DO MODELS GENERATE REALISTIC SIMULATIONS OF THE NORTH ATLANTIC SST?

Timothy DelSole[1] and **Michael K. Tippett**[2]

[1]George Mason University, Fairfax VA
[2]Columbia University, New York, NY

What is the Problem?

Many climate studies compare model output to observations, yet few of these studies use statistically rigorous criterion for deciding if a model simulation is consistent with observations. Without such a criterion, we see very different behaviors across models without knowing which model is realistic. Although objective criteria do exist, they often do not account for serial correlations, which is pervasive in climate data, and even fewer criteria can be applied to multivariate data.

Our Approach

We propose a rigorous statistical method for comparing model simulations and observations that accounts for correlations in space and time. The basic idea is to fit each multivariate time series to a Vector Autoregressive (VAR) Model, and then test if the parameters of the two models are equal. A Vector Autoregressive (VAR) process $\mathbf{z}_t$ is generated by a model of the form

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

where $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are constant $S \times S$ matrices, $\boldsymbol{\mu}$ is a constant vector that controls the mean, and $\boldsymbol{\epsilon}_t$ is a Gaussian white noise process with covariance matrix $\boldsymbol{\Gamma}$. We refer to $\boldsymbol{\epsilon}_t$ as the *noise* and $\mathbf{A}_1, \ldots, \mathbf{A}_p$ as the *AR parameters*. Given

$$\left\{\mathbf{z}_1^{(1)}, \ldots, \mathbf{z}_{N_1}^{(1)}\right\} \quad \text{and} \quad \left\{\mathbf{z}_1^{(2)}, \ldots, \mathbf{z}_{N_2}^{(2)}\right\},$$

we assume each comes from a VAR(p) process and test the hypothesis

$$H_0: \quad \mathbf{A}_1^{(1)} = \mathbf{A}_1^{(2)}, \ldots, \mathbf{A}_p^{(1)} = \mathbf{A}_p^{(2)}, \quad \boldsymbol{\Gamma}^{(1)} = \boldsymbol{\Gamma}^{(2)}.$$

The test statistic is called the *deviance* and defined

$$D = \log\left(\frac{|\hat{\boldsymbol{\Gamma}}_0|^{\nu_1+\nu_2}}{|\hat{\boldsymbol{\Gamma}}_1|^{\nu_1}|\hat{\boldsymbol{\Gamma}}_2|^{\nu_2}}\right),$$

where $\hat{\boldsymbol{\Gamma}}_1$ and $\hat{\boldsymbol{\Gamma}}_2$ are estimates of $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$, $\hat{\boldsymbol{\Gamma}}_0$ is an estimate of the noise covariance matrix under H_0, and ν_1 and ν_2 are appropriate degrees of freedom. D behaves as a measure of the difference in VAR models.

$$D \sim \chi^2_{MS+S(S+1)/2} \quad \text{if } H_0 \text{ is true}$$

Large values of D indicate rejection of H_0.

Application to North Atlantic SST (NASST)

- **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.
- **Variable:** Annual-mean NASST projected onto the first 7 eigenvectors of Laplace's equation over the Atlantic between $0-60°$N.
- **Removal of Forced Variability:** We assume the response to human and natural forcings is well represented by a second-order polynomial over 1854-2018 (other approaches were explored but not included in this poster).
- **Validation:** Time series from the earlier half (1854-1935) are compared to time series in the later half (1936-2018).

Main Result

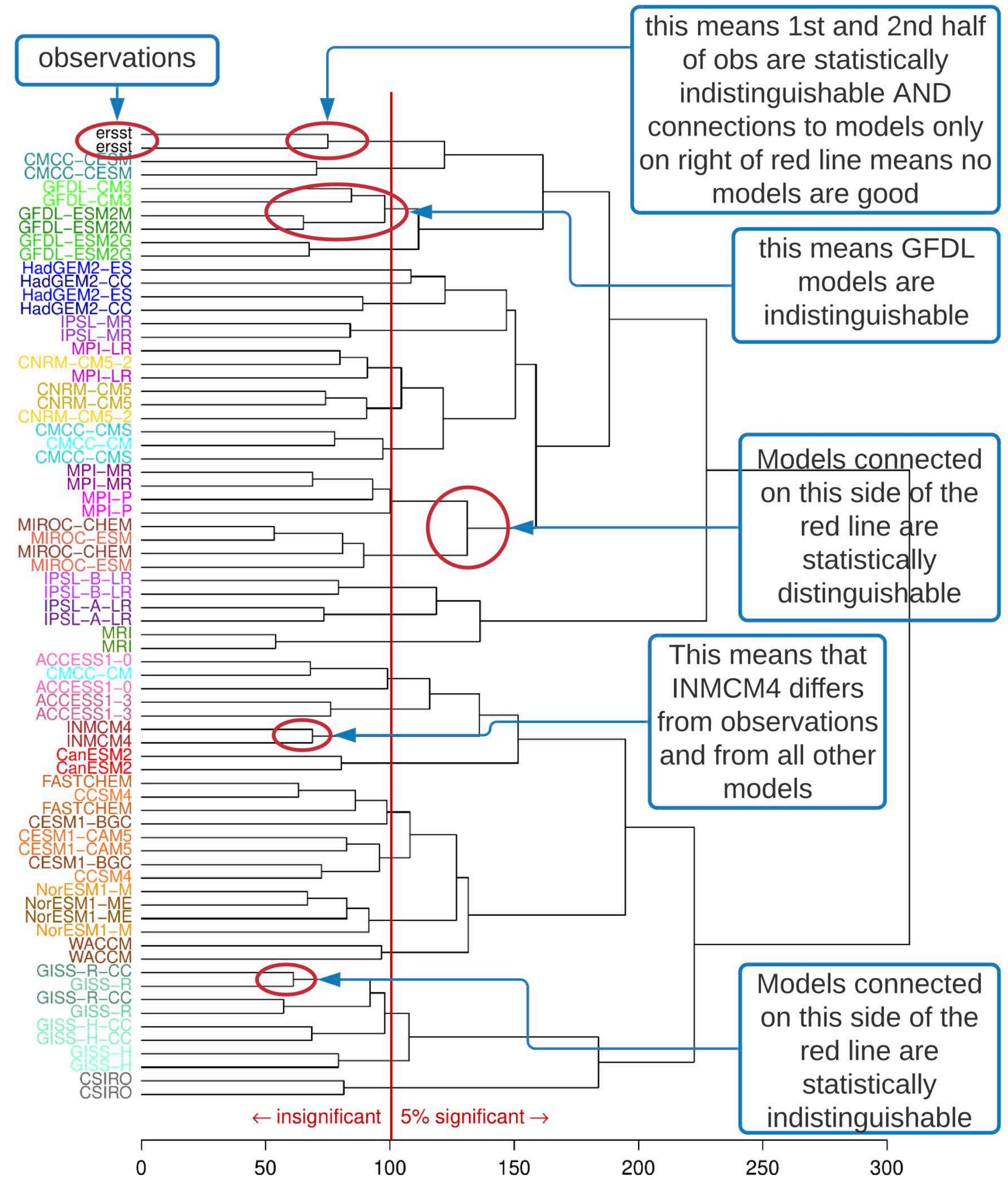

Fig. 1: Dendrogram derived from the deviance matrix between all pairs of VAR(1) models estimated from the first and second halves of the 1854-2018 period (the specific year is not relevant for pre-industrial control simulations). The clusters are agglomerated according to the complete-linkage clustering, which uses the maximum deviance between elements of each cluster. The VAR models contain seven Laplacian eigenfunctions and a second-order polynomial in time is removed. The vertical red line shows the 5% significance threshold for a significance difference in VAR model.

Acknowledgements

This research was supported primarily by the National Oceanic and Atmospheric Administration (NA20OAR4310401).

Conclusions

- Overall, CMIP5 models do not generate realistic simulations of North Atlantic SST (i.e., all connections to ERSST in fig. 1 lie to right of red line).
- Some CMIP5 models differ significantly from all other models (e.g, MRI, INMCM4), indicating that these models not only are inconsistent with observations, but also inconsistent with other CMIP5 models.
- Models from the same institution tend to be indistinguishable from each other (e.g,. GISS, NCAR, MPI, CMCC).
- The dendrogram in fig. 1 is new in that it is based on a dissimilarity measure whose significance test accounts for dependencies in space and time.
- A plot of deviance versus Equilibrium Climate Sensitivity (fig 2) has a significant negative slope, indicating that models that best simulate observations tend to have larger ECS.

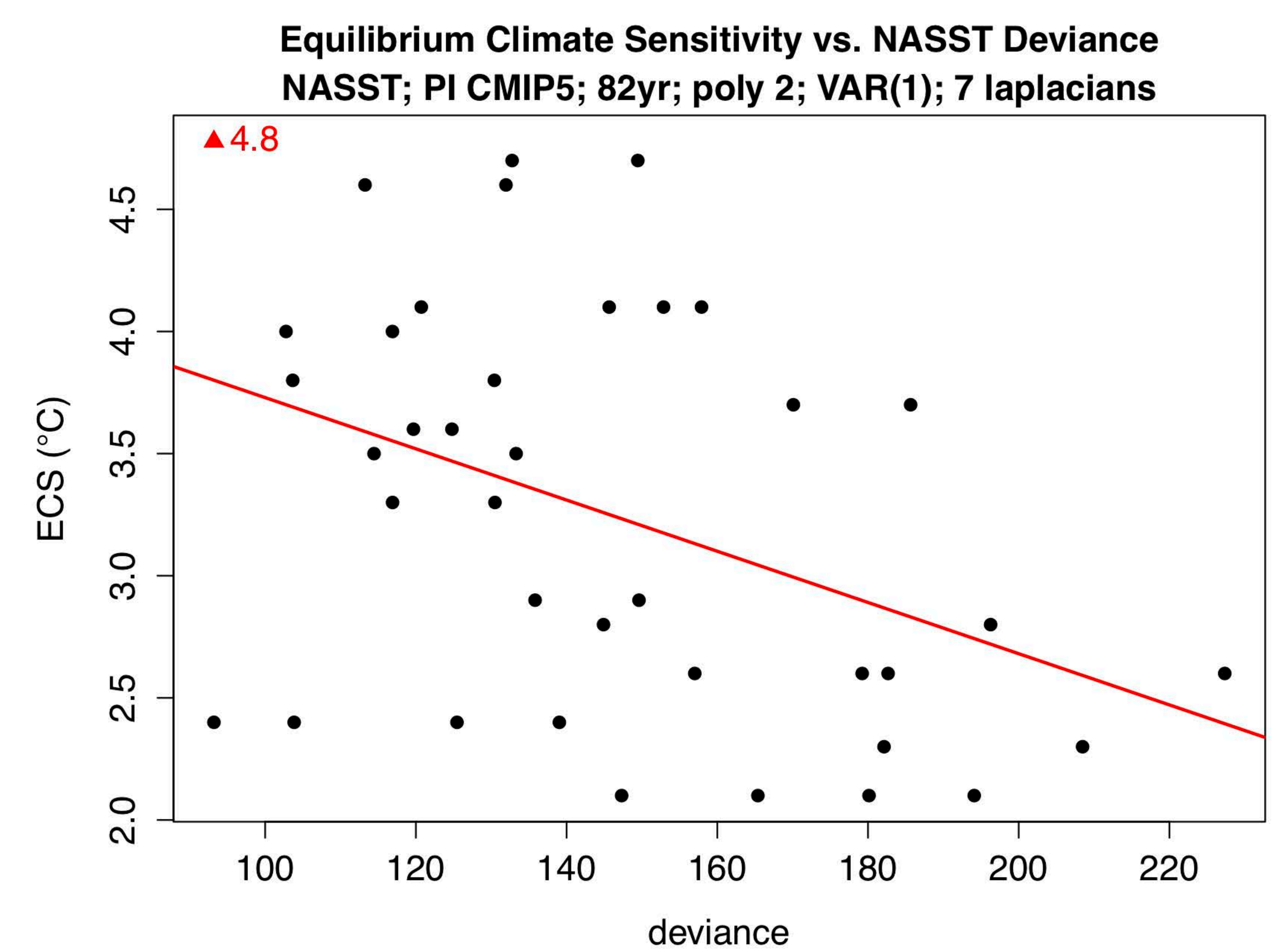

Fig. 2: Deviance versus Equilibrium Climate Sensitivity of CMIP5 models. The deviance is computed for NASST separately for the first and second halves of the 1854-2018 period, which yields two points per CMIP5 model for a total of 72 points. ECS is derived from table 9.5 of IPCC AR5 WG1. The red line shows the least-squares line fit, and the red triangle at the top shows the intercept of the best-fit line.

Papers

[1] T. DelSole and M. K. Tippett. "Comparing climate time series – Part 1: Univariate test". In: *Adv. Stat. Clim. Meteorol. Oceanogr.* 6.2 (2020), pp. 159–175. DOI: 10.5194/ascmo-6-159-2020. URL: https://ascmo.copernicus.org/articles/6/159/2020/.

[2] T. DelSole and M. K. Tippett. "Comparing climate time series – Part 2: Multivariate test". In: *Adv. Stat. Clim. Meteorol. Oceanogr.* submitted (2021).