

Motivation

Running climate models costs a lot! \$\$\$\$ How can we maximise the benefit of a model run? How can we collaborate?

What is COSIMA?

- Consortium for Ocean and Sea Ice Modelling in Australia
- Formed in 2012 (\sim 10 people); now >150 people
- Culture is key. Shared models, data and tools build community.
- Developed ACCESS-OM2 (MOM5-CICE5) global ocean-sea ice model suite at 1°, 0.25°, and 0.1° resolution
- Adopted in climate models (ACCESS-CM2, ACCESS-CM2-025) and Australian Bureau of Meteorology ocean forecasts (Bluelink)
- 550TB of shared output data, easily explored and analysed; 200 users Underpinned >70 papers since 2019 (>1100 total citations)
- >50 ongoing research projects



ACCESS-OM2 global ocean-sea ice model suite at 1°, 0.25°, and 0.1° lateral resolution

Collaborative ocean modelling within COSIMA

- Model runs are very expensive & output data is valuable for many research questions
- For example, a global simulation at 0.1° with biogeochemistry: ~15TB memory
- >12,000 cores (>250 Cascade Lake nodes × 48 CPUs)
- Up to 12hr (\sim 150,000 core hours) per simulated year
- Runs for months on Gadi (Australia's HPC) to simulate 60 years
- Costs $\mathcal{O}(10^7)$ core hours
- Generates ~40TB of output
- Few groups in Australia have resources to run these models, so in COSIMA we cooperate:
- **Piggyback:** save outputs for multiple projects in each model run
- Share all data freely on Gadi and via THREDDS: >550TB of shared data so far
- How to make this useful to people?

The COSIMA community

COSIMA supports a diverse set of users:

- Most want to analyse output, not run models • so we seek requests for diagnostics to output from runs, and make data discoverable, accessible and computable via the COSIMA Cookbook
- Many want to run standard executables with configuration changes (perturbation experiments)
- so we supply **12 standard configurations**, with input files and precompiled executables
- A few want to make code changes
- easy! git clone downloads all model source code, with 1-line build \implies contribute source code improvements via pull requests with automated tests
- Core team of developers
- model development, maintenance, performance, porting, bug fixes, PR reviews, etc

All of COSIMA's work is open source!





github.com/COSIMA

A community-driven cookbook of recipes for analysing global ocean-sea ice model output by the **Consortium for Ocean and Sea Ice Modelling in Australia**

COSIMA *is* a Community



COSIMA Annual Workshop 2022 in Hobart, Tasmania; only 2 out of >150 people are funded by COSIMA

Easy data discovery with Database Explorer

- Names of variables for the data you need
- Experiments that saved that variable

Datab Experiment	ase Explorer	Filter by: and experiments that saved them					DataArray using the CC	
01deg_jra55v13_ryf9091 01deg_jra55v13_ryf9091_5Kv 01deg_jra55v13_ryf9091_qian_wp 01deg_jra55v13_ryf9091_tides_fixed 01deg_jra55v140_iaf 01deg_jra55v140_iaf_cycle2		Keyword Variable			The loaded DataArray is			
		All models Filter variables: Search: start typing Stritty_m surface pot temp Image: Stritty_m			Filter variables:		The selected experiment	
						311.1	01deg_jra55v140_iaf	
						Ocean only		
		surface_pot_t	emp_max	>			surface	
		 surface_salt surface_temp surface_temp_max surface_temp_min sw_edges_ocean sw_heat Hide coordinates Hide restarts 			«		eta_nonbouss net_sfc_heating salt_surface_ave sfc_hflux_coupler sfc_hflux_pme surface_salt surface_temp surface_temp_min temp_surface_ave total_net_sfc_heating total_ocean_hflux_coup total_ocean_hflux_prec wt	
Load Ex	xperiment	Filter						
Experimen Descriptio Notes: Notes: Contact: No. files: Created:	<pre>ht:01deg_jra55v140_iaf n:0.1 degree ACCESS-OM2 global described in Kiss et al. (2020), conditions are WOA13v2 potent salinity. Run with JRA55-do v1.4 no heat transfer. 61-year spin u Source code: https://github.cor International (http://creativeco users of this or other ACCESS-O [http://doi.org/10.5194/gmd-1 thank the Consortium for Ocear the ACCESS-OM2 suite of mode undertaken with the assistance supported by the Australian Gov can add them to our list: https: history: https://github.com/CO http://cosima.org.au/index.php from 1 Jan 1987 (run 116) onw mean 3d bih_fric_u, bih_fric_v, v, wt (not yet synced to ik11) A sea_level - monthly snapshot 3 Andrew Kiss <andrew.kiss@anu 26168 2020-06-09T00:00:00.0000000</andrew.kiss@anu </pre>	model configuration un https://doi.org/10.519- ial temperature (NB: sl 4.0 interannually-varyin p, from 1 Jan 1958 to n/COSIMA/access-om2 mmons.org/licenses/by 0M2 model code or outg 3-401-2020] (b) includ o-Sea Ice Modelling in A ls available at https://g of resources from the l vernment." (c) let us kn //scholar.google.com/c SIMA/01deg_jra55_iaf/ /2020/07/29/data-avai ard: - daily mean sea i u_dot_grad_vert_pv - dditional output from 1 d salt, temp, u, v, vert_ i.edu.au>	nder interannual forcin 4/gmd-13-401-2020, b nould have been conse in forcing with all solid 1 Jan 2019. License: Attribution-N -nc-nd/4.0/legalcode) but data: (a) consider a acknowledgement sustralia (COSIMA; http: http://doc.om/COSIMA/actional tow of any publications tations?hl=en&user=in tree/01deg_jra55v140 lable-0-1-1958-2018-a ce area and volume, b daily mean 3d conserv Jan 2012 (run 216) o _pv and vorticity_z	g. The oput with rvative runoff Conditiciting Kit such a p://www. ccess-or I Infrast s which hVqu_4. _iaf Ou access-or y catego ative te nward:	configuration is based on that many improvements. Initial temperature) and practical converted to liquid runoff with mercial-NoDerivatives 4.0 ons of use: We request that ss et al. (2020) s the following: "The authors v.cosima.org.au) for making n2. Model runs were ructure (NCI), which is use these models or data so we AAAJ Run configuration and cputs described here: om2-iaf-run/ Additional output ory (aicen, vicen) - monthly mperature, practical salinity, u, - monthly snapshot 2d		Practical Salinity Practical Salinity Hide coordinates Hide restarts Load Loaded data with cc.querying.getvar xarray.DataArray 'surfa Array Bytes 4.43 GB	
							Shape(114, 2700, 34 Count5738 Tasks	

Data: discoverable, accessible, computable

- **Discoverable**: allow everybody to find what they need without knowing technical details
- Accessible: easy to obtain and understand data, with provenance connecting it to all details of model configuration that created it
- **Computable**: systems to enable rapid calculation on huge datasets
- Don't want to download 10s of TB
- Can't process on a local machine anyway; won't fit in
- memory
- xarray: python package to use netCDF metadata to combine and subset individual files without loading data until it actually needed: lazy evaluation

- dask: python package to process big computations in parallel without needing to fit it all in memory

• Spatial & temporal resolution and time period of data Code needed to access the required data



Analyze data in-place

Analyze data in-situ and in parallel using:

- 100s of experiments
- 20-100 output variables in each experiment
- *u*, *v*, *w*, temperature, salinity, sea surface height, 10 biogeochemical tracers, & many online-calculated diagnostics
- Large experiments may contain >100,000 files, tens of TB
- Each experiment may have 100s of run directories (due to queue limits)
- Each run directory may contain hundreds of output NetCDF files

In summary: quite confusing!

Climate models produce *a lot* of data! No need to know where each file is! cosima-cookbook remembers for you!



COSIMA Cookbook Python package



During a "Finding Nemo"-themed hackathon where we updated the COSIMA Recipes; Sep. 2023 **COSIMA** Recipes Shared, community-contributed notebooks for model data analysis

- speed in 1-2 days!



cosima.org.au

What the COSIMA ocean data looks like

Abundance of output!

COSIMA Cookbook & Recipes

xarray dask

An experiment database that knows where all the files live

• Open-source Python package github.com/COSIMA/cosima-cookbook • Framework for indexing and querying ocean-sea ice model output • SQLite database of metadata from all experiments, updated nightly • Users don't need to know file names or directory structure • Calls xarray and returns a dask dataset with the requested output • **GUI database explorer** to browse available experiments and variables

• Excellent resource for new researchers/students to get them up to

• Community can learn from and teach one another

• github.com/COSIMA/cosima-recipes

Acknowledgments

• Australian Research Council grants LP160100073 & LP200100406 • Computing and storage resources supplied by Australia's National Computational Infrastructure (NCI)

 Supported by ACCESS-NRI, enabled by the Australian Government through the National Collaborative Research Infrastructure Strategy