Reproducibility of the Results of Climate Modeling Experiments

Dr. Richard Loft
Director, Technology Development
National Center for Atmospheric Research

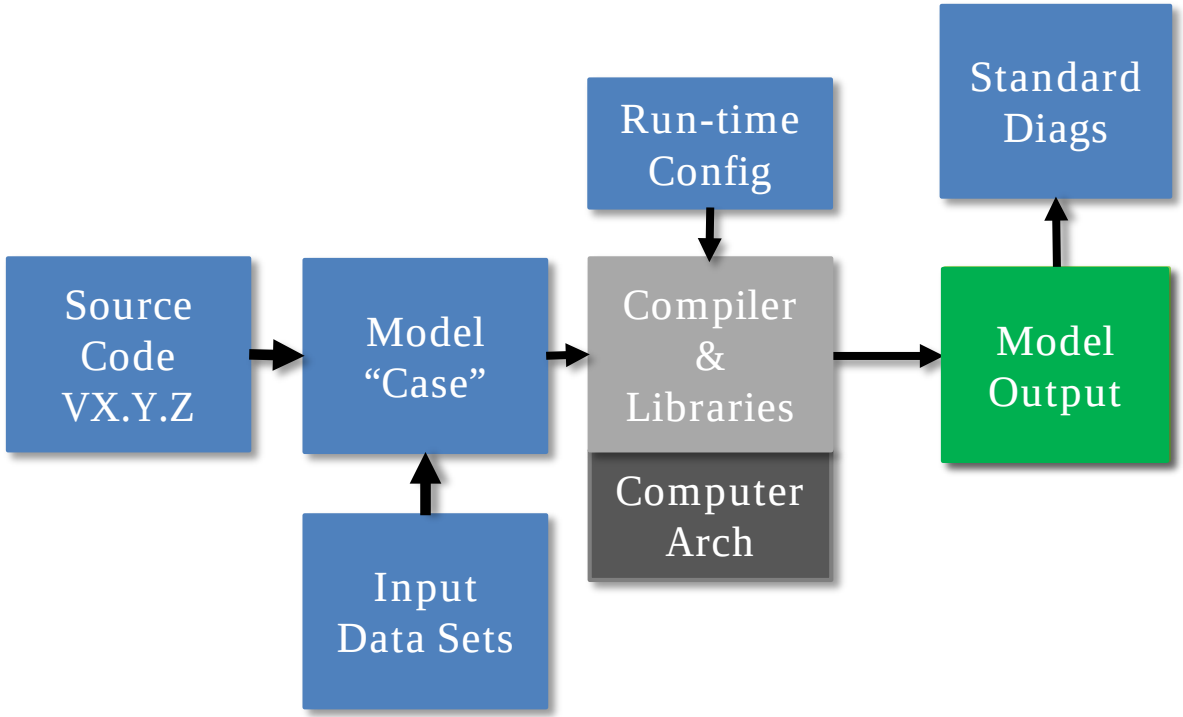
Computational reproducibility: achieves bitwise identical results from the same computation when given equivalent inputs.

- **Why does climate modeling need this?**
  - Sensitivity to initial conditions of the nonlinear governing equations.
- **Why is it hard?**
  - Discrete nature of the floating point representation of a continuum of real numbers means operations are commutative but not associative.
  - Must preserve the order of operations to achieve bitwise identical results.

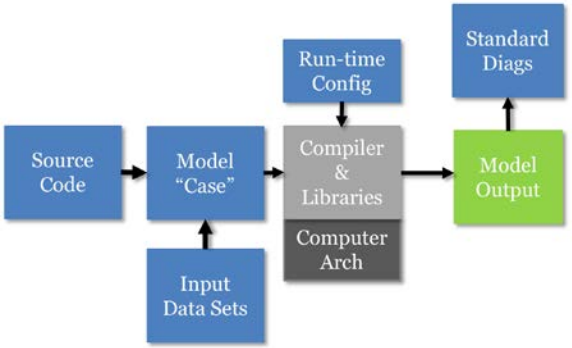
Computational reproducibility: issues

- **Currently, many features of computer architectures are at odds with reproducibility:**
  - Computation: fused multiply-add
  - Memory details: CPU vector unit, GPU warp, cache line size
  - Parallelism: number of processors, network properties (e.g. topology or in-flight operations).
- **In the future, even more problematic reproducibility trends:**
  - Reproducibility of machine learning algorithms (e.g. in training neural networks)
  - Stochastic computing devices (i.e. ones harnessing inherent noise in electronic circuits)
  - Reduced/variable/mixed precision computing devices
  - Undetected soft errors (e.g. cosmic rays striking computer components)

Climate Modeling pipeline



CESM Experimental Reproducibility



- **Configure** The CIME (https://github.com/ESMCI/cime) **Case Control System** (CCS) in CESM permits users to create, configure and build an experimental case with just 3 commands.
- **Customize and Document** The resulting experimental sandbox (i.e. **CASEROOT**) permits users to customize and document their experiments.
- **Experimental provenance is provided** by enabling users to store these experimental configurations in a database.
- **Workflow provenance** Standard diagnostic packages also can be attached to a CASEROOT and run - thereby creating reproducible end-to-end workflow capability.

Climate Model Reproducibility (CESM)

Runtime Config

Results should not depend on the details of how the experiment is conducted. This requirement is routinely tested for every snapshot that is created of the model system.

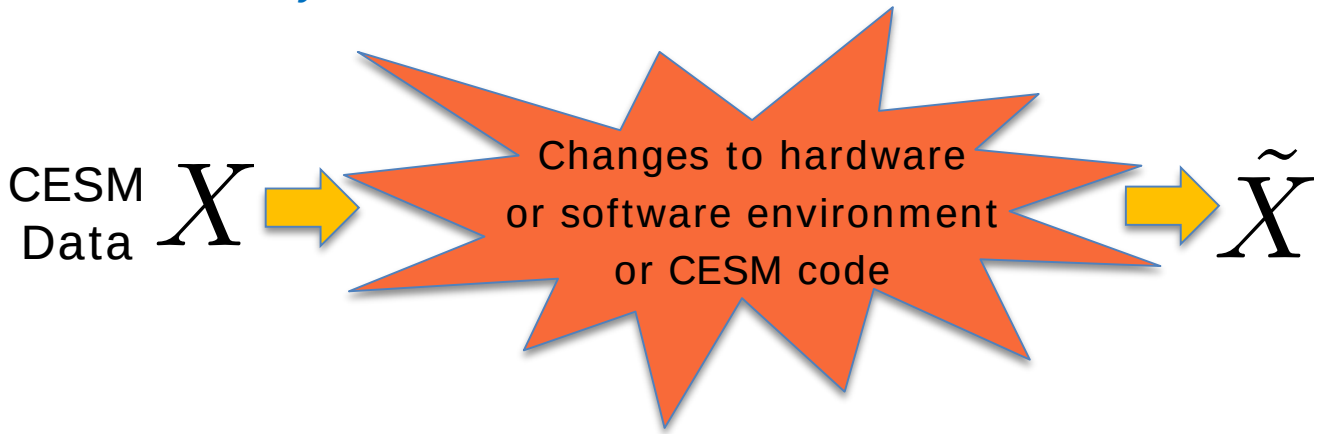
Aspects of Reproducibility

- **Restart:** Model simulations must have bit-for-bit restart functionality
- **Threading:** Changing the thread count in any threadible component must not change answers.
- **MPI Tasks:** Certain CESM components (e.g. atm, land) answers are invariant w.r.t. MPI task count. This is not true for all CESM components (e.g. ocn, sea ice) however.
- **Component Layout:** For a fixed MPI tasks/threading configuration of the coupled system, answers should not vary if the components are run concurrently, sequentially or in a mixed concurrent/sequential layout.
- **I/O:** Answers must not depend on the choice of IO backend, i.e. parallel or serial (e.g. pnetcdf v.s. netcdf)

Software Quality Assurance for CESM

Computer Arch

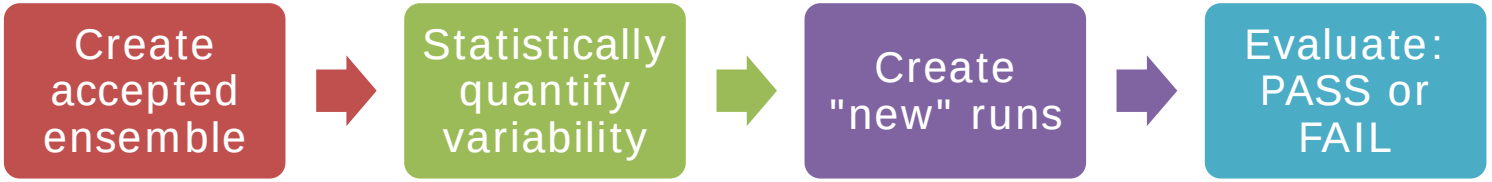
Goal: Insure that changes during the CESM development life cycle do not *adversely* affect the results!



Question: If $X \neq \tilde{X}$, is the new result correct?

Does the new data represent the same climate?
Is it *statistically distinguishable* from the original?

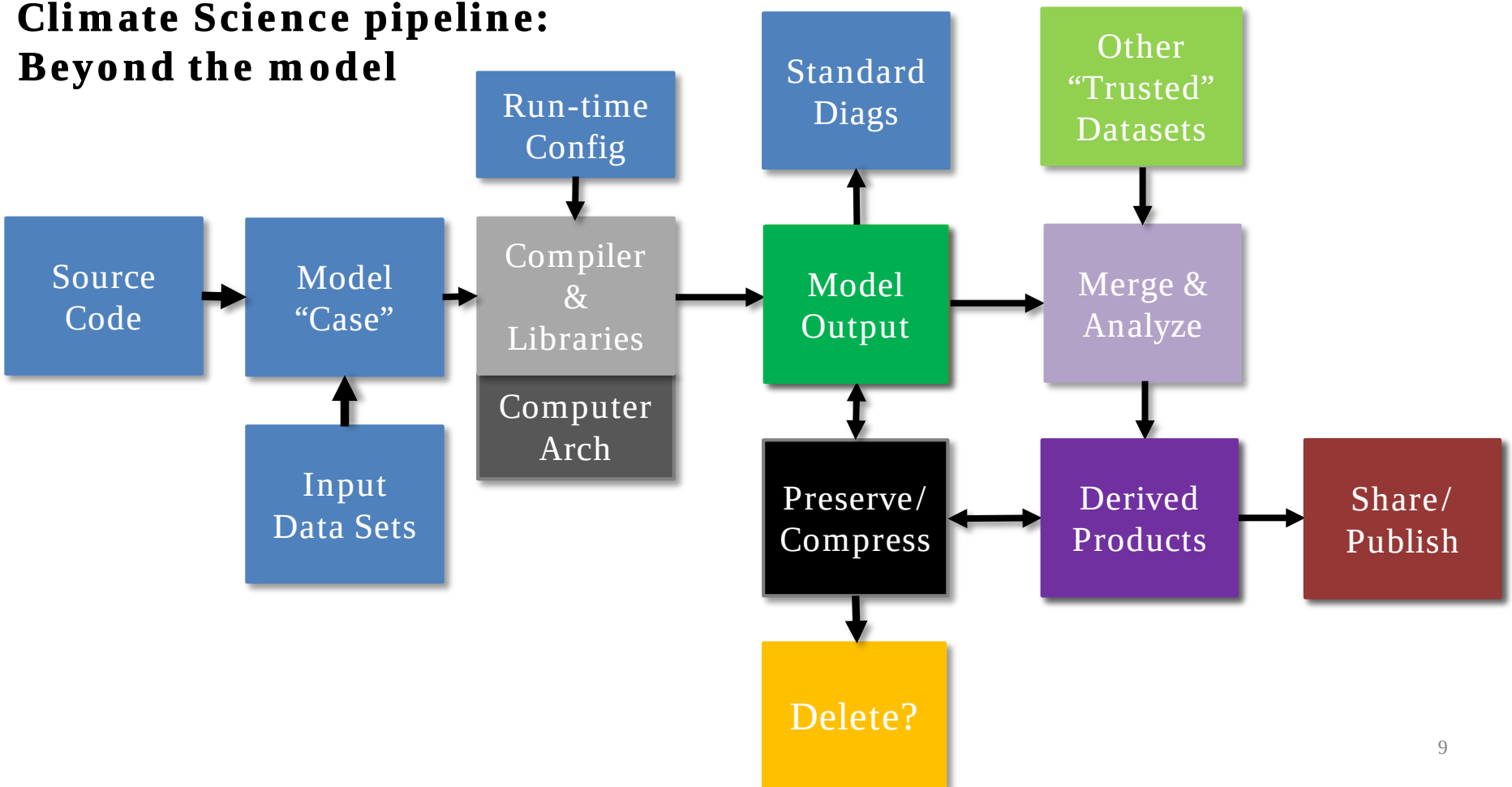
CESM Ensemble Consistency Test (ECT)



Highlights:

- **Detects changes in relationships between variables (PCA)**
  - even when individual variable distribution looks right!
- **Objective, automated, user-friendly, fast**
  - climate science expertise not required
- **Allows "letting go" of bit-for-bit reproducibility**
  - facilitates optimization, multiple hardware/software platforms

Climate Science pipeline: Beyond the model



Lossy Data Compression (Work in progress)

Preserve/ Compress

- **Motivation**
  - Increasing resolution and computational power lead to more and more data. *And there is no end in sight!*
  - Can we use lossy compression to reduce climate storage needs …*quickly, and without (negatively) impacting science results?*
- **Opportunity**
  - An average 5x compression factor observed across climate variables using ***fpzip***
- **Complications**
  - Max compressibility characteristics of variables differ a lot
  - Different compression algorithms better suited to certain types of variables
  - Ideal to use a set of methods tailored to each variable.
- **Can lossy compression be made reproducible?**

Compression Metrics: evaluating information loss

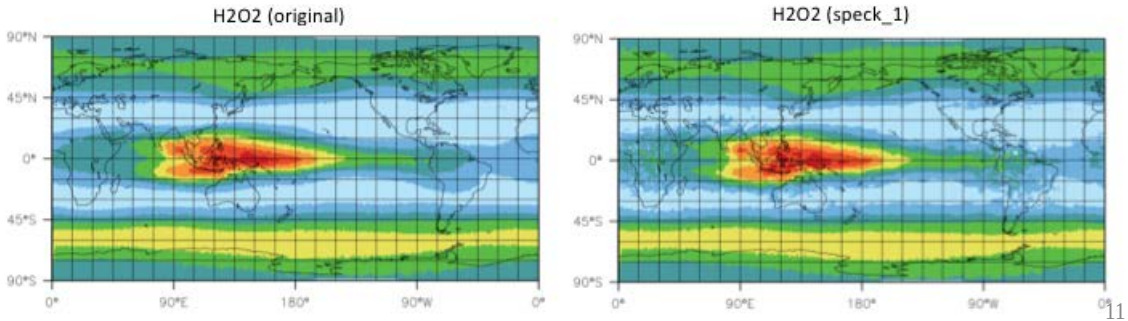
- Statistical test suite to measure different aspects of data
- Not ensemble-based (use only the fields themselves)

(1) Pearson correlation coefficient

(2) Kolmogorov-Smirnov (K-S) test

(3) Spatial relative error

(4) Structural similarity index (SSIM)



Ingredients for a Reproducible Analytics Platform

Merge & Analyze

- **Discoverable**
  - You can't reproduce what you can't find.
  - e.g. **schema.org/dataset** proposed in 2012
- **Citable** data and tools via DoI's.
- **Peer Reviewed**
  - e.g. Peer reviewed **Jupyter notebooks**?
  - Upwelling of interest in this topic (Pérez, Mietchen).
- **Parallel**
  - Analyses needs to be parallel to handle big climate datasets.
  - e.g. **PanGeo** project (xarray, DASK)
  - Can a parallel analytics system be reproducible?

Thanks! Questions?