

Winter School 2019

Ewan Short

June 27, 2019

1 Todd Lane - Fundamentals

Models allow reproducible experiments. Hinting at a need for people who can edit code. Reanalyses are MODELS not observations. Clouds and precipitation are weaknesses of reanalysis products.

“Plug and play” approach with different model components not ideal - need a “unified” approach where all the parametrisations etc are internally consistent.

“Forced euler equations” name of our meteorology/climate primitive equations. Forcings (F_x etc terms) come from both physics and parametrisations.

Equations solved using both spectral and grid based methods. Have to switch constantly between spectral methods and grid methods - land properties etc. stored on a grid. Regional models all grid based models! “Physics” and parametrisations typically require a grid.

Difficult to use grids at poles - EC reduces number of grid points at poles - use filter to slow down model. Other models use a hexagonal grid. MPAS model for example - dynamic core of NCAR model. Requires scale aware physics schemes. Physics schemes are all heavily “tuned”.

20 years ago, most supercomputer centres were for weather prediction. Tendency not just to increase resolution, but to add complexity. Competition between complexity and resolution in model development.

Two way nesting also used - finer model upscaled fed into coarser model. UM can only do one-way nesting. Two way nesting prevents examination of the effects of resolution. Can’t change vertical resolution with nesting?

Improvement in skill in SH due to satellite data! Accounts for NH and SH coming together. Errors associated with initial condition error, physics, numerical errors, boundary condition errors (e.g. larger scale models producing errors in nested grids.)

Periodic boundary conditions are “exact” - i.e. invisible to the flow. Don’t have to worry about boundary conditions in global models. We don’t want reflective boundary conditions (except at lower boundary). Open boundary conditions good for regional models - hard to remove signal entirely. Open boundary conditions are not exact! Affect the flow.

Vertical boundary conditions. “Damping” - add an extra term to vertical momentum equation, e.g. Rayleigh friction $-\alpha w$. Key idea is that sponge layer solution includes non-physical terms - cannot be used for physical insight. Sponges also consume resources!

Recommended configuration for WRF includes sponges that are too shallow!

Spin-up issues in nested models - fine scale features need to “spin-up”. Eddy time-scale in finer model. Also to do with interpolation scheme needed when downscaling boundary conditions. Cannot downscale vertical coordinate in many models (e.g. WRF) because coordinate depends on physics (e.g. mass).

Parametrisations are weakest part of our models! Hard to observe small scale processes. Parametrisations usually deterministic. Surface friction needs to be parametrised - as does gravity wave drag etc. “Philosophy” between parametrisations can be quite different.

Reynolds averaging basis of all parametrisations of dynamic processes. Decompose $u = \tilde{u} + u'$ where \tilde{u} is the resolved part of the flow and u' is the unresolved part. Sub into equations of motion. Reynolds stress terms are parametrised. Important to differentiate between parametrising physical processes (i.e. physics not captured by momentum equations) and parametrisations of terms that are unresolved in momentum equations at given resolution.

K theory involves approximation

$$\frac{\partial \langle u'u' \rangle}{\partial x} \approx \frac{\partial}{\partial x} \left(-K \frac{\partial \tilde{u}}{\partial x} \right).$$

Smagorinsky closure used in most models - positive K for particular Richardson numbers.

K theory is an example of “first order closure”. Zero-order closure assumes $\langle u'u' \rangle = 0$. Can do higher order closures using more terms. Not constrained by observations outside boundary layer - little point in atmosphere models.

Convective mass flux $Mc = \rho \sigma w_c$ where σ is the cloud fraction.

All parametrisations assume scale separation between resolved flow and process being parametrised - i.e. assumes grid boxes much larger than scale of process being parametrised. This assumption fails for higher resolution models!

Wingard (2004) nice paper on grey zone - “terra incognita”. As resolutions increase - parametrised processes are partially resolved - should we parametrise?

What is “effective model resolution”? Can’t resolve features at grid spacing - need features to be multiples of grid spacing. See Skamarock (2004). Energy density of atmosphere follows power density spectrum. Deviation in WRF resolves approx. 7 times grid spacing.

You need sub 100 m grid spacing to properly resolve deep convection - would need even smaller to resolve shallow convection! Bryan (2003 MWR). 1st order differencing strongly damped. 2nd order method diffusive. General result!

First order approximation to advection equation actually a second order approx to

$$\frac{Du}{Dt} = \frac{-\Delta x}{2} \frac{\partial^2 \phi}{\partial x^2}.$$

Important to understand model construction.

1.1 Questions

1. What determines how many nests are used, and the structure of those nests? Is it mostly just trial and error? E.g. in Wapler plot, why use four grids instead of just two? Due to spin-up

issues when going to higher resolution. Also due to interpolation issues when interpolating boundary conditions down to higher resolution grid. Mostly trial and error - no comprehensive theory.

2. What causes artificial reflection?

3. Given there's so much tuning - why bother with physical basis? Zeb? Answer confusing. Craig says can still impose physics on a larger scale.

2 Nina Ridder - Young Earth System Scientists (YESS)

Supports young researchers. International network.

3 Anna Ukkola - Land Surface Modelling

Land surface models included from the very beginning - mid 1960s! By 80s and 90s vegetation and photosynthesis included. 90s 20s, carbon cycle and interactive vegetation added. Land takes up 25% of atmospheric CO₂.

"Memories" of different components - Mariotti (2018). Between 2 weeks and 2 months land surface import source of predictability. Ocean most important at longer than 2 month scales.

Lyons (2002) nice figure on effect of land-use type.

Land surface models built on fundamental leaf stomata processes.

"Arctic amplification of temperatures"? Albedo changes due to decreasing sea-ice and snow leads to positive feedback. In Finland Decembers are ≈ 5 degrees hotter!

Shallow boundary layers of water due to more latent heat. Feedbacks to atmosphere. Less rainfall, less moisture, more sensible heating, higher boundary layer. Koster (2004). Feedbacks important in Mediterranean. In models, feedbacks are very uncertain. Europe 2003 heatwave amplified by land-surface drying. Ukkola (2006) compares model treatment of these feedbacks.

Water cycle: water balance equation

$$P = E + R + \frac{dS}{dt}.$$

where S is storage. Between 60-80% of precipitation evaporated. Transpiration 65 to 80% of E globally. Surface run-off generated by infiltration excess or saturation excess. ET modelled used Penman-Monteith equation

$$\lambda E = \frac{sR_n + D\rho_a c_p g_h}{s + \gamma g_h / g_w}.$$

Radiation and wind speed potentially more important than heating for evaporation.

Most models just simulate vertical moisture transport through soil - not horizontal. Darcy's law and Richards equation used for this.

Carbon cycle processes. Photosynthesis and respiration roughly in balance - slightly more carbon uptake. Allocation of carbon between roots, leaves etc. uncertain. In the model, photosynthesis

taken as minimum between enzyme and light limitations. Land won't take up as much carbon as CO₂ increases! C4 plants already photosynthesising at slower rate. Also coupled with water cycle!

Land-surface models diverge the most in representation of vegetation. Broadleaf, needle-leaf shrub etc. Some models use "big leaf" - just scale up single leaf processes. Others use demographic models which include competition for light, water etc. Australian models don't include fire. Also models don't let vegetation wilt or die. Models don't include urban environments.

3.1 Questions

1. How do snow and atmospheric models interact?
2. Is CO₂ greening a myth?

4 Andy Hogg - Introduction to Modelling the Ocean

Models can also be used to interpret observations! How do ocean models differ from atmospheric models? Air is compressible - seawater is nearly incompressible. Ocean equation of state - cannot use ideal gas law. Ocean: density function of salinity etc. Salinity a little bit like humidity - effects density. Salt constant over model. Solve for temperature and salinity - get density out of equation of state.

Thermocline steep change in temperature.

Ocean models often use tripole grid! Two poles in NH! Fronts etc occur on monthly time scales, but 10 to 100 km spatial scales.

Horizontal momentum equations include $\nu \nabla^2 u$ term (viscosity)? Need also conservation of heat $\frac{dT}{dt} = \kappa \nabla^2 T + Q_T$, and salinity equations $\frac{dS}{dt} = \kappa_S \nabla^2 S + Q_s$, and nonlinear equation of state $\rho = \rho(P, S, T)$. Also need boundary conditions, topography. Also initial conditions - Andy Hogg assumes zero velocity, imposes temperature and salinity.

Singularities at poles creates small spatial cells which require small time steps! If don't care about arctic, waste a lot of computational resources. So use tripole grid. Schopf (2005) and Murray (1996). If want to resolve coastal bathymetry can use irregular grid FEM - but then how do we take zonal averages etc? Hogg doesn't use irregular grids. Most parametrisations have resolution dependant physics - so may need scale aware schemes for FEM.

Vertical grid can be z , σ , ρ based. ρ based grid huge advantage in general - but fails in mixed layer because isopycnals vertical! New grid system will be Arbitrary Lagrangian-Euler. Hybrid grid - z near surface, terrain following near bottom and isopycnal interior. MOM6 uses this system. Time step limited by vertical CFL - but with AEL grid there is no vertical time step!

4.1 Questions

1. Why does density change close to land in conceptual model? Isopycnals slope upwards in southern ocean. DSW (dense surface water), HSSW (high salinity surface water). Water cooled by Antarctic continental shelf due both to temperature (cooling) and sea-ice formation which

rejects salt, increasing salinity of resulting liquid water. Similar reasons but less pronounced in Arctic. Isopycnals vertical in mixed layer. Also evaporation at equator slightly increases salinity and density.

5 Andy Hogg - The ACCESS-OM2 Model

Need to do a better job of model development in Australia. CSIRO believes university model development too short term. So Hogg etc built collaborations with CSIRO etc. Should we develop our own models? Zeb - control. Roseanna - calibrate to our hemisphere. Need a range of models. We need to have experience and expertise in the community - we need capacity. Success of model development depends on key individuals with experience.

Divide model development into two parts - 1) code, 2) configurations.

WRF code has undergone decades of community support. Probably uses old software standards. Configurations undergo more rapid development time frames. We need *configuration* expertise! Australia cannot write a new ocean model that is competitive with what's out there.

Need to balance cores between component models. Balance so that MOM5.1 is never waiting. Only 60 years of atmospheric forcing - either repeat non-anomalous day or repeat entire 60 years twice. Latter case has massive sawtooth!

No sea-ice modellers in Australia? [What about Luke Bennetts?]

Finer resolutions can be worse because tuning well established in coarser cases.

5.1 Questions

1. What is the base of ACCESS-OM2? MOM5.1, CICE 5.1.2, YATM JRA55-do, OASIS3-MCT-2.
2. Have to spin up from scratch because can't use coarser resolution - eddy processes etc fundamentally different.

6 Katrin Meissner - Earth System Models

Parallelising couplers between atmosphere, ocean, land, chemistry, sea ice, etc very difficult. Atmosphere generally largest component of ES models.

Case study 1: Paleocene eocene thermal maximum. Large acidification event. Massive release of carbon. Horses evolve in eocene, primates also. GW between 5 and 8 degrees. Deep water temperatures from 4-5 degrees.

Questions: How much carbon was released? Where did it come from?

Good for understanding climate sensitivity, fast release carbon reservoirs, thresholds. Understand ability of ecosystems to adapt (or not).

Acidification (lower pH) less of an issue than lower carbonate levels, as animals build shells out of carbonate.

First simulations of this event used MOM2 (ocean only, no coupling). Also has been used - LOSCAR (carbon box model, no dynamics, Penman and Zachos (2018)). cGenie more complex, has 3D dynamics. UVic too complex?

Need to understand to what extent ocean anoxic.

Ventilation events due to diffusive-convective oscillations. Are our models too diffusive?

Oschlies (2018) models don't reproduce bio-geo-chemical processes.

Huber and Sloan (2001).

Meissner (2014) models struggle with SST in arctic.

ESMs don't do climate sensitivity well. Fischer (2018). Gottschalk (2019).

Meissner doesn't trust models in high emission scenarios after 2050. Permafrost, clouds, ocean ecosystem etc. Clouds important!

Todd: How fast do particles fall? Katrin: We do not know! Days to months.

7 Jo Brown - Paleoclimate Modelling

Use the same atmosphere-only or coupled atmosphere-ocean models that are used to simulate current and future climate (e.g. CMIP models). We assume laws of physics are the same [but we've acknowledged models are tuned to death!]

Paleoclimate model boundary conditions. Orbital parameters - calculated from known astronomical equations. Obliquity, eccentricity and precession.

8 Rachel Law - ACCESS

National effort starting 2005. Mostly written in fortran. Supported by NESP and NCRIS (can ACCESS be seen as "national infrastructure!") Partnership with MetOffice for UM, ocean from GFDL, ice from Los Alamos.

Atmospheric component: Unified Model (UM). Configuration more important than code version.

Land: CABLE (Australian community model), CASA-CNP (biogeochemistry). Replaces most UM land scheme (MOSES or JULES). CABLE deals with canopy, soil, atmospheric fluxes etc.

Ocean: NOAA/FDL MOM4p1 or MOM5. Ocean biogeochemistry (WOMBAT). Fairly simple - 1 category of phytoplankton etc.

Sea-Ice: LANL CICE4.1 (ESM), CICES...

Aerosol and Chemistry: CLASSIC or GLOMAP-mode aerosol scheme. Can "nudge" this model toward reanalysis. E.g. can nudge to C02 at Cape Grim.

Coupler OASIS3-MCT is sued for UM-CICE and CICE-MOM coupling. A lot of fields need to be passed. Coupling system not trivial.

Spin up much slower for coupled models! Atmosphere-only order 10^0 years, AO-coupled order 10^2 years for deep ocean. Carbon cycle order 10^3 years!

Law et al. GMD, (2017).

Different types of experiments: "What if" experiments. What if could suddenly remove C02 from

atmosphere? Control simulations can be used to assess model drift. Understand natural variability of model. Test sensitivity of different resolutions.

“Equilibrium Climate sensitivity” - Gregory (2004).

Diagnosed variables can be faulty! Bug in diagnosed wind speed. 3 bugs identified.

9 Julie Arblaster - Model Intercomparison

Fyfe et al 2016. CMIP6 - cmip6workshop19.sciencesconf.org/

Various necessary comparisons to be included in CMIP - gradual CO₂ change, abrupt 4x CO₂ change etc.

input4MIPS infrastructure for forcing data. ESMVar tools? Routine MIP analysis done automatically?

CMIP6 data on NCI oiXXX something.

New science: CMIP6 models much warmer than CMIP5 models! Abromowitz and Arblaster wrote piece on CLEX website. ECS of IPCC based not only on CMIP6, but also paleoclimate records etc. “Gregory method”. CESM1: linear, CESM2: elbow.

1. Why should Gregory relationship be linear?

10 Christian Jakob - Model Evaluation

Huge field. First question - what is model used for? Evaluation not about “proving” whether model “right” or “wrong”. Models can be used as research tools to “understand” things.

Cannot use a regional model to do a 7 day weather forecast. Not good as a community in CHECKING model fit for purpose - e.g. to understand ENSO need to know whether model even has ENSO, resolves it for the right reasons etc. What are the *causes* of the model errors. Most people neglect this last point.

Tuning scary for climate models.

Qualitative - “Subjective assessment” Catto et al. 2014.

CMIP5 tropical wave analysis. Klocke et al. 2011 (climate models don’t simulate current climate well! Therefore - why have confidence if future projections?), Hall and Qu, 2006.

Mean rainfall in climate still have errors (same since 1990!)

With RMSE - smooth fields do better than rough fields. This is why multi model means give better RMSE. Taylor diagram a measure of RMSE.

Climate community uses “observation simulations”. Bodas-Salcedo et al., 2011.

Model needs to predict current climate - i.e. PDF of weather!

Application oriented - model evaluation examples. Govekar (2014). How well do models simulation 3D structure of clouds.

Regime analysis - cluster analysis, Self organising maps, EOFs. Mason 2015. Williams 2013.

At CMIP conference - 80% talks will be on model errors.

Need to be in touch with model developers DURING RESEARCH PROCESS.

11 Scott Wales - Computing for Climate Models

Alternate grid types - cubed sphere, Voronoi tessellation. Spectral methods not used because they need FFT to go from physical to spectral space.

After compiling - programming language gets converted to “assembly code.” Generally best to leave low level optimisations to compiler.

Memory in a computer is 1 dimensional! Compiler has to change layout of model fields for example into a one dimensional array.

Parallelisation: modern climate models can use up to thousands of cores. Raijin has 4500 nodes.

Decomposition: Spatial domains broken up into chunks. Different cores can either deal with shared memory (OpenMP) or

With shared memory (OpenMP) ordering important. GPUs still using a shared memory system.

Distributed memory (MPI). Memory is separate, then there’s a communication step. “Halo” around each field.

Coupling: different grids in different models (ocean, atmosphere etc). For ACCESS, automatically coupled using program OASIS.

Optimisation: IO Strategies - reading data from disk slower than reading data from memory. Parallelisation output faster than outputting once after all processes completed - i.e. use IO on all processes. Alternatively use special IO processes. key idea, don’t want subsequent processes waiting on IO.

Scaling: Supercomputer resources finite! Experiment with short model runs. Optimise (SU cost x walltime) for different numbers of CPUs to find the best efficiency for your experiment.

ARMAP profiler on Raijin. ARM DDT debugger also on Raijin.

Storage of model output: Only take fields you need! Specially for high-resolution models.

Data Publishing Findable - DOI, metadata, Accessible - Standard formats, Interoperable - Common vocabularies, Reusable - Provenance, licence (FAIR)

Using “standard names” in netcdf files good idea. Files can be read over the internet with THREDDS/OpenDAP!

Grant of 70 million to NCI! Claiming they will be in top 25 in the world. 2GB per CPU.

Jobs prioritised by memory etc etc.

MDSS “put” and “get” for long term tape storage. 400 GB SSD on each compute node.

CLEX Resources shared between projects: v45, w35, w40, w42, w48 etc. Communicate with group members for jobs 1 KSU or higher.

ncimontory - graph of how users have used storage.

CLEX CMS: wiki: climate-cms.wikis.unsw.edu.au chat line: arccss.slack.com cws_help@nci.org.au

NCI: docs: opus.nci.org.au

cws_help@nci.org.au

1. What is a rank? Essentially a task handed by a single CPU (MPI term).
2. How to estimate memory usage?

References