UNIVERSITY OF CAMBRIDGE

Institute of Computing for Climate Science

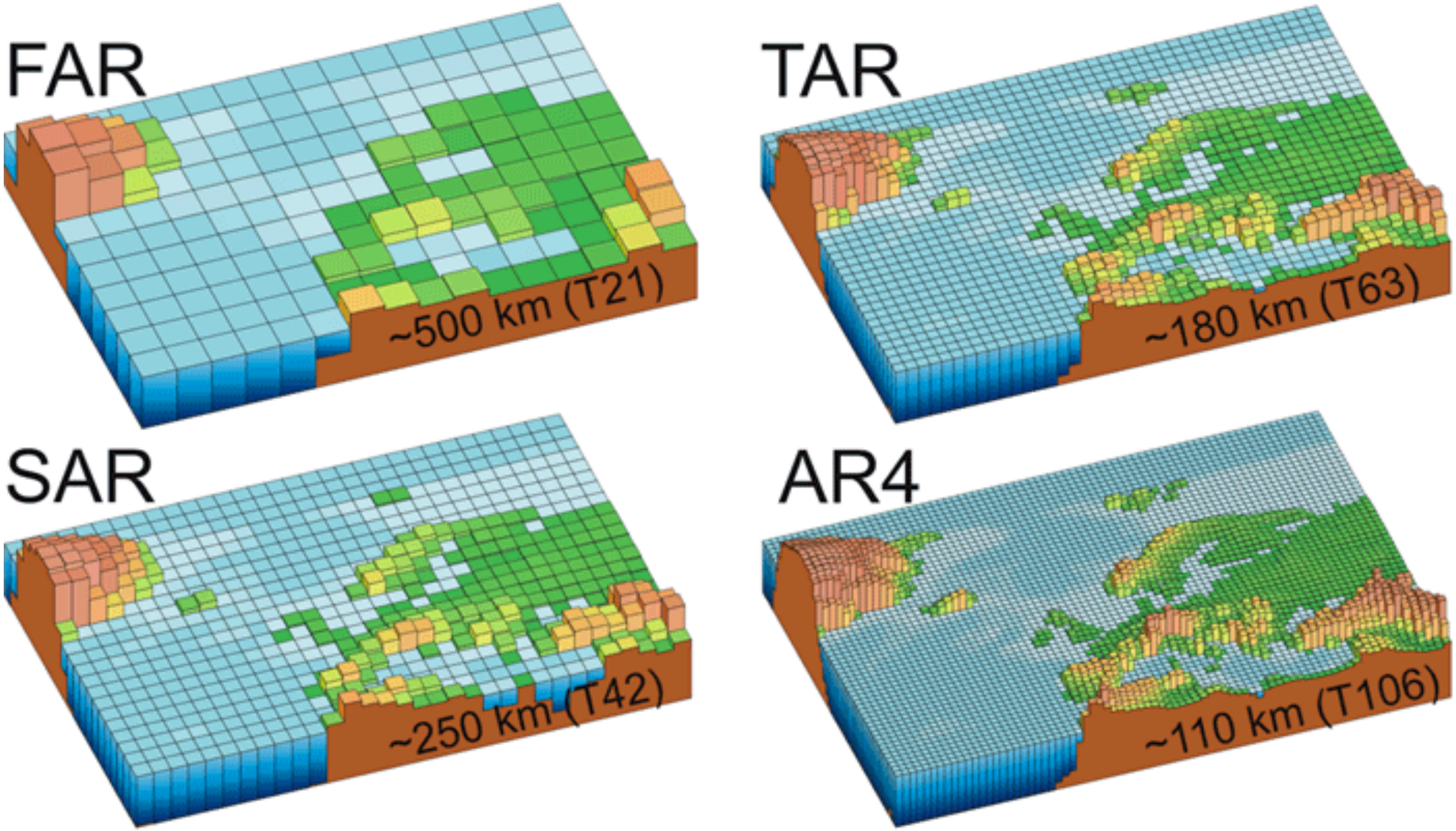
ICCS: The what, why, and how.

Dominic Orchard
Department of Computer Science and Technology

SCHMIDT FUTURES

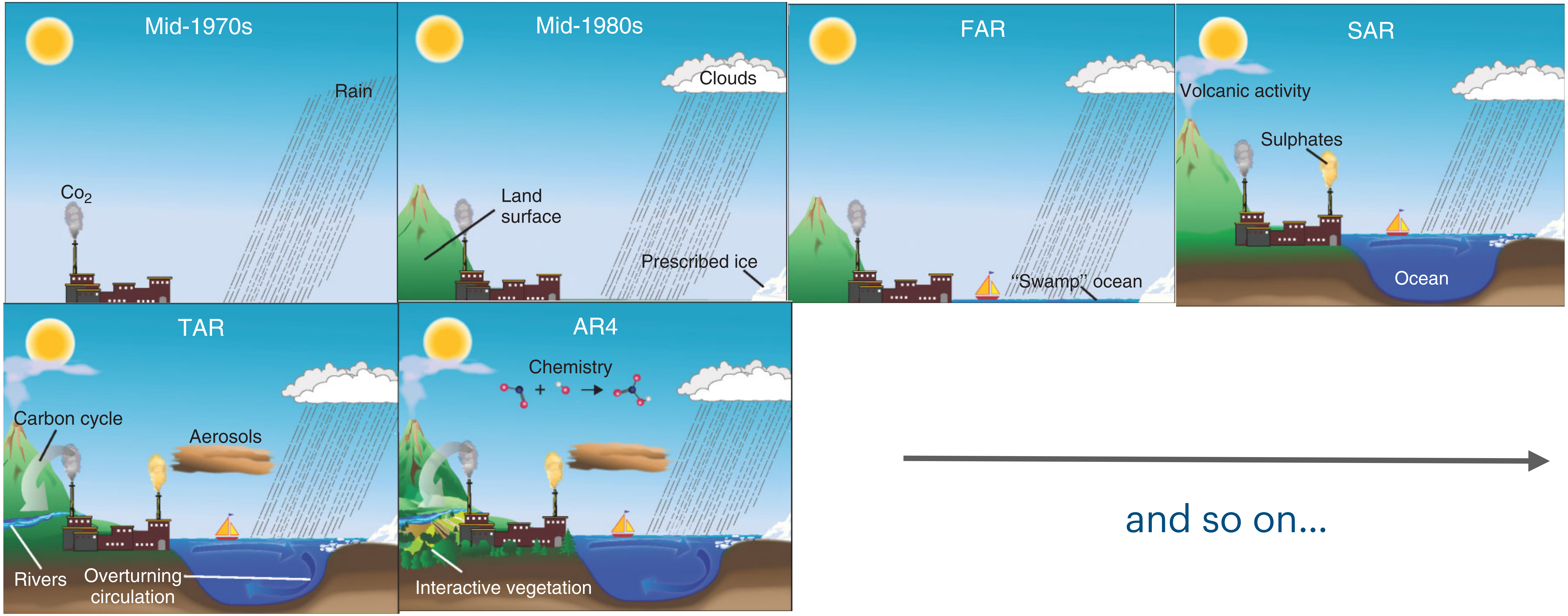
1850-2021 (Ed Hawkins “Warming stripes”)

Increasing resolution



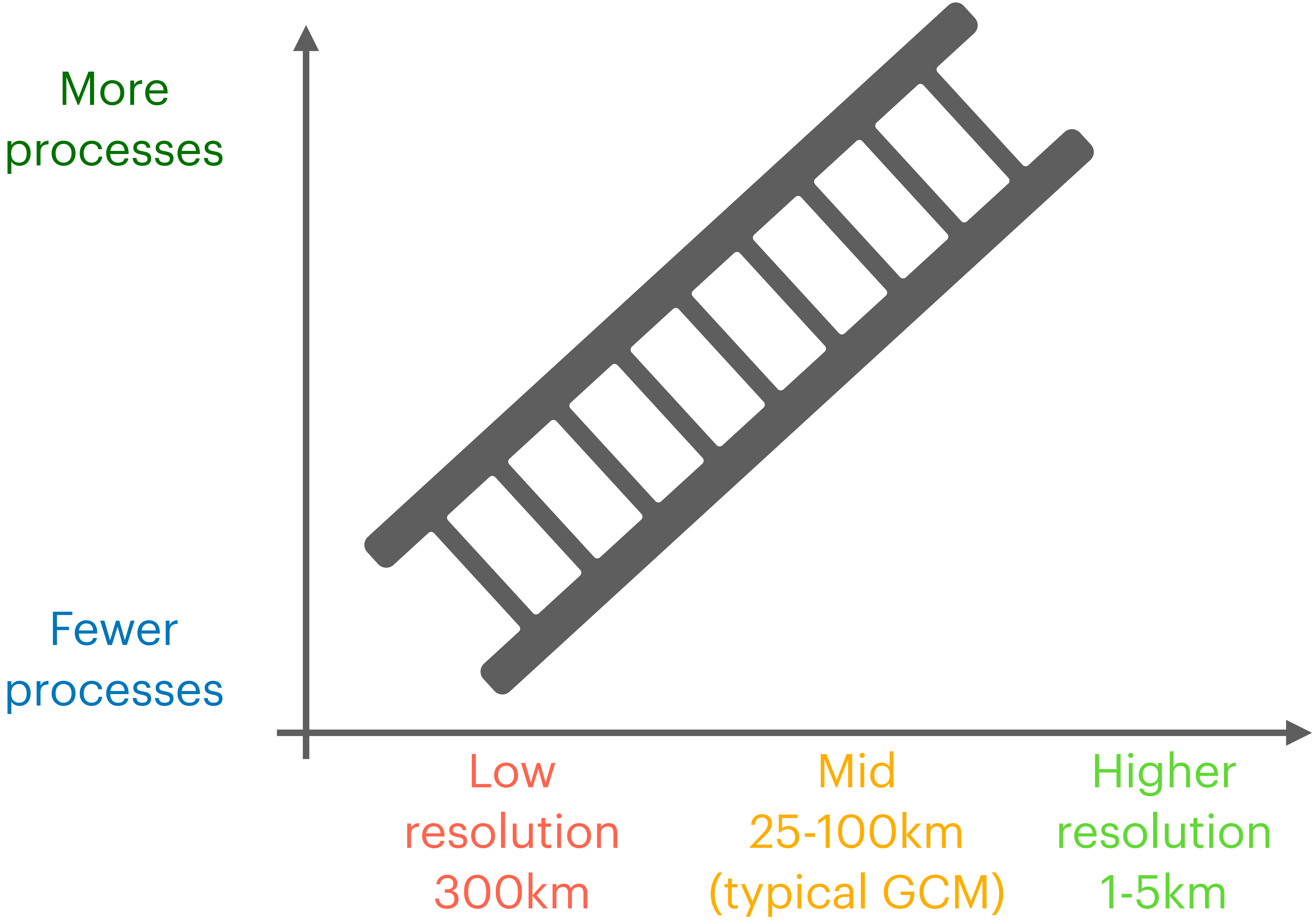
graphics from 4th IPCC report (2007)

Increasing process complexity

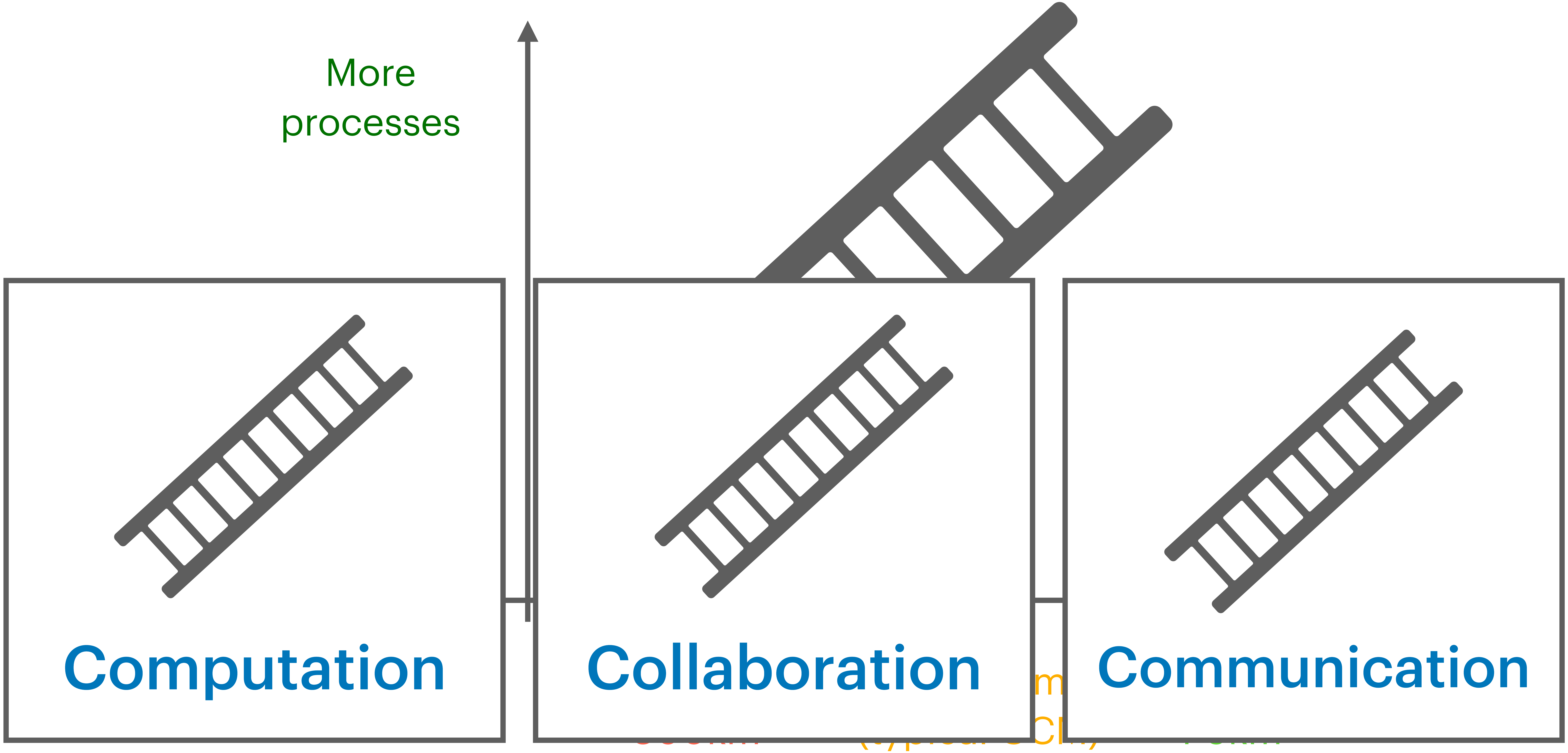


and so on...

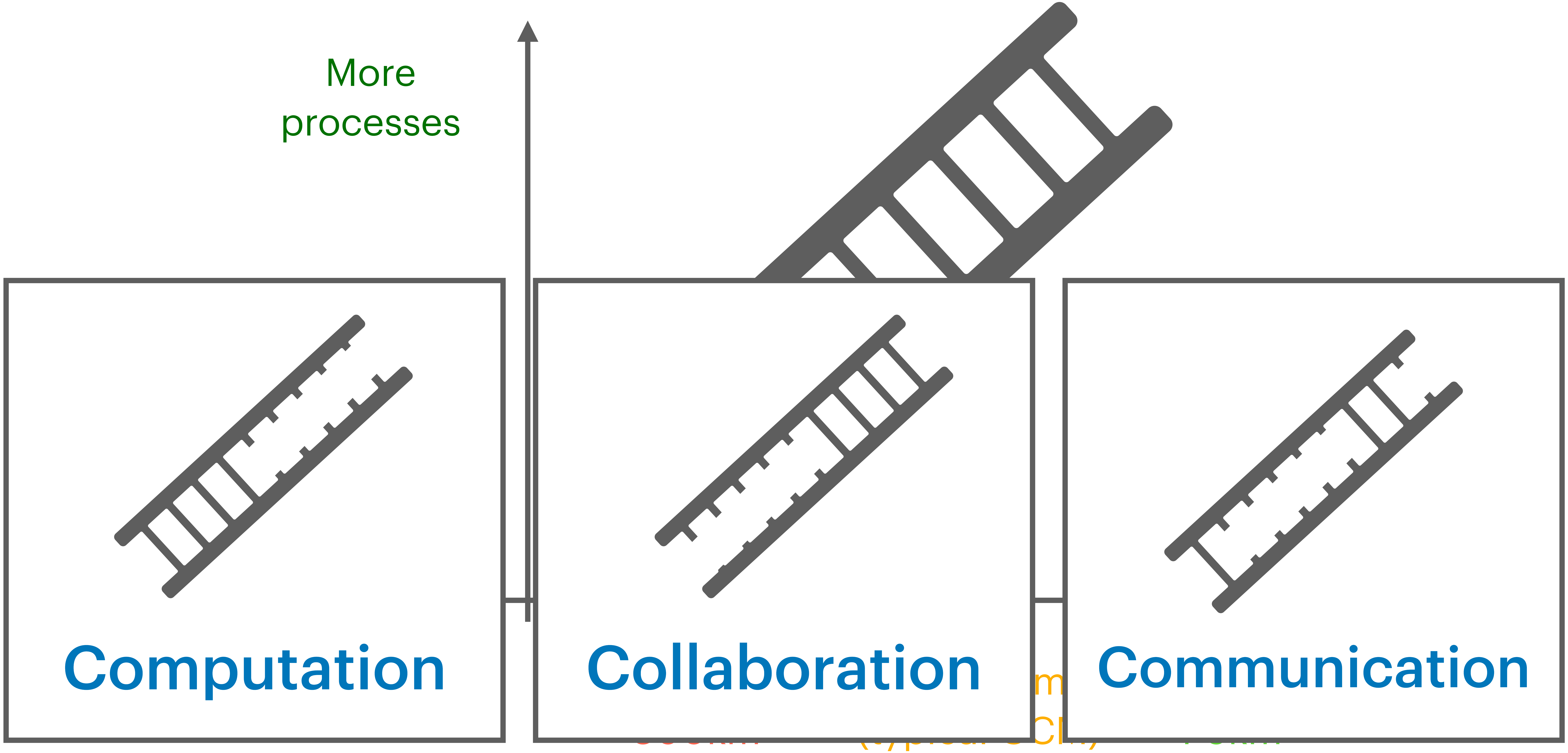
Better prediction: “climbing the ladder” (Charney)



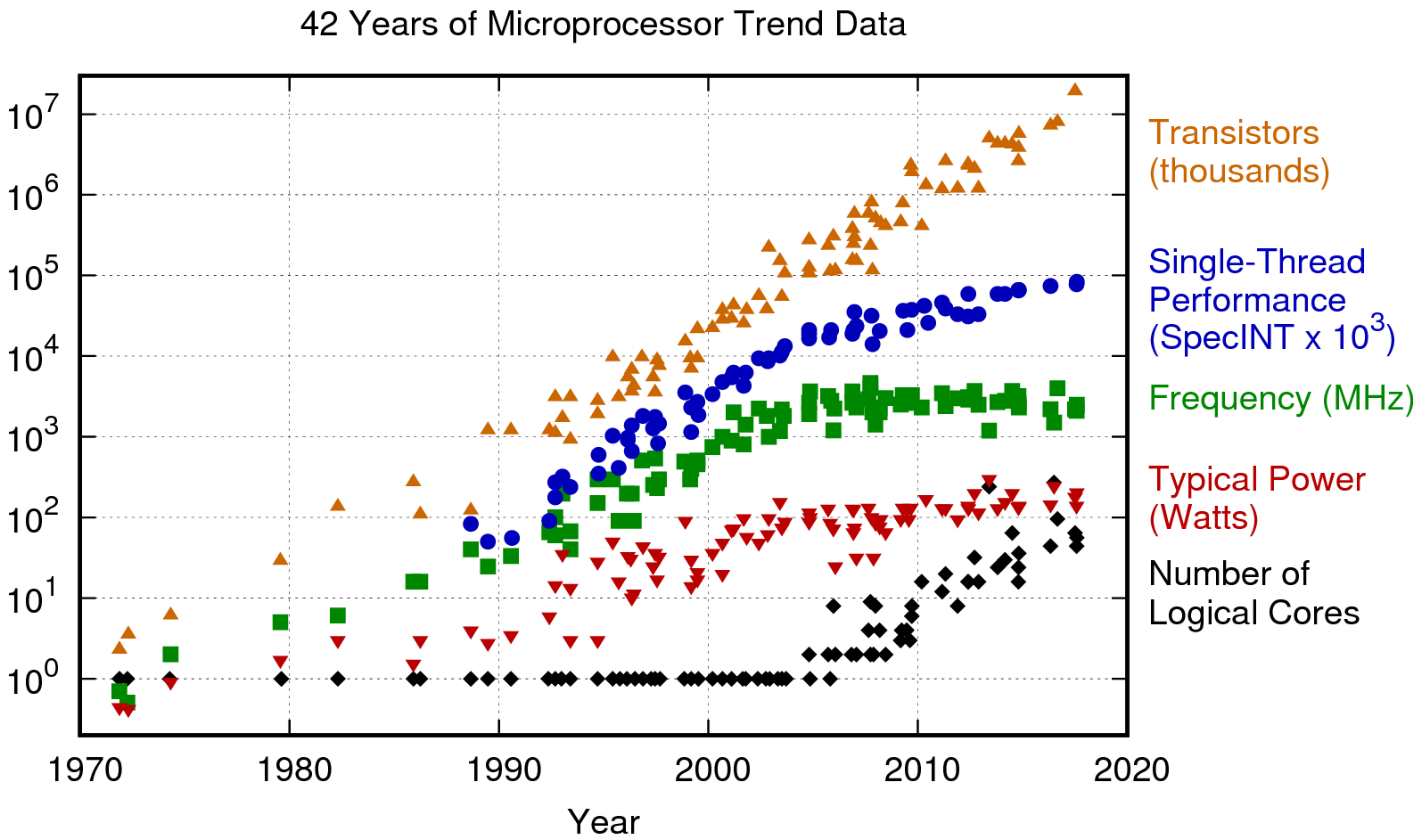
Better prediction: “climbing the ladder” (Charney)



Better prediction: “climbing the ladder” (Charney)

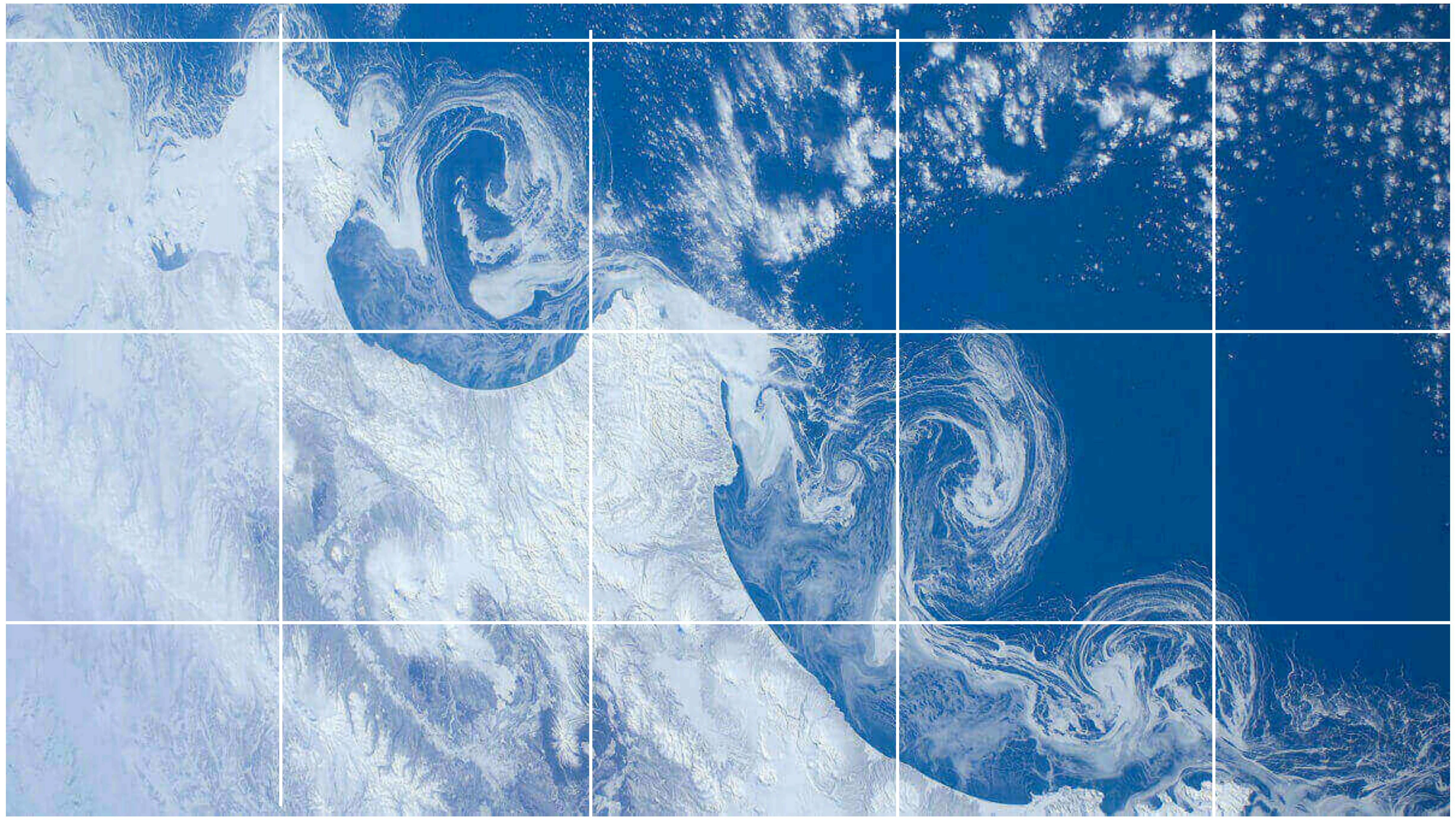


1. Scaling computational performance

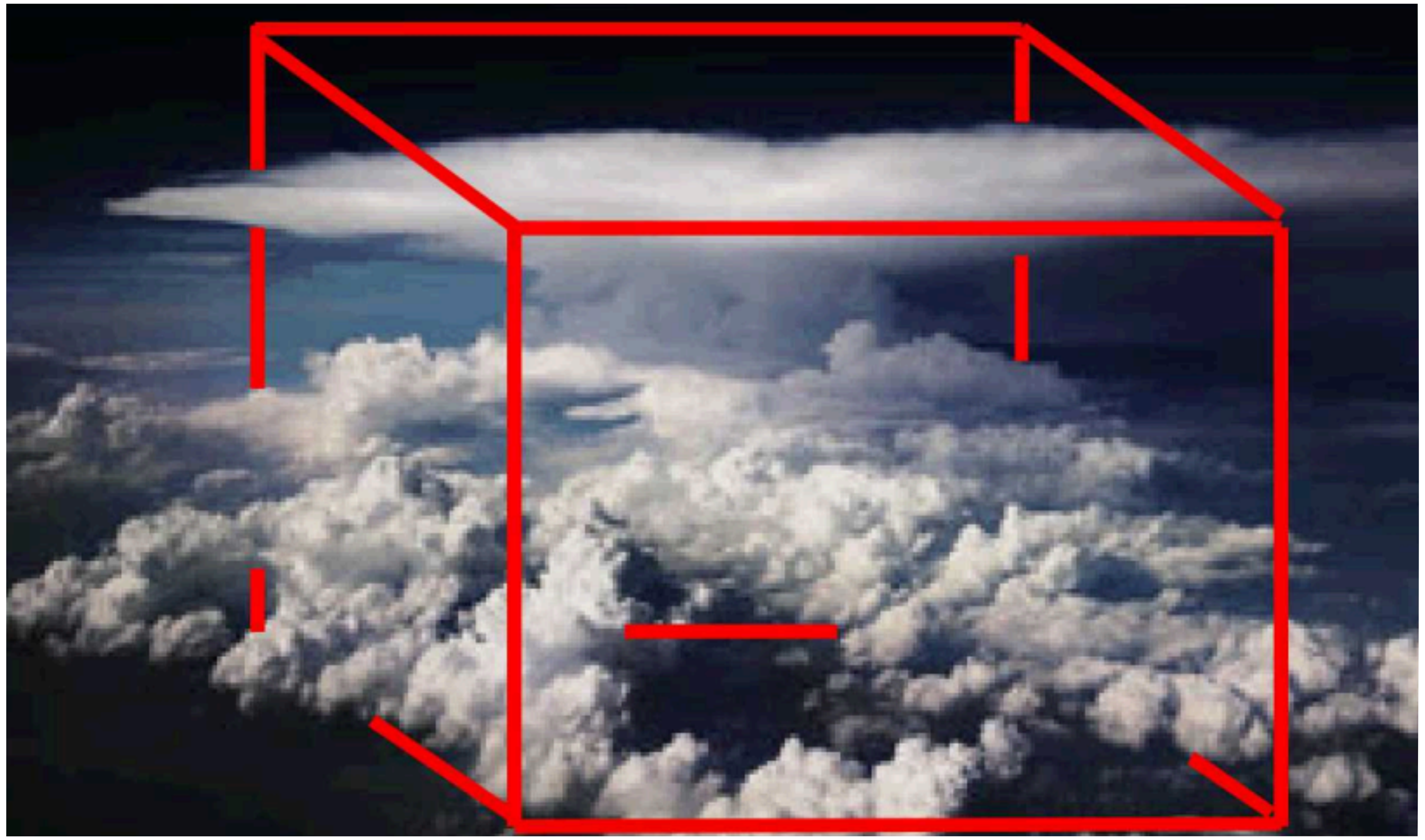


Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

Approximating sub grid processes



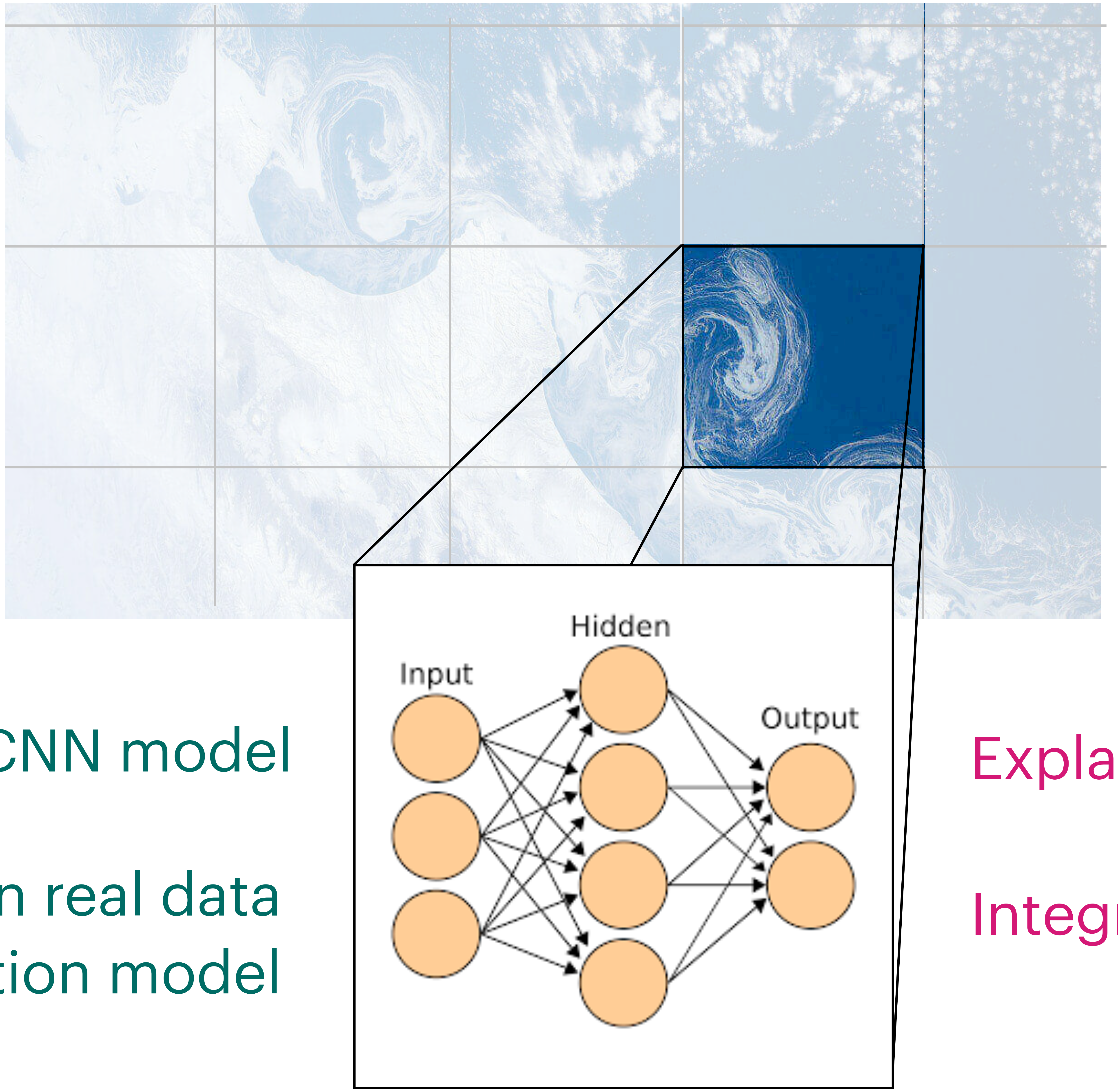
NASA / Wikimedia Commons



Hillman et al. 2020

Uncertainty / error *vs.* expense

Solution: Data-driven subgrid closures



CNN model

Train on real data
or high-resolution model

Explainability?

Integration into GCM?

2. Scaling collaborations

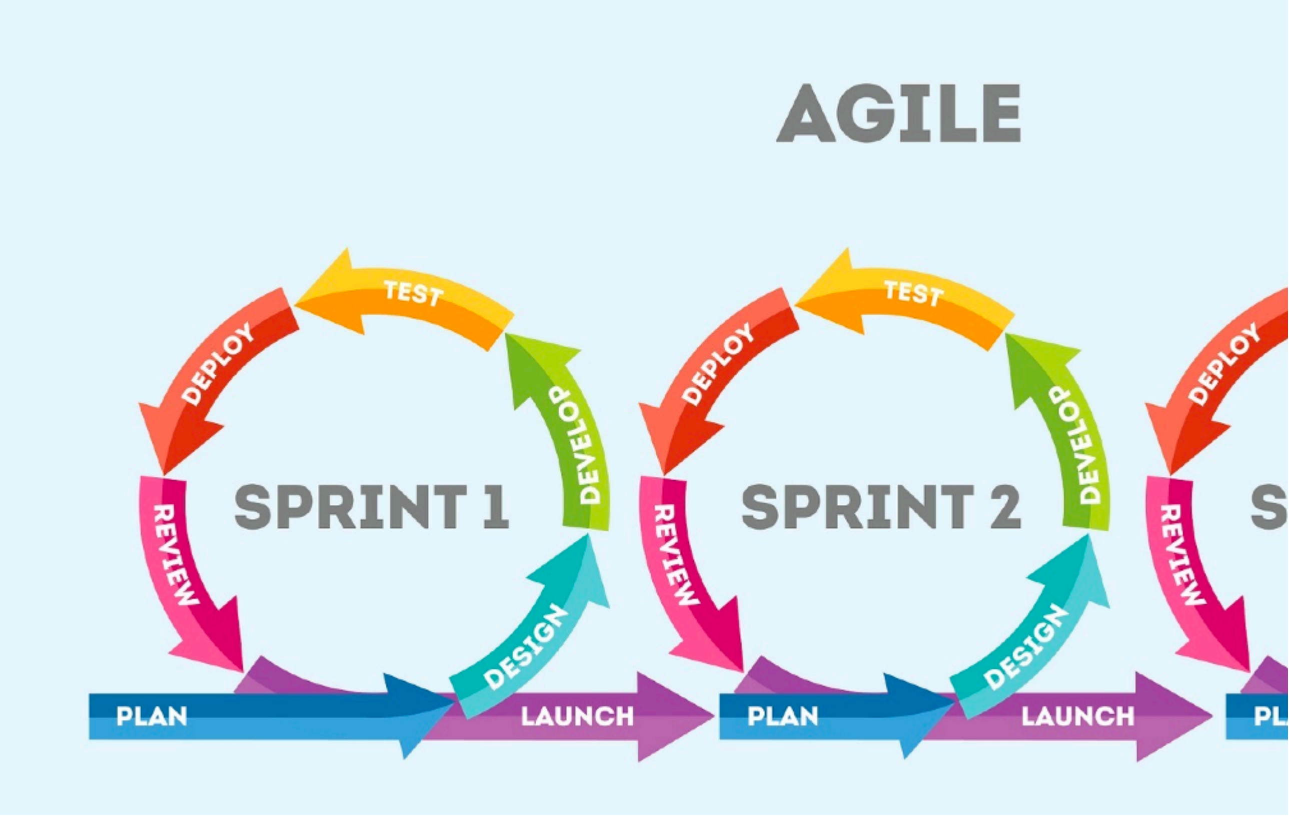
The Two Complexities

Inherent

Accidental

Solution: Software engineering tools & techniques

Processes



Version control & public curators

Build systems & containers

Debugging

Profiling

Testing and verification

Structural and culturual/sociological change

Software Sustainability Institute

BETTER SOFTWARE BETTER RESEARCH

RSE

Society of Research Software Engineers

The TURING WAY

3. Scaling communication

Environmental Data Science (2022), 1: e11, 1–28
doi:10.1017/eds.2022.10

CAMBRIDGE UNIVERSITY PRESS

APPLICATION PAPER

A sensitivity analysis of a regression model of ocean temperature

Rachel Furner[1,2,*], Peter Haynes[1], Dave Munday[2], Brooks Paige[3], Daniel C. Jones[2] and Emily Shuckburgh[4]

[1]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, United Kingdom
[2]British Antarctic Survey, Cambridge, United Kingdom
[3]UCL Centre for Artificial Intelligence, Computer Science, University College London, London, United Kingdom
[4]Department of Computer Science and Technology, University of Cambridge
*Corresponding author. E-mail: raf59@cam.ac.uk

Received: 14 January 2022; **Revised:** 09 June 2022; **Accepted:** 21 July 2022

Keywords: Data science; interpretable ML; model sensitivity; oceanography; regression model

Abstract

There has been much recent interest in developing data-driven models for weather and climate predictions. However, there are open questions regarding their generalizability and robustness, highlighting a need to better understand how they make their predictions. In particular, it is important to understand whether data-driven models learn the underlying physics of the system against which they are trained, or simply identify statistical patterns without any clear link to the underlying physics. In this paper, we describe a sensitivity analysis of a regression-based model of ocean temperature, trained against simulations from a 3D ocean model setup in a very simple configuration. We show that the regressor heavily bases its forecasts on, and is dependent on, variables known to be key to the physics such as currents and density. By contrast, the regressor does not make heavy use of inputs such as location, which have limited direct physical impacts. The model requires nonlinear interactions between inputs in order to show any meaningful skill—in line with the highly nonlinear dynamics of the ocean. Further analysis interprets the ways certain variables are used by the regression model. We see that information about the vertical profile of the water column reduces errors in regions of convective activity, and information about the currents reduces errors in regions dominated by advective processes. Our results demonstrate that even a simple regression model is capable of learning much of the physics of the system being modeled. We expect that a similar sensitivity analysis could be usefully applied to more complex ocean configurations.

Impact Statement

Machine learning provides a promising tool for weather and climate forecasting. However, for data-driven forecast models to eventually be used in operational settings we need to not just be assured of their ability to perform well, but also to understand the ways in which these models are working, to build trust in these systems. We use a variety of model interpretation techniques to investigate how a simple regression model makes its predictions. We find that the model studied here, behaves in agreement with the known physics of the system. This works shows that data-driven models are capable of learning meaningful physics-based

```fortran
module simulation_mod
  use helpers_mod
  implicit none

  contains

  subroutine compute_tentative_velocity(u, v, f, g, flag, del_t)
    real u(0:imax+1, 0:jmax+1), v(0:imax+1, 0:jmax+1), f(0:imax+1, 0:jmax+1),  &
         g(0:imax+1, 0:jmax+1)
    integer flag(0:imax+1, 0:jmax+1)
    real, intent(in) :: del_t

    integer i, j
    real du2dx, duvdy, duvdx, dv2dy, laplu, laplv

    do i = 1, (imax-1)
      do j = 1, jmax
        ! only if both adjacent cells are fluid cells */
        if (toLogical(iand(flag(i,j), C_F)) .and.                              &
            toLogical(iand(flag(i+1,j), C_F))) then

          du2dx = ((u(i,j)+u(i+1,j))*(u(i,j)+u(i+1,j))+                      &
                  gamma*abs(u(i,j)+u(i+1,j))*(u(i,j)-u(i+1,j))-               &
                  (u(i-1,j)+u(i,j))*(u(i-1,j)+u(i,j))-                        &
                  gamma*abs(u(i-1,j)+u(i,j))*(u(i-1,j)-u(i,j)))               &
                  /(4.0*delx)
          duvdy = ((v(i,j)+v(i+1,j))*(u(i,j)+u(i,j+1))+                       &
                  gamma*abs(v(i,j)+v(i+1,j))*(u(i,j)-u(i,j+1))-               &
                  (v(i,j-1)+v(i+1,j-1))*(u(i,j-1)+u(i,j))-                    &
                  gamma*abs(v(i,j-1)+v(i+1,j-1))*(u(i,j-1)-u(i,j)))           &
                        /(4.0*dely)
          laplu = (u(i+1,j)-2.0*u(i,j)+u(i-1,j))/delx/delx+                   &
                  (u(i,j+1)-2.0*u(i,j)+u(i,j-1))/dely/dely

          f(i,j) = u(i,j) + del_t*(laplu/Re-du2dx-duvdy)
        else
          f(i,j) = u(i,j)
        end if
      end do
    end do
```

3. Challenge: conflation of concerns in code

Abstract model

Solution strategy

Prediction calculation

Prediction calculation
Abstract model
Solution strategy

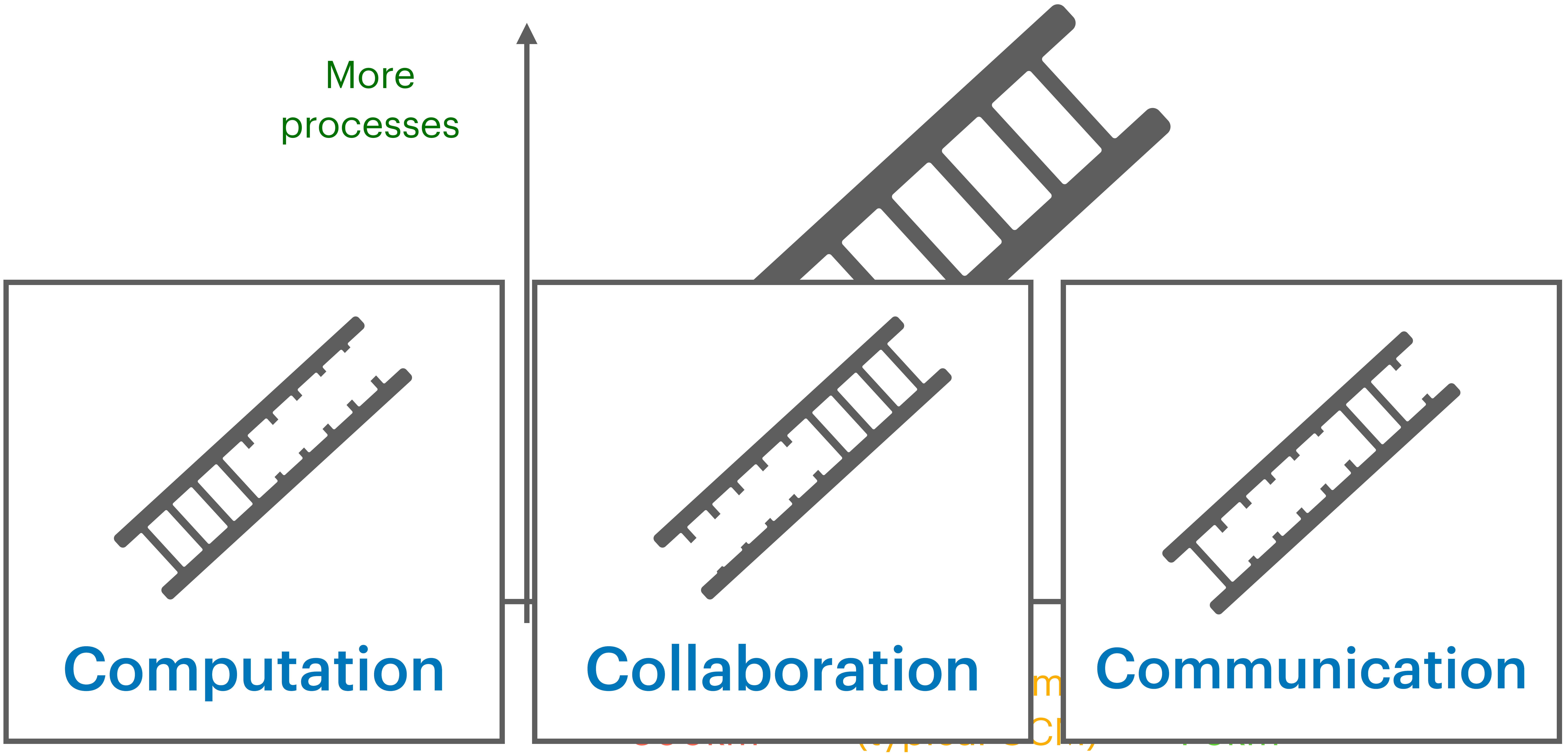
3. Challenge: conflation of concerns in code

- Extra technical documentation
- Clear systems design
- High modularity

But language support possible: more research needed

Is a future language tailored to science possible?

Better prediction: “climbing the ladder” (Charney)



Emily Shuckburgh

Cambridge Zero
+ CST

CAM
ZERO$_2$

Colm Caulfield

Department of Applied Maths and Theoretical Physics

Chris Edsall

University Information Services

Dominic Orchard

Department of Computer Science & Technology

Marla Fuchs

DAMTP

UNIVERSITY OF CAMBRIDGE
Institute of Computing for Climate Science

Emily Shuckburgh

Colm Caulfield

Chris Edsall

Dominic Orchard

Marla Fuchs

Artificial Intelligence

Mathematics

Software Engineering

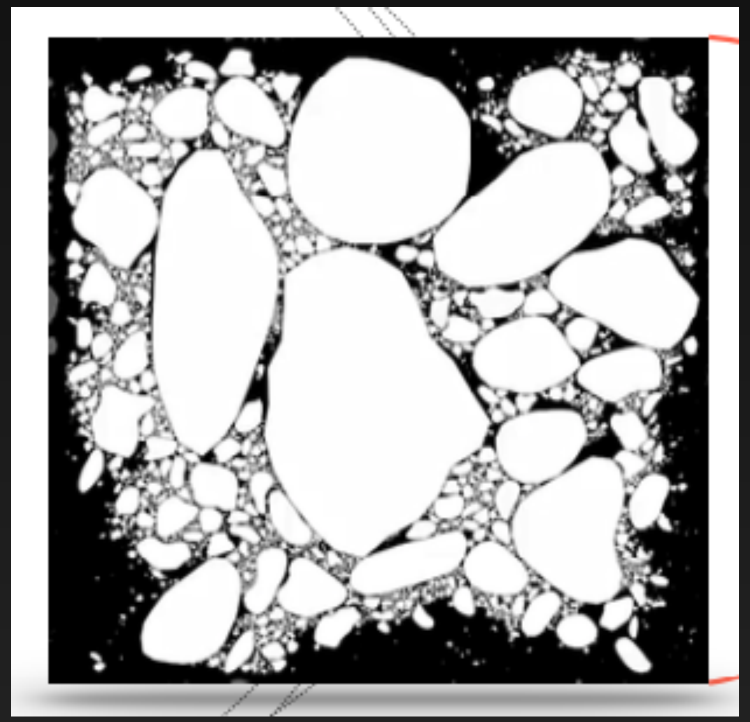
Data Science

Computer Science

Programming Langauges

https://www.schmidtfutures.com/our-work/virtual-earth-system-research-institute-vesri/

DataWave



SASIP

LEMONTREE

M^{2}LInES

Postdoctoral Advanced Fellowships

Laura Cimoli

2x more hiring

Research Software Engineers

Kacper Kornet

Simon Clifford

Ben Orchard

Matt Archer

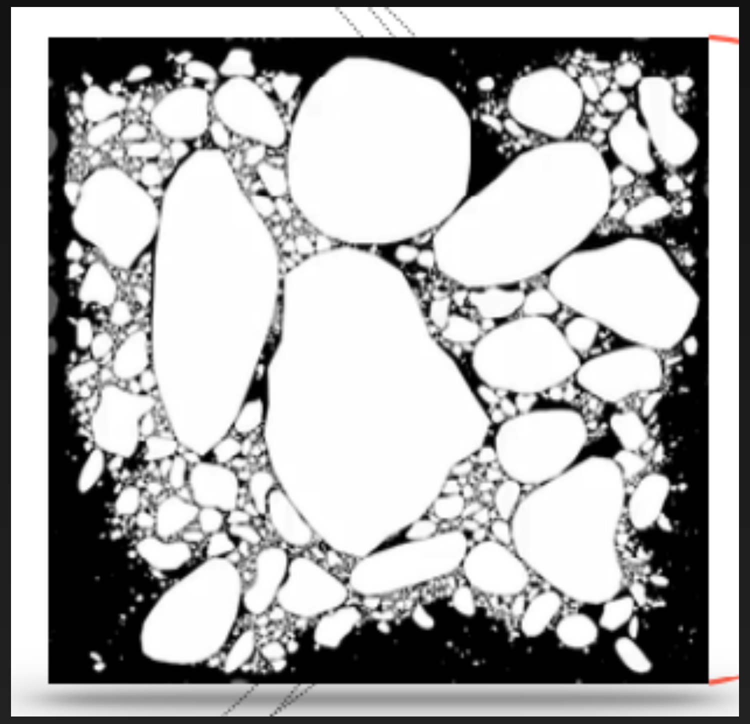
Jack Atkinson

Alexander Smith

Jim Denhom

More on the way…

DataWave



SASIP

LEMONTREE

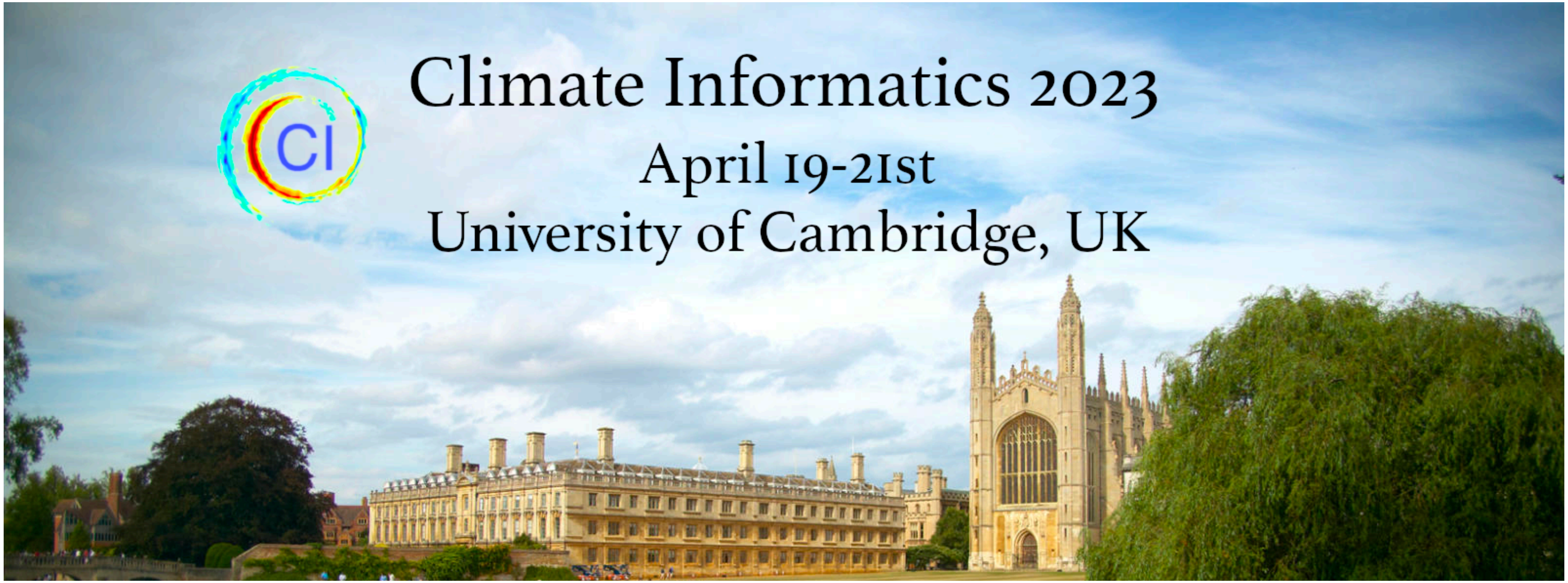
M^{2}LInES

UNIVERSITY OF CAMBRIDGE

Institute of Computing for Climate Science

@Cambridge_ICCS

https://cambridge-iccs.github.io


CI
Climate Informatics 2023
April 19-21st
University of Cambridge, UK