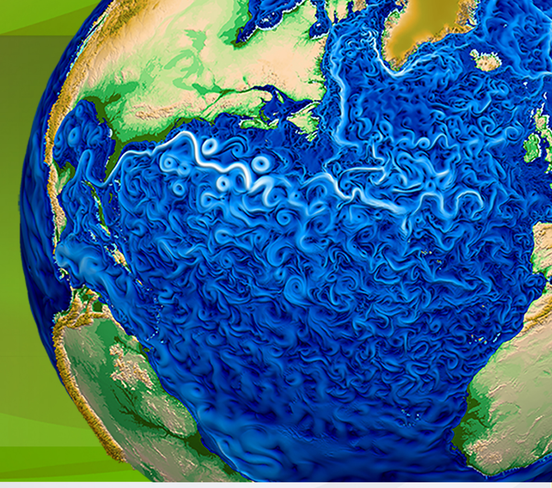
Climate reproducibility testing with EVE

Joseph H. Kennedy, Salil Mahajan, Katherine J. Evans, Peter Caldwell



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

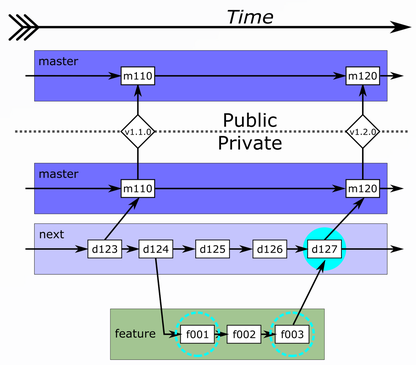
Concurrent to development
- Integrates into the development cycle
  - Useable, portable, flexible, extensible
  - Run frequently (easy) to continuously (scriptable)
- Minimal time to solution

Granular
- Functions → processes → components → model

Informative
- Clear context
- Detailed analysis
- Appropriate metrics

Shareable
- Discussions across many institutions

ACME's development cycle

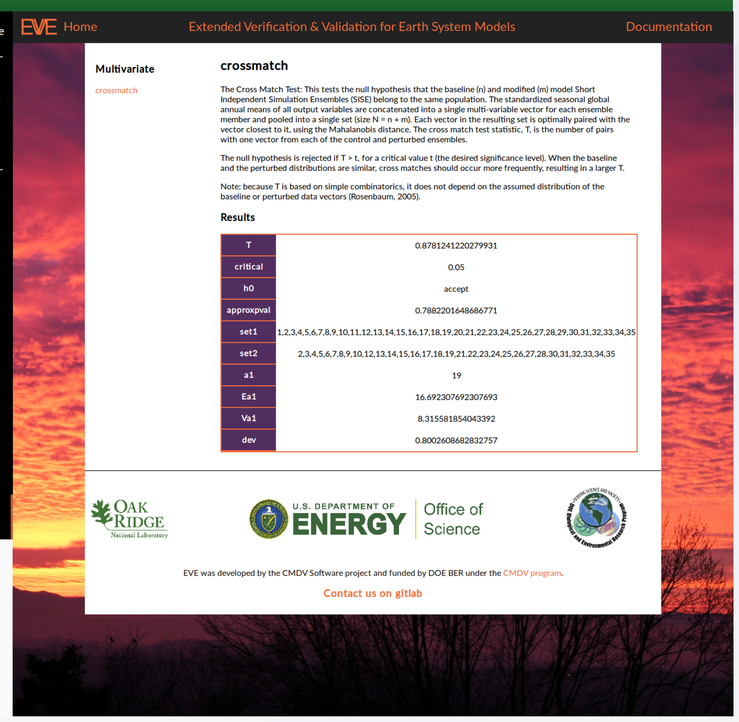


Developers' test suite
- Smaller set of tests; developers preferred machine
- Only tests feature changes
- Should exercise only feature relevant components

Integrators' test suite
- Larger suite of tests; multiple machines
- Tests integration of feature into next
- Should exercise whole model for unintended issues

Using EVE

```
(EVE-LIVVdev) fjk@pc0101123:~/Documents/Code/EVE/eve$ ./test -e exte
nsions/crossmatch.json
-------------------------------------------------------------------------
                    _____ _____ _____
                   |  ___|  ___|  ___|
                   | |__ | |__ | |__
                   |  __||  __||  __|
                   |_____|_____|_____|

   Extended Verification and Validation for Earth System Models
-------------------------------------------------------------------------

  Current run: 2017-06-02 14:11:12
  User: fjk
  OS Type: Linux 4.4.0-78-generic
  Machine: pc0101123

-------------------------------------------------------------------------
  Previous output data found in output directory!
   Backing up data to:
   /home/fjk/Documents/Code/EVE/eve/vv_2017-06-02_20170602_125042
-------------------------------------------------------------------------
-------------------------------------------------------------------------
   Beginning extensions test suite
-------------------------------------------------------------------------

     Cross Match Test:
       Null hypothesis: Accept
       Critical value: 0.05
       Test statistic: 0.8781241220279931

-------------------------------------------------------------------------
   Extensions test suite complete
-------------------------------------------------------------------------

-------------------------------------------------------------------------
Done!  Results can be seen in a web browser at:
   /home/fjk/Documents/Code/EVE/eve/vv_2017-06-02/index.html
-------------------------------------------------------------------------
(EVE-LIVVdev) fjk@pc0101123:~/Documents/Code/EVE/eve$
```



Above: EVE's command line interface was used to run the crossmatch test, one of the multivariate climate reproducibility tests that will be available. A general summary of the test results is displayed, and the path to the generated web output is given.

Right: EVE's web output showing detailed results of the crossmatch test, including a description of the test.

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*

ACME Accelerated Climate Modeling for Energy

For additional information, contact:

Joseph H. Kennedy
Computational Scientist in Glaciology

Oak Ridge National Laboratory
(206) 714-3375
kennedyjh@ornl.gov

climatemodeling.science.energy.gov/acme

U.S. DEPARTMENT OF ENERGY