Wikiprint Book

Title: Estimate the CO2eq computation footprint

Subject: lgcmg_doc - Doc/Footprint

Version: 21

Date: 07/06/24 06:59:23

Table of Content

Reducing our environmental footprint

Running simulations has a significant environmental impact both in terms of computing and mass storage. Determining the carbon footprint of a simulation is not an easy task as it depends on several factors.

Estimate the CO2eq computation footprint

To provide a rough estimate we may consider that running a **100 year long** climate simulation with IPSL-CM6A-LR requires **150 000 core-hours** and produces about **150 kg CO2eq**. Multiplied by the number of simulations performed (for example 80 000 years of simulation were done for cmip6) this constitutes a non-negligible fraction of the IPSL carbon footprint. This is why we should collectively aim to reduce the environmental footprint of climate modeling.

Adopting good computing practices can contribute towards that goal. Here we attempt to provide such good practices when it comes to using the IPSL climate models. The environmental footprint comes from the computing itself but also from the mass storage hence it is important to target both aspects in our modeling activity.

Good practices

We encourage you to go through the questions / steps below before running a long computationally-expensive set of climate experiments. Feedbacks on this guide are welcomed.

Which model is needed to answer my research question ?

Is a full-fledged climate model needed to answer my research question? Could a simple, less computationally-expensive model (such as OSCAR for present-day and future climate or iLOVECLIM for paleoclimates) provide me the answer?

If the IPSL climate model is what you need to do your research, then it is important to minimize the number of simulations that you have to run.

Are there experiments that already exist to answer my research question?

Are there experiments (e.g., from CMIP6 on /bdd/CMIP6 on spirit) that already exist and that can be used to answer my research question? Maybe these experiments are not exactly what I need but their preliminary analysis can speed up and narrow down the design of the experiments I need to run.

Is my experimental design well thought through?

Do I have an estimate of the signal-to-noise ratio that I am looking for in my experiments? If my experimental design requires running an ensemble of simulations, do I have an estimate of the ensemble size required and can this ensemble size be reduced by a better experimental design (e.g., by increasing the forcing)?

Notify the configuration manager

Before using a model configuration, you should always discuss with the configuration manager. This will ensure that your study is well thought through and prepared.

Can I share my results to help another researcher?

Maybe my climate experiments can be useful to someone else at IPSL. Discuss your experimental design around and pool with other scientists at IPSL to run a joint set of experiments with all the climate model output you need.

It is also important to minimize the risk of having to rerun simulations because of missing diagnostics or an incorrect experimental design.

Consider the necessary diagnostics

You should check that you have all the diagnostics you expect will be needed to analyze the results later on. As high-frequency diagnostics slow down the model and use a lot of mass storage, you should limit the high-frequency diagnostics to what is required. It may be appropriate to output diagnostics at a high-frequency resolution only for a sub-period of the simulation.

Check everything before launching the production run

First, run short simulations in TEST or DEVT mode to check that you have all the diagnostics you need and that the model is doing what you expect it to do. Always compile in `prod` mode for the production experiments to get most optimization out of the Fortran compiler.

If you plan to launch an ensemble of simulations: start with only one member, wait to have outputs and check them before starting all other members. Indeed it is easier and less consuming to clean and redo a short simulation instead of a full ensemble.

Two pairs of eyes are better than one

If in doubt, you may ask a colleague to double check your experimental setup with you.

Don't wait for the end of the simulation before checking

Your model experiments are now running.

You may check during the run that the simulations are doing what you expect them to do. You may also check that their computational cost is in line with what you expect. If the computational cost is beyond what you expect (see ■this page), and if you didn't add many diagnostics, it may be that you introduced a not-so-well-optimized new piece of code.

> Note: despite all your attention, your experiments turn out to be bugged, the experimental design was inappropriate or some diagnostics are missing. These things happen, do not blame yourself too much, analyze your errors and you may rerun the simulations with the corrected experimental design or setup! High research quality remains what we aim for.

Share your results with other researchers

Your model experiments are done and it is time to analyze them.

When it comes to analyzing your results, you may avoid duplicating the output. If you run your analysis on the IPSL spirit cluster, then you can make your model output visible through thredds (see this page).

Decrease the number of inodes

When you have finished the analysis, it is a good idea to keep only the part of the model outputs that has been useful or may be useful in the future. You may delete test simulations that are obsolete and archive the rest (using cc_pack, tar, zip) on the storedir to diminish the environmental impact of mass storage.