Breakout group summaries

Land breakout group summary

[Link to folder with land group notes and table of needs](#)

- Tools currently used for model evaluation workflows
    - Bespoke scripts in:
        - § Python
        - § NCL
        - § ILAMB
        - § Benchcab
        - § Ferret (for looking at netcdf files)
    - TLDR: tools are currently all over the place!
- What metrics/variables are needed?
    - Albedo
    - Tas
    - Pr
    - Tasmax/Tasmin (difficult to model in current models, but important)
    - Soil moisture
    - GPP
    - Energy closure
    - Water closure
    - Latent heat
    - Sensible heat
    - Ground heat
- 3 cases of runtime diagnostics
    - We are building something new, and just want the model to start running
    - Model running now, we want to evaluate multiple years as sanity check
    - Now we have a model we can trust, have we set up the forcings correctly?
- Metrics should be decided for both short and long timescales
- We need to decide which online diagnostics we want to know to determine what variables to output from the model as it runs
- Fields in CABLE go through multiple stages (EM, postprocessing,…) – runtime diagnostics should be helpful in understanding where in the chain the error comes from
- Drawback of both ESMValTool and ILAMB is that they only take in CMORized output
    - Some ongoing work to accept CABLE output directly into scripts
- Summary:
    - Desire for ability to do ACCESS model evaluations on the fly, without the need to CMORize
    - Land community already uses ILAMB to a certain degree

- o Discussed 3 types of runtime diagnostics
- o Discussed specific metrics needed
- o Discussed current tools used – tools currently all over the place (everyone doing their own thing)
- Wrote down several identified Needs/Issues, and made them into GitHub issues

Ocean breakout group summary

[Link to ocean group notes document, which includes table of desired metrics and table of needs](#)

- Ocean diagnostics are often tricky because the datasets are very large
- Started by looking through an already-curated list of evaluation metrics collected by the COSIMA community on this [Hive Forum post](#)
  - o We ended up creating another table that lists all metrics from the post, and also other metrics that group members added or that are included in existing ESMValTool recipes
- Discussed how the community currently runs these metrics: by writing their own tailored software scripts using:
  - o Python, Xarray, Dask
  - o Ferret (a language built specifically by and for the ocean-atmosphere community)
    - § Note there is a python wrapper called PyFerret
  - o NCL
- Difference between running diagnostics on the ocean-only model first, and then again when running in coupled mode
  - o There are additional metrics that would need to be done in coupled mode (e.g. ENSO frequency, AMOC strength)
- For most ocean evaluation metrics, at least 100 years of model run is needed (longer is always better, but rarely doable due to computational cost)
- Variable output frequencies:
  - o Most climate diagnostics only require monthly output
  - o If want higher frequency output (e.g. daily) 3-d fields, can reduce the vertical layers to reduce data output size
- Current known issues in ACCESS models
  - o Sea ice extent in summer is ~20-30% of what we see in observations – i.e., the modeled sea ice extent in summer is far too low
  - o ENSO appears too frequently – on a ~2.5-year cycle
    - § Some hope that MOM6 will help fix this, but may also stem from atmosphere.
- In ESMValTool currently, there are limited ocean recipes, and so the ocean community will require a lot of ocean-tailored recipes to be added.

- Existing recipes are very poorly (if at all) documented, particularly within the recipe itself. Understanding what the recipe does requires going back to ESMValTool documentation and sifting through the relevant preprocessing steps (and even that documentation isn't always well-described).
- Note: the current ocean recipes on Gadi were all taken from a single, giant recipe on ESMValTool, and hence why the names of the yaml files are not very specific (e.g. "map1.yml")
- Concerns about using ESMValTool from the ocean/ice communities:
  - Need to CMORize data to use – not useful for on-the-fly diagnostics of ACCESS model
  - Based on Iris, while ocean community uses Xarray
  - Some concerns as to whether other ocean modeling communities have adopted ESMValTool – if not, then the community will not gain international collaboration by adopting the tool
  - Ocean/ice community already has good set of scripts (via COSIMA recipes) to do some ACCESS model evaluation, and are hesitant to move all those workflows into a new, less flexible system
- Some other desires that came up in discussion
  - Better documentation of the recipes
  - Ability to use ESMValTool on non-CMORized data to be able to use with ACCESS as it runs
  - Flags to be able to run comparisons on subsets of CMIP5 or CMIP6 models, so users don't have to know specific CMIP models to call each time a script is run

Atmosphere breakout group summary

- The group discussed how ESMValTool could be used in the community
  - Integrating preprocessing and ESMValTool to existing workflow
  - Systematic basic evaluation methodology
- Important aspects of simulations:
  - Mean precipitation
  - Improving temporal and spatial resolution
  - Basic checks for large-scale circulation features
    § Existence of Hadley + Walker cells
    § Does the MJO exist?
    § Polar jet
    § ICTZ
  - **Benchmarking of specific phenomena is important so we can select models that do well or get rid of the bad ones**
- Types of uncertainty

-    o Variable to observations
-    o Variable between models
- For downscaling, need 3-hourly 10m winds and 1-hourly MSLP (helpful for storm surge community)
  -    o Knowing biases and variability that the models represent
  -    o Parameters to describe distributions of high-res variables, even if high-res variables can't be saved
- For extremes, these are grid point anomalies
  -    o Need subdaily fields
- **Need to keep track of uncertainties (standard deviation) throughout workflow for every model**
- Other diagnostics mentioned:
  -    o Maps of maxima + minima wind speed + precipitation (for extremes)
  -    o Gradients in continuous variables (e.g. SST)
  -    o Energy + moisture budgets are they conserved?
  -    o Vertical integrals – calculated on model levels before they are thrown out (to save space when doing budget calculations)
  -    o Distributions of sub-monthly data (eg PDFs of monthly rainfall are different from PDFs of daily data)
- **We want daily + subdaily data please!**
- The group created a physical table of desired metrics (an image of the table is included in the Google Drive folder)

General diagnostics/Coupled breakout group summary
[Link to general/coupled group folder with notes and spreadsheet containing list of prioritized needs, desired metrics, and desired observational datasets](#)

- Discussed types of model evaluations that have been done in the past, including
  -    o Energy balance
  -    o Surface air temperature and precipitation
  -    o Thermocline in ESM1.5 and CMIP6
  -    o Ocean variables – will give you ~300 temperatures in one single time series file
    -     § Takes time to run, particularly if run across all ensembles
- We have the sanity checks in place now, but could we reuse them again and again? Can't store data forever – need to decide which key points to save timeseries of.
- NCAR compares their last model with the new model – is that something we should focus on?
- Should we focus on the variability around Australia?
- How do we define what is wrong?

- o Drift and sea level should be less than a certain value, maybe take linear regression?
- o We should have a tool that constantly monitors the time series
- What about paleo model runs? These can be useful to run, but may not be practical at this stage.
- How can we monitor what is going on in real time and compare against observational datasets?
  - o For atmospheric variables, ERA5 (but be careful, as it has huge biases)
  - o Global surface temperature
  - o SST
- Other datasets being used
  - o GPCP for global precipitation
  - o HeadSLP
- Day 2: group looked through existing ESMValTool recipes and metrics
- **The model dev teams have automated scripts for eg checking drift, but it would be nice to have a centralized recipe for others to use**
- We need to think more about what acceptable ranges are
- Difficult to compare across different variables and units, and % change means different things for different metrics
  - o Best way is to normalize by CMIP MMM (?)
- PCMDI has examples of metrics – maybe we can use ESMValTool recipes to automate this
- Group created a nice spreadsheet with three aspects:
  - o Prioritized list of issues/needs
  - o List of desired metrics/recipes
  - o List of observational datasets that they use