Hybrid physical+data-driven climate modelling

M2 internship 2024

Overview

The numerical models used for climate studies combine a solver for the atmospheric flow and simplified representations of small-scale processes such as turbulence and convection that affect on average the resolved flow. The latter are called physical parameterizations and have been so far built from a mix of first principles and phenomenological laws. Recently, data-driven approaches for the construction of physical parameterizations have shown a lot of promises and also severe limitations, especially regarding stability of long simulations, lasting hundreds of years.

The aim of the internship is to explore the combination of the DYNAMICO solver, a massively parallel finite volume/finite difference solver on a spherical Voronoi mesh, running on CPUs and GPUs [1,2,3], with data-driven physical parameterizations, building on the recently released ClimSim dataset [4]. Open questions include : are the hybrid models stable over long time scales ? Do they produce a reasonable climate ? Are they computationnally efficient on CPU, GPU ?

Proposed work

Programming will involve a mix of Python, Fortran, and possibly Julia [5]. The planned programme has several steps with some overlap:

1. Appropriation of tools and methods :
    - reproduction of the ClimSim examples
    - execution of standard DYNAMICO runs
2. Extend the ClimSim models (and re-train them) so that they predict wind. Indeed ClimSim includes a large dataset of atmospheric wind, temperature, and moisture, as well as a few pre-trained models. However, although the data does include wind, the models do not predict the wind, which is required for using them in simulations.
3. Combine data-driven models with DYNAMICO : because inference will be done in Python/TensorFlow, some work may be needed to ensure that the DYNAMICO entrypoint is callable from Python/Julia and that DYNAMICO can call back into Python/Julia. This will leverage Fortran 2003 standard interoperability with C.
4. Run the hybrid models. Assess their long-time stability. For the stable models, evaluate the produced climate and computational cost on CPU and GPU.

Working environment

The work will take place at LSCE/CEA under the supervision of Ségolène Crossouard, in collaboration with Thomas Dubos and Yann Meurdesoif, and others. The trainee will have access to the IPSL/ESPRI computing cluster, which includes CPU and GPU nodes. All the code involved in the work is distributed under an open source licence and versioned with git.

The ideal candidate

- has a background in scientific computing and possibly machine learning
- is familiar with Python and git, as well as a compiled language (C++/Fortran/Julia)
- is technically autonomous, rigorous and systematic in his or her work, as well as its presentation and discussion
- is interested in the application of scientific computing to climate challenges

Contact

- Ségolène Crossouard : segolene.crossouard@lsce.ipsl.fr
- Thomas Dubos : dubos@lmd.polytechnique.fr
- Yann Meurdesoif : yann.meurdesoif@lsce.ipsl.fr

Références

1. Dubos et al. (2015)
2. DYNAMICO
3. Tropical Cyclones in Global High-Resolution Simulations using the IPSL Model
4. ClimSim: An open large-scale dataset for training high-resolution physics emulators in hybrid multi-scale climate simulators
5. Julia language