Climate Statistics
Reproducibility: Updates

Salil Mahajan, Joe Kennedy, Kate Evans, Xylar Asay-Davis
Climate Statistics Reproducibility: Introduction

• **Goal:** Test the null hypothesis that two simulation ensembles belong to the same population

• **Approach:** Multivariate two-sample equality of distribution tests:
  - Energy Test: Evaluate a distance based metric (Szekely and Rizzo, 2005)
  - Kernel Test: Evaluate a kernel function based metric (Gretton et al. 2012)
  - KS Testing Framework: Evaluate each variable separately (Mahajan et al. 2015)

• **Implementation:**
  - Short simulation ensembles (1yr runs, >30)
  - Using global annual means (~150 variables)
  - i.e., 150 dimensions, 30 samples (high dimension, low sample size)

Illustration: Energy Test
Conducted Power Analysis:

- Determine false negative rates (Type II error rate):
  - rate of falsely accepting a null hypothesis
- Determine the required ensemble size for detecting a given magnitude of change
- F1850C5 compset E3SMv1
- ne4 resolution
- Three suites of ensemble simulations:
  - zm_c0_ocn (deep convection scheme, 0.0071 – 0.045)
  - dcs (cloud microphysics, 100 - 400)
  - rhminl (cloud macrophysics, 0.80-0.8975)
- 100-member ensembles for each discrete value of the tuning parameters
- Resampling techniques to determine power

**Climate Statistics Reproducibility: Updates**

*Mahajan et al. 2019, PASC'19*
Climate Statistics Reproducibility: Updates

- **KS Testing Framework** included in E3SM:
  - CIME MVK test to run control and perturbed ensembles
  - **EVV**:
    - Post processing
    - Run test
    - Publish detailed results on a shareable website
  - CIME publish results on cdash

- Future plans:
  - Expand power analysis for other tuning parameters
  - Incorporate Energy and Kernel tests in EVV
  - Provide bootstrapping and power analysis capabilities in EVV
  - Begin conducting ocean simulation ensembles
    - Identify best strategy
  - Port tests for ocean model