Moving beyond bit-for-bit

Both internal (code) and external (machine) changes can affect a climate model’s solution to a particular simulation

There are three types of changes:
1. Technical changes that continue to produce bit-for-bit identical solutions
2. Non-identical changes that produce a statistically similar solution
3. Changes that lead to a different solution

Only type 3 changes requires in-depth analysis of the changes, but there is no current capability to distinguish between type 2 and type 3 changes

We will enhance ACME’s testing infrastructure to provide a robust climate reproducibility testing capability

Characteristics of successful testing

Concurrent to development
- Integrates into the development cycle
  - Useable, portable, flexible, extensible
  - Run frequently (easy) to continuously (scriptable)
  - Minimal time to solution

Granular
- Functions → processes → components → model

Informative
- Clear context
- Detailed analysis
- Appropriate metrics

Shareable
- Discussions across many institutions

Integrating with ACME

Three main steps for each test:
1. Launch the test
   - Add a new ensemble test type to CIME
   - Strategy for each type of climate reproducibility test may be needed
2. Post process the test ensemble
   - Launch automatically when tests finish
   - Integrate with CIME and/or ACME post processing
3. Analyze the test results
   - Quickly tell if tests pass/fail
   - Detailed info on fail to help developers find bugs

Questions to be answered:
- When should they be run and by whom?
  - Developers’ vs integrations’ test suite
- What are the costs?
  - Both computational and personnel
- Where are the tests applicable?
  - What types of issues can be identified
- What do the developers need and want?
  - Successful tests are used