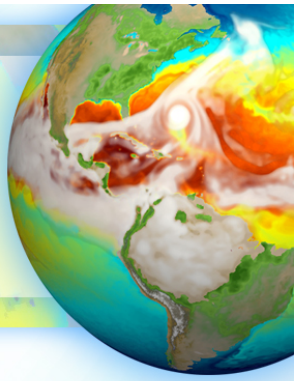
Infrastructure Group
Phase 2 review



Rob Jacob and Jill Zhang

(on behalf of the entire Infrastructure Group)

E3SM Review

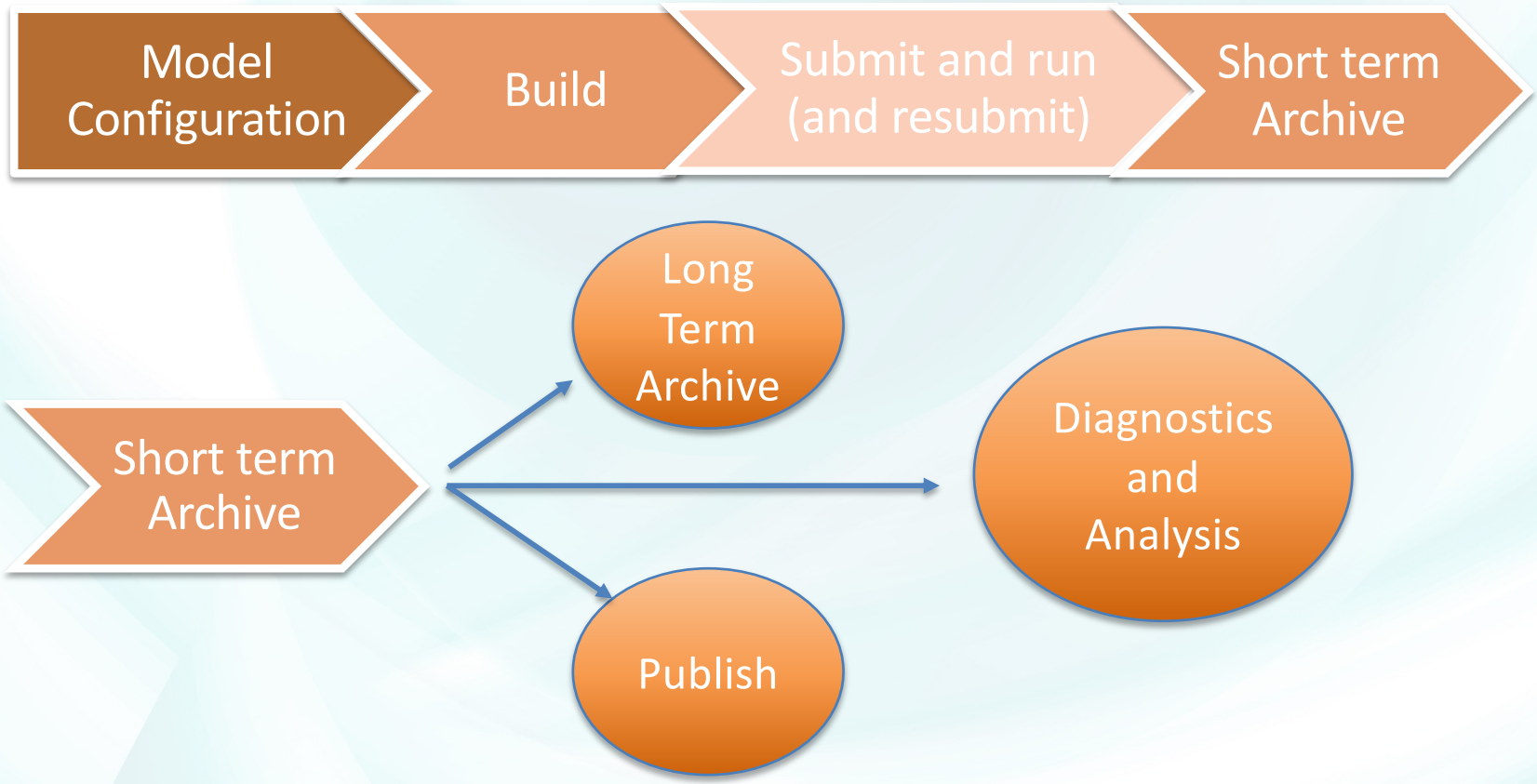
November 9, 2020

Infrastructure Group responsibilities

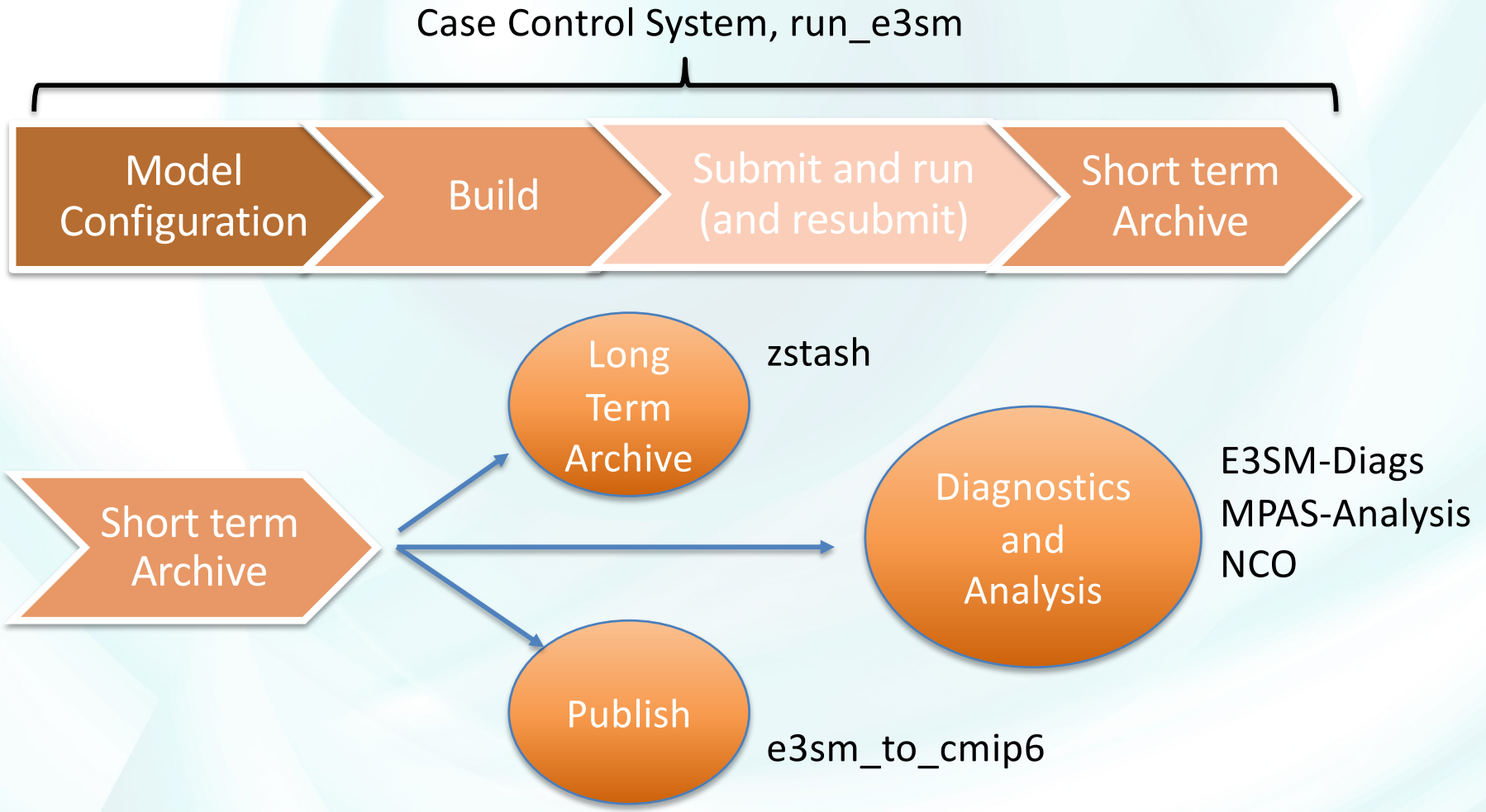
- Develop, maintain and support software that is needed for E3SM but is **not** part of the main prognostic models.
  - Configure, build, test software, driver/coupler
  - diagnostics, analysis, workflow tools
- Manage data sets (input and output)
- Define, document, manage the process and procedures used in software development within the E3SM Project.

Everything we do should help make the model development, simulation and analysis happen.

Typical ESM workflow



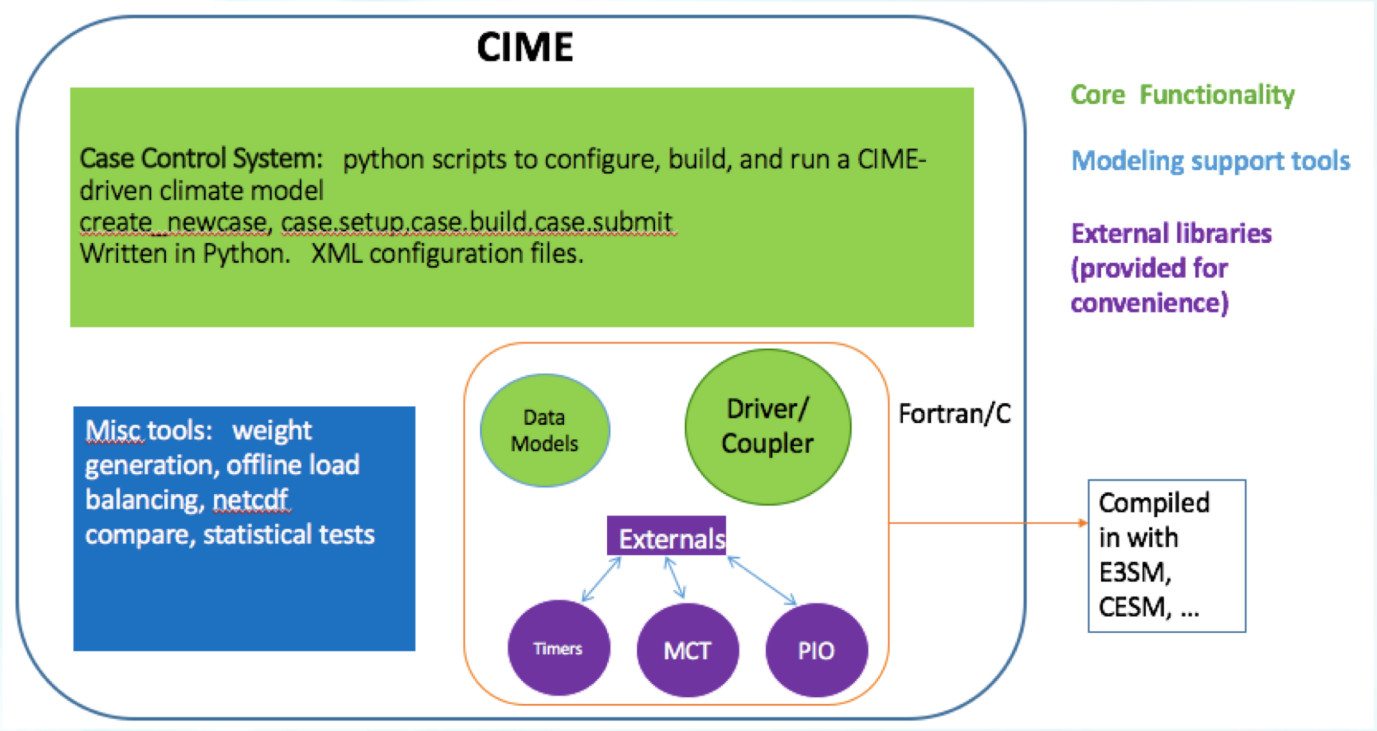
IG provides or co-develops tools for every phase



Outline

- What is the Infrastructure Group (just did that)
- Phase 2 accomplishments in configure/build/test/driver/coupler (CIME)
- Phase 2 accomplishments in diagnostics/analysis/workflow
- Phase 2 data management
- External collaborations

Common Infrastructure for Modeling the Earth



CIME packages together all the parts of a climate model that are not the main simulation components.

- Configure, build, test, driver/coupler in one repository
- Developed jointly with CESM Software Engineering Group.

The CIME Case Control System

CCS is the workflow control code for creating an experiment in a CIME-enabled model like E3SM (and CESM).

- Written in python with XML files for configuration
- Allows configuring, building and running an experiment

with a handful of commands

  - Create_newcase: make the basic case
  - ./case.setup: create batch scripts for submission to queue
  - ./case.build: build the model
  - ./case.submit: submit to the queue on HPC system
  - These can be made part of a script (e.g. run_e3sm)
- **In phase 2**
  - Sped up ./case.build (2x in some cases) by replacing Make with CMake and rewriting some old perl in python.
  - Introduction of CMake also increases the maintainability. Shorten spinup time for new SE's who are familiar with CMake

E3SM develop and test process

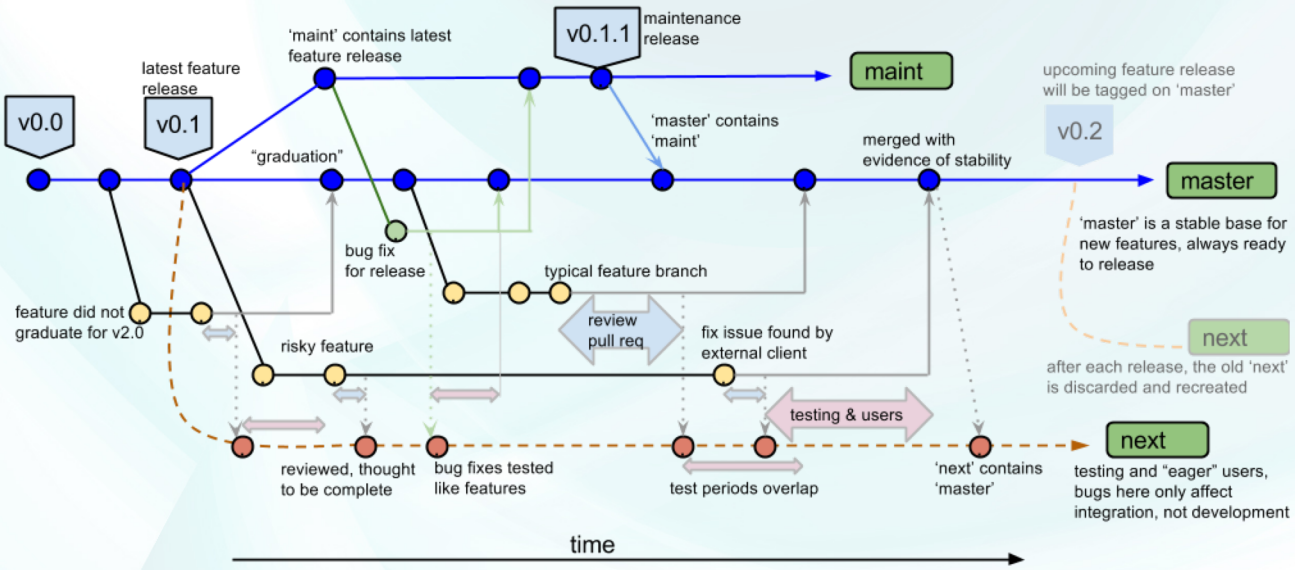
- How do you allow dozens of developers to work on E3SM:
  - At the same time?
  - At multiple institutions?
  - On multiple machines?
  - Without breaking the code?
  - Without slowing down progress?
  - To make changes with confidence?

Git, gitworkflow, Github, Testing, Testing and Testing.

- Assertion: E3SM emphasis on and support for testing is unique in the global climate modeling community.
- IG defines, documents and maintains this system

E3SM development

- Git: Distributed version control system scales to large numbers of users/machines
- Gitworkflow: process for making changes with git branches (used by Linux kernel)
- Github.com: coordinates development on a single git repo. Web interface to git commands enhances gitworkflow. Developers make a “Pull Request” to integrate a feature branch. Integrators handle merging to next/master.



In phase 2:

- Increased available documentation
- Added more integrators (for atmosphere; need more for land, ocean)

E3SM test system

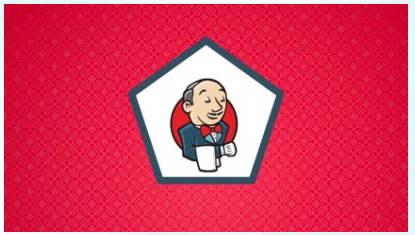
CIME provides about 30 system test types. Combine test types with different model configurations (fully coupled, atmosphere-only) and resolutions used by E3SM to form a single test.


```
> ./create_test SMS_Ld2.ne30_oECv3_ICG.A_WCYCL1850S_CMIP6
```

Group tests in to **test suites**

- e3sm_developer: run in an hour or so by individual developer on their development cluster (or maybe a big workstation)
- e3sm_integration: run overnight on version of the day (VOTD).
- e3sm_highres: Run high-resolution configs on high-node-count platforms that can accommodate them.
- e3sm_prod: just the production configurations.

Use Jenkins service to automatically run all test suites on our supported platforms



E3SM test system

- Results reported every day to Cdash dashboard (another widely used standard)

		Configure		Build		Test			
Site	Build Name	Error	Warn	Error	Warn	Not Run	Fail	Pass	Start Time
anvil	e3sm_prod_next_intel	0	0	0	0	0	3	0	4 hours ago
compy	e3sm_integration_next_pgi	0	0	2	0	0	7^{+2}	80_{-2}	22 hours ago
compy	e3sm_bgcprod_maint-1_1_intel	0	0	0	0	0	0	4	21 hours ago
cori-knl	e3sm_prod_next_intel	0	1	0	0	0	1	2	Sep 24, 2020 - 00:46 EDT
cori-knl	e3sm_prod_maint-1_0_intel	0	0	0	0	0	0	3	Sep 24, 2020 - 00:05 EDT
mappy	e3sm_developer_master_gnu	0	0	1	0	0	0	45	16 minutes ago
mappy	e3sm_developer_next_gnu	0	0	1	0	0	0	45	22 hours ago
sandiatoss3	e3sm_integration_next_intel	0	0	1	0	0	4	83	21 hours ago
sandiatoss3	e3sm_integration_master_intel	0	0	1	0	0	4	83	21 hours ago
sandiatoss3	homme_integration_next_intel	0	0	0	0	0	1_{-1}	1^{+1}	23 hours ago
sandiatoss3	homme_integration_master_intel	0	0	0	0	0	0	2	1 hour ago

- **In Phase 2:**
  - Added more test suites (bgcprod, gpu, climate-significance)
  - Added fast unit tests (in SCREAM)
  - Added another machine (compy) and compiler (pgi)
  - Always trying to balance coverage, test length, available HPC resources.

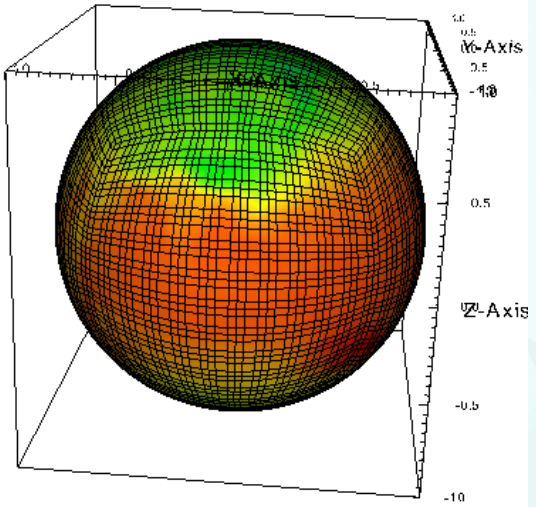
E3SM machine support

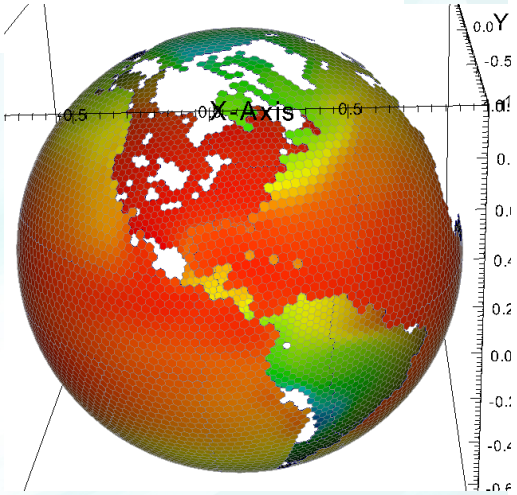
The Infrastructure Group includes machine Points of Contact: individuals who help diagnose problems on specific machines. Elevate to system support if necessary. Track testing.

- Cori-knl at NERSC Noel Keen
  - 9688 Intel Xeon Phi "Knights Landing" nodes
- Cori-haswell at NERSC Noel Keen
  - 2388 Intel Xeon Haswell nodes
- Compy at PNNL Bibi Mathew
  - 460 Intel Skylake nodes; new in phase 2
- Theta at ALCF Jason Sarich
  - 4392 Intel Xeon Phi "Knights Landing" nodes
- Summit at OLCF Min Xu
  - 4608 IBM Power 9 (2) and NVIDIA V100 (6) nodes
- Anvil at ANL Jason Sarich
  - 240 Intel Broadwell nodes
- *Coming soon*: *Chrysalis* POC TBD
  - 512 AMD EPYC nodes

Phase 2 coupler/driver progress

- Slowly replacing MCT with MOAB: Mesh Oriented datABase.
  - Online interpolation weight generation through complete mesh representation.
  - Lower memory
  - Faster communication routines.
- Pivoted to support tri-grid
- Made offline tool for weight generation: mbtempest
  - Uses MOAB mesh intersection, TempestRemap weight calculation, MPI-parallel.
- Co-developing task mapping with NGD Soft-Alg task
- Will be an option in v2 release.
- MOAB is library funded by other offices. (ASCR, NE). Some of others used: Albany, Kokkos, PETSc, ADIOS

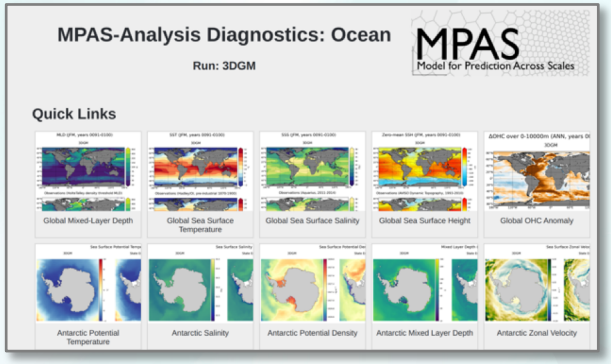


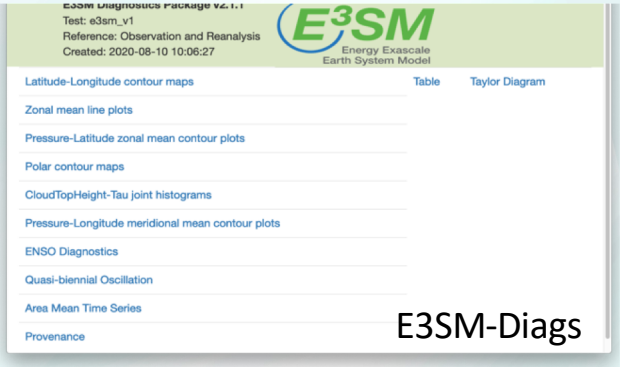


Analysis and Diagnostics Tools

Tools included in standard post-processing workflow:

- **NCO:** Regidding, Climatology Generation and Timeseries Extraction, support **all E3SM components**
- **E3SM-Diags:** Diagnostics package for **Atmosphere**, extending the support to **Land** and **River**
- **MPAS-Analysis:** Diagnostics package for **Ocean** and **Sea-ice**
- **Zstash*:** Long term archiving tool
- **E3sm_to_cmip***: convert output to **CMIP** format (new)
- **E3SM-Unified:** Conda package for distributing all the above and more (RGMA CMIP6 hackathon)





E3SM-Diags

More tools used in E3SM evaluation /connected to E3SM:

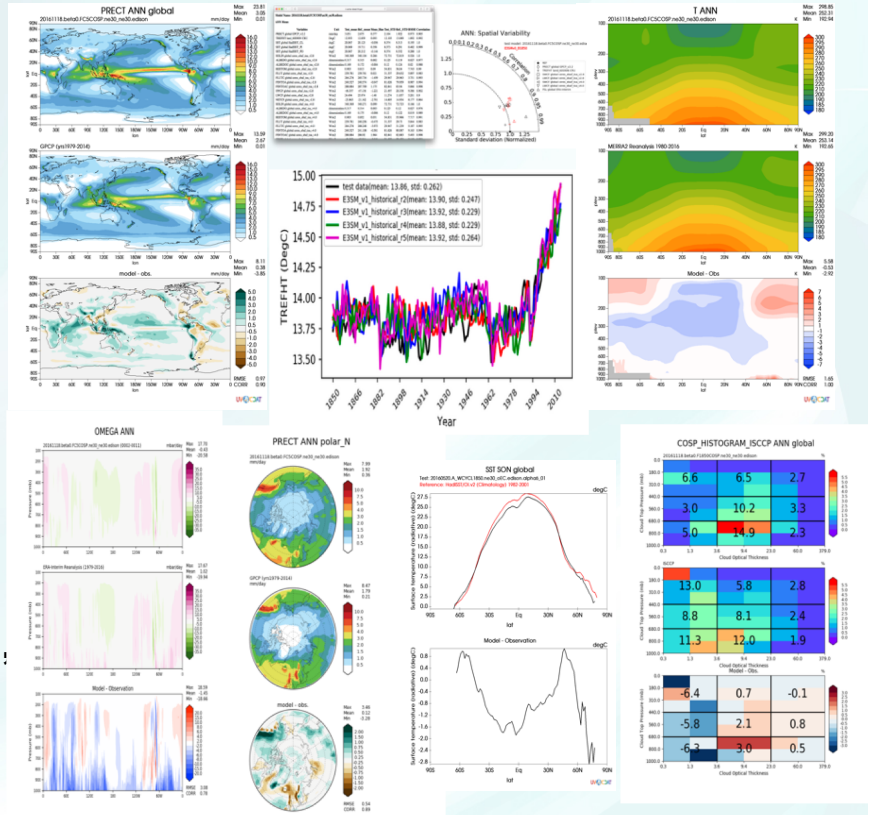
PMP ILAMB ARM ARM-Diags AMWG CVDP LIVVkit ESMValTool Earth System Model eValuation Tool

E3SM-Diags: E3SM Diagnostics Package

Comprehensive E3SM atmosphere diagnostics, analysis data and toolkit

- A **modern, Python-based** diagnostics package for evaluating earth system models
  - Flexibly add new obs/variables/diagnostics, modify figures.
  - Easy installation, configuration and execution, with provenance saved
  - Uses multi-processing
  - Up-to-date analysis data

Other tools have similar development philosophy (python, configurable, extendable, up-to-date datasets, DOE datasets)



Phase 2 Progress Highlights

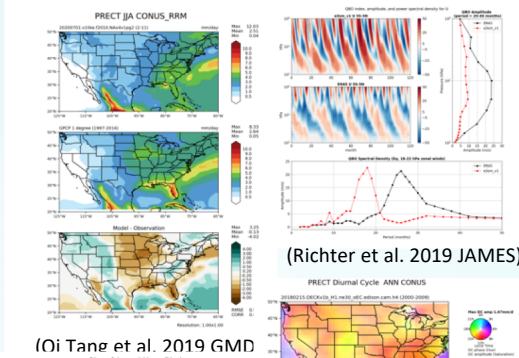
New features with focus on **RRM**

- RRM support for **North America** (potentially other regions) with **high resolution** atmosphere observation and reanalysis data.
- New types of analysis for diagnosing and improving biases as well as understanding processes relevant to **coastal North America**.

More diagnostics for **higher spatial and temporal** data

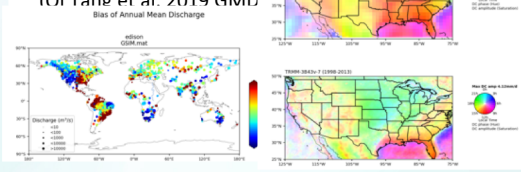
- **Ultra-high resolution grids** support in NCO.
- Sub-monthly capability for **diurnally cycle** evaluation
- Diagnostics on **ENSO and Quasi-biennial Oscillation**
- High-density **streamflow gauge** data for hydrologic processes
- New **ocean analysis** on eddy kinetic energy, time series of volume transport and ocean basin means of temperature and salinity

Diags harvested from our papers



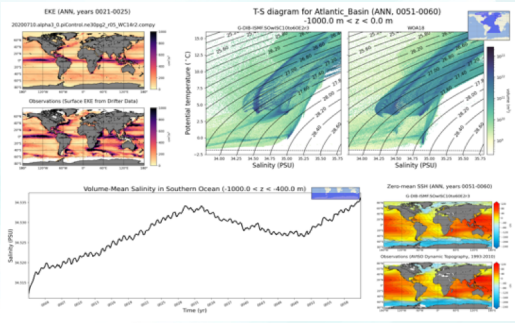
(Richter et al. 2019 JAMES)

(Qi Tang et al. 2019 GMD



(Golaz et al. 2019 JAMES)
(Caldwell et al. 2019 JAMES)

(Xie et al. 2019 JAMES)



Examples of new ocean diagnostics (MPAS-Analysis)

Phase 2 Progress Highlights

E3SM UNIFIED ENVIRONMENT
CONDA

Performance Improvement

- **All** tools support single-node **task parallelism** (NCO already has MPI support)
- Preliminary investigation of **parsl** for multi-node task parallelism for diagnostics packages

Portability and Provenance

- **GPU-offloading** to appropriate loops of the NCO regridder.
- **E3SM-Unified** conda package continues to be the most-used portability solution. A container to test E3SM-Unified and a container is available for E3SM-Diags.
- Provenance (config file, commands, data) are saved for **reproducing diagnostics**.

Community Support

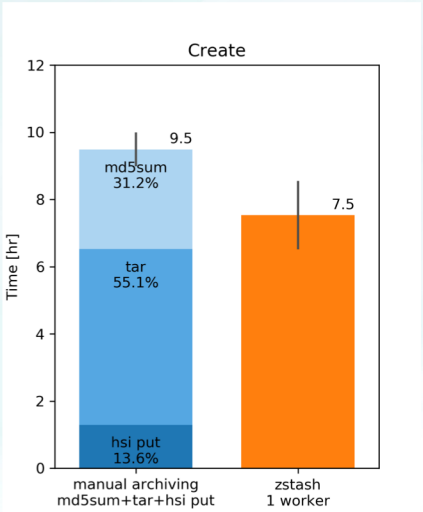
- Open development on GitHub
- **Tutorials and Documentation** are geared to new external users.

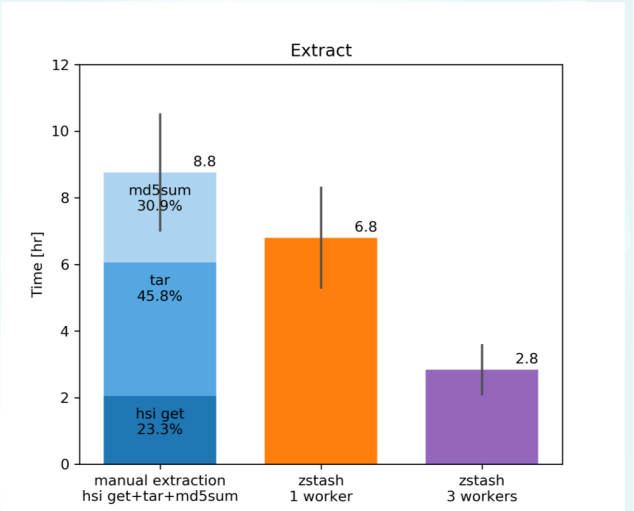
zstash: HPSS long-term archiving tool

Developed from a prototype to a production software during Phase 2

- A Python tool for long-term HPSS archiving of E3SM simulations with features:
  - **Standard tar files** generated locally before transfer to HPSS
  - **Checksums (md5)** are computed *on-the-fly* during archiving and **verified** on extraction
  - Metadata stored in a sqlite3 index **database,** enabling **faster retrieval** for target files
  - **Parallel extraction and verification** for increased performance

  - **All E3SM v1 production simulations have been archived on NERSC HPSS using zstash**
  Will continue to be standard for v2 sims.





Performance data for a 4 TB archive consisting of more than 13,000 files. Mean and range of three realizations on NERSC's Data Transfer Nodes (dtn).

Unplanned Phase 2 effort: Data Wrangling

- Implemented spinning-disk E3SM Archive (at LLNL) to unify data access (~1PB on-hand)
  - Manual retrieval of NERSC HPSS zstashed archives conducted “up front”
  - Non-standard archive structures and content “mapped” for automated access
  - Data validation and issue remediation (missing timesteps) conducted well-prior to publication requests
- Processing for CMIP submission
  - CMIP6 processing is now (mostly) automated and parallelized
    - New “e3sm_to_cmip” script created
  - Support Ecosystem and external projects on CMIP processing/publication
- Supports “On Demand” publication of well-prepared datasets

Automation: Once the datasets have been copied to the LLNL warehouse, much of the extraction and publication has been automated. Additional planned improvements will close the loop and allow for automated post-processing and publication.

The NERSC HPSS (1st centralized archive)	The LLNL E3SM Data Achieves	The LLNL E3SM Warehouse	Publication of Warehouse
[Production runs are archived using **zstash**] **Globus** Transfer to NERSC Verify integrity of archives using **zstash verify** Update archive path and publication specs (Procedures executed by simulation POCs)	[Notification of new datasets] **Globus** transfer from NERSC Verify integrity of archives using **zstash verify** **Archive mapping** for variably-structured archives to facilitate survey and automation	Data extraction using **ztash extract** Validation and fixing Time coordinate if needed Post-processing: ➢E3SM project: regidding/Climo/Time-series using **ProcessFlow** and Generate mapfiles ➢CMIP6: **e3sm_to_cmip**	[Publication Authorized] Move files to publication directory path Finalize mapfiles Run **esgf-publisher** ESGF Node Check Inventory status update

Data Publication Progress

(as of Oct. 24th 2020)

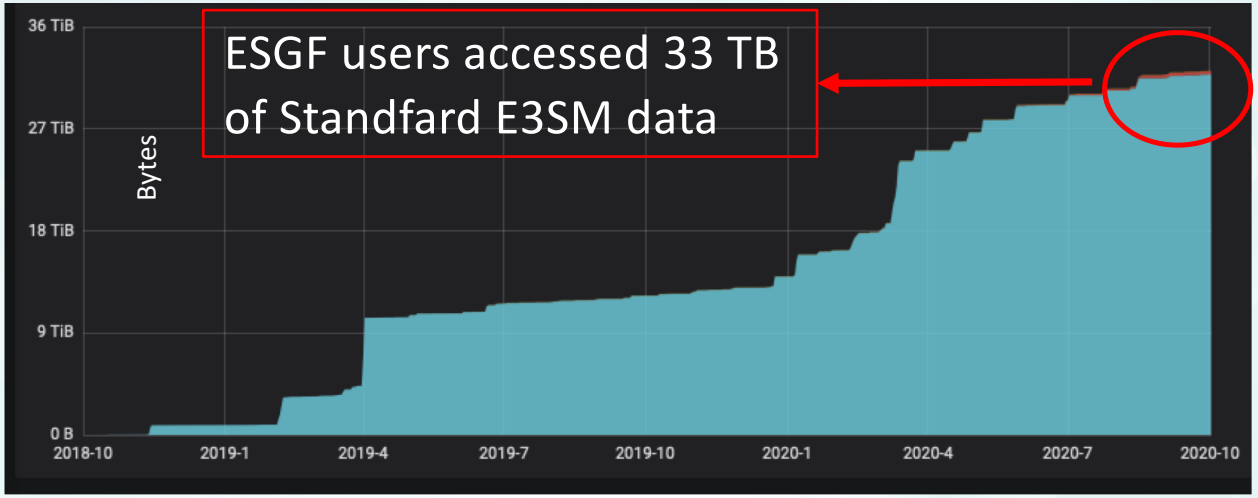
CMIP6

- 17 simulations
- 105 variables per simulation
- 1808 datasets
- 39,550 files
- 6.55TB of data

Publications include BGC, Cryosphere, DECK, and HighResMIP simulation campaigns.

Standard

- 38 simulations
- 386 datasets
- 510,471 files (300,000+ this year)
- 360TB of data



Coordination with DOE-funded external projects

- We set the procedures other projects should follow to get their code in to E3SM (laying groundwork for coordination)
  - On the few external-produced PRs that have been submitted, integrators have had to work at varying levels to get code in to master.
  - Our extensive documentation is now public.
  - Tutorials are geared to new external users.
- Open development and testing help external projects follow code.
- Requests for features in diagnostics from external projects are honored time permitting.

Coordination with DOE-funded external projects

- Examples of other projects we coordinate with (not complete)
  - ECP: I/O developers in ECP and E3SM are the same people. Developing SCORPIO for joint needs.
  - SciDAC CANGA: jointly-funded staff will coordinate incorporation of new interpolation techniques, use E3SM for testing time integration.
  - ESGF: Consulting on Globus role in ESGF re-design
  - SciDAC-Prospect: Adding NCO features for LivvKit
  - Multiscale: library-ize TempestRemap for use in MOAB, NCO
  - PCMDI: Helping define specifications for CMIP6; Coordinating obs data with PMP
  - ARM: co-funded staff developing ARM-diags and adding to E3SM-Diags.
  - RUBISCO: incorporating ILAMB in to diagnostics workflow

Contributions to v2 and v3

- Development process defined in v1 has worked well in v2.
  - We will continue to expand/refine our development process for future versions.
  - Support development with integrators and machine POCs
- Diag packages will keep incorporating features needed for v2 analysis.
  - Will work on v3 as those are defined.
  - Support actionable metrics in diagnostics packages

Summary

- IG will continue to provide crucial software and support for E3SM science, both development and production.