Virtual CP4C Workshop Notes
April 21, 2022 (1st Day)

CP4C project overview

Time: 12:40 pm
Facilitator: Paul Kushner

- Paul provided an overview about the Collaborative Platform for CanESM (CP4C) project.
- The project developments started from March 2021 and the initial seed funding from the ECCC.
- CP4C is a *research platform* to increase Canada's capacity in development, application, and analysis of Earth System Models (ESMs) using Canada's Earth System Model, CanESM.

Main comments/ questions from participants:

- Paul Myers: He and his team spent a lot of time post-processing NEMO (ocean model) data into usable format. Coupling of CanAM and CanNEMO.

- Jeremy Fyke: It seems like, at face value, a lot of work in this (awesome) initiative will lead to a platform that does have some significant conceptual similarities to the CESM framework. But Neil mentioned that there are significant differences. What are these? I'm asking mostly because if there is conceptual overlap, that could be a great thing, because then the CP4C project could learn a lot by looking to CESM lessons-learned (perhaps?).
    - Paul K: What could be taking shape is an effort with relatively more leadership on model development directions within ECCC and collaborative agreements with a community extramural to ECCC. Featuring lots more opportunities for training and specific research projects than have existed with CanESM before. However, there are lessons to learn from the way CESM is managed which we should pay attention to.
    - Neil: The most obvious difference is the mandate and resources. UCAR/NCAR have an order of magnitude more staff than CCCma and have a clear mandate and funding to produce and fully support a community model. At CCCma, we have neither these resources nor mandate. However, we do have the willingness to collaborate, and the ability to shape our modelling tool-chain to facilitate this collaboration, which is exactly what we are doing. In many ways, being smaller is actually better - it makes us far more agile than the larger centres.
      Also, there are lessons to be learned from the experience of CLASSIC.

- Neil T.: As you further develop CanESM, will there be efforts to maintain a modelling hierarchy, including configurations with simplified physics, to aid with controlled model experimentation?
    - Paul K: This would be lovely to see but in the short term what we're trying to do is aim for more flexible model configurations, such as a slab-ocean model capability for CanESM,

which has fallen by the wayside. But an aquaplanet version of this with CanAM physics would be great to see…

- Neil S: At a basic level there is some of this. The most obvious example is the ability to run the fully coupled ESM vs say on an AMIP-type simulation. Taking this a step further, as we heard from Jason, the physics is increasingly being setup with namelist switches. In our old system supporting many different model configurations was quite hard. Under IMSI, it is going to be much easier. If people put effort into developing some simplified physics configuration, it should be easy to reproduce and share moving forward.

Collaborative CanESM

Time: 13:00
Facilitator: Neil Swart

- Neil gave a recap abut the CanESM history and strategic plan; CanESM status and development plan and progress on developing a collaborative version of CanESM

Main comments/ questions from participants:

- Philippe Lucas-Picher: What is the time horizon when CanESM6 using GEM will be ready and available? - *At least years for that, GEM is not even used internally.*
- Vivek Arora: Concerns about the quick pace of updates to the infrastructure for the model developers. How are devs supposed to keep up with the constant change? - *A lot of the changes that happened come from some catching up from years without updates. It will never be completely stable but will become more steady and slowdown in the near future.*
- Yi Huang: What physical parameters are found to be important? What is the initiative to port GEM to compute Canada platform? - *Coupling that into ESM, within CCCma*.
- John: History match, basic approach for automatic tunning. – *It is hard to see sensitivity to one parameter because they come in a collection, but the convention is an important parameter*
- Vivek Arora: As a developer of CanESM, if the system changes under the hood so quickly, do we expect the downstream projects to change fast as well? - *We have to check with CCCma, sharp learning curve currently to learn the previous implementation. Hope to reach a flat plateau so users can catch up on updates. That will be better for everyone.*
- Vivek Arora: That's a lot of work to support people inside CCCma and also outside CCCma. - *Paul K.: There can't be an obligation to support every user. There is a sense of expectation, and we are used to that which helps us to decide what to support. It is a work in progress, and there is no simple answer.*

GEM and CanRCM

Time: 13:20
Facilitator: John Scinocca

- John presented a about the Regional Climate Model CCCma, its physics characteristics , and advantages. Additionally, he explained about how the global GEM configuration has been tested and evaluated with CanAM5 physics.

Main comments/ questions from participants:

- Paul: Is [bias correction a] temporary measure? - *Bias exists in any model created. What is the model for? If it improves prediction, then use it. Doing bias correction in the regional model. It improves the model and is a very useful tool.*

CLASSIC

Time: 13:40
Facilitator: Vivek Arora and Joe Melton

- Vivek and Joe presented about the land component of CanESM named now as "CLASSIC" and its size capabilities. More technical features, including software content and tenderization sales documentation of the code, have been implemented, making it a complex model. As CanESM turns into a more community usable tool, CCCma welcomes and encourages the academic community to evaluate climate-vegetation interactions, land-atmosphere coupling, and impact of climate change on physical and biogeochemical land surface processes

Main comments/ questions from participants:

- Kirsten: In the new CanESM approach, how will the community outside of CCCma be able to use the framework? how is a new development ported to CCCma? - *Neil: We have a process for integrating changes to the main CanESM branches. It should seamlessly become available. Sometimes there is a little bit of delay, but it would basically be the same code.*

- Vivek: CCCma needs the community to do science with the model as it is the applications of the model that are really meaningful. “We have the option to run offline” If there is any update to the parameterisation, we can run offline to evaluate the change to see if it breaks the code. Run the control run and make a comparison to see if there is any surprise.
- Vivek: We can keep making the model more complex, but it is the application of the model to evaluate the important climate impact. If the community can help with applications, it will be great.
- Paul: There is a need for more downstream work than upstream work. These models are completely under-analyzed. To define which part has the highest priority is important.
- Philip: Agree with the analysis emphasis. Do we have a relationship with [an organisation] for resources?
- Neil: We are doing that by some aspect applying Compute Canada cloud (not commercial cloud), but we could do both.
- John: [Difficulty associated with development. Someone can be a real expert and developed some parameterisations. But if the person leaves the organisation, we are just left with the code. That’s why it is important to have the ECCC / CCCma relationship between developer and scientist to have someone taking care of some parameterisations.]

Analysis 4 Development (A4D) and CCCma published data

Time: 14:30
Facilitator: Michael Sigmond and James Anstey

- Michael and James presented about the Analysis for Development (A4D), intending to establish a comprehensive process through which CanESM output is analyzed in a systematic and ongoing manner to improve future model versions

Main comments/ questions from participants:

- Philip Austin [comment]: Kerchunk allows getting cloud-ready data from NetCDF. There has been a technical breakthrough, [something about NetCDF], game-changer; we do not need to rewrite all the NetCFD on the commercial cloud, easy to move data
- Paul: I like the issue trackers. Making a more formal platform for collaboration is a healthy direction.
- Philip A.: Would be interested to know who is interested in the working groups. There are records of past meetings; the sciences are on GitLab and are only available internally. We have considered making it more publicly available.
- James Anstey: MetOffice has PEGs (Process Evaluation Groups), which would be interesting to implement here. So far, the working groups have been an internal thing. A suggestion to make it more publicly available in the long-term.

CanAM development

Time: 14:50
Facilitator: Jason Cole

- Jason talked about the what the Can-Am atmospheric physics working group has been working on around the CCCma Atmospheric CanAM, its resent parameterization development, and the future development and needs. The long-term goal is for Can AM to be capable of modelling atmospheric processes, including aerosols and chemistry, from the surface to the stratosphere

Main comments/ questions from participants:

- A community model would be beneficial to look at the models in different ways
- Philip: Can you run with the bounce model? Are you using a version free to us? -Yes. Tried to compile on laptop / compute Canada with intel compiler
- Paul K: As the model gets more complex, the need for increased collaboration is growing and perhaps to bring in external

CanAM parallel sequence performance on Compute Canada platforms

Facilitator: Jack Wong and Paul Kushner
Time: 15:10 pm

- Jack presented about his work and experience around the CanESM Porting and Performance on a Compute Canada / Alliance Platform using IMSI, determining the runtime statistics, and getting the main findings on Niagara. CP4C is under active development, and everyone can contribute through GIT. The Jupiter Notebook used for this analysis will be shared on Git repository

Main comments/ questions from participants:

- Is there performance statistics on just CanAM? that's something that people might want to do - *Currently, this test is just done for CanESM. (Neil)- We can easily make it to just run CanAM*

Ensemble Consistency Test

Time: 15:30
Facilitator: Haruki Hirasawa

- Haruki talked about the ensemble consistency test on CanESM in order to verify Earth System model code changes and ports to these new machines outside of the program in Canada. Within future works include Run port verification for “bare-metal” CanESM5 port, implement ECT/FDR using IMSI diagnostic output, and test a larger suite of parameter perturbations

Main comments/ questions from participants:

- John: That can be a very useful tool. Optimizations on models can change the model behaviour. If you start from an initial state and just nudge and remove the nudging, it would be a large signal. Would it be useful to run large ensembles instead of just a single run? - *it's definitely an adaptable method. You could definitely take it in any number of directions in terms of trying to improve the performance of the test.*
- Neil: ECT is using 400 ensembles, are you using the same? - *I think ECT is using 250 as well*

Reflections Day 1

Facilitator: Paul Kushner
Time: 16:00

Main Takeaways:

- Climate modelling enterprises are getting bigger and bigger. There are more and more tests that a small team can handle. The time is right to come up with a mechanism to enhance collaboration.
- CP4C is a good part of where we can shape it together. It will work as a really effective tool if we plan it collaboratively.

1. Documentation
   - Projects are growing in scope, such as A4D if the list of desired diagnostics or goals for that effort were written down somewhere. There might be a group of people outside of CCMA where your current network could actually jump in.
   - Documentation is critical so that people can join the initiatives. Part of the effort of CP4C is to obtain cloud resources to prototype that idea.
2. Cloud Capacity
   - Taking advantage of cloud technology to do more analysis of CanESM. Part of our effort in the CP4C proposal to complete Canada was to obtain cloud resources, partly to prototype that kind

of activity that we have. If we could have some guaranteed resources to do some prototyping, it's an analysis that could then translate into something else, something larger later on.

3. Time and Resources
    - Based on comments from speakers and participants, resources and time are limitations. However, it is critical to identify priorities.

On the atmospheric side, there are two big challenges that have been identified

4. Technical challenges
    - Get GEM into CanESM, need to couple GEM into the rest of the ESM System. That will be a lot of technical issue. Would take some time to be available
5. Make GEM publicly available
    - Current model components we have absolute control over. GEM is an open-source model, so there is a challenge around how easy it would be to make it publicly available in terms of easily compiling and running it elsewhere.
    - This is a multi-year effort that will be involved in that whether we're talking about external people, being able to run, just an atmosphere model, whereas the global atmosphere regional atmosphere for the couple models.
6. Data Management
    - Earth's Observation data will be needed to bring into the models as they become more statistical. It is an opportunity to think more carefully within Canada about how we make observational model comparisons. That is an aspect of model validation that would be great to discuss for some time.
    - There is a potential opportunity to help develop the setup metrics with the broader community. People would know every time a new model version is created; it's going to be tested against those observational metrics to see how the tuning that's in with that.
    - There are a lot of scopes for a 3-year timescale. We can add downstream with what you are doing. There are a lot of projects that just take the model as it is (i.e. student projects).
    - There is a lot of work on the core physics of the models
    - The more metrics you have, the better the character of that quality. The uncertainty and, of course, the model limitations and structural engineers say it can't satisfy all of those at once reminders and model set of parameters and satisfies this group, we can tell you beforehand. Make an effort to Identify a core set of metrics.