Do Models Generate Realistic Simulations?

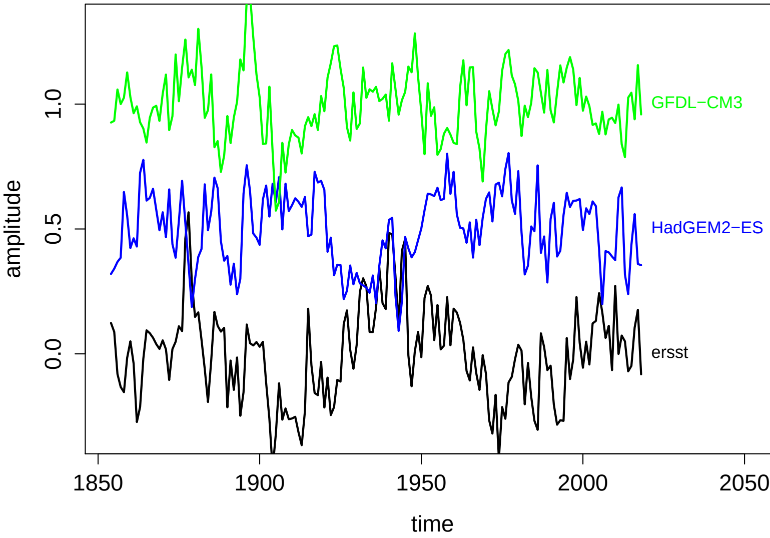
Timothy DelSole

George Mason University, Fairfax, Va and
Center for Ocean-Land-Atmosphere Studies

December 6, 2021

Collaborators: Michael K. Tippett

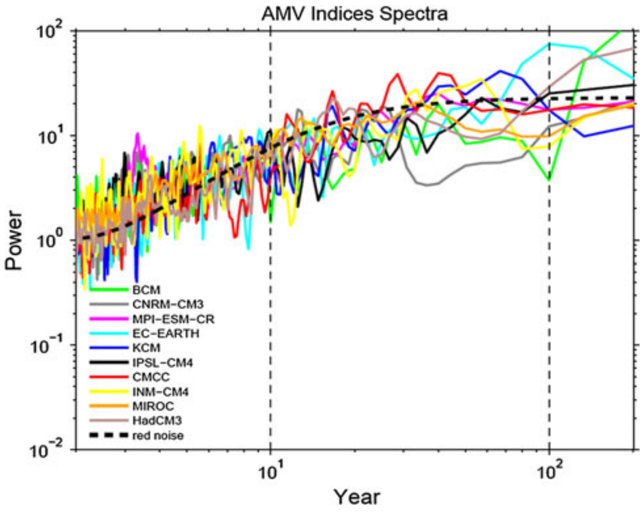
AMV indices



CMIP6 pre-industrial control runs. 2nd-order polynomial removed

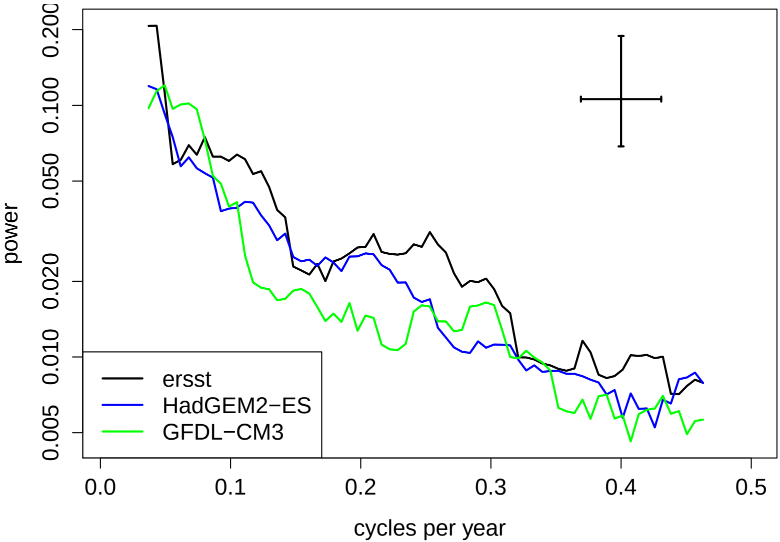
How Do You Compare Serially Correlated Time Series?

Power Spectra



Ba et al. 2014, Climate Dynamics

Power Spectra of AMV indices

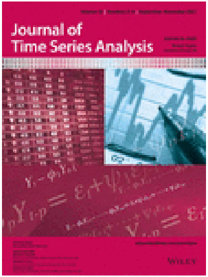


TESTS FOR COMPARING TWO
ESTIMATED SPECTRAL DENSITIES

By D. S. Coates and P. J. Diggle

Department of Statistics, University of Newcastle upon Tyne

Abstract. This paper was motivated by a problem in the gas industry and describes a number of periodogram-based tests of the hypothesis that two independent time-series are realizations of the same stationary process. Non-parametric tests analogous to the maximum periodogram ordinate and cumulative periodogram tests for white noise are compared with a likelihood ratio test based on a postulated quadratic model for the log spectral ratio. The latter is found to be generally more powerful against alternatives in which the two series are realizations of different low order AR processes. The operation of the likelihood ratio test is illustrated by two sets of data, the classic Beveridge wheat price series and a set of data supplied by British Gas.



- If two power spectra are equal, then their ratio is a constant.
- A constant power spectra indicates white noise.
- There already exists (ingenious) spectral tests for white noise.
- Tests for white noise can be adapted to tests of equal power spectra.

1986, J. Time Series Analysis vol 7, No 1, p7 doi:10.1111/j.1467-9892.2009.00616.x

TESTING EQUALITY OF STATIONARY AUTOCOVARIANCES

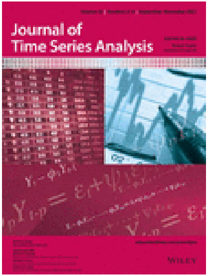
By Robert Lund*, Hany Bassily* and Brani Vidakovic†

*Clemson University and †Georgia Institute of Technology

First Version received August 2007

Abstract. This article studies tests for assessing whether two stationary and independent time series have the same dynamics – specifically, whether the autocovariances of both series coincide at all lags. Frequency domain statistics previously proposed for this purpose are reviewed. A time domain statistic is then developed and investigated. The performance of these statistics are compared. Multivariate versions of the results are constructed. The methods are applied in the analysis of temperatures and precipitations from Atlanta and Athens, Georgia. Our interest here is driven by the need to identify a good climatological reference series for a given station.

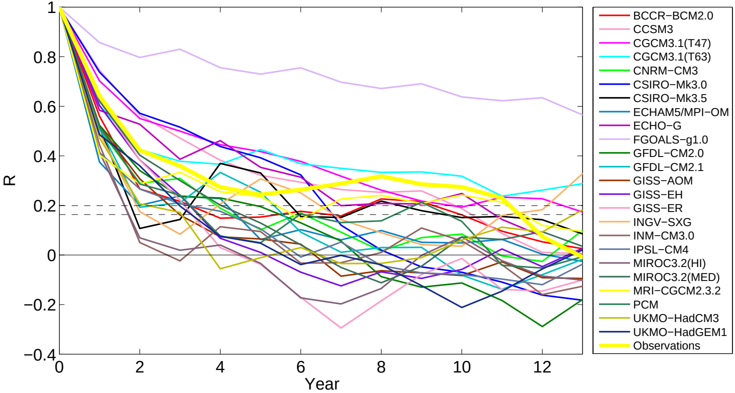
Keywords. Multivariate series; periodogram; spectral density.



- Lund et al. (2009) reviews tests of differences in power spectra.
- Difference-in-spectra tests have low statistical power; i.e., only large differences in spectra can be detected.

Autocorrelation Function

Autocorrelation function for annual mean AMO index



Medhaug and Furevik, 2011, Ocean Sciences Discussion

doi:10.1111/j.1467-9892.2009.00616.x

TESTING EQUALITY OF STATIONARY AUTOCOVARIANCES

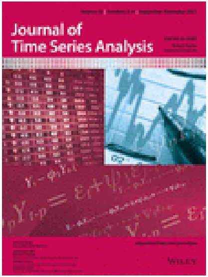
By Robert Lund*, Hany Bassily* and Brani Vidakovic†

*Clemson University and †Georgia Institute of Technology

First Version received August 2007

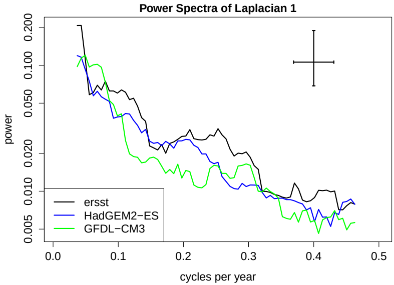
Abstract. This article studies tests for assessing whether two stationary and independent time series have the same dynamics – specifically, whether the autocovariances of both series coincide at all lags. Frequency domain statistics previously proposed for this purpose are reviewed. A time domain statistic is then developed and investigated. The performance of these statistics are compared. Multivariate versions of the results are constructed. The methods are applied in the analysis of temperatures and precipitations from Atlanta and Athens, Georgia. Our interest here is driven by the need to identify a good climatological reference series for a given station.

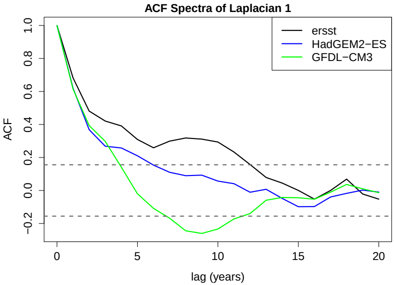
Keywords. Multivariate series; periodogram; spectral density.



- Lund et al. (2009) develop a test for equality of autocovariances.
- This test requires estimating a covariance matrix, but the proposed estimate is not guaranteed to be positive definite.
- In practice, we find that the matrix often is non-positive definite, hence it cannot be applied.

Climate time series generally produce distinctive shapes for the autocorrelation and power spectrum.





Autoregressive models of order p: AR(p)

We assume time series X_t and Y_t come from the models

$$X_t = \phi_1^X X_{t-1} + \cdots + \phi_p^X X_{t-p} + \gamma_X + \epsilon_t^X$$
$$Y_t = \phi_1^Y Y_{t-1} + \cdots + \phi_p^Y Y_{t-p} + \gamma_Y + \epsilon_t^Y,$$

where

$$\epsilon_t^X \overset{iid}{\sim} \mathrm{GWN}(0, \sigma_X^2) \quad \text{and} \quad \epsilon_t^Y \overset{iid}{\sim} \mathrm{GWN}(0, \sigma_Y^2).$$

$$\boxed{\phi_1^2, \ldots, \phi_p^X, \gamma_X, \sigma_X^2}$$

Are the parameters of the AR(p) models equal?

Are the parameters of the AR(p) models equal?

$$\begin{aligned} \text{null hypothesis } H_0: &\quad \phi_1^X = \phi_1^Y, \dots, \phi_p^X = \phi_p^Y, \sigma_X^2 = \sigma_Y^2 \\ \text{alternative hypothesis } H_A: &\quad \text{at least one parameter differs} \end{aligned}$$

γ_X and γ_Y are unrestricted, to forgive biases.

- An AR(p) model uniquely specifies the ACF and power spectra.
- Equality of AR(p) models implies equality of ACFs and of spectra.
- If two time series could have come from the same AR(p) model, then I will say they are statistically indistinguishable.

Likelihood Ratio Test (Bias-Corrected)

$$\boxed{\text{deviance } D = \log\left(\frac{\hat{\sigma}_0^{2\nu_X+2\nu_Y}}{\left(\hat{\sigma}_X^{2\nu_X}\right)\left(\hat{\sigma}_Y^{2\nu_Y}\right)}\right)}$$

$$\begin{aligned}
\hat{\sigma}_X^2 &: \quad \text{unbiased estimate of } \sigma_X^2 \\
\hat{\sigma}_Y^2 &: \quad \text{unbiased estimate of } \sigma_Y^2 \\
\hat{\sigma}_0^2 &: \quad \text{unbiased estimate of } \sigma^2 \text{ under } H_0 \\
\nu_X &: \quad \text{degrees of freedom for } X_t \\
\nu_Y &: \quad \text{degrees of freedom for } Y_t
\end{aligned}$$

D vanishes if and only if $\hat{\sigma}_X^2 = \hat{\sigma}_Y^2$ and $\hat{\phi}_j^X = \hat{\phi}_j^Y$, and is positive otherwise

$$
\begin{aligned}
X_t &= \phi_1^X X_{t-1} + \cdots + \phi_p^X X_{t-p} + \gamma_X + \epsilon_t^X \\
Y_t &= \phi_1^Y Y_{t-1} + \cdots + \phi_p^Y Y_{t-p} + \gamma_Y + \epsilon_t^Y,
\end{aligned}
$$

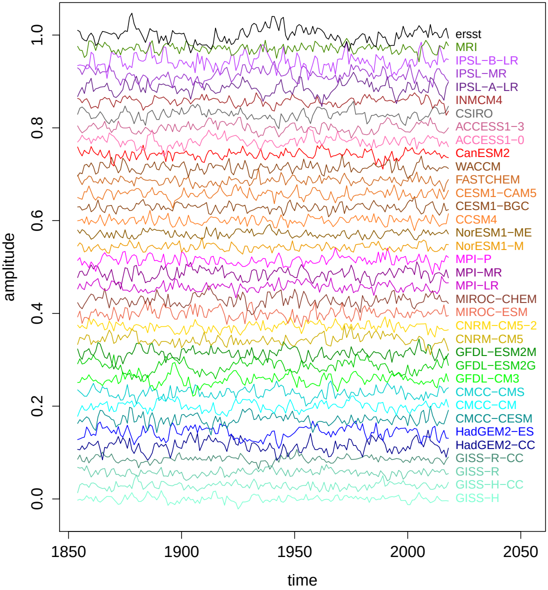
This test uses <u>prewhitened</u> variances, rather than variances directly.

$$
\hat{\epsilon}_t^X = X_t - \left(\hat{\phi}_1^X X_{t-1} + \cdots + \hat{\phi}_p^X X_{t-p} + \hat{\gamma}_X\right)
$$

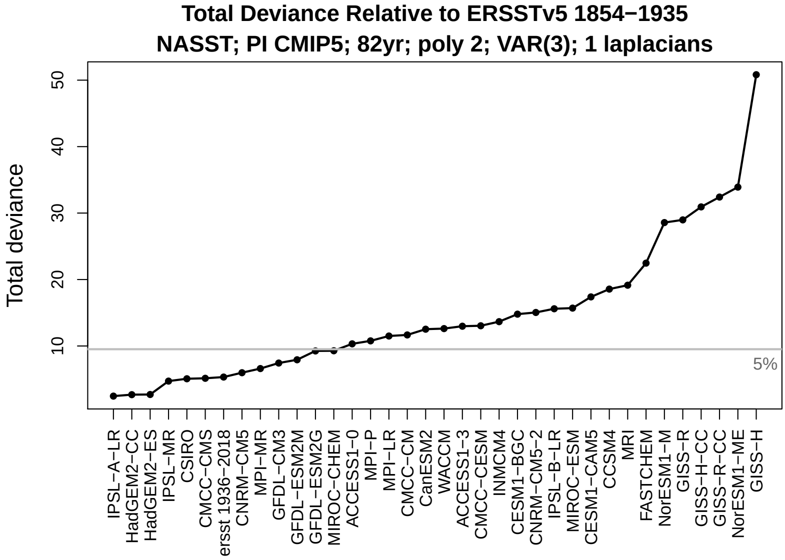
For large sample size, $\hat{\epsilon}_t^X$ is approximately white noise.

Application

- **Variable:** AMV index: annual-mean SST averaged over the Atlantic between $0-60^{\circ}$N.
- **Model Simulations:** Pre-industrial control simulations of SST from phase 5 of the Coupled Model Intercomparison Project (CMIP5).
- **Observations:** the 165-year period 1854-2018 from ERSSTv5.
- **Removal of Forced Variability:** Response to human and natural forcings assumed to be removed after regressing out second-order polynomial over 1854-2018 (other approaches were explored but not included in this talk).
- **p selection:** AICr selects $p=1$ for most CMIP5 models, suggests $p=3$ is adequate for all but two CMIP5 models. We use AR(3).



amplitude
time
1850
1900
1950
2000
2050
0.0
0.2
0.4
0.6
0.8
1.0
ersst
MRI
IPSL−B−LR
IPSL−MR
IPSL−A−LR
INMCM4
CSIRO
ACCESS1−3
ACCESS1−0
CanESM2
WACCM
FASTCHEM
CESM1−CAM5
CESM1−BGC
CCSM4
NorESM1−ME
NorESM1−M
MPI−P
MPI−MR
MPI−LR
MIROC−CHEM
MIROC−ESM
CNRM−CM5−2
CNRM−CM5
GFDL−ESM2M
GFDL−ESM2G
GFDL−CM3
CMCC−CMS
CMCC−CM
CMCC−CESM
HadGEM2−ES
HadGEM2−CC
GISS−R−CC
GISS−R
GISS−H−CC
GISS−H



Total Deviance Relative to ERSSTv5 1854−1935
NASST; PI CMIP5; 82yr; poly 2; VAR(3); 1 laplacians
Total deviance
10
20
30
40
50
5%
IPSL−A−LR
HadGEM2−CC
HadGEM2−ES
IPSL−MR
CSIRO
CMCC−CMS
ersst 1936−2018
CNRM−CM5
MPI−MR
GFDL−CM3
GFDL−ESM2M
GFDL−ESM2G
MIROC−CHEM
ACCESS1−0
MPI−P
MPI−LR
CMCC−CM
CanESM2
WACCM
ACCESS1−3
CMCC−CESM
INMCM4
CESM1−BGC
CNRM−CM5−2
IPSL−B−LR
MIROC−ESM
CESM1−CAM5
CCSM4
MRI
FASTCHEM
NorESM1−M
GISS−R
GISS−H−CC
GISS−R−CC
NorESM1−ME
GISS−H

For simulations of AMV, more than half the models are inconsistent with observations.

For simulations of AMV, more than half the models are inconsistent with observations.

How do they differ?

$$X_t = \boxed{\phi_1^X X_{t-1} + \cdots + \phi_p^X X_{t-p}} + \gamma_X + \boxed{\epsilon_t^X}$$

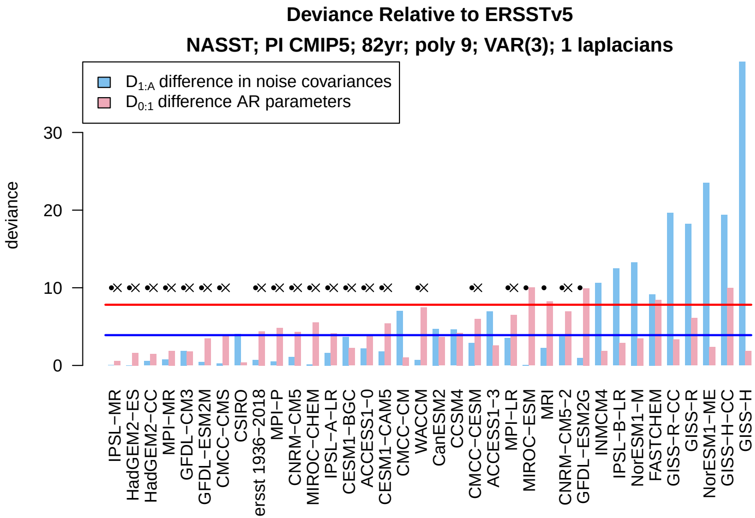
AR parameters $\phi_1^X, \ldots, \phi_p^X$

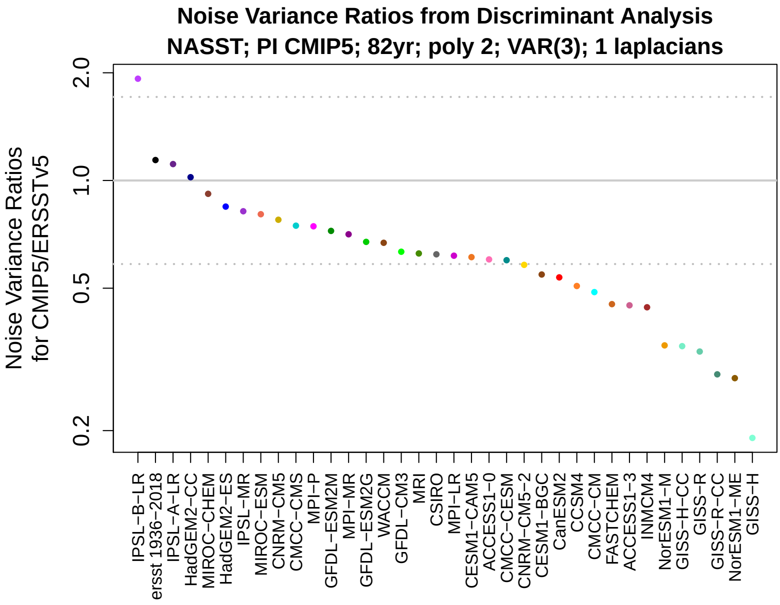
noise parameters σ_X^2

- Difference in noise parameters?
  - Implies differences in prewhitened variances.
  - Implies differences in one-step prediction errors.
- Differences in AR parameters?
  - Implies differences in memory.
  - Implies differences in predictability.
  - Implies differences in “dynamics”

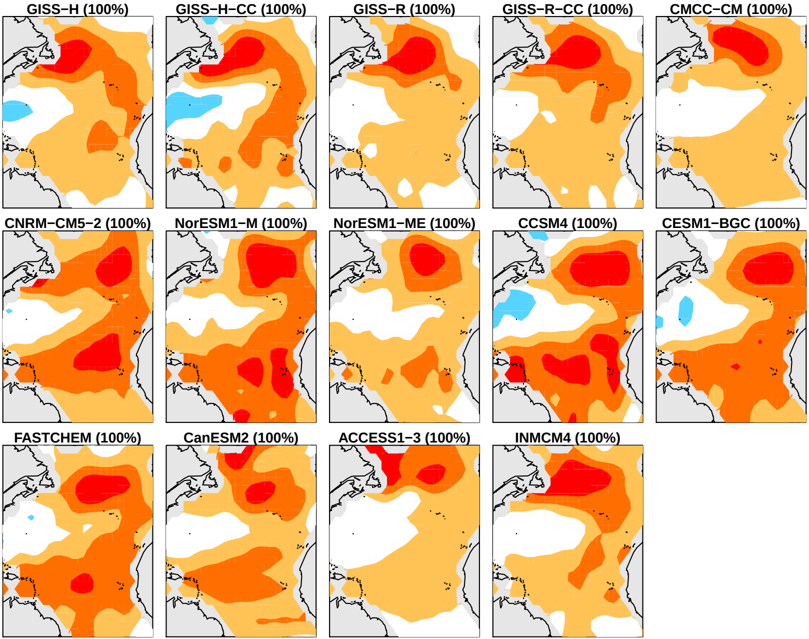
$$\underset{\text{total}}{D} = \underset{\text{difference in noise}}{\color{blue}{D_{\text{noise}}}} + \underset{\text{difference in AR parameters}}{\color{red}{D_{\text{P}}}}$$

Under the null hypothesis, $D_{\textbf{noise}}$ and $D_{\textbf{P}}$ are independent and have chi-squared distributions.


Deviance Relative to ERSSTv5
NASST; PI CMIP5; 82yr; poly 9; VAR(3); 1 laplacians
$D_{1:A}$ difference in noise covariances
$D_{0:1}$ difference AR parameters
deviance
0
10
20
30
IPSL−MR
HadGEM2−ES
HadGEM2−CC
MPI−MR
GFDL−CM3
GFDL−ESM2M
CMCC−CMS
CSIRO
ersst 1936−2018
MPI−P
CNRM−CM5
MIROC−CHEM
IPSL−A−LR
CESM1−BGC
ACCESS1−0
CESM1−CAM5
CMCC−CM
WACCM
CanESM2
CCSM4
CMCC−CESM
ACCESS1−3
MPI−LR
MIROC−ESM
MRI
CNRM−CM5−2
GFDL−ESM2G
INMCM4
IPSL−B−LR
NorESM1−M
FASTCHEM
GISS−R−CC
GISS−R
NorESM1−ME
GISS−H−CC
GISS−H

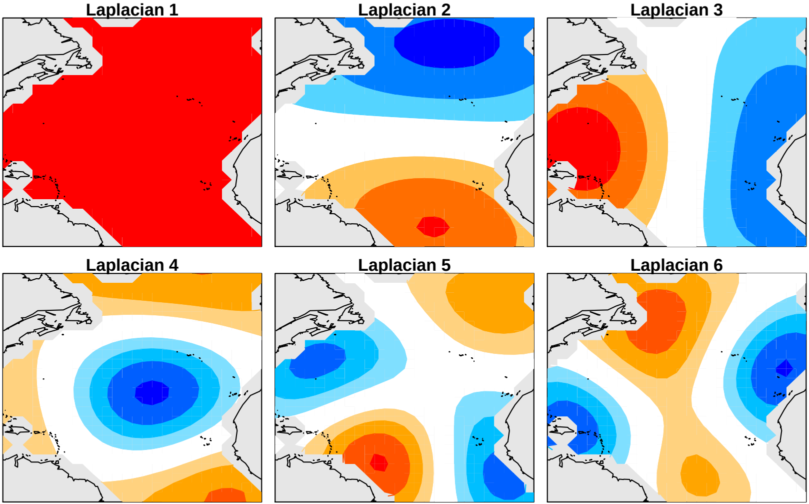


Noise Variance Ratios from Discriminant Analysis
NASST; PI CMIP5; 82yr; poly 2; VAR(3); 1 laplacians
Noise Variance Ratios
for CMIP5/ERSSTv5
2.0
1.0
0.5
0.2
IPSL−B−LR
ersst 1936−2018
IPSL−A−LR
HadGEM2−CC
MIROC−CHEM
HadGEM2−ES
IPSL−MR
MIROC−ESM
CNRM−CM5
CMCC−CMS
MPI−P
GFDL−ESM2M
MPI−MR
GFDL−ESM2G
WACCM
GFDL−CM3
MRI
CSIRO
MPI−LR
CESM1−CAM5
ACCESS1−0
CMCC−CESM
CNRM−CM5−2
CESM1−BGC
CanESM2
CCSM4
CMCC−CM
FASTCHEM
ACCESS1−3
INMCM4
NorESM1−M
GISS−H−CC
GISS−R
GISS−R−CC
NorESM1−ME
GISS−H


GISS−H (100%)
GISS−H−CC (100%)
GISS−R (100%)
GISS−R−CC (100%)
CMCC−CM (100%)
CNRM−CM5−2 (100%)
NorESM1−M (100%)
NorESM1−ME (100%)
CCSM4 (100%)
CESM1−BGC (100%)
FASTCHEM (100%)
CanESM2 (100%)
ACCESS1−3 (100%)
INMCM4 (100%)

Is the Relation Between AMV and Other Patterns Realistic?

Laplacian Eigenfunctions over the Atlantic



DelSole and Tippett, 2015, *J. Climate*

Multivariate Generalization

Vector Autoregressive Model (p)

$$\mathbf{z}_t = \mathbf{A}_1 \mathbf{z}_{t-1} + \cdots + \mathbf{A}_p \mathbf{z}_{t-p} + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t,$$

- For $p = 1$, this is equivalent to Linear Inverse Model (LIM).
- Deviance statistic is analogous to univariate case, except with variances replaced by determinants of covariance matrices.

Order p and number of Laplacians selected using Mutual Information Criterion (MIC)

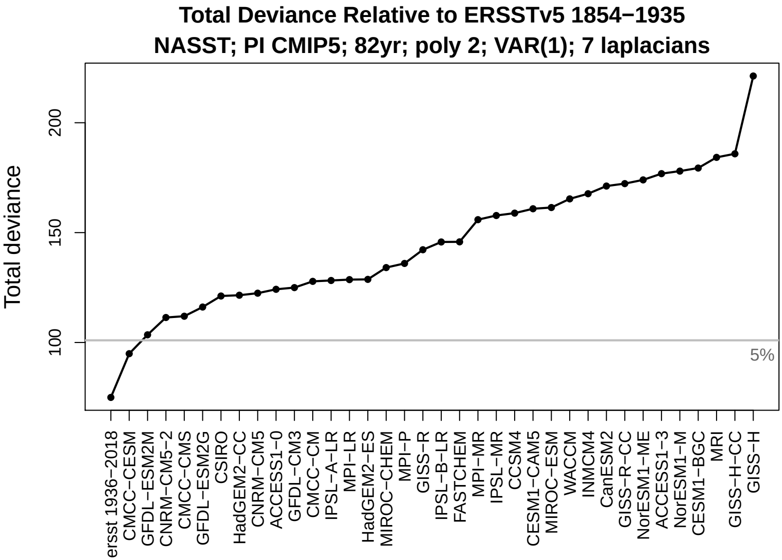
This criterion also can be used to select

- number of EOFs in LIM
- number of EOFs in CCA

A VAR(1) with 7 Laplacians is adequate for most models.

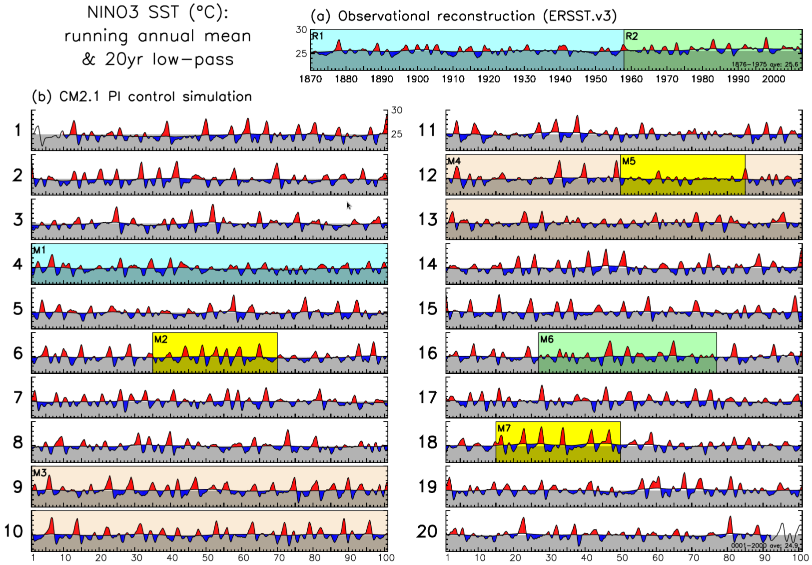
DelSole and Tippett, 2021, *Stat*

Laplacian Eigenfunctions over the Atlantic

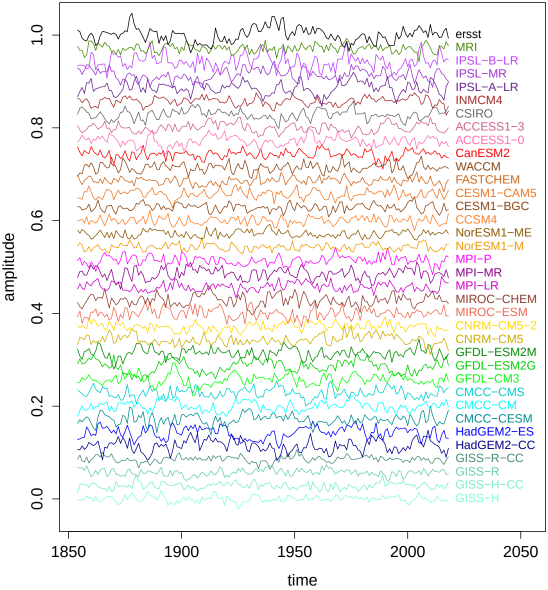


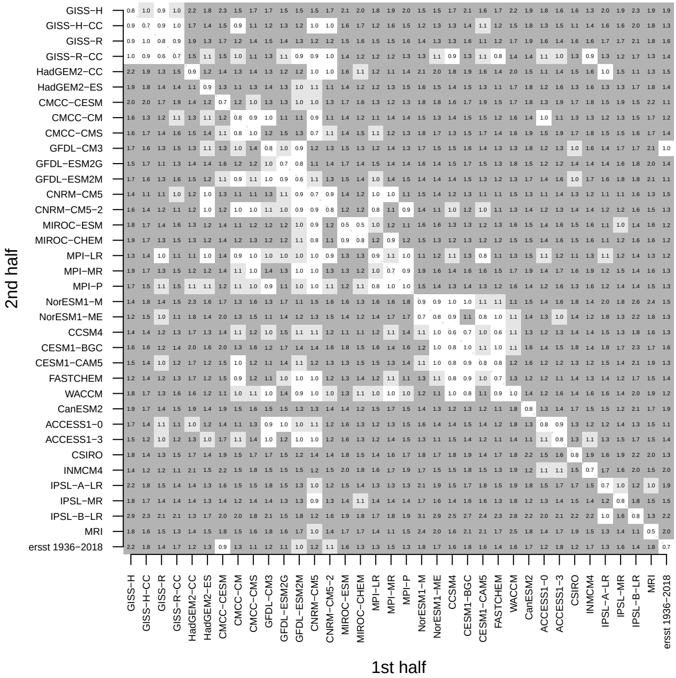
By including smaller scale variability ($\sim 2000km$), virtually all models are unrealistic.

Epochs



Wittenberg (2009) https://extranet.gfdl.noaa.gov/ atw/vgrp/enso/#diversity


ersst
MRI
IPSL−B−LR
IPSL−MR
IPSL−A−LR
INMCM4
CSIRO
ACCESS1−3
ACCESS1−0
CanESM2
WACCM
FASTCHEM
CESM1−CAM5
CESM1−BGC
CCSM4
NorESM1−ME
NorESM1−M
MPI−P
MPI−MR
MPI−LR
MIROC−CHEM
MIROC−ESM
CNRM−CM5−2
CNRM−CM5
GFDL−ESM2M
GFDL−ESM2G
GFDL−CM3
CMCC−CMS
CMCC−CM
CMCC−CESM
HadGEM2−ES
HadGEM2−CC
GISS−R−CC
GISS−R
GISS−H−CC
GISS−H
amplitude
0.0
0.2
0.4
0.6
0.8
1.0
time
1850
1900
1950
2000
2050


GISS−H
GISS−H−CC
GISS−R
GISS−R−CC
HadGEM2−CC
HadGEM2−ES
CMCC−CESM
CMCC−CM
CMCC−CMS
GFDL−CM3
GFDL−ESM2G
GFDL−ESM2M
CNRM−CM5
CNRM−CM5−2
MIROC−ESM
MIROC−CHEM
MPI−LR
MPI−MR
MPI−P
NorESM1−M
NorESM1−ME
CCSM4
CESM1−BGC
CESM1−CAM5
FASTCHEM
WACCM
CanESM2
ACCESS1−0
ACCESS1−3
CSIRO
INMCM4
IPSL−A−LR
IPSL−MR
IPSL−B−LR
MRI
ersst 1936−2018
2nd half
1st half

In What Ways Do the Statistics Differ?

$$\mathbf{z}_t = \boxed{\mathbf{A}_1\mathbf{z}_{t-1} + \cdots + \mathbf{A}_p\mathbf{z}_{t-p}} + \boldsymbol{\mu} + \boxed{\boldsymbol{\epsilon}_t}$$

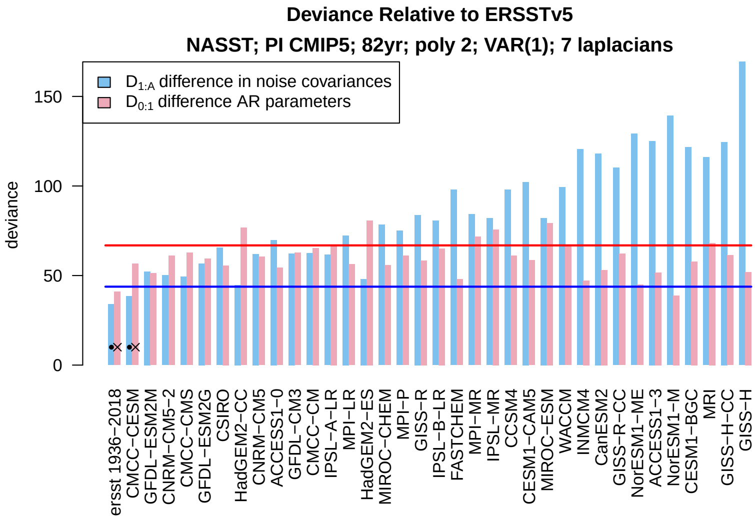
$$\boldsymbol{\epsilon}_t \sim \mathrm{GWN}(\mathbf{0}, \boldsymbol{\Gamma})$$

AR parameters	$\mathbf{A}_1, \ldots, \mathbf{A}_p$
noise parameters	$\boldsymbol{\Gamma}$

- Difference in noise parameters?
  - Implies differences in prewhitened variances.
  - Implies differences in one-step prediction errors.
- Differences in AR parameters?
  - Implies differences in memory.
  - Implies differences in predictability.
  - Implies differences in "dynamics"

$$\underset{\text{total}}{D} = \underset{\text{difference in noise}}{D_{\text{noise}}} + \underset{\text{difference in AR parameters}}{D_{\text{P}}}$$

Under the null hypothesis, D_{noise} and D_{P} are independent and have chi-squared distributions.


Deviance Relative to ERSSTv5
NASST; PI CMIP5; 82yr; poly 2; VAR(1); 7 laplacians
D1:A difference in noise covariances
D0:1 difference AR parameters
deviance
150
100
50
0
ersst 1936−2018
CMCC−CESM
GFDL−ESM2M
CNRM−CM5−2
CMCC−CMS
GFDL−ESM2G
CSIRO
HadGEM2−CC
CNRM−CM5
ACCESS1−0
GFDL−CM3
CMCC−CM
IPSL−A−LR
MPI−LR
HadGEM2−ES
MIROC−CHEM
MPI−P
GISS−R
IPSL−B−LR
FASTCHEM
MPI−MR
IPSL−MR
CCSM4
CESM1−CAM5
MIROC−ESM
WACCM
INMCM4
CanESM2
GISS−R−CC
NorESM1−ME
ACCESS1−3
NorESM1−M
CESM1−BGC
MRI
GISS−H−CC
GISS−H

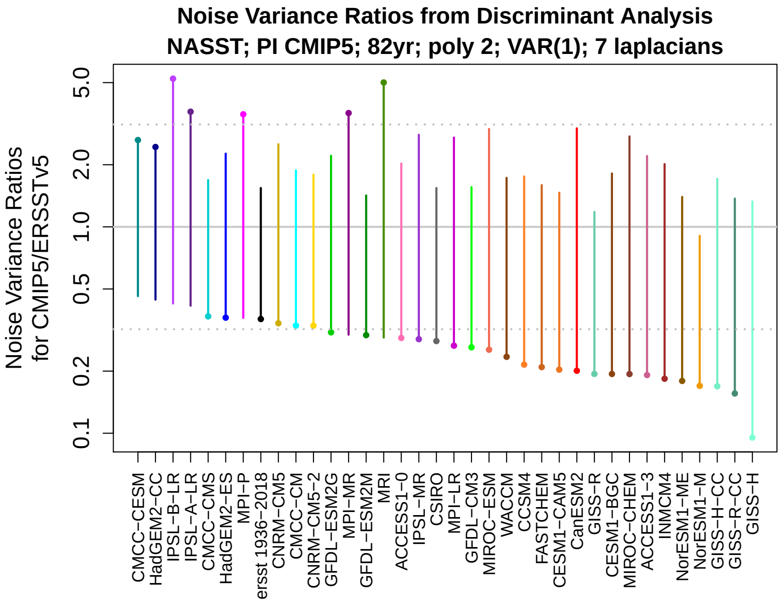
$$\text{Significant } D_{\text{noise}} \quad \Longrightarrow \quad \boldsymbol{\Gamma}_X \neq \boldsymbol{\Gamma}_Y$$

Find linear combination that optimizes differences in noise variances

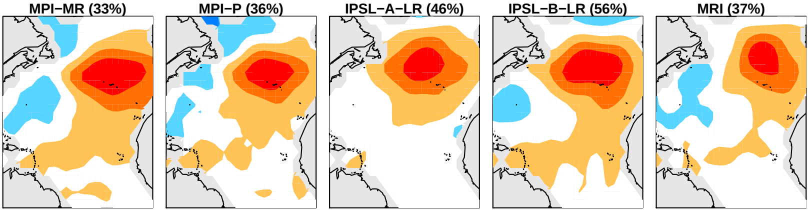
- This is equivalent to maximizing the ratio of variances.
- You get an uncorrelated set of components that decomposes D_{noise} into independent components.

$$D_{\text{noise}} = D_{\text{noise}}^{(1)} + \cdots + D_{\text{noise}}^{(M)}$$

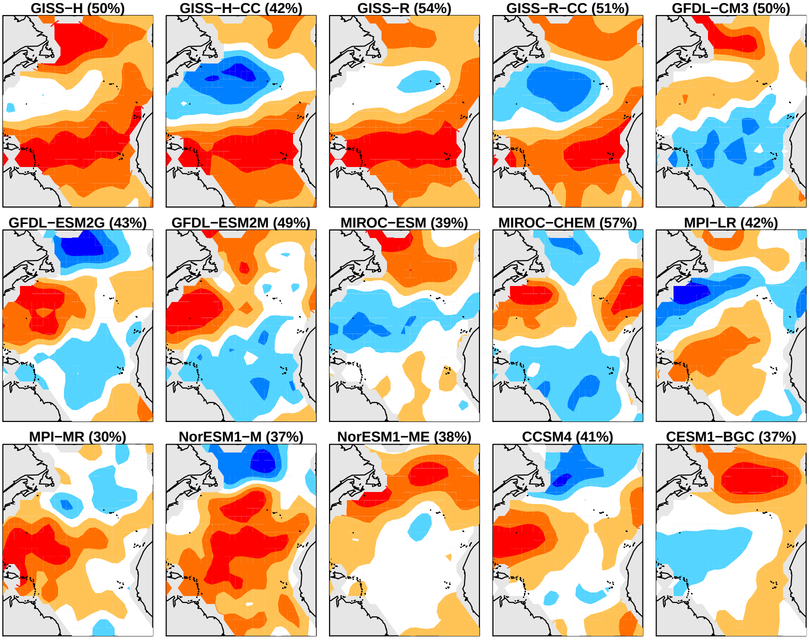
- This is called Covariance Discriminant Analysis (CDA) and generalizes PCA to describe differences in covariance matrices.


Noise Variance Ratios from Discriminant Analysis
NASST; PI CMIP5; 82yr; poly 2; VAR(1); 7 laplacians
Noise Variance Ratios
for CMIP5/ERSSTv5
5.0
2.0
1.0
0.5
0.2
0.1
CMCC−CESM
HadGEM2−CC
IPSL−B−LR
IPSL−A−LR
CMCC−CMS
HadGEM2−ES
MPI−P
ersst 1936−2018
CNRM−CM5
CMCC−CM
CNRM−CM5−2
GFDL−ESM2G
MPI−MR
GFDL−ESM2M
MRI
ACCESS1−0
IPSL−MR
CSIRO
MPI−LR
GFDL−CM3
MIROC−ESM
WACCM
CCSM4
FASTCHEM
CESM1−CAM5
CanESM2
GISS−R
CESM1−BGC
MIROC−CHEM
ACCESS1−3
INMCM4
NorESM1−ME
NorESM1−M
GISS−H−CC
GISS−R−CC
GISS−H

Components with too much noise variance



Components with too little noise variance



$$\text{Significant } D_{\mathsf{P}} \quad \Longrightarrow \quad \{\mathbf{A}_1^X, \ldots, \mathbf{A}_p^X\} \neq \{\mathbf{A}_1^Y, \ldots, \mathbf{A}_p^Y\}$$

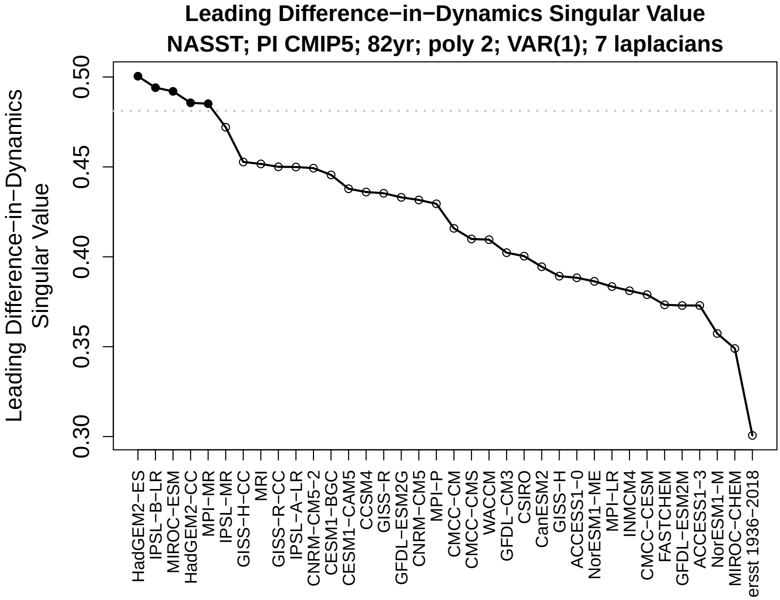
Find linear combination that maximizes $D_{\mathbf{P}}$.

Find linear combination that maximizes D_P.

- Yields a set of uncorrelated components that decomposes D_P.

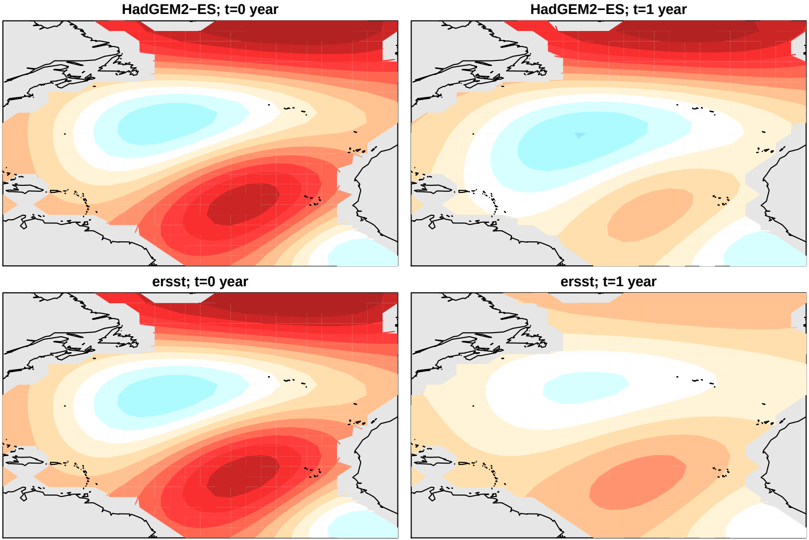
$$D_\mathsf{P} = D_\mathsf{P}^{(1)} + \cdots + D_\mathsf{P}^{(M)}$$

- Equivalent to GSVD of the difference in AR parameters.
- GSVD finds the initial condition that maximizes the difference in response of the VAR(p) dynamics.
- We call this Difference-in-Dynamics SVD.

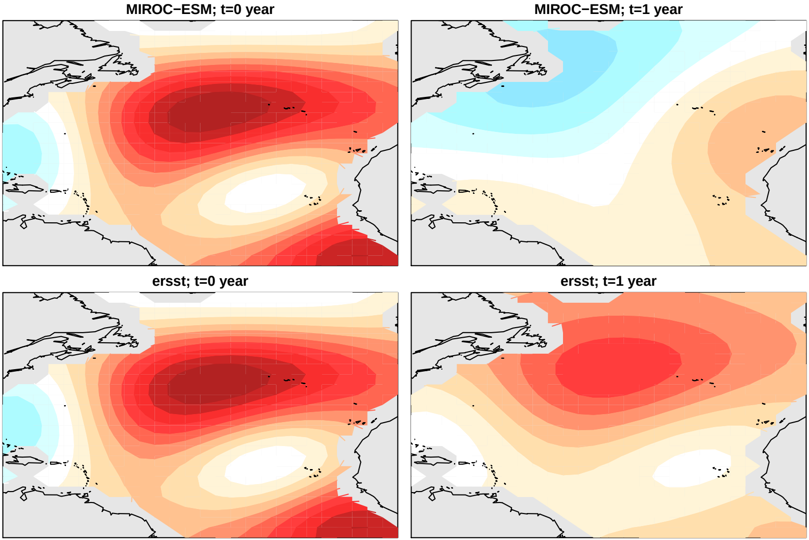


Leading Difference−in−Dynamics Singular Value
NASST; PI CMIP5; 82yr; poly 2; VAR(1); 7 laplacians
Leading Difference−in−Dynamics Singular Value
0.30
0.35
0.40
0.45
0.50
HadGEM2−ES
IPSL−B−LR
MIROC−ESM
HadGEM2−CC
MPI−MR
IPSL−MR
GISS−H−CC
MRI
GISS−R−CC
IPSL−A−LR
CNRM−CM5−2
CESM1−BGC
CESM1−CAM5
CCSM4
GISS−R
GFDL−ESM2G
CNRM−CM5
MPI−P
CMCC−CM
CMCC−CMS
WACCM
GFDL−CM3
CSIRO
CanESM2
GISS−H
ACCESS1−0
NorESM1−ME
MPI−LR
INMCM4
CMCC−CESM
FASTCHEM
GFDL−ESM2M
ACCESS1−3
NorESM1−M
MIROC−CHEM
ersst 1936−2018

$$\mathbf{z}_t = \boxed{\mathbf{A}_1 \mathbf{z}_{t-1}} + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t$$

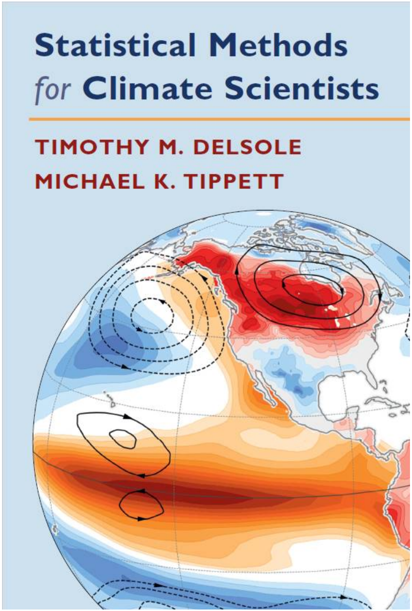


$$\mathbf{z}_t = \boxed{\mathbf{A}_1 \mathbf{z}_{t-1}} + \boldsymbol{\mu} + \epsilon_t$$



Summary

- We propose a rigorous statistical method for comparing simulations and observations that accounts for correlations in space and time.
- The test statistic D measures difference in VAR processes.
- A new criterion called **Mutual Information Criterion (MIC)** is used to select variables and maximum lag in VAR models
- For annual-mean North Atlantic SST, virtually all models are unrealistic after smaller-scale ($\sim 2000km$) information is included.
- Discrimination techniques can be used to optimally diagnose differences in noise statistics and differences in AR parameters.
- Difference-in-dynamics SVD shows some climate models produce one-year predictions with the wrong sign over large spatial scales.
- DelSole and Tippett, 2020, 2021a, 2021b, *Advances in Statistical Climatology, Meteorology and Oceanography* (ASCMO)



Cambridge University Press
December 2021