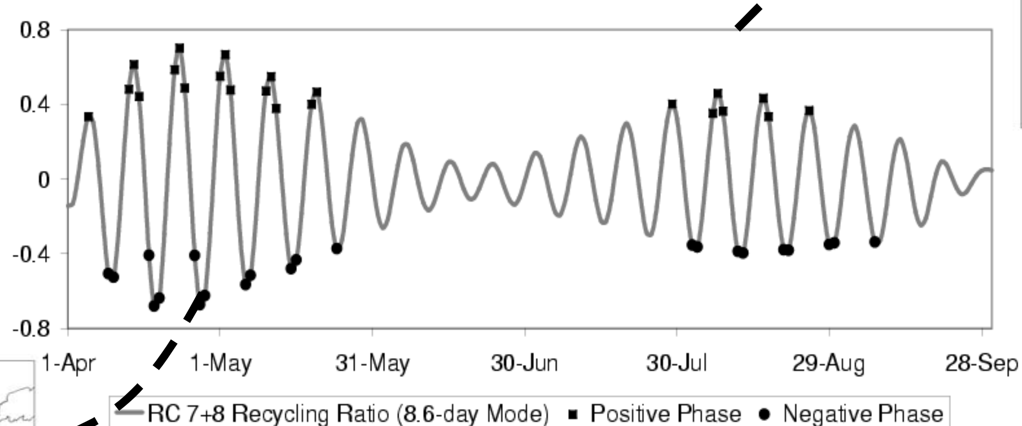
Midwestern
United States

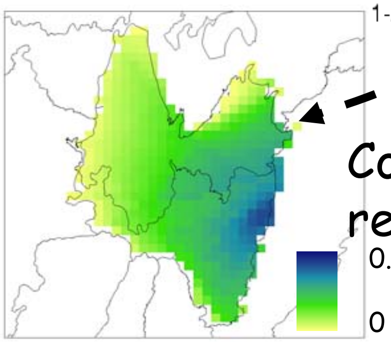
To understand what these modes look like in the "real world" we can visualize the representative days.

Year 1994 Modes 7+8

Composite of High recycling days



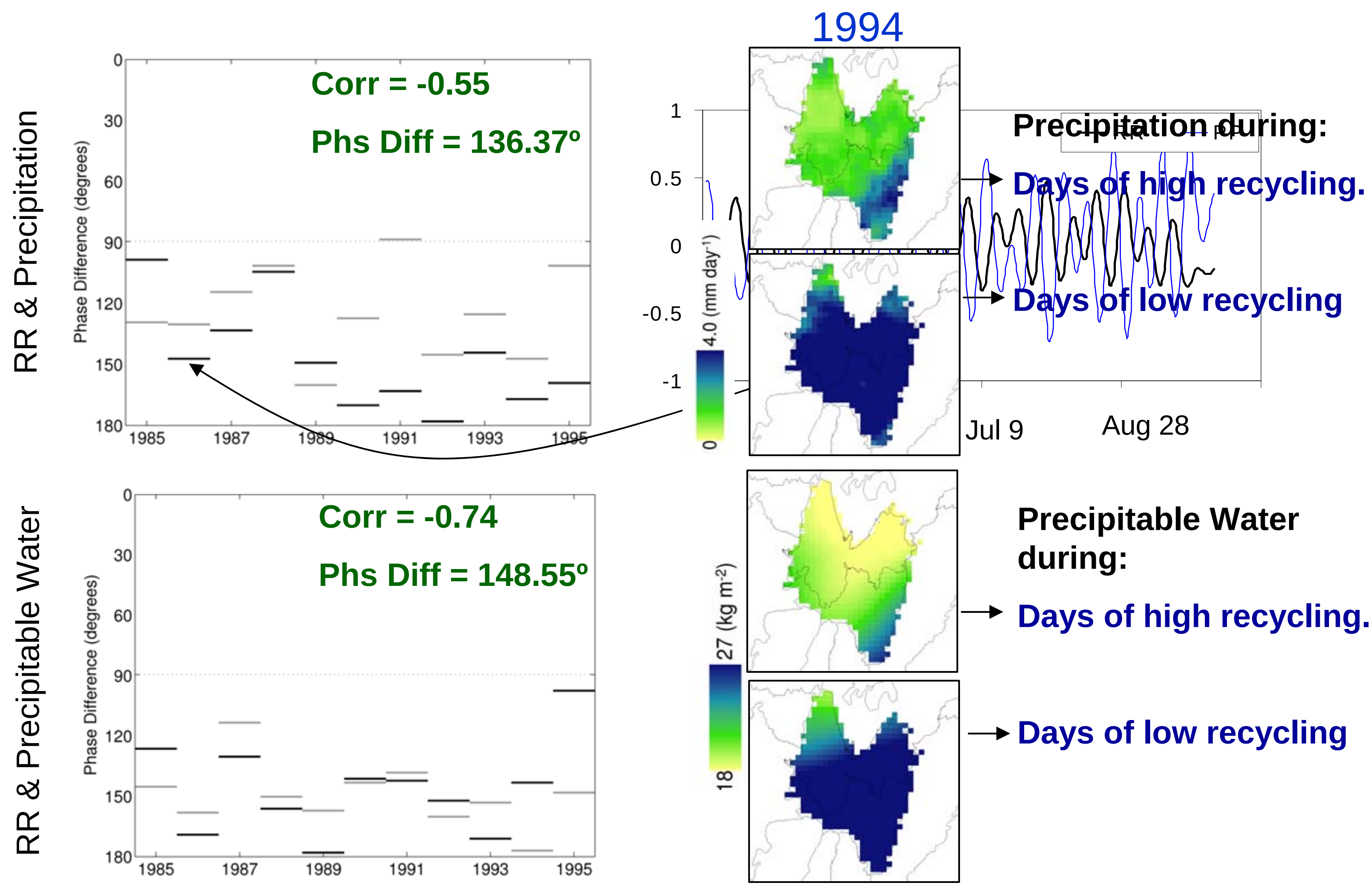


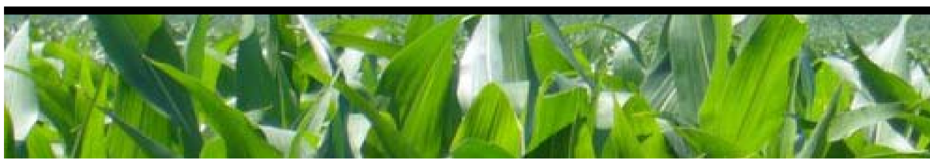


Composite of Low recycling days

Midwestern United States

At short timescales the recycling ratio is enhanced when precipitation and precipitable water are low.

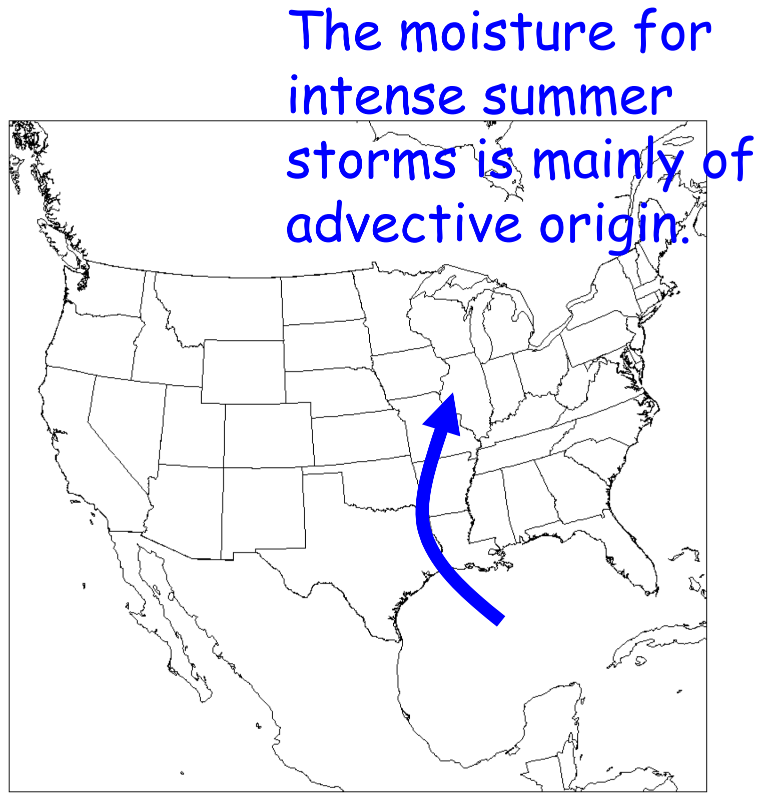


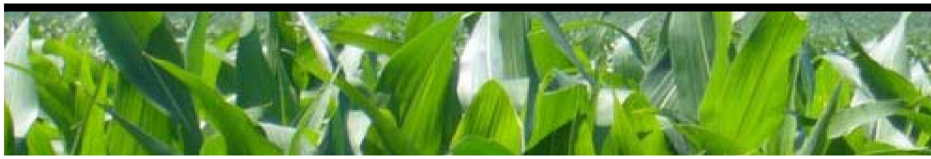


Midwestern United States

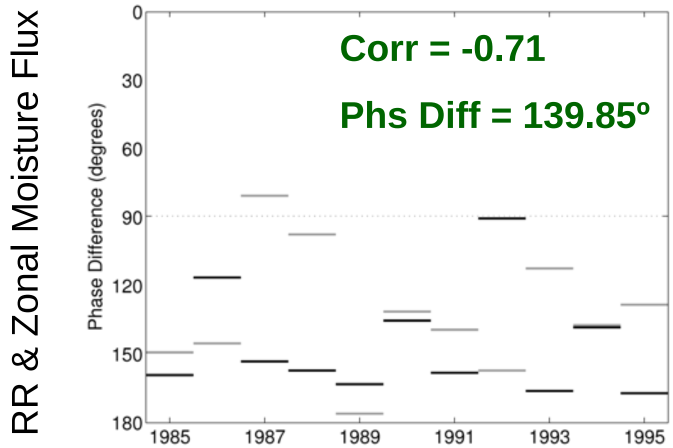
At short timescales the recycling ratio is enhanced when precipitation and precipitable water are low.

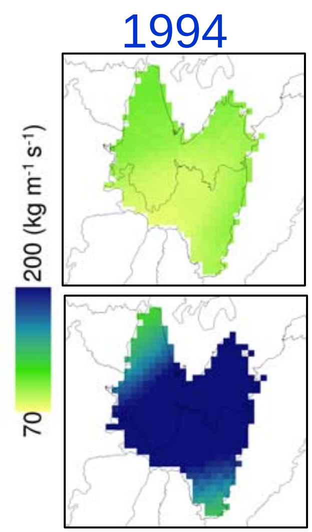
Reason:
Recycling only becomes important when advected precipitation decreases.





At short timescales the recycling ratio is enhanced when winds and moisture fluxes are low.

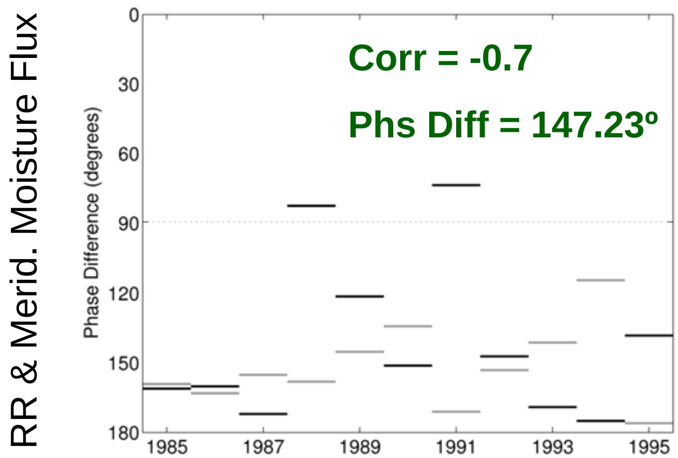


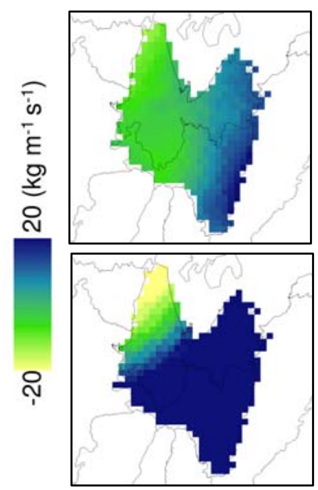


Zonal Moisture Flux during:

→ Days of high recycling.

→ Days of low recycling





Meridional Moisture Flux during:

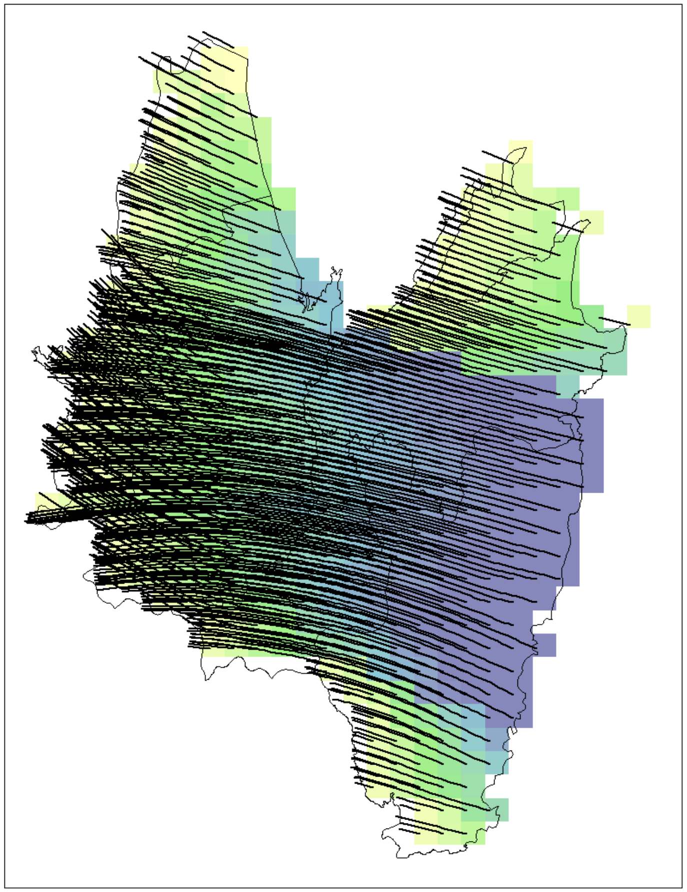
→ Days of high recycling.

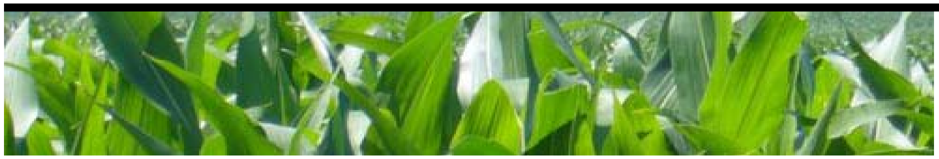
→ Days of low recycling

At short timescales the recycling ratio is enhanced when winds and moisture fluxes are low.

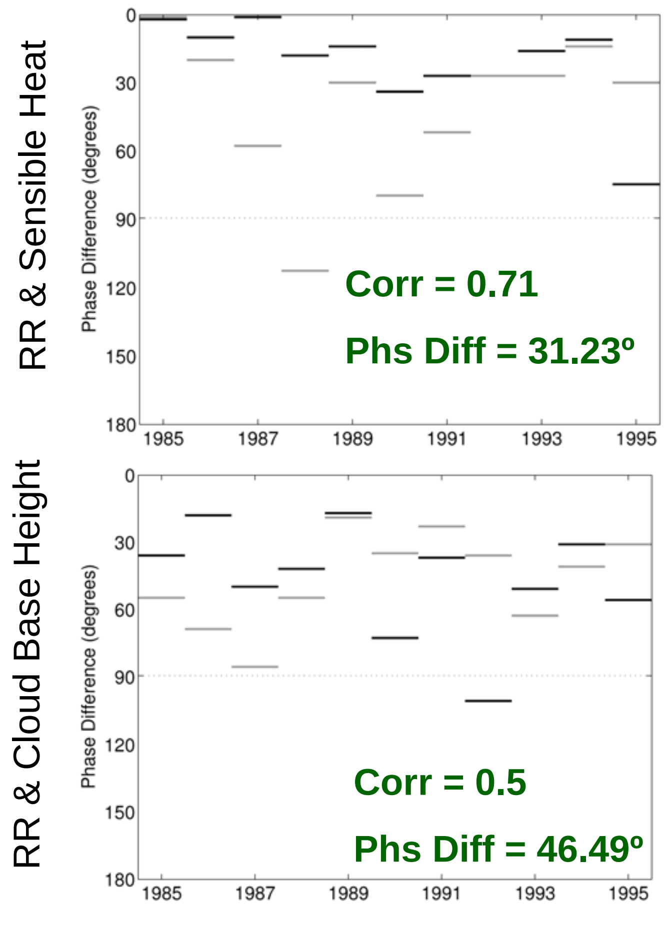
Reason:

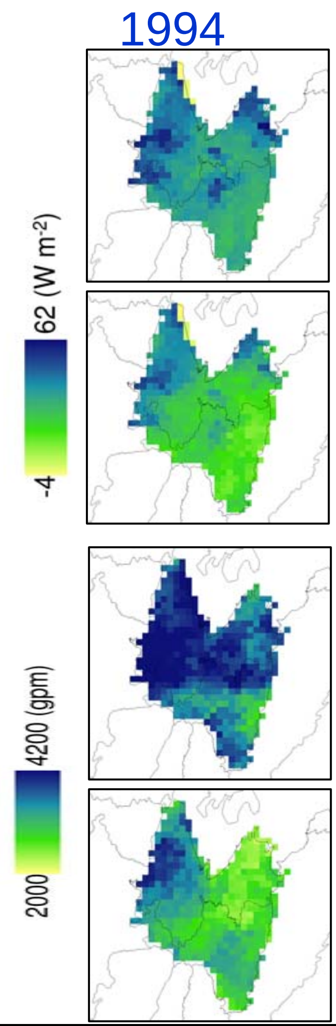
The air has more time to traverse the region and pick up evapotranspired moisture.





At short timescales the recycling ratio is enhanced when sensible heat and cloud base height are high.





Zonal Moisture Flux during:

→ Days of high recycling.

→ Days of low recycling

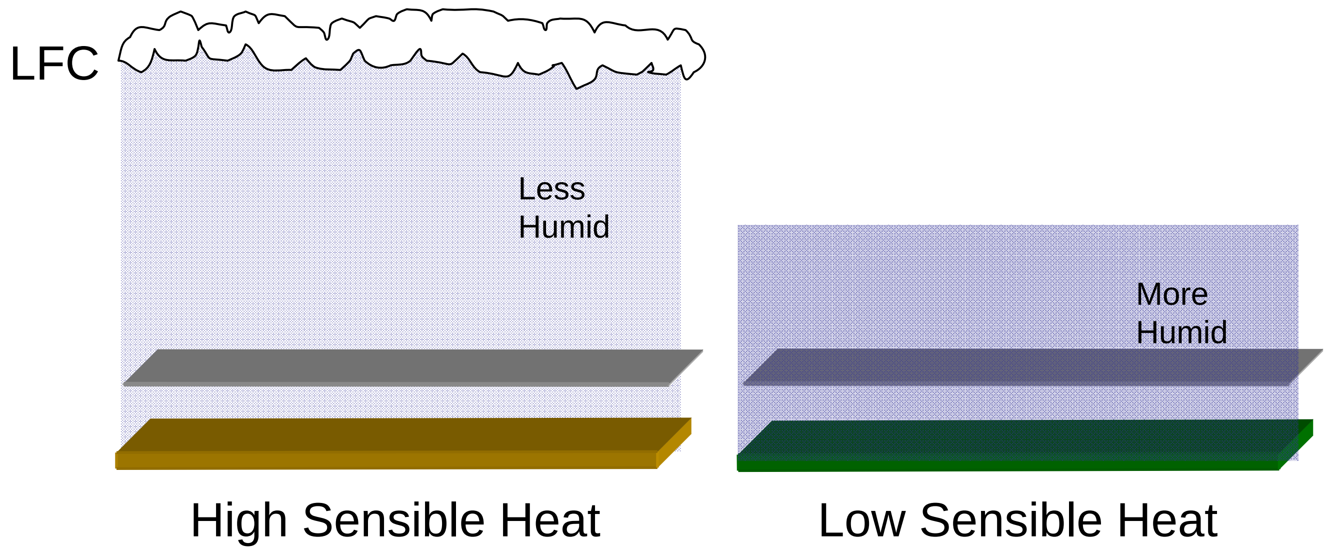
Meridional Moisture Flux during:

→ Days of high recycling.

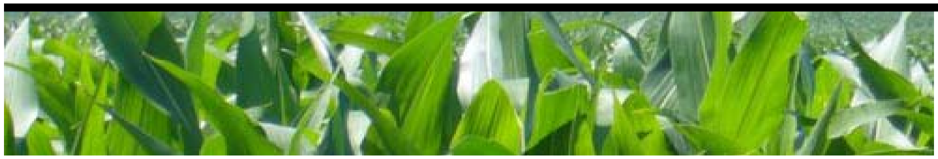
→ Days of low recycling

Why? Because the dry soil will have a much faster growth of the boundary layer.

Reason:
Sensible heat flux increases BL depth and promotes convective cloud formation.



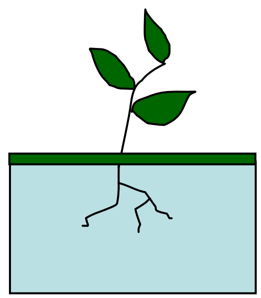
In some cases, the atmosphere over dry soil (high sensible heat) will reach the level of free convection before the moist soil.



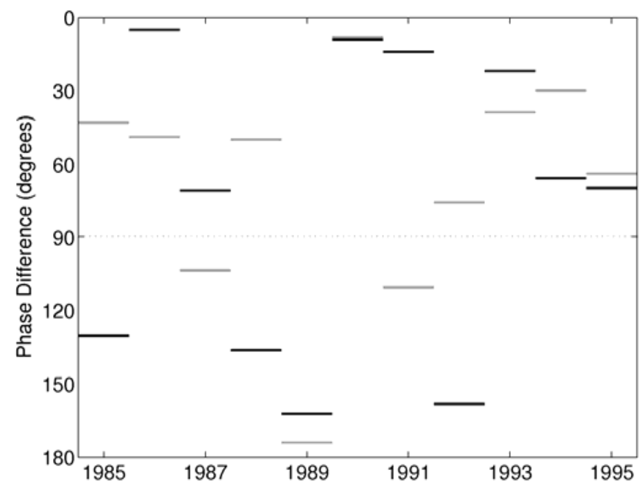
Contrary to what one might expect, at timescales < 40 days evapotranspiration variability is NOT related to recycling variability .

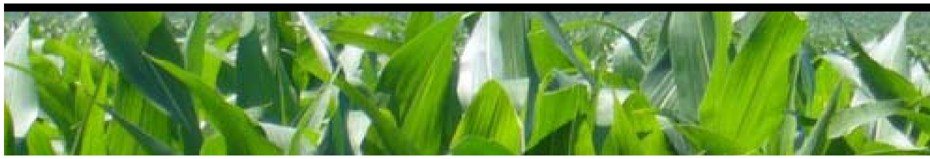
In non-drought years : In this moisture abundant region, soil moisture anomalies will not stress vegetation

➡ Evapotranspi-ration will not be affected.



➡ Recycling will not be affected.

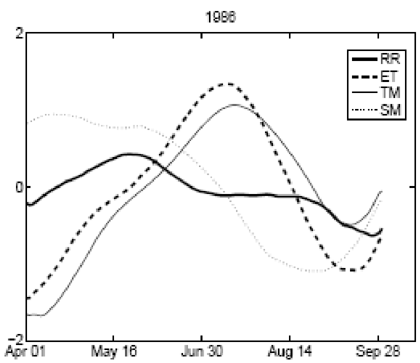




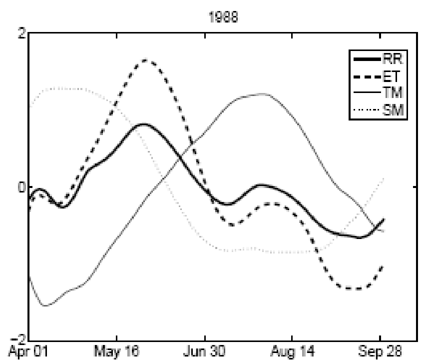
Midwestern United States

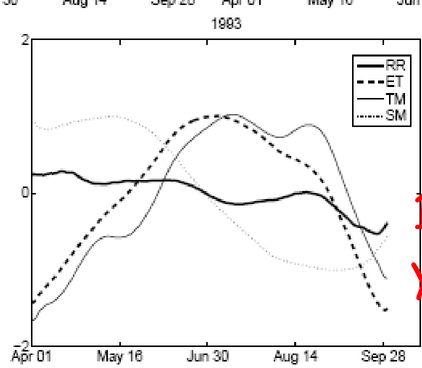
At long timescales (>40 days), the recycling ratio is related to evapotranspiration only during dry years.

1986 normal year

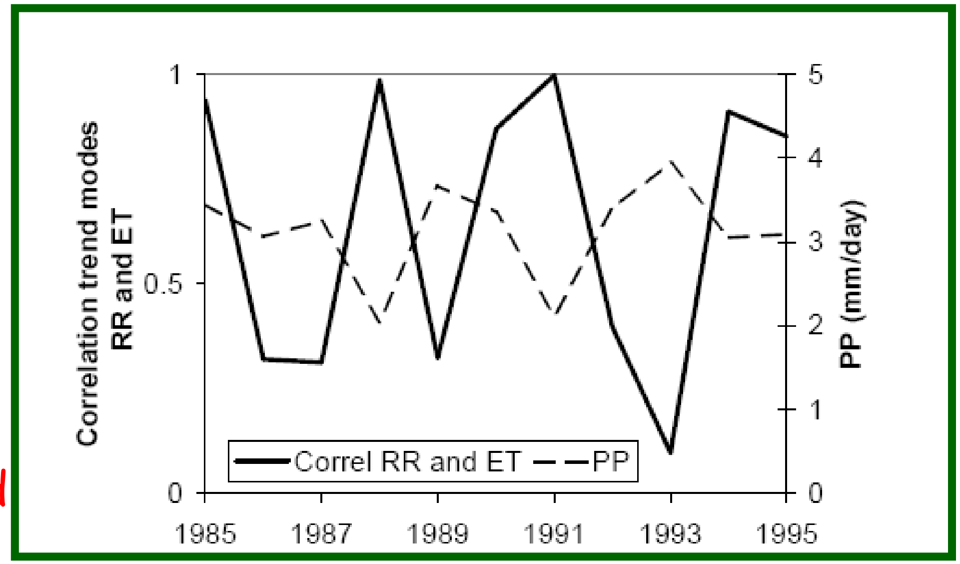


1988 drought year

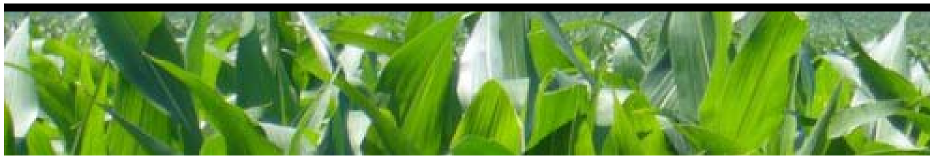




1993 flood year



This result is quite surprising!



Midwestern
United States

In the Midwest precipitation recycling acts as a mechanism for ecoclimatological stability through local negative feedbacks.