2024 CESM Tutorial
:
Intro to Lab: Simple XML Modifications

Hui Li
Project Scientist II
CGD, CCR

Aug 5-9, 2024

This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

Outline

- **Overview of model control files (env_*.xml files)**
  - What are model control files
  - How to modify an env_*.xml file
- **Controlling Run Length**
  - Starting and Stopping a run
  - Restarting a Run
  - Changing Run Length
  - Using Timing Files
- **Modifying the Type of Run**
  - Hybrid, branch, startup
- **Modifying Physics Timestep**
- **Exercise overview**

What are model control files (*.xml)?

Yesterday, we learned the basic workflow of running your first CESM run:

1. Create New Case
2. Case Setup
3. Case Build
4. Case Submit

```
cases/b.day1.0> ls -1
archive_metadata
Buildconf
case.build
case.cmpgen_namelists
case.qstatus
case.setup
case.submit
check_case
check_input_data
env_archive.xml
env_batch.xml
env_build.xml
env_case.xml
env_mach_pes.xml
env_mach_specific.xml
env_run.xml
LockedFiles
pelayout
preview_namelists
preview_run
README.case
SourceMods
Tools
xmlchange
xmlquery
```

User Customizable case XML files

What are model control files (*.xml)?

- We control how we compile and run the model with ***env_*.xml*** files.
- These files are created with ***create_newcase*.**
- The most commonly used model control file is the "**env_run.xml**". We use this file to control the run length, run type, etc., based on our experiment design. We will practice this in today's lab session.

There are multiple `env_*.xml` files in the `$CASEROOT` directory:

- `env_archive.xml`: specifies rules for short term archive script `case.st_archive`
- `env_batch.xml`: specifies batch specific settings used in `case.submit` script
- `env_build.xml`: specifies build information used in the `case.build` script
- `env_case.xml`: set by create_newcase and **cannot** be modified
- `env_mach_pes.xml`: specifies PE layout on NCAR HPC for components and used by `case.run` script
- `env_mach_specific.xml`: specifies machine specific information used in `case.build` script
- `env_run.xml`: sets run time information (such as length of run, number of submissions, ...)

How to modify and search env_*.xml files?

Editing:

When modifying “xml” files, we **highly recommend** using the tool, **xmlchange.** However, the user is free to use their editor of choice, i.e. **vi or emacs.**

Evaluate your understanding

If you want to manually resubmit an initial case that previously had a `CONTINUE_RUN` value of `FALSE`, how do you change it to `TRUE`?

▼ Click here for the solution

Use **`xmlchange`** to modify the variable value with the command:

```
./xmlchange CONTINUE_RUN=TRUE
```

How to modify and search env_*.xml files?

Searching:

To find xml variables in your case directory, we recommend usng the tool **xmlquery.** Note: You need to be in your case directory to execute these commands.

For example, to find out the run type of your job, search for xml variable `RUN_TYPE`:

```
./xmlquery RUN_TYPE
```

This will return the default `RUN_TYPE` value:

```
RUN_TYPE: hybrid
```

For help, type *./xmlchange --help*
type *./xmlquery --help*

Controlling run length using env_run.xml

- The length of your model run is controlled by several run time variables in the **env_run.xml** file. These variables may be modified at the initialization of the model run and during the course of the model run.

- We will learn about customizing runtime settings to control starting, stopping, restarting and continuing a model run, and practice how to specify the run length of your long simulations.

Controlling run length using env_run.xml

Set runtime limits with **STOP_OPTION** and **STOP_N**

Run length options can be set using **STOP_OPTION** and **STOP_N** variables in env_run.xml.

- **STOP_OPTION ==>** sets the run length time interval type, i.e. nmonths, ndays, nyears.
- **STOP_N ==>** sets the number of intervals (set by STOP_OPTION) to run the model during each submission within the specified wallclock time.

For example:
If you want to run a simulation for 6 months during the job submission, you will need to set:

```
./xmlchange STOP_N=6
./xmlchange STOP_OPTION='nmonths'
```

Note: STOP_N and STOP_OPTION control the length of the run **per job submission**. A typical simulation is comprised of many job submissions. This is because you can only stay in the computer queue for a limited time.

Controlling run length using env_run.xml

Continue a run

Important concepts: "Initial run" and "continue run"

When a CESM model run is first initialized, it is called an **initial run**.

The variable ***CONTINUE_RUN*** *in env_run.xml* is a flag indicating if the current model run is an "**initial run**" or a "**continue run**".

➢ For an initial run, *CONTINUE_RUN* must be set to ***FALSE***

➢ If the model continues a run, *CONTINUE_RUN* is set to ***TRUE***

For example, say we submitted our initial job with CONTINUE_RUN = FALSE (because it was just initialized). If the run has been finished and everything looks good, and we want to continue the run for another month, what do we do?

We will need to use xmlchange to change **CONTINUE_RUN = TRUE** and submit the run again to carry on running the model. The model will use the restart files to continue our run with a bit-for-bit match as if it had never been stopped.

Controlling run length using env_run.xml

Setting run length for long simulations

Number of submissions and run length

- Recall that we can use **STOP_N** and **STOP_OPTION** to control the run length of each batch job submission.
- A typical long model simulation (say you want to run the model for 100 years) is comprised of many job submissions. This is because we have limited batch wallclock time for each job submission. For example, on Derecho, the regular queue wallclock limit is 12 hours.
- We can specify ***the number of times to resubmit the run*** *using the* **RESUBMIT** variable in **env_run.xml** to complete the long run.

Note: If RESUBMIT > 0, your scripts will automatically change CONTINUE_RUN = TRUE after completion of the first submission for all subsequent submissions into the queue.

Controlling run length using env_run.xml

Setting run length for long simulations

Question:

The tutorial version of the CESM model on Derecho simulates ~10 model years per wallclock day. The maximum wallclock request is 12 hours. If you want to run the model for 100 years, what values should be set for STOP_OPTION, STOP_N and RESUBMIT?

Answer:

Assume we want to use the full 12 hours for each job submission.

The model runs 10 years / wallclock day, which means that 12 hours would give us 5 years per job submission.

STOP_OPTION = nyears, STOP_N = 5 , RESUBMIT = 19

Initial run of 5yrs + (19 resubmits * 5 years per job) = 100 years

Modifying the type of run using env_run.xml

CESM has three "types" of initial runs:

STARTUP: All model components are initialized from basic default initial conditions. The coupler does NOT need an initial file.

HYBRID: - The atmosphere is initialized from initial condition files generated by a user-specified CESM simulation

- The land, runoff, ocean and ice are initialized from restart files generated by a user-specified CESM simulation.

- No coupler file is needed

- Initial conditions and restart files use the same reference case and reference date.

BRANCH: All model components are initialized from restart files generated by a user-specified CESM simulation.

Modifying the type of run using env_run.xml

- The xml variable ***RUN_TYPE*** determines the initialization type.

- Note that the ***RUN_TYPE setting is only important for the initial run of a production run*** when the CONTINUE_RUN variable is set to FALSE.
- After the initial run, the CONTINUE_RUN variable is set to TRUE, and the model restarts exactly using input files in a case, date, and bit-for-bit continuous fashion.

- For runs that are initialized as hybrid or branch runs, we will need restart/initial files from previous model runs (as specified by the variables, $RUN_REFCASE and $RUN_REFDATE). See the tutorial material for more details!

Exercise overview

In today's lab session, we will do three exercises to better understand xml modifications.

- In Exercise 1, we will practice runtime variable modifications.
- In Exercise 2, we will create a "branch" run and modify the ocean coupling frequency.
- In Exercise 3, we will create a hybrid run and modify the atmosphere physics timestep.