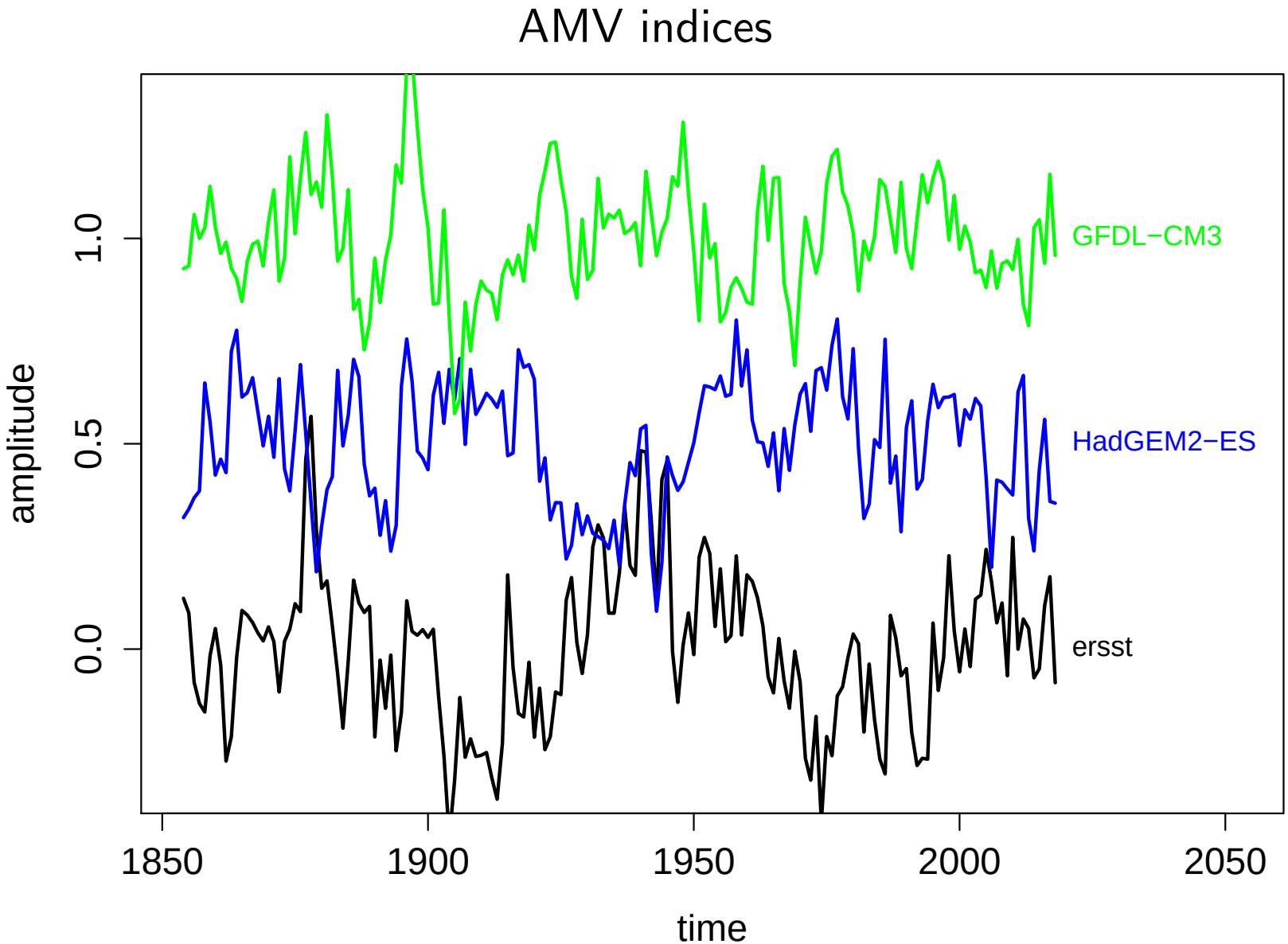
Do Models Generate Realistic Simulations?

Timothy DelSole

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

October 26, 2021

Collaborators: Michael K. Tippett



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

How Do You Compare Serially Correlated Time Series?

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

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

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

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

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

$$\text{null hypothesis } H_0: \quad \phi_1^X = \phi_1^Y, \ldots, \phi_p^X = \phi_p^Y, \sigma_X^2 = \sigma_Y^2$$

$$\text{alternative hypothesis } H_A: \quad \text{at least one parameter differs}$$

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

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

$\hat{\sigma}_X^2$: unbiased estimate of σ_X^2

$\hat{\sigma}_Y^2$: unbiased estimate of σ_Y^2

$\hat{\sigma}_0^2$: unbiased estimate of σ^2 under H_0

ν_X : degrees of freedom for X_t

ν_Y : degrees of freedom for Y_t

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

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

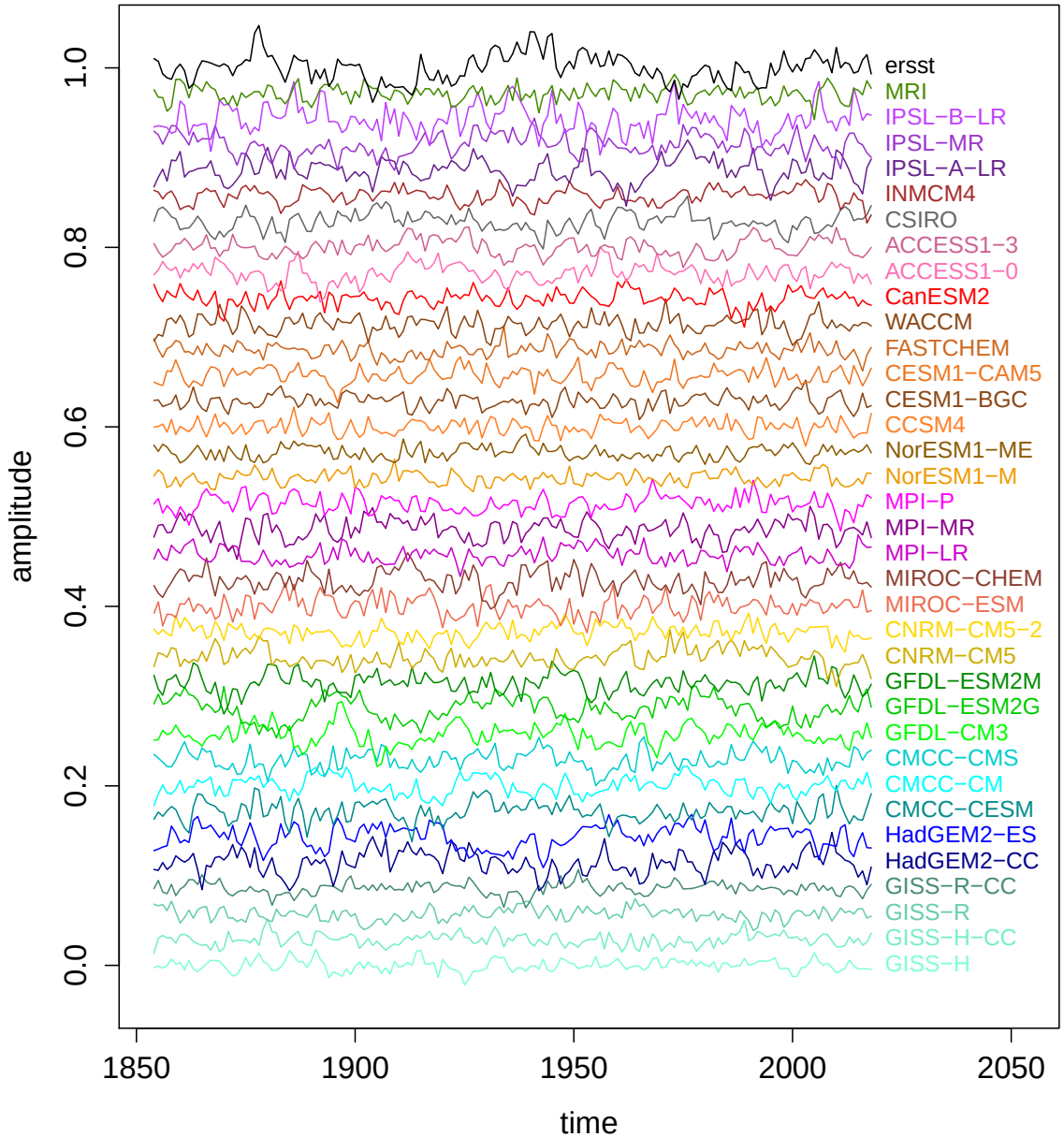
This test uses prewhitened variances, rather than variances directly.

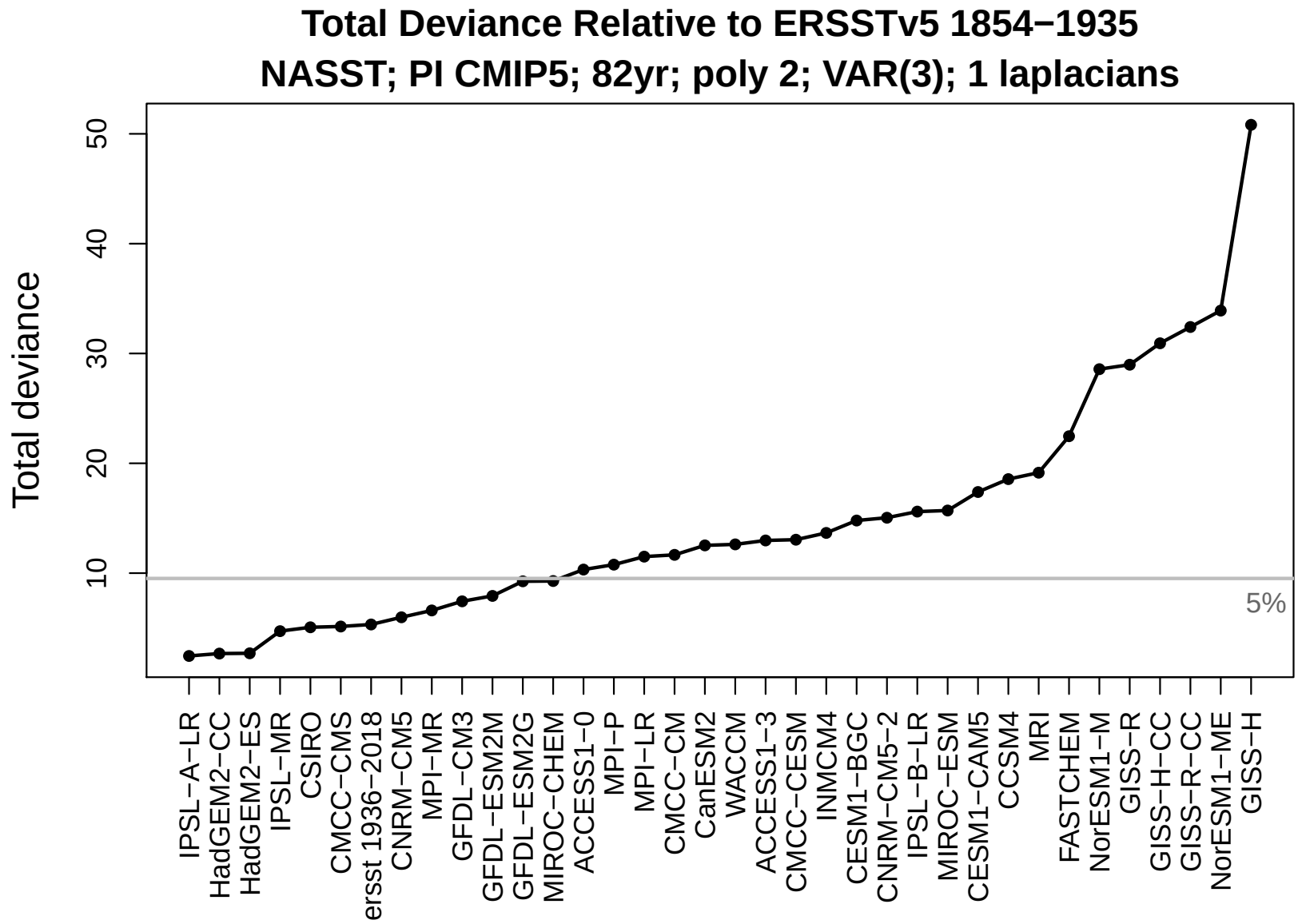
$$\hat{\epsilon}_t^X = X_t - \left(\hat{\phi}_1^X X_{t-1} + \cdots + \hat{\phi}_p^X X_{t-p} + \hat{\gamma}_X\right)$$

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

Application

- **Variable:** AMV index: annual-mean SST averaged over the Atlantic between $0 - 60°$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).
- **Validation:** Time series from the earlier half (1854-1935) are compared to time series in the later half (1936-2018).


amplitude
0.0
0.2
0.4
0.6
0.8
1.0
1850
1900
1950
2000
2050
time
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

In terms of the AMV index, more than half the models are inconsistent with observations.

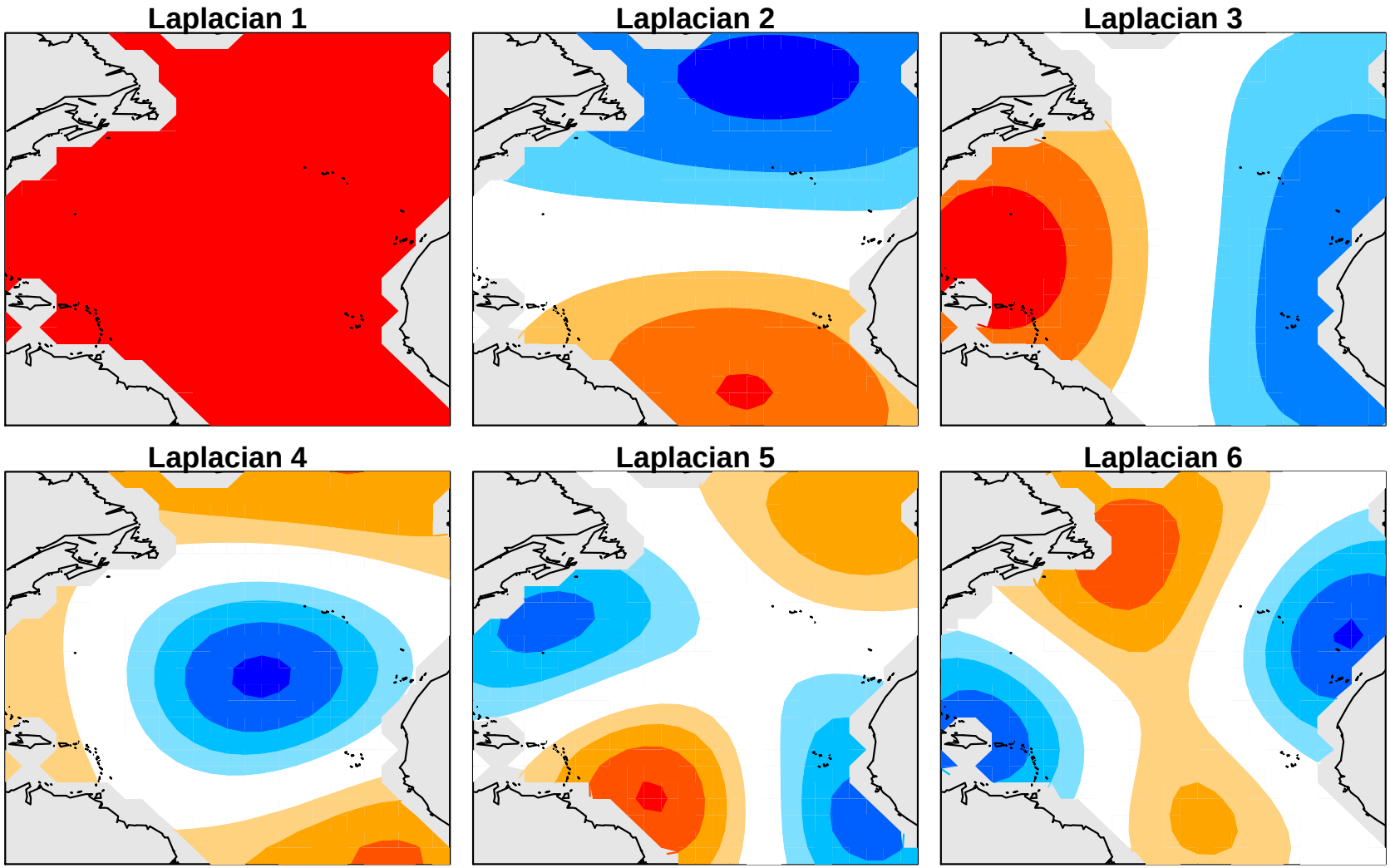
Multivariate Generalization

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.

Laplacian Eigenfunctions over the Atlantic



DelSole and Tippett, 2015, *J. Climate*

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

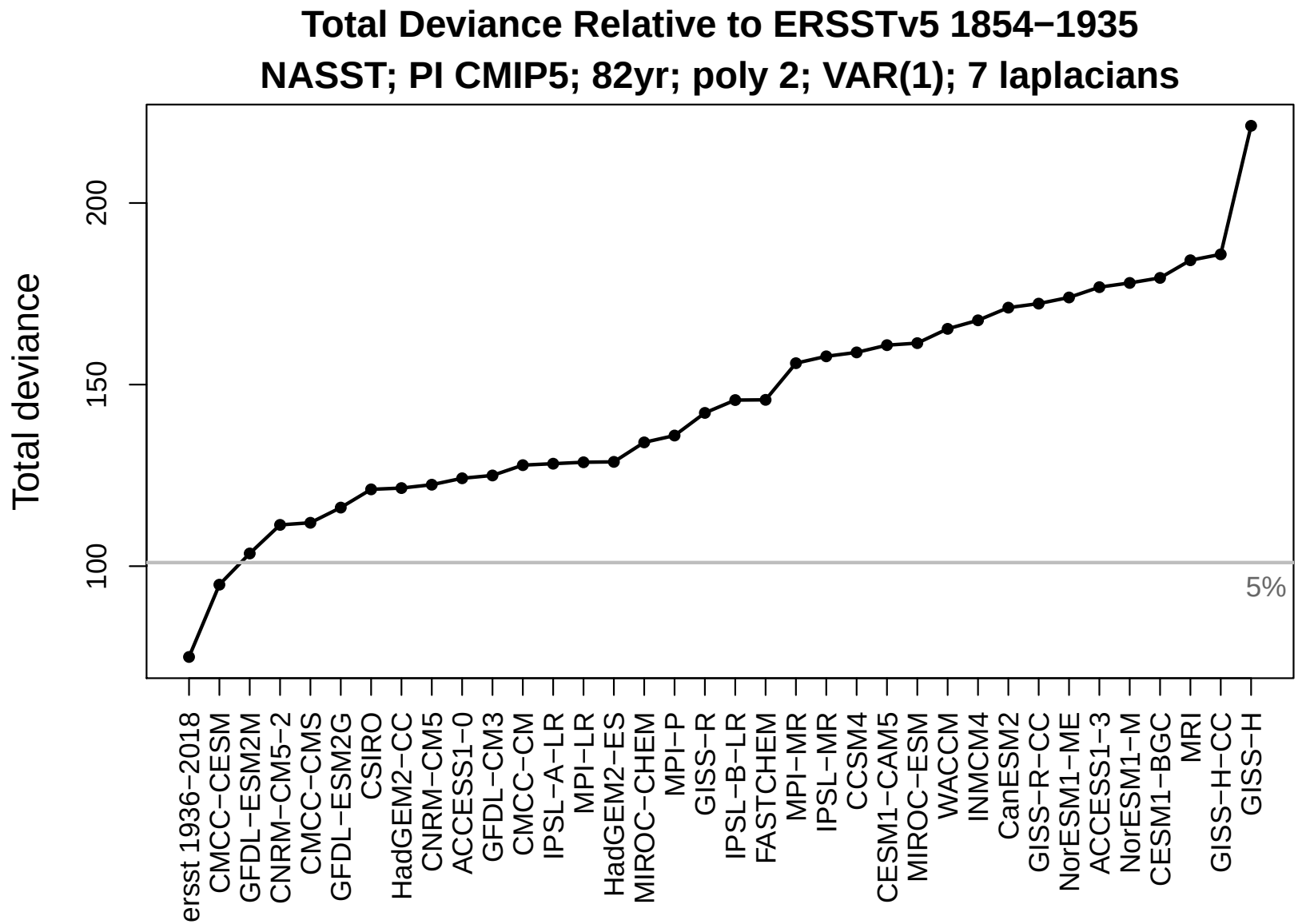
This criterion also can be used to select

- 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.

In What Ways Do the Statistics Differ?

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

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

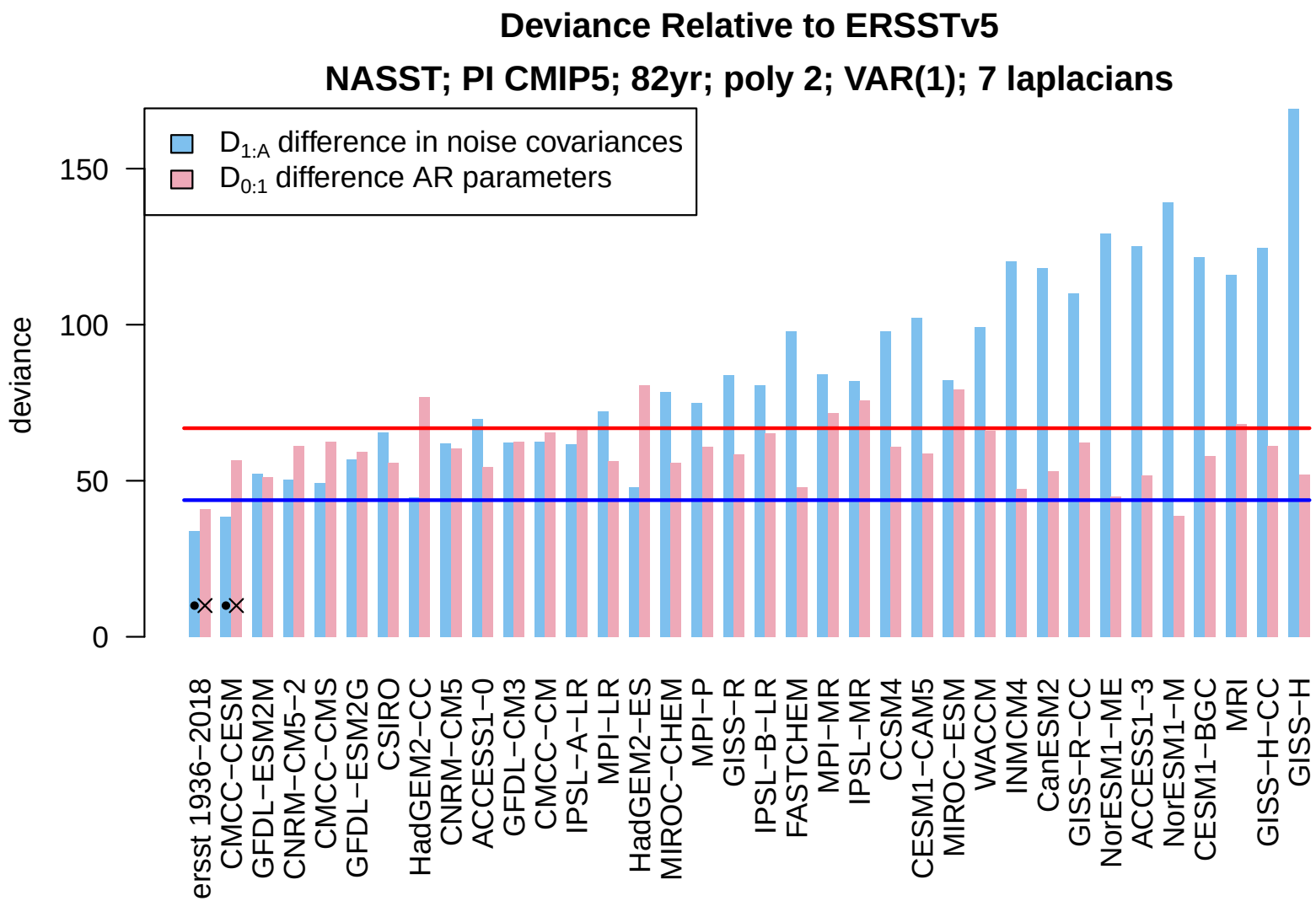
noise parameters $\boldsymbol{\Gamma}$

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

- 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"

$$\underset{\text{total}}{D} = \underset{\text{difference in noise}}{D_{1:A}} + \underset{\text{difference in AR parameters}}{D_{0:1}}$$

Under the null hypothesis, $D_{1:A}$ and $D_{0:1}$ are independent and have chi-squared distributions.


Deviance Relative to ERSSTv5
NASST; PI CMIP5; 82yr; poly 2; VAR(1); 7 laplacians
$D_{1:A}$ difference in noise covariances
$D_{0:1}$ 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

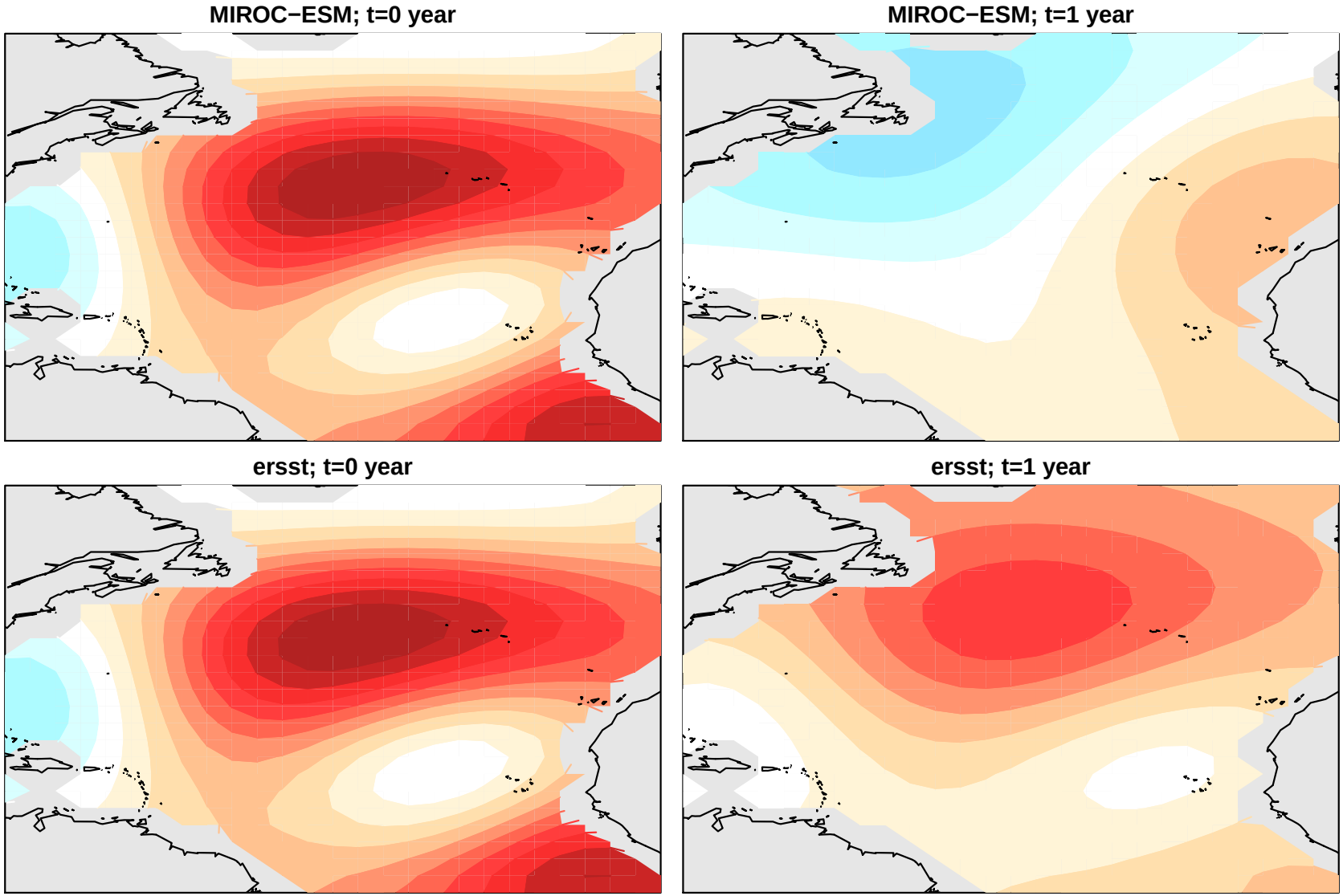
By applying discriminant analysis:

- Noise differences can be further decomposed.
- Differences in AR parameters decomposed using Generalized SVD.

By applying discriminant analysis:

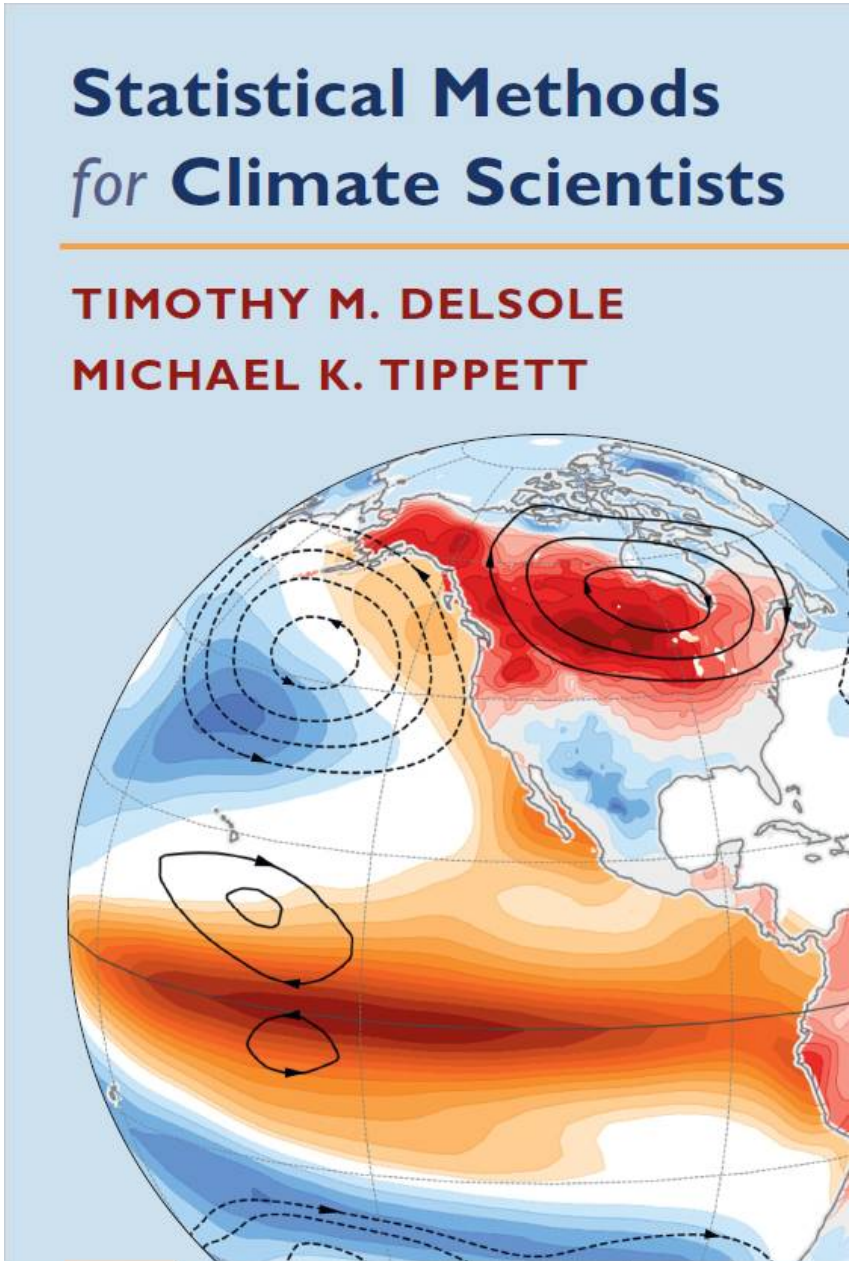
- Noise differences can be further decomposed.
- Differences in AR parameters decomposed using Generalized SVD.

no more time...


MIROC−ESM; t=0 year
MIROC−ESM; t=1 year
ersst; t=0 year
ersst; t=1 year

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