

Running and Analyzing E3SM: Water Cycle Group's Approach

Ryan Forsyth (he/him/his)

Chris Golaz (he/him/his)

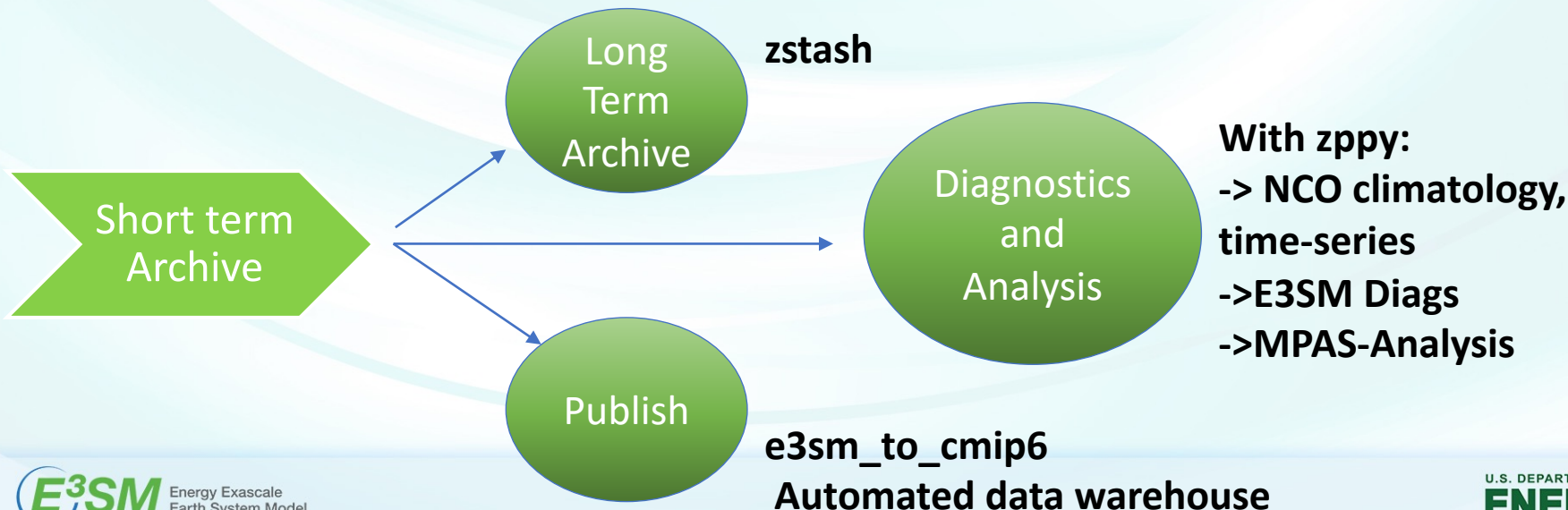
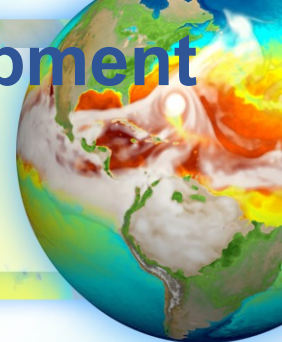
E3SM All Hands Presentation

2021-08-19

This work is performed under the auspices of the U. S. DOE by Lawrence Livermore National Laboratory under contract No. DE-AC52-07NA27344. It is supported by the Energy Exascale Earth System Model (E3SM) project, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. IM Release LLNL-PRES-825860

E3SM Tools for Every Phase of Model Development

CIME Case Control System, `run_e3sm`



Step-by-Step Guide

- We now have a guide for running the model and analyzing the results: <https://acme-climate.atlassian.net/wiki/spaces/ED/pages/2309226536/Running+E3SM+step-by-step+guide>
- This is a living document – it is easily changed to reflect changes in the process. Confluence keeps a page history to see older versions.

Summary

- Useful Aliases
 - The aliases provided in this section may be useful.
- Configuring the Model Run – Run Script
 - The run-script remains mostly the same between runs. Users can change configuration parameters in it though.
- Running the Model
 - It is recommended to run a few short tests before starting a production simulation.
- Short Term Archiving
 - To avoid having so many output files in one directory, we can archive output into appropriate subdirectories.
- Performance Information
 - PACE is used to display performance information from a run.
- Re-Submitting a Job After a Crash
 - If a job crashes, it is easy to re-submit after making a fix.
- Post-Processing with `zppy`
 - `zppy` is a tool that brings multiple post-processing tools together – with `zppy`, users can easily run E3SM Diags and MPAS-Analysis as well as generate global time series plots.
- Documenting the Model Run
 - Model runs should be documented on its own Confluence page providing run details and results.
- Long Term Archiving with `zstash`
 - `zstash` can be used to archive results for the long-term.

Configuring E3SM

- Configuring the Model Run – Run Script
 - It's easy to customize a run script by changing just a few parameters!
 - Make sure all your changes are done within the script to keep provenance.

```
# Simulation
readonly COMPSET="WCYCL1850"
readonly RESOLUTION="ne30pg2_EC30to60E2r2"
readonly CASE_NAME="v2.LR.piControl"
readonly CASE_GROUP="v2.LR"

# Code and compilation
readonly CHECKOUT="20210702"
readonly BRANCH="bd2ef1265d5738c624968ae27d427a4d5b27bddc" # master as of 20210702
readonly CHERRY="( "7e4d1c9fec40ce1cf2c272d671f5d9111fa4dea7" "a5b1d42d7cd24924d0dbda95e24ad8d4556d93f1" ) # PR4349
readonly DEBUG_COMPILE=false

# Run options
readonly MODEL_START_TYPE="hybrid" # 'initial', 'continue', 'branch', 'hybrid'
readonly START_DATE="0001-01-01"

# Additional options for 'branch' and 'hybrid'
readonly GET_REFCASE=TRUE
readonly RUN_REFDIR="/lrcrc/group/e3sm/ac.golaz/E3SMv2/v2.LR.piControl/init"
readonly RUN_REFCASE="20210625.v2rc3c-GWD.piControl.ne30pg2_EC30to60E2r2.chrysalis"
readonly RUN_REFDATE="1001-01-01" # same as MODEL_START_DATE for 'branch', can be different for 'hybrid'

# Set paths
readonly CODE_ROOT="${HOME}/E3SMv2/code/${CHECKOUT}"
readonly CASE_ROOT="/lrcrc/group/e3sm/${USER}/E3SMv2/${CASE_NAME}"

# Sub-directories
readonly CASE_BUILD_DIR=${CASE_ROOT}/build
readonly CASE_ARCHIVE_DIR=${CASE_ROOT}/archive

# Define type of run
# short tests: 'XS_2x5_ndays', 'XS_1x10_ndays', 'S_1x10_ndays',
# 'M_1x10_ndays', 'M2_1x10_ndays', 'M80_1x10_ndays', 'L_1x10_ndays'
# or 'production' for full simulation
readonly run='production'
```

See <https://github.com/E3SM-Project/SimulationScripts/blob/master/archive/v2/run.v2.LR.piControl.sh>

Running E3SM

- Running the Model
 - First, run a couple of small (pre-defined) tests to be sure the model actually runs.
 - Check that the results from the small tests are all BFB (bit-for-bit). This will ensure that the model is BFB for restarts and a few different layouts.
 - This is typically already tested by the nightly E3SM tests, but it's good practice to double check (especially if new options have been turned on).
 - Submit the production simulation.

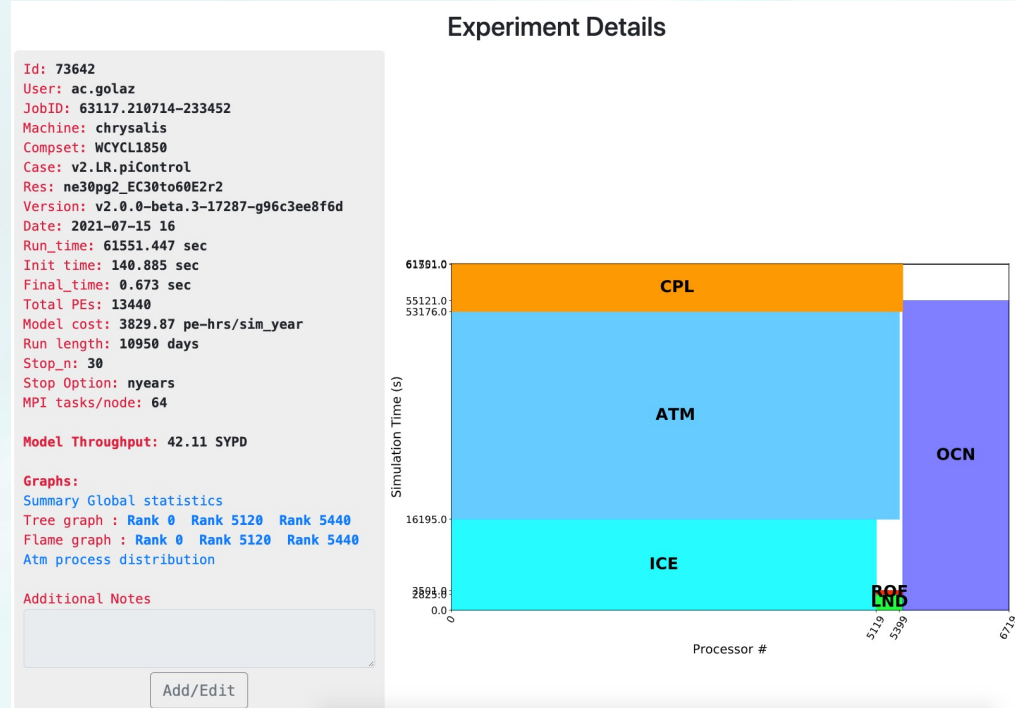
Looking at Results

- Looking at Results
 - You can look at logs for each component of the model: atmosphere, coupler, sea ice, land, ocean, river runoff. These logs are compressed to `.gz` files.
- Short Term Archiving
 - Short term archiving neatly organizes files in an `archive` directory separate from the `run` directory.
 - Without short term archiving, you have 10,000 to 100,000 files, which slows down even `ls`.
 - Each component of the model has a subdirectory in the `archive` directory.
 - Each component of the model has gzipped log files in `archive/logs`.

```
(base) :) /lrc/group/e3sm/ac.golaz/E3SMv2/v2.LR.piControl> ls archive
atm  cpl  ice  lnd  logs  ocn  rest  rof
```

Looking at Results

- Performance Information
 - PACE (<https://pace.ornl.gov/>) provides detailed performance information.
- Re-Submitting a Job After a Crash
 - The guide includes instructions for re-submitting a job after the error is fixed.



See <https://pace.ornl.gov/exp-details/73642>

zppy

- Post-Processing with zppy
 - zppy is a new post-processing tool for E3SM: <https://e3sm-project.github.io/zppy/build/html/main/index.html>
 - It's easy to customize a zppy run by changing parameters in a single configuration file!
 - Zppy can launch jobs for various post-processing tasks: **climatology**, **time-series**, **E3SM Diags**, **MPAS-Analysis**, and **global time series plots**.
 - Each job has a log file and a status file.
 - If you rerun zppy, only jobs that have failed will be rerun.

zppy: Analysis and Diagnostics Tools

- Main tools included in zppy workflow:
 - **NCO**: Regridding, Climatology Generation and Timeseries Extraction, supports **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**
 - **E3SM Unified**: Conda environment that includes all the above (and more) packages
 - Deployed on all E3SM supported machines (**Acme1, Andes, Anvil, Chrysalis, Compy, Cooley, Cori**)
 - zppy is included in E3SM Unified v1.5.0.
 - zppy can run the above packages using this environment.
- Upcoming:
 - Enable new sets being developed in these packages.
 - Add ILAMB package for land variable diagnostics.

E3SM Diagnostics Package v2.1.1
Test: e3sm_v1
Reference: Observation and Reanalysis
Created: 2020-08-10 10:06:27

E3SM
Energy Exascale
Earth System Model

Latitude-Longitude contour maps Table Taylor Diagram

Zonal mean line plots

Pressure-Latitude zonal mean contour plots

Polar contour maps

CloudTopHeight-Tau joint histograms

Pressure-Longitude meridional mean contour plots

ENSO Diagnostics

Quasi-biennial Oscillation

Area Mean Time Series

Provenance

MPAS-Analysis Diagnostics: Ocean
Run: 3DGM

MPAS
Model for Prediction Across Scales

Quick Links

Global Mixed-Layer Depth

Global Sea Surface Temperature

Global Sea Surface Salinity

Global Sea Surface Height

Global OHC Anomaly

Antarctic Potential Temperature

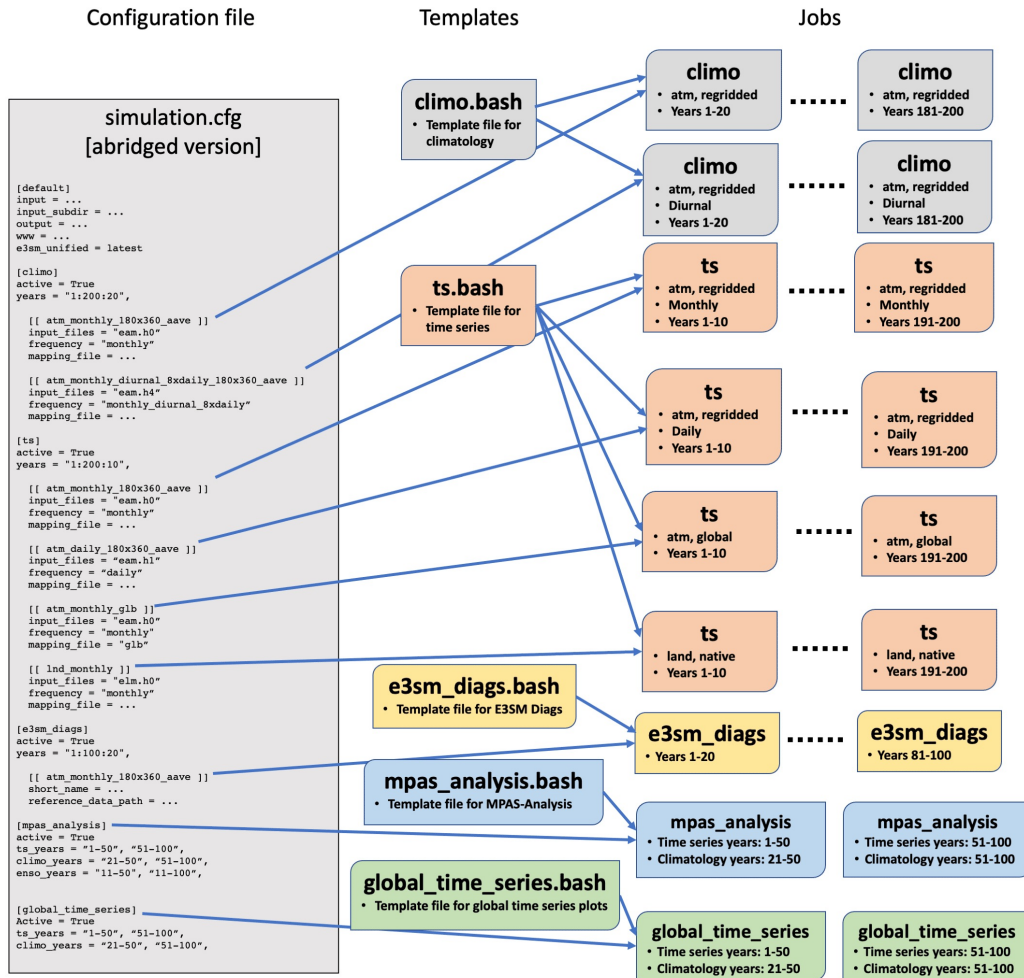
Antarctic Salinity

Antarctic Potential Density

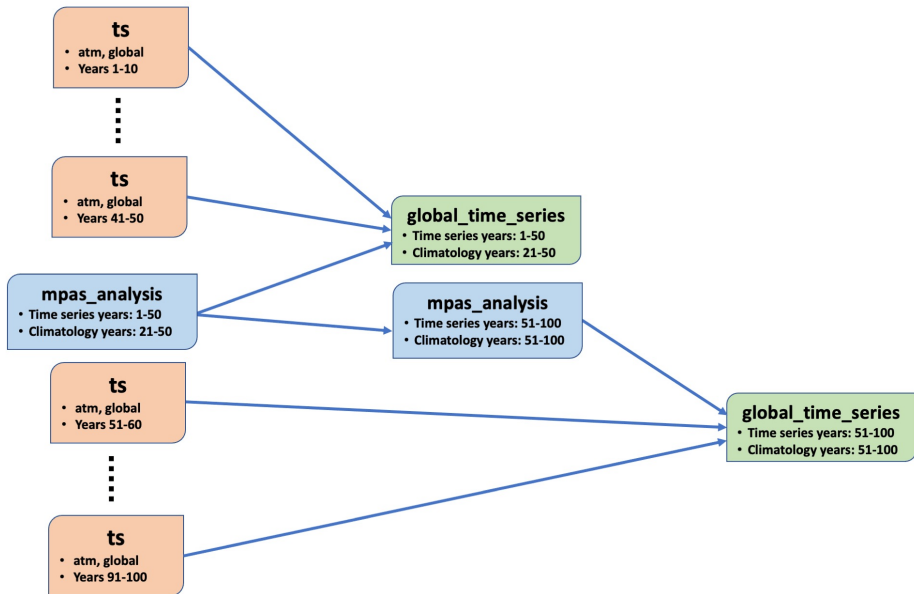
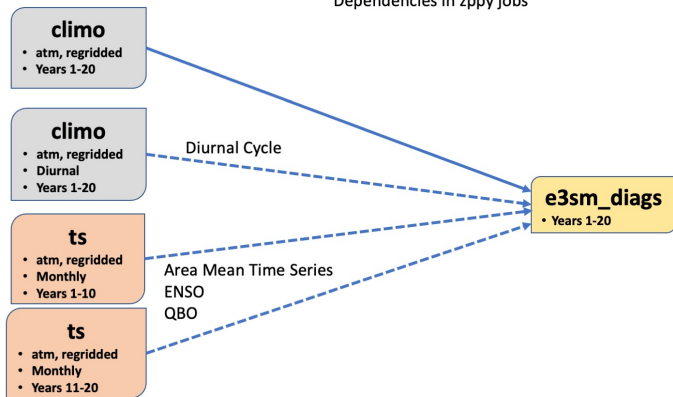
Antarctic Mixed Layer Depth

Antarctic Zonal Velocity

Jobs launched by zppy



zppy



Documenting

- Documenting the Model Run

- Every model run (or ensemble) should be documented on Confluence.
- Each job associated with the run should be listed along with the corresponding PACE links, number of nodes, and simulated years per day.
- The global time series plots should be included. Links to E3SM Diags and MPAS-Analysis plots should also be provided.

Production simulation: 0101

Job	Years	Nodes	SYPD	Notes
60640.210709-141750	1850-01-01 - 1880-01-01	80	29.68	
60933.210710-144953	1880-01-01 - 1910-01-01	80	31.72	
61647.210711-233830	1910-01-01 - 1940-01-01	80	31.67	
62510.210712-223657	1940-01-01 - 1970-01-01	80	31.52	
62763.210714-124300	1970-01-01 - 2000-01-01	80	31.58	
63337.210715-164446	2000-01-01 - 2015-01-01	80	31.47	

Long Term Archiving

- Long Term Archiving with zstash

- If a simulation is valuable, it should be long-term archived with zstash
- zstash is an HPSS long-term archiving solution for E3SM: <https://e3sm-project.github.io/zstash/build/html/master/index.html>
- Key features of zstash:
 - Files are archived into standard **tar files** with a user **specified maximum size**.
 - Tar files are first created locally, then transferred to HPSS.
 - **Checksums (md5)** of input files are computed *on-the-fly* during archiving. For large files, this saves a considerable amount of time compared to separate checksumming and archiving steps. Checksums are also computed *on-the-fly* for tars.
 - Checksums and additional metadata (size, modification time, tar file and offset) are stored in a **sqlite3 index database**.
 - **Database enables faster retrieval** of individual files by locating in which tar file a specific file is stored, as well as its location (offset) within the tar file.
 - **File integrity** is verified by computing checksums on-the-fly while **extracting** files.
- Globus can be used to transfer the zstash archive files to NERSC HPSS.

Key Takeaways

- Declarative, not imperative: define *what* you want done, not *how* to do it
 - Set parameters in the beginning of the run script; no changes to the actual logic.
 - Set parameters in zppy configuration file; zppy will determine dependencies and launch the appropriate jobs.
- One stop shop: everything you need in one place
 - The step-by-step guide provides detailed instructions on all parts of running and analyzing E3SM.
 - A model run's Confluence page provides all information about a model run, from each job's information on PACE to the generated E3SM Diags and MPAS-Analysis plots.
 - zppy automates multiple post-processing tasks, with more planned for future versions.

Questions?

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.