Environmental Modelling at both large and small scales: How simulating complexity leads to a range of computing challenges.

Bryan Lawrence
(and a cast of thousands)

Types of models: "Global Climate Model" (GCM)

Fully Coupled.

All components interact via two-way fluxes of relevant quantities.

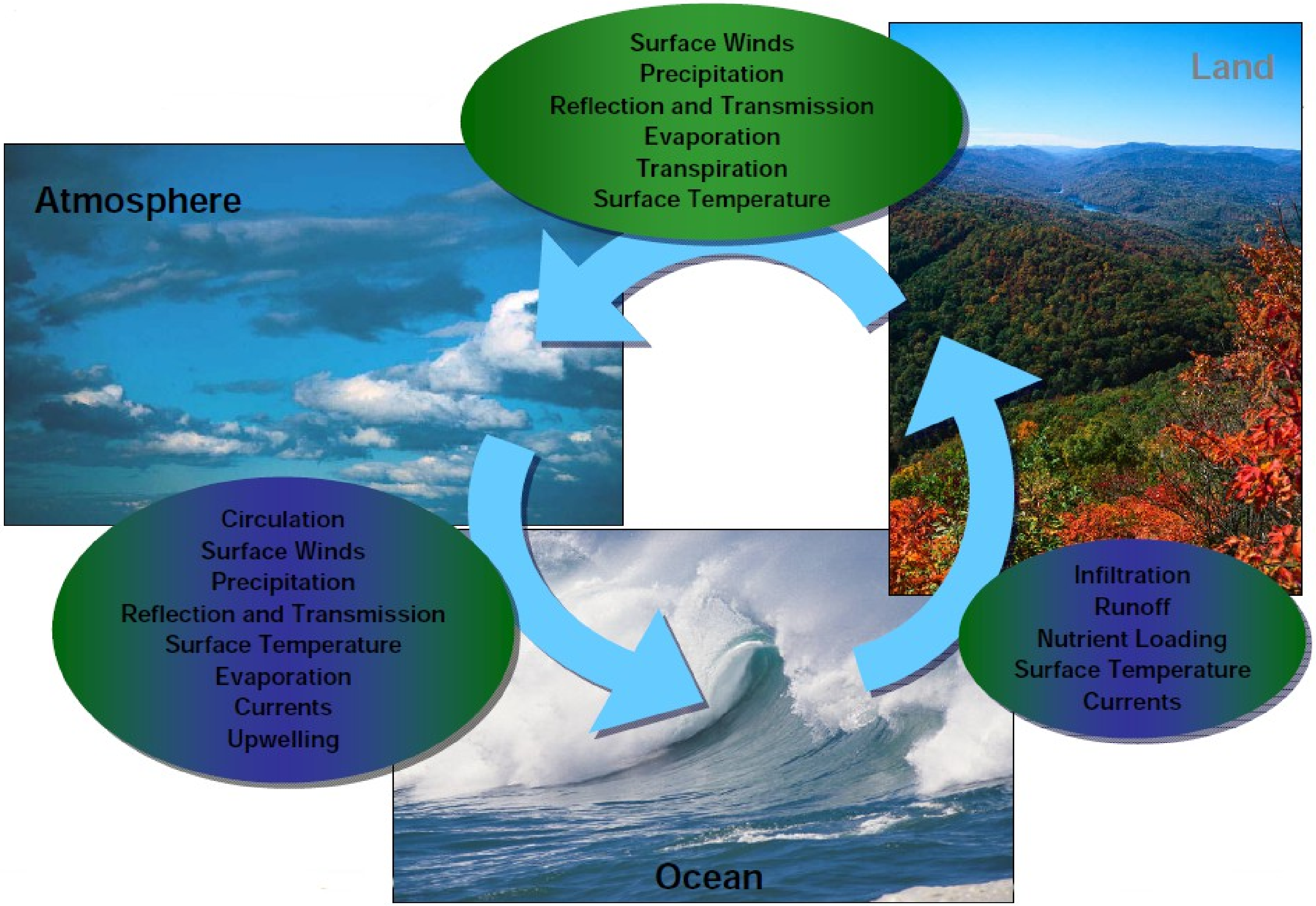
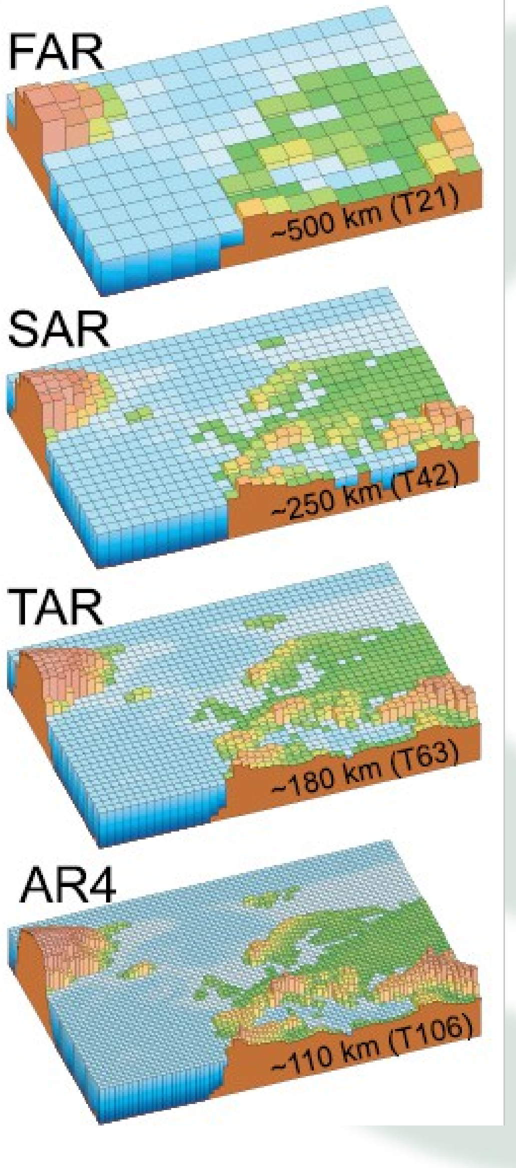
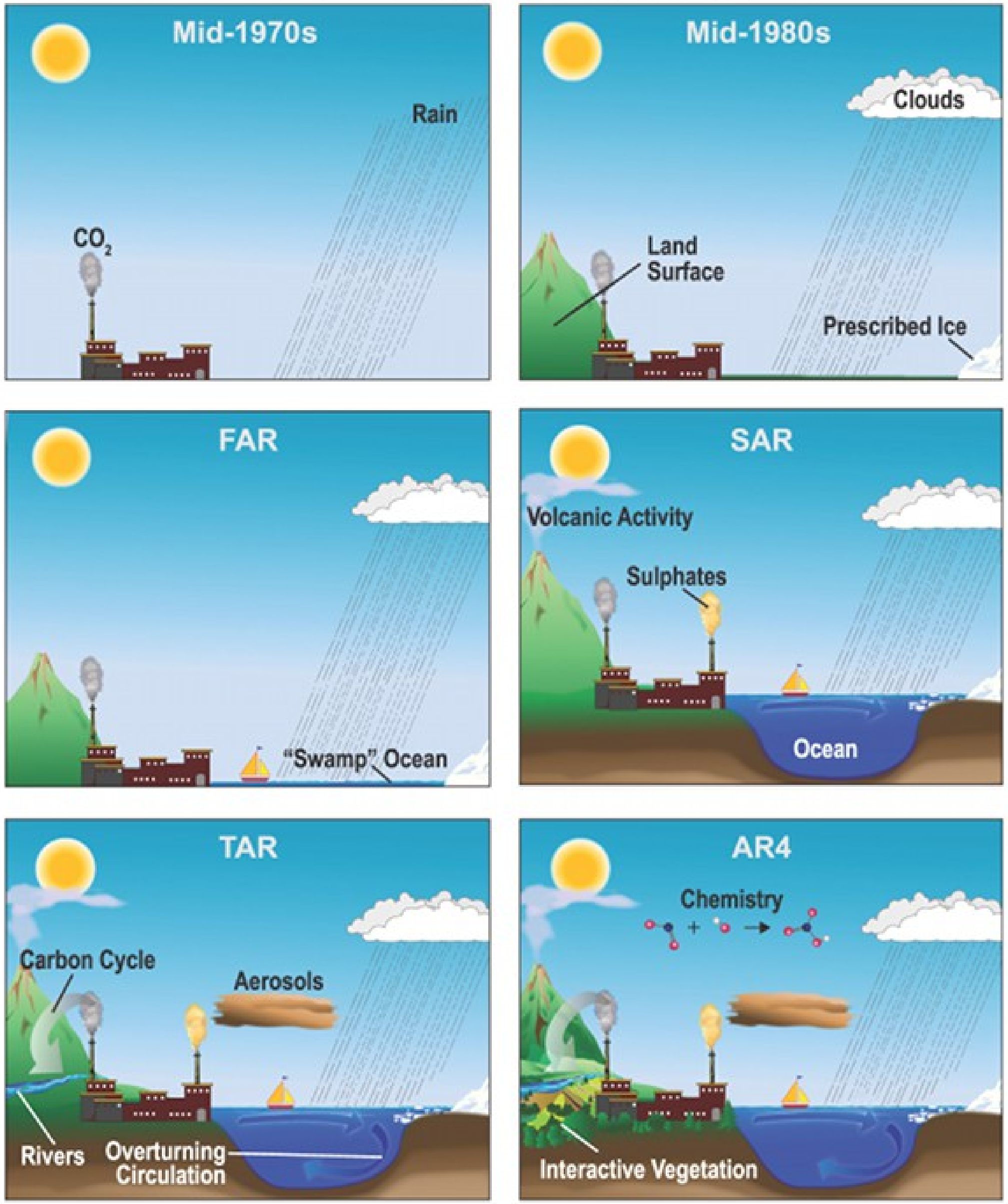


Image: from
J. Lafeuille, 2006

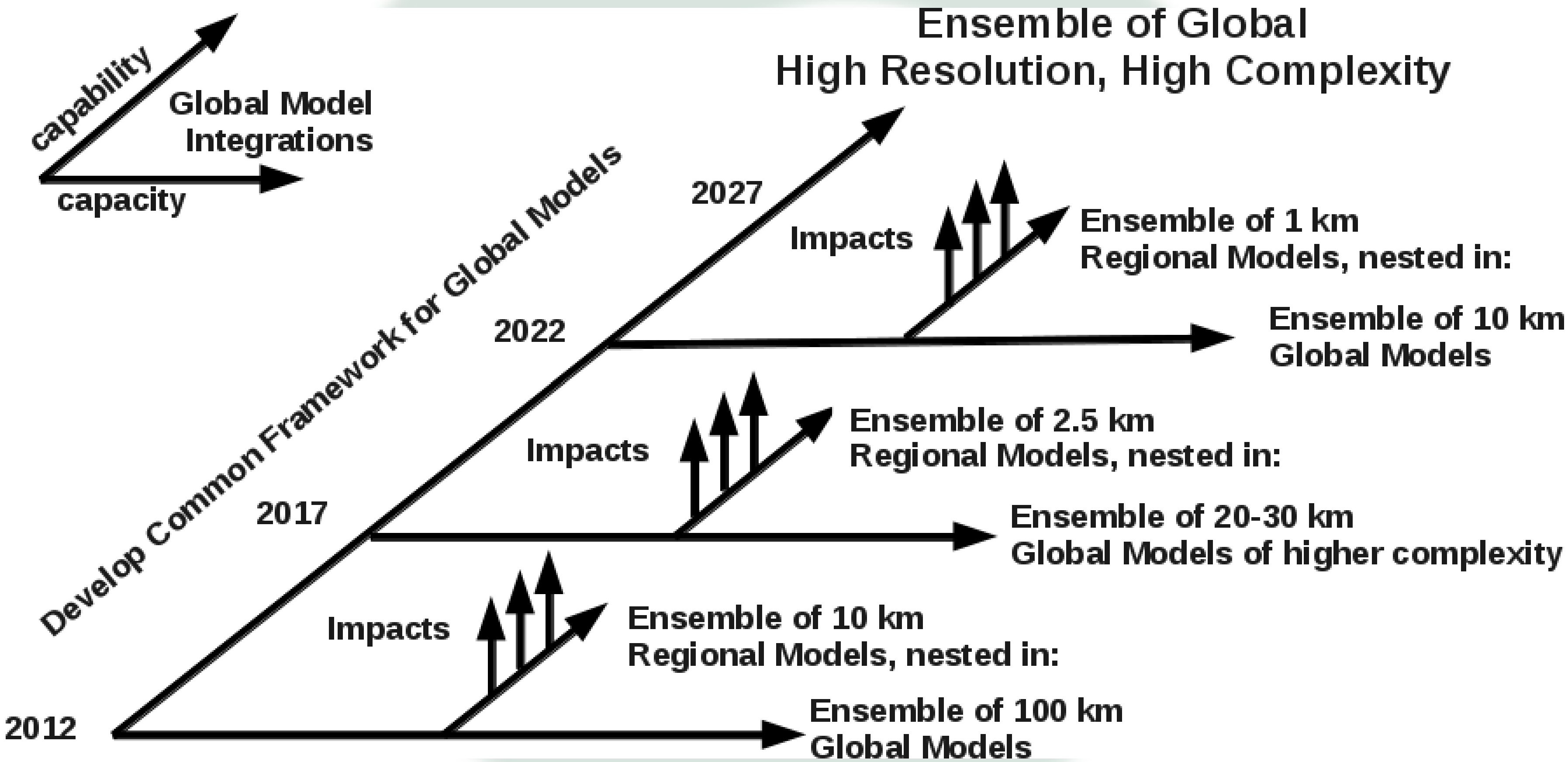
The world in global climate models

FAR:1990
SAR:1995
TAR:2001
AR4:2007
AR5:2013

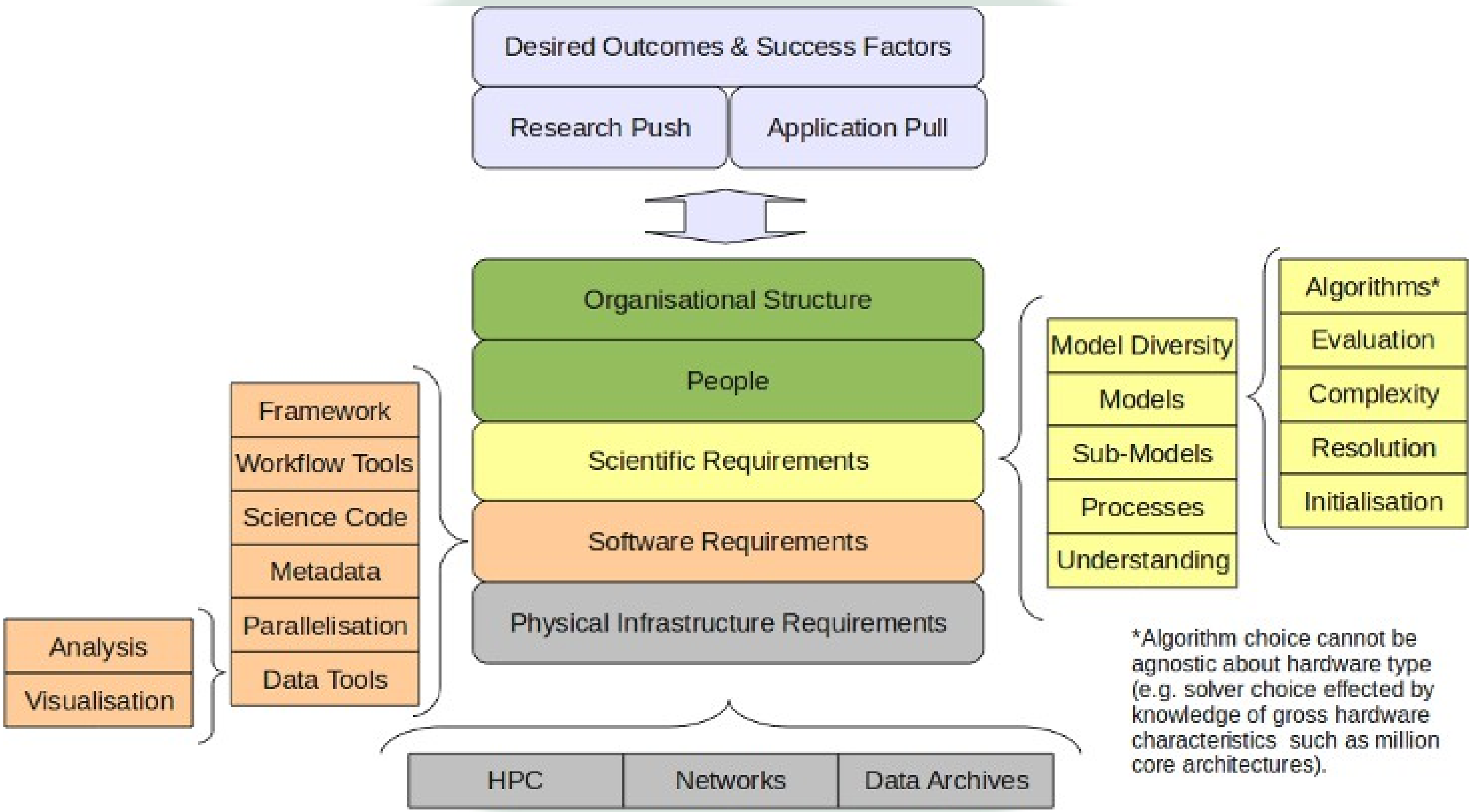




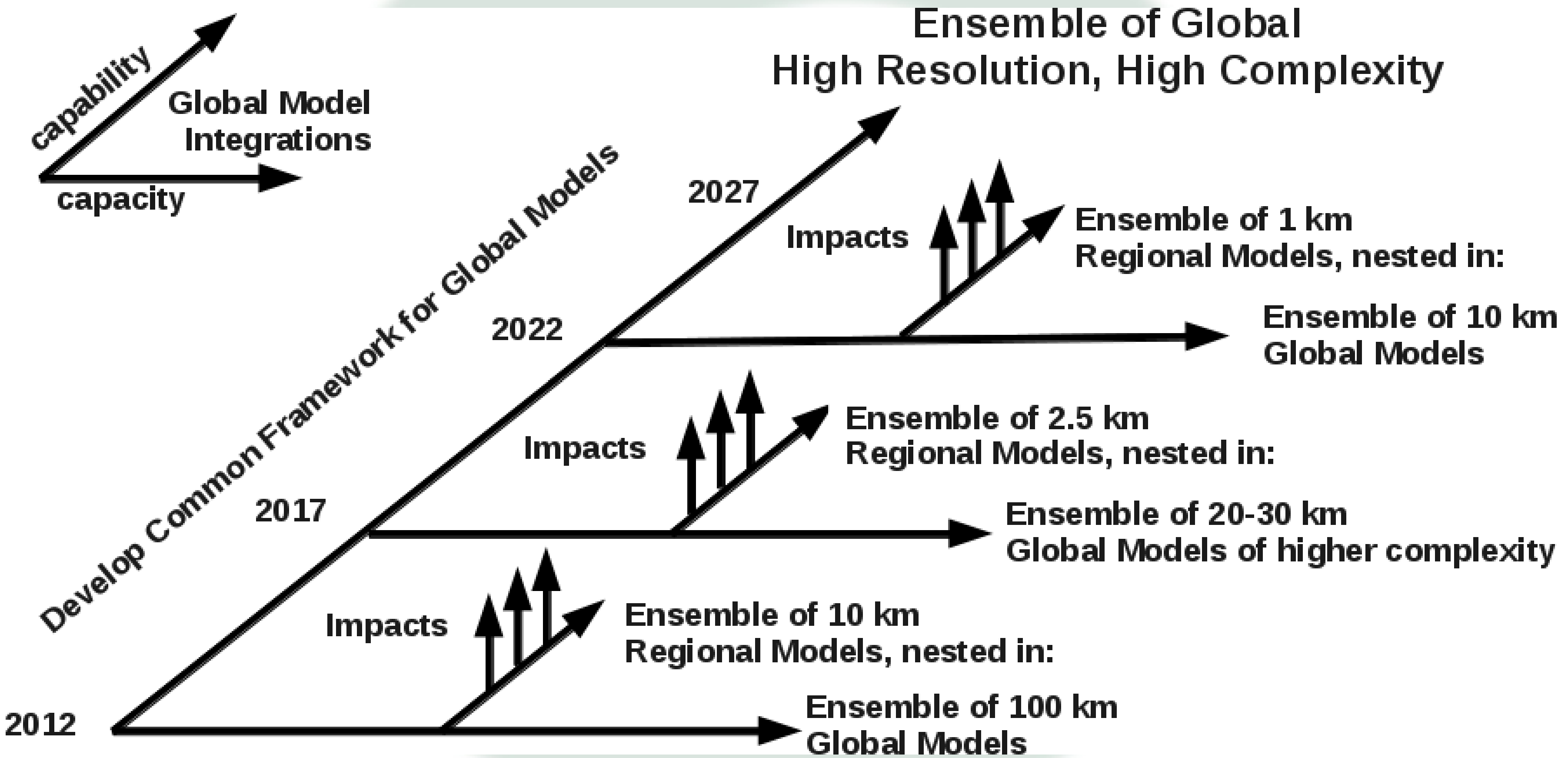
Interacting models and scales



What's needed?



Interacting models and scales



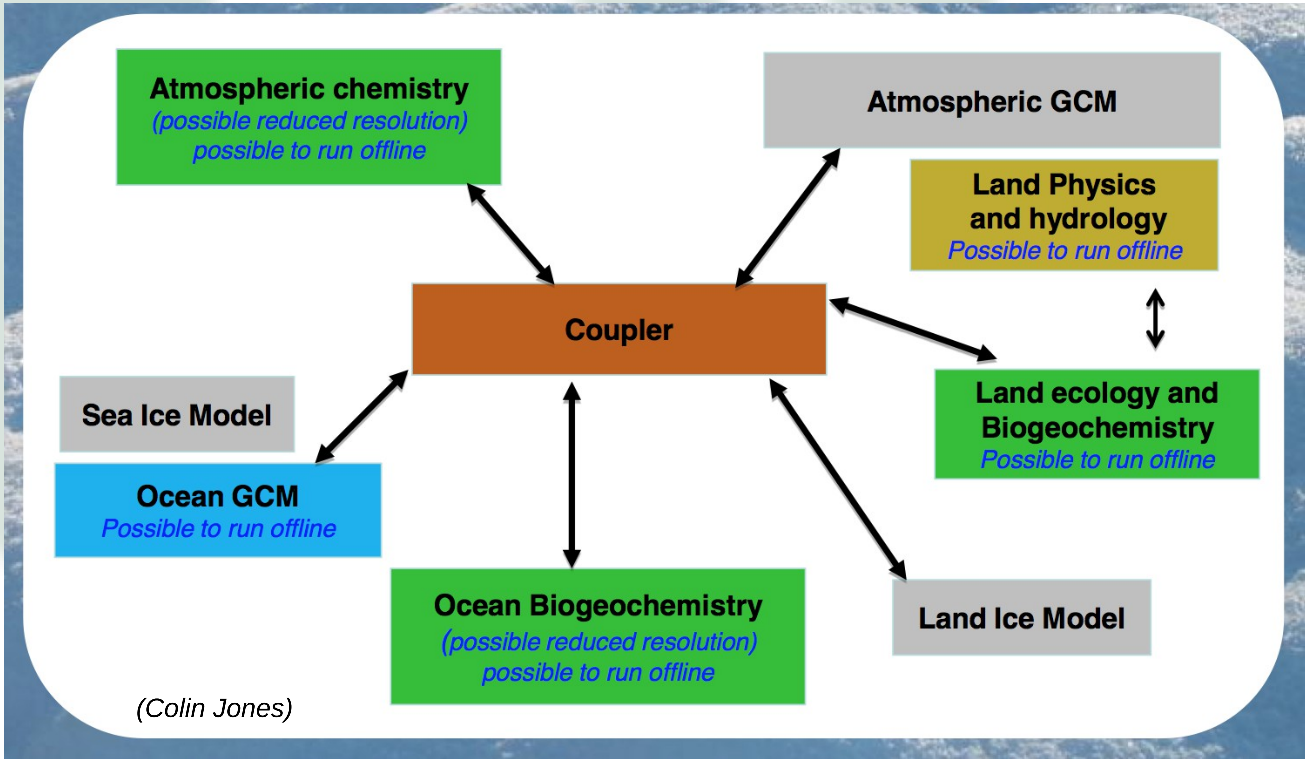
Consider two examples from two ends of the spectrum

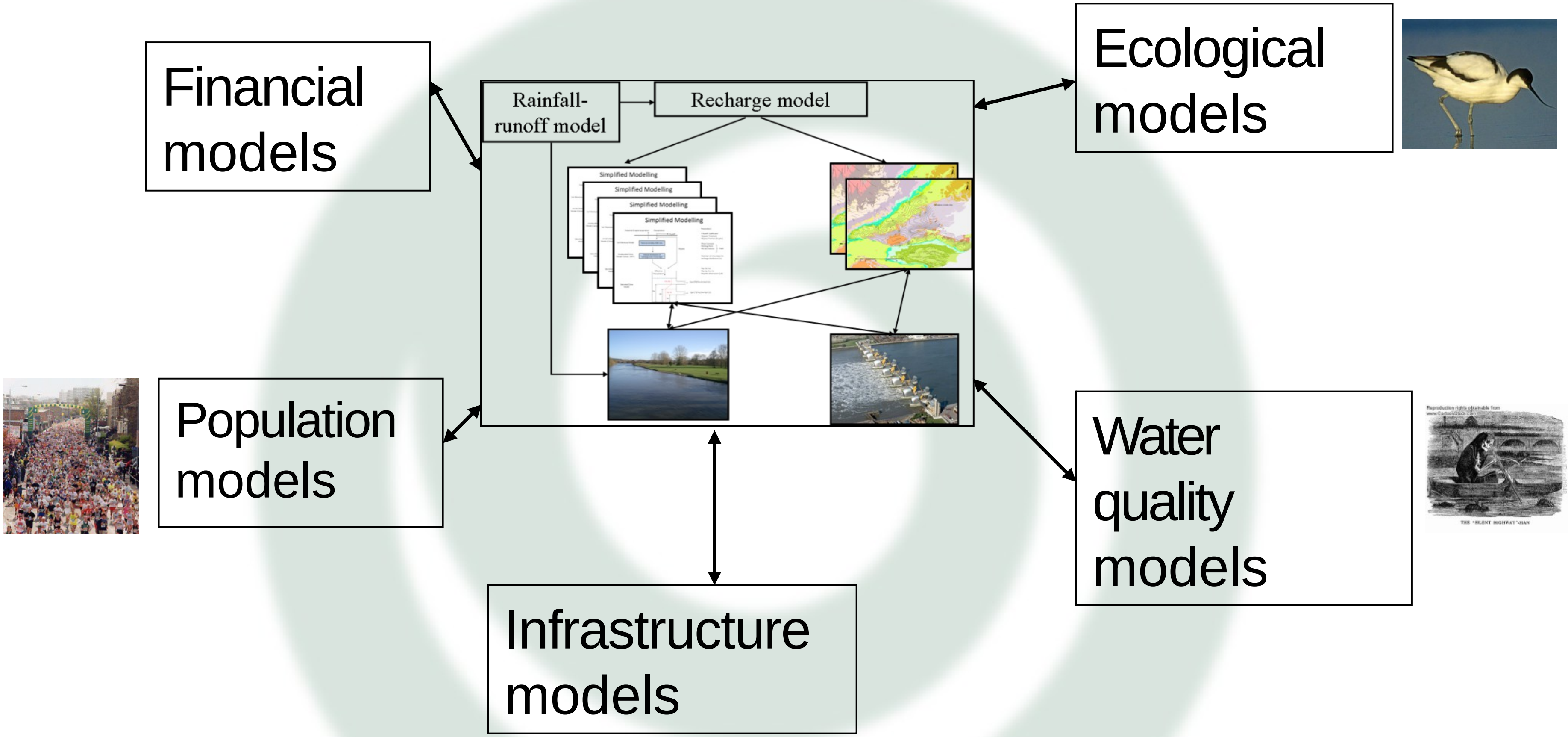
(2) Complexity - UKESM1

UK Earth System Model!

- Joint Project between the Met Office and the NERC community, led by Colin Jones (NCAS, University of Leeds, based in the MOHC).
- Two overriding objectives: develop a world-leading ESM *and* grow a community around it!

Aiming to change the computational structure to have a centralised coupler with two way coupling of key processes.





Slide courtesy of Andrew Hughes, British Geological Survey

One way coupling

ESM+ "Impact" Model: Antarctic Krill fishery

Antarctic krill:
Gross biomass production: 342–536Mty^{-1}
Adult human biomass: 287Mt
Sustainable catch limit: 5.61Mty^{-1}
Current global fisheries landings: 80Mty^{-1}

BAS designed statistical model of Krill growth: Driven by observed or simulated SST and chlorophyll!

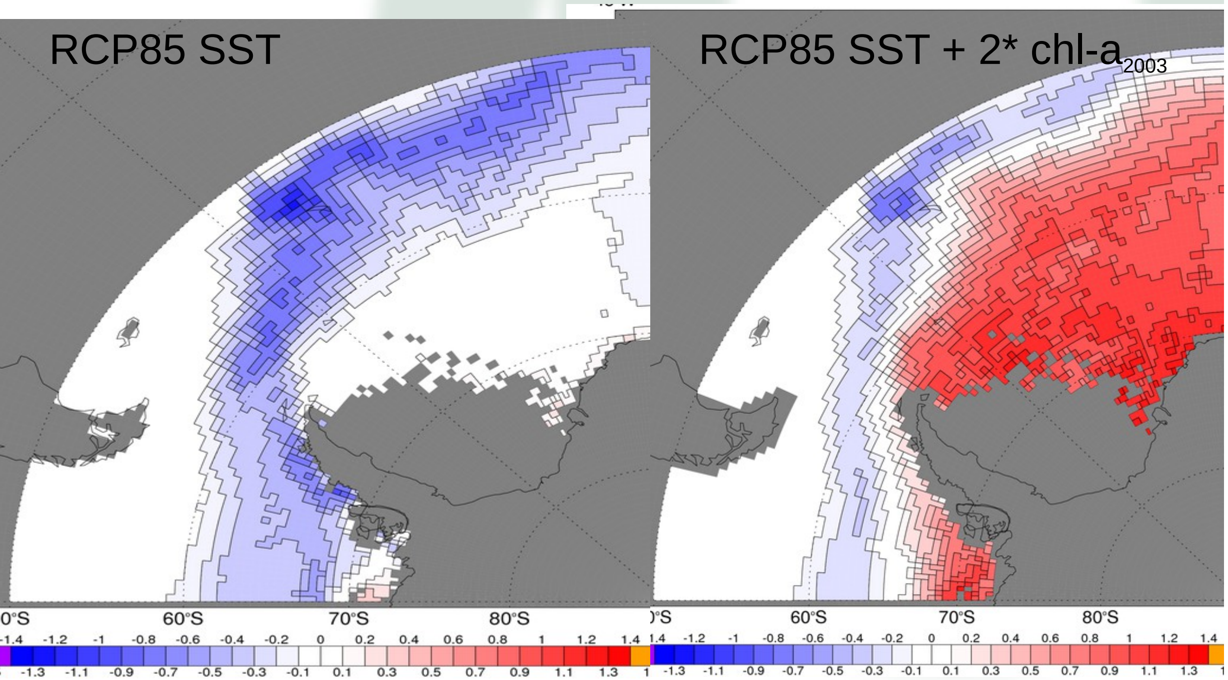
Simeon Hill, Tony Phillips, Angus Atkinson

ESM+ “Impact” Model: Antarctic Krill fishery

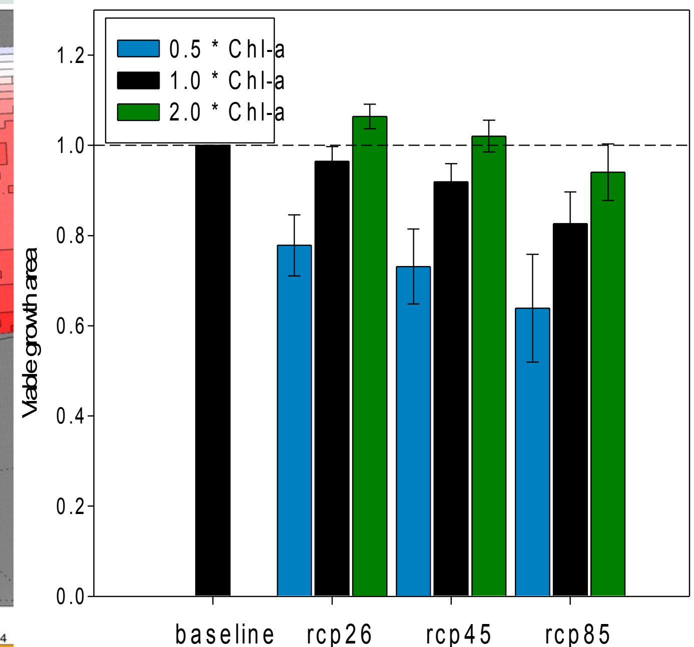
Antarctic krill:
Gross biomass production: 342–536Mty^{-1}
Adult human biomass: 287Mt
Sustainable catch limit: 5.61Mty^{-1}
Current global fisheries landings: 80Mty^{-1}

BAS designed statistical model of Krill growth: Driven by observed or simulated SST and chlorophyll!

Simulated change in Krill growth habitat by 2100
(Average warming in critical 90° sector of southern ocean=1.3C



Changes in viable habitat
Using CMIP5 projections



Simeon Hill, Tony Phillips, Angus Atkinson

A major challenge:

Interacting communities &
Interacting Codes!

The “Coupling” Problem

(Lots of other challenges, some of which I'll discuss tomorrow ...)

Coupling Requirements

1) **Two sides of the interface need to provide the right variables.**
   - And have they been modelled "sufficiently" well?
   - (This is about our scientific confidence in the individual models.)
2) **Can the exchange be modified explicitly**
   - Solution is stable if the future state of variables in either component can be calculated from past states in the other.
   - (This can be problematic! New approaches: service models!)
3) **Are the variables on the same grid.**
   - Or can they be made to be so.
   - (We know we can solve this one, but maybe not at exascale)

A representative sample of coupling technologies

1) Direct/Bespoke

2) ESMF
   - A framework
   - Single Executables

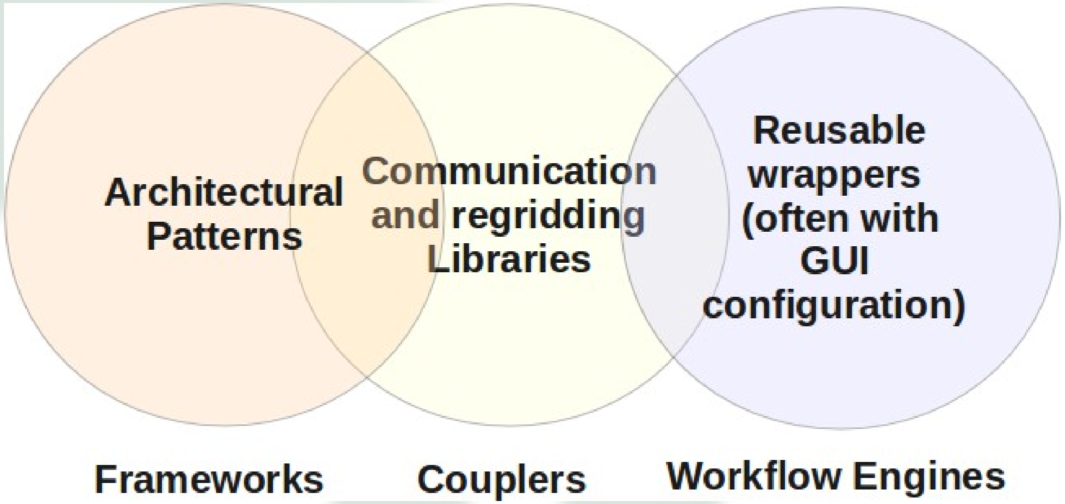
3) OASIS
   - A coupler
   - Multiple Executables

4) OpenMI
   - A limited framework

5) CSDMS
   - minimally intrusive framework (Basic Model Interface) + library implementation of the Common Component Architecture (CCA)

6) Kepler
   - Workflow Management (coupling via files)

7) *BFG (Bespoke Framework Generator)*
   - *Metadata driven coupling a la carte*



Segue from the Science to the Technology

We begin with communities, and their models.

We progress to interacting communities, and interacting models.

- Generally one community modelling paradigm dominates how that is done! The “top-model”, often an atmosphere dynamical core (or it's driver) …
- Almost immediately we start to see a code divergence, as the coupled version differs from the standalone version.

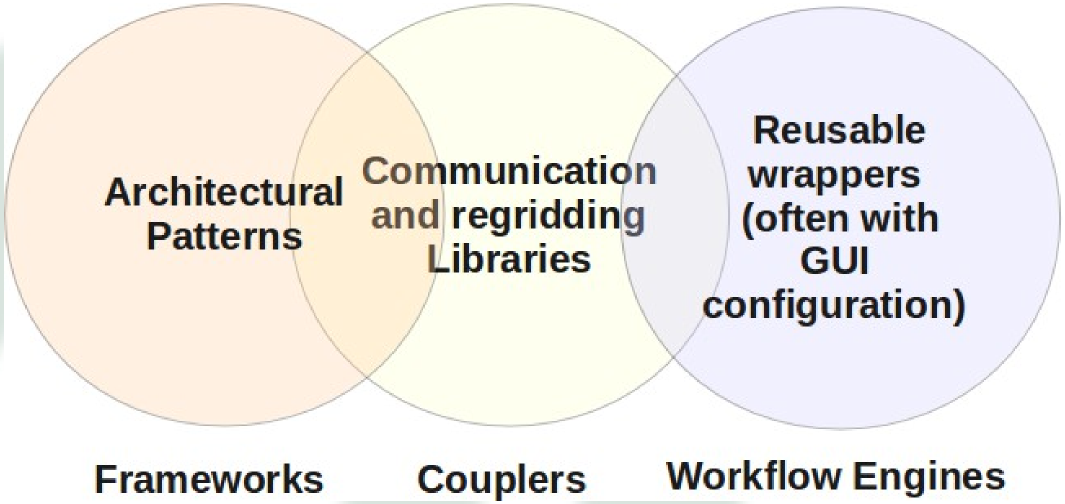
We know that not all communities are going to be able to interact by direct two-way coupling via a “top-model”.

- This simply doesn't scale, socially, or technically.
- But we don't always know what things we can neglect in terms of feedback. We need to experiment.

Two use cases to consider:

- Can we mitigate against that code divergence?
- Can we simplify the interfaces to support experimentation?

A representative sample of coupling technologies



1) Direct/Bespoke
2) ESMF
   – A framework
3) OASIS
   – A coupler
4) OpenMI
   – A limited framework
5) CSDMS
   – minimally intrusive framework (Basic Model Interface) + library implementation of the Common Component Architecture (CCA)
6) Kepler
   – Workflow Management (coupling via files)
7) *BFG (Bespoke Framework Generator)*
   – *Metadata driven coupling a la carte*

Different implications for code and workflow

Questions to consider: How usable (and reusable) are these approaches? In particular, how intrusive/invasive is the approach?

If the methodology is difficult to approach, intellectually, or in terms of the implementation, it can be difficult for all communities involved in coupling to have equal knowledge & that's not good for the science!

If the methodology is intrusive, this might have real implications for the necessity for having multiple versions of the component models.

- Two forms of intrusiveness to consider:
  - the need for refactoring (changing and/or reordering code), and
  - Sheer volume of code inserted/needed/to-be-comprehended
- (Hidden dependencies on other component code and behaviour … can't be avoided, but can it be minimised?)

All of these apply to the entire work flow, not just the running model! Need to consider debugging, evaluation, post-processing data formats etc.

Avoiding Top Model Coupling Paradigms: The role of generative approaches.

From a science perspective: there is no such thing as a component model! From my/your viewpoint my/your “component” model is a top model:

- Ideally I/you want to be coupling other components into my/your model.
- Consider the land surface case, running at (lower) resolution in an ESM, and at (higher) resolution being the top model (coupling precip via files) and complex ground water models …
  - Inevitably using different coupling paradigmsin those two directions!
- It simply cannot be good (efficient) science to maintain two code stacks. Much better to generate the coupling from one code stack.

Obvious role for generative tools like BFG.

- (I have yet to fully understand the possibilities of CSDMS/BMI ...)

Need: Better Workflow Tooling for "Coupling"!

It's not just about the runtime!

- Comprehending the code!
- Development
- Debugging
- Documenting
- Validating
- Evaluating

Comment from meeting a year ago: " communities interacting.. it's easy to get output, it's hard to know if it's correct … "

This talk in a sentence! (Except maybe it's not so easy to take the first step.)

All these things *currently* require "artisans" not "engineers" and certainly not "scientists". That has to change, and the tooling needs to facilitate all these things!

Tightly Coupling versus Data Coupling

Take the Krill example previously: statistical sub-process model driven both by parameters which can be simulated in an ESM (Sea Surface Temperature) and Chlorophyll (which is not yet).

It could be integrated/coupled within an ESM system, but it might be unreasonably expensive to do the hypothesis testing required (e.g. varying chlorophyll) by using the whole model for each sub-experiment!

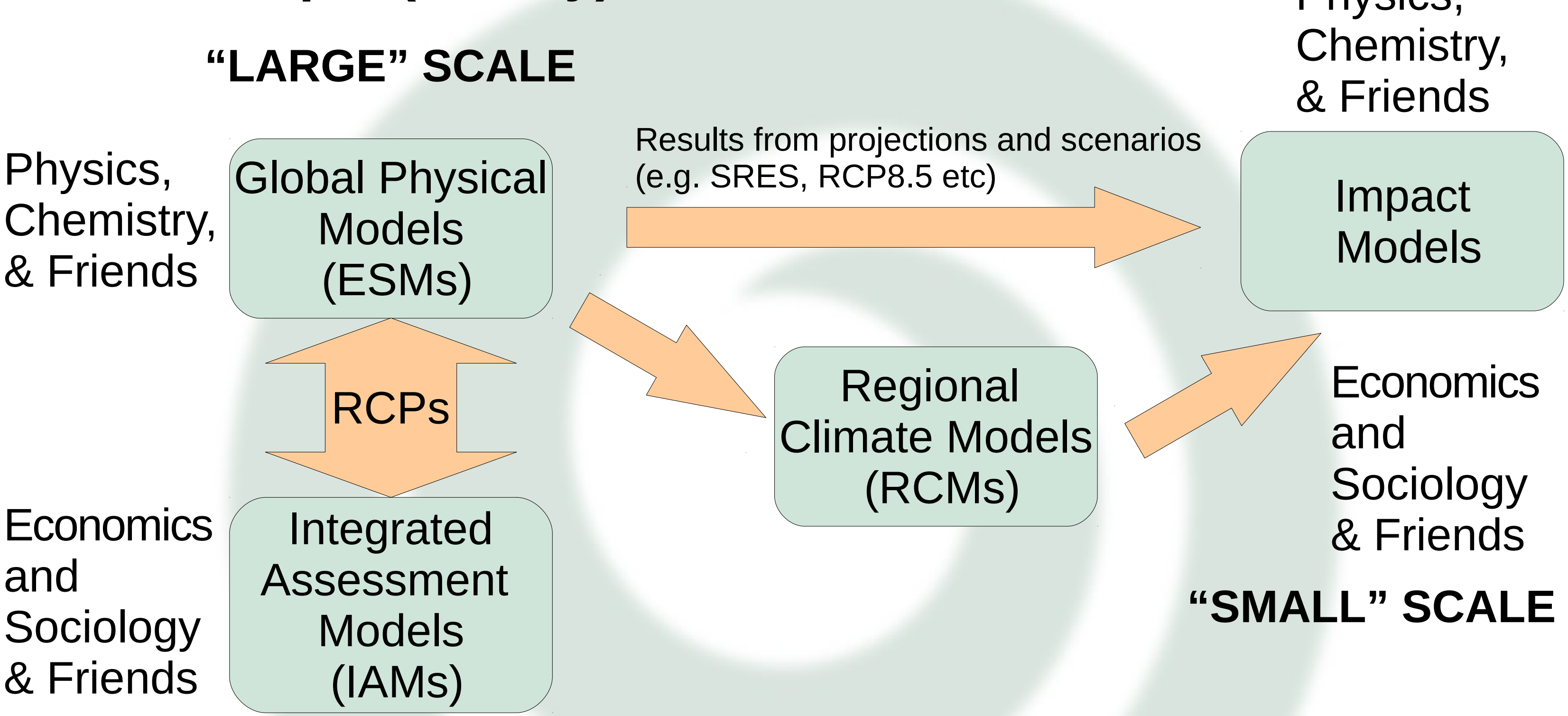
It certainly becomes unreasonable if we think we are going to do this for every "small-scale" impact problem!

However, we can integrate the plans (experiments) and the data (interfaces) to expedite joint science! Full in-simulation "coupling" might or might not follow!

Important criteria for "full coupling": Is there a science use case which demonstrates "two-way" coupling on timescales we can afford to simulate?

Scale Interaction and Impact: two strategies!

(1) The status quo (mostly)



“Coupling” (the arrows), mostly achieved by using data from one model as boundary/initial conditions for the other!

Scale Interaction and Impact: two strategies!

(2) Adding a risk based paradigm

Physics,
Chemistry,
& Friends

Impacts
Models

“SMALL”
SCALE

Vulnerabilities:
e.g. Length of sequence of drought days
e.g # of degree days

Economics
and
Sociology
& Friends

Analysis
(risks
changing?)

Analysis leads to policy response

ESMs
(and IAMS/RCPs)

“LARGE”
SCALE

Projections/
Scenarios

Economics
and
Sociology
& Friends

Physics,
Chemistry,
& Friends

“Coupling” (the arrows), mostly achieved by activities identifying risks and response!

Vulnerability Assessment: Example

Catastrophe Modelling: Assesses the vulnerability of insurance companies to financial loss from natural hazards including extreme weather and climate events.

- Based on sets of plausible events (e.g. hurricanes, wind storms). Conceptually easy to assess the financial impact for a given event.
- Relies heavily on short historical records (generating large sampling uncertainty) & upon increasingly dodgy assumptions about the stationarity of climate!

Non-trivial to go from ensembles of climate predictions to reliable likelihoods, since our existing ensembles cannot be assumed to cover all dimensions of uncertainty! However that's where we need to go, but this is not a talk about uncertainty!

Summary

Coupling is a technical solution to BOTH the science requirements and the shape of the scientific community.

Optimising for any one of those alone (or just for performance) is likely to result in short time wins at the expense of long term victory.

There is no one right solution for all communities and all problems.

- Generative techniques (e.g. BFG) or really simple framework support (e.g. CSDMS BMI) will be part of dealing with that!
- Sometimes it's better to frame the problem to avoid coupling at all … (e.g the catastrophe modelling example).

Having those points in mind when we develop our coupling toolboxes should increase their utility.

Probably smart not to assume that our explicit coupling paradigms are going to survive in a higher resolution and more highly concurrent exascale world.