Earth System Models – From Equations to Exascale Supercomputers

Peter Dueben

Head of the Earth System Modelling Section

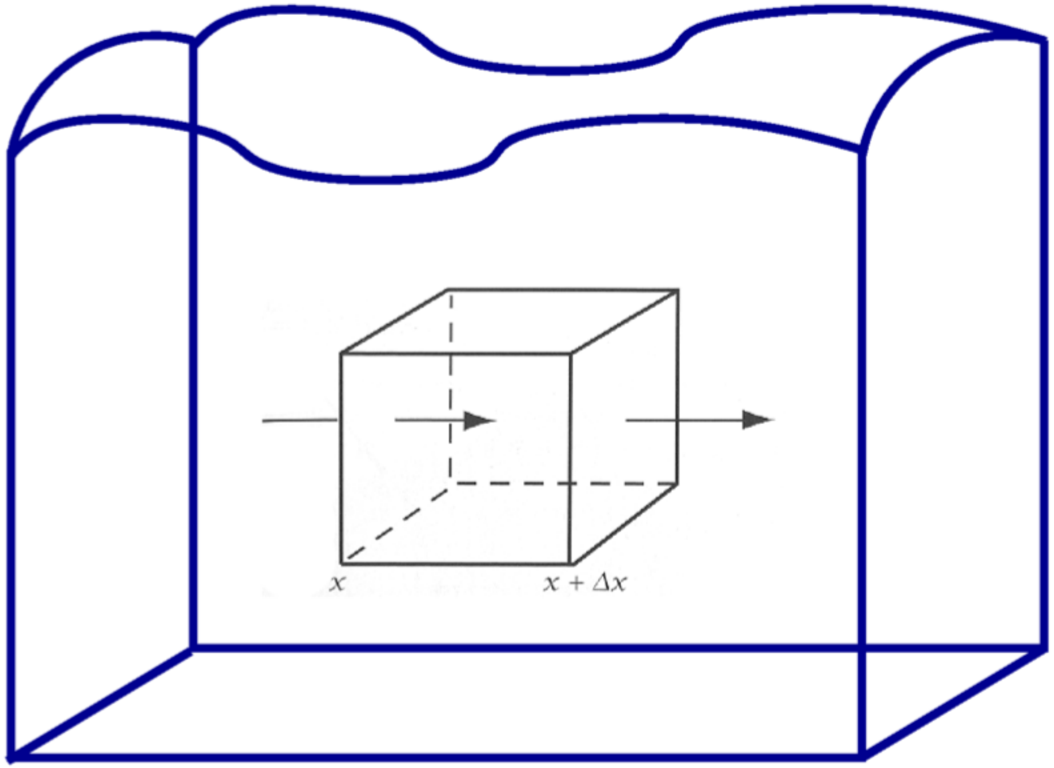
ECMWF | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

The strength of a common goal

Outline – From Equations to Exascale Supercomputers

1. **Deriving equations**
2. **From equations to models**
3. **From models to supercomputing**
4. **State of the art: How good are we? And what's next?**

How to derive the equations?



Let's consider a volume of a fluid with a specific density ρ(x,y,z,t) and velocity **u**(x,y,z,t)

Resume:
We obtain the continuity equation of mass by evaluating mass conservations

world → continuous math description

The total mass inside the volume is given by

$$M = \int_V \rho \, dV.$$

The change of mass in the volume is given by:

$$\frac{dM}{dt} = \frac{d}{dt}\int_V \rho \, dV = \int_V \frac{d\rho}{dt} \, dV. \tag{1}$$

We can also evaluate the change of mass by looking at fluxes through the boundaries:

$$\frac{dM}{dt} = -\int_S \rho \mathbf{v} \, d\mathbf{S} = -\int_V \nabla \cdot (\rho \mathbf{v}) \; dV. \tag{2}$$

(1) and (2) together form the mass continuity equation

$$\int_V \frac{d\rho}{dt} \, dV + \int_V \nabla \cdot (\rho \mathbf{v}) \; dV = 0.$$

If we shrink the volume to an infinitesimal small area ($\lim_{\Delta x \to 0}$, $\lim_{\Delta y \to 0}$, $\lim_{\Delta z \to 0}$) we end up with the differential form of the continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0.$$

We know the equations, so what's the problem?

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{\nabla p}{\rho} + \nu \nabla^2 \mathbf{v} + \frac{\mathbf{F}}{\rho} - 2\Omega \times \mathbf{v}$$

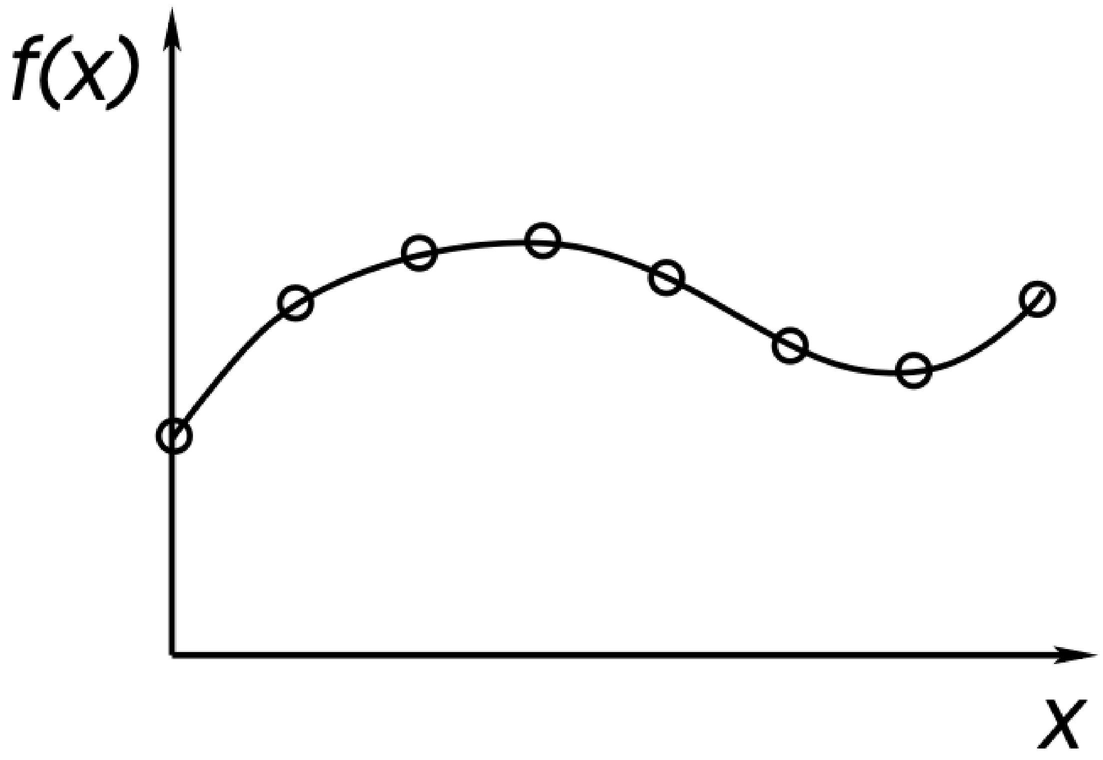
$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0$$

The equations are non-linear and we cannot solve them…

How do we still make weather predictions?

world → continuous math description → discretised equations

Finite difference method



$$\frac{\partial f(x_0, t)}{\partial x} \approx \frac{f(x_0 + \Delta x, t) - f(x_0 - \Delta x, t)}{2\Delta x},$$

$$\frac{\partial f(x_0, t)}{\partial x} \approx \frac{f(x_0 + \Delta x, t) - f(x_0, t)}{\Delta x},$$

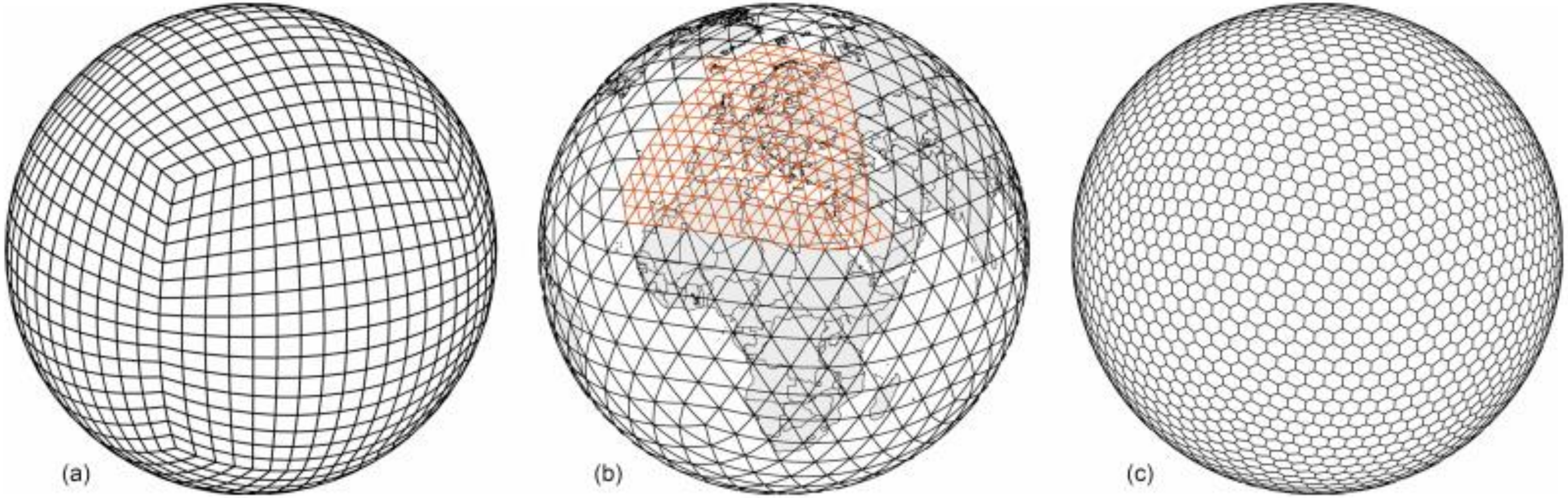
$$\frac{\partial f(x, t_0)}{\partial t} \approx \frac{f(x, t_0) - f(x, t_0 - \Delta t)}{\Delta t}.$$

We discretise our function f(x) at specific grid points f(0), f(Δx), f(2Δx)…

Derivatives are described by differential quotients
→ There are plenty of different discretisation schemes

We need to discretise in both space and time

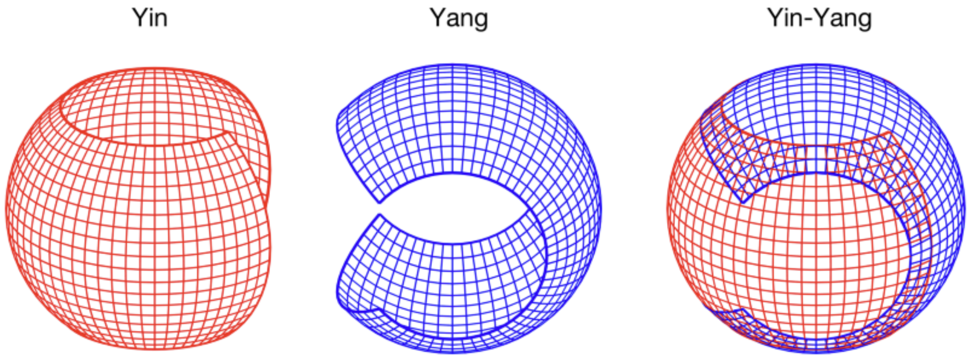
Popular grids



Cubed sphere

Icosahedral (triangular)

Icosahedral (hexagonal)



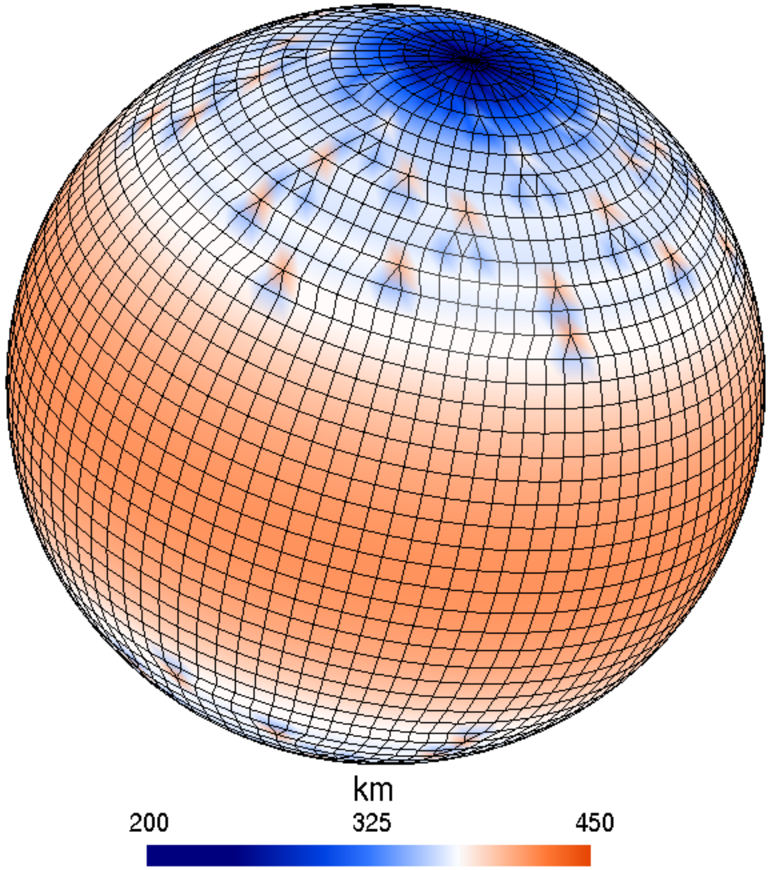
See also Annual Seminar 2020, ECMWF
https://www.ecmwf.int/en/learning/workshops/annual-seminar-2020

Slide from Nils Wedi

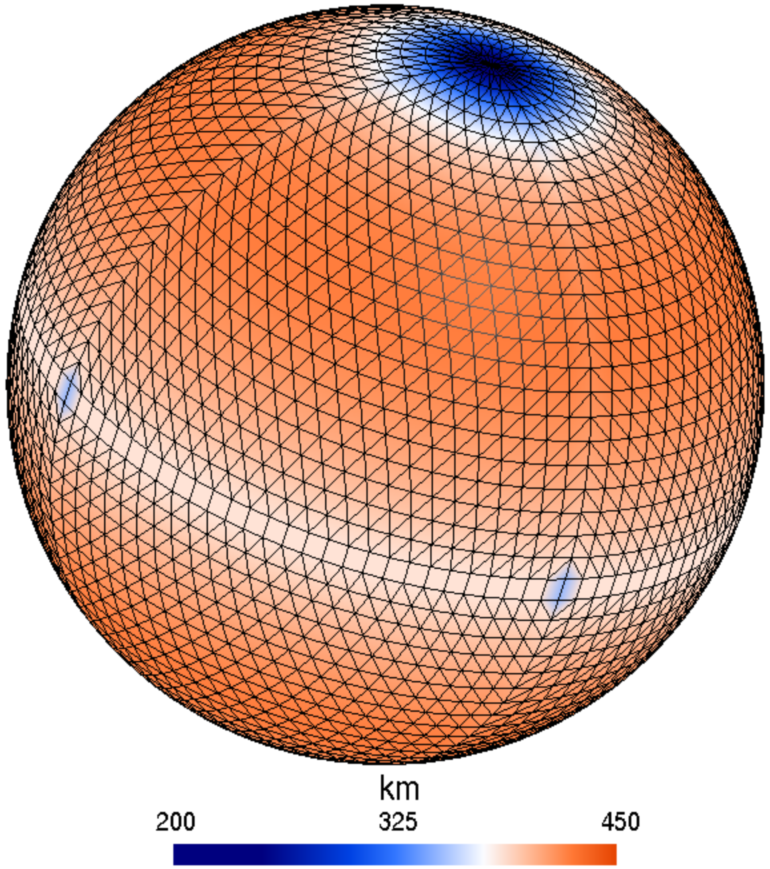
A cubic octahedral grid

What is a uniform grid ?

A further ~20% reduction in gridpoints
=> ~50% less points compared to full grid



N24 reduced Gaussian grid



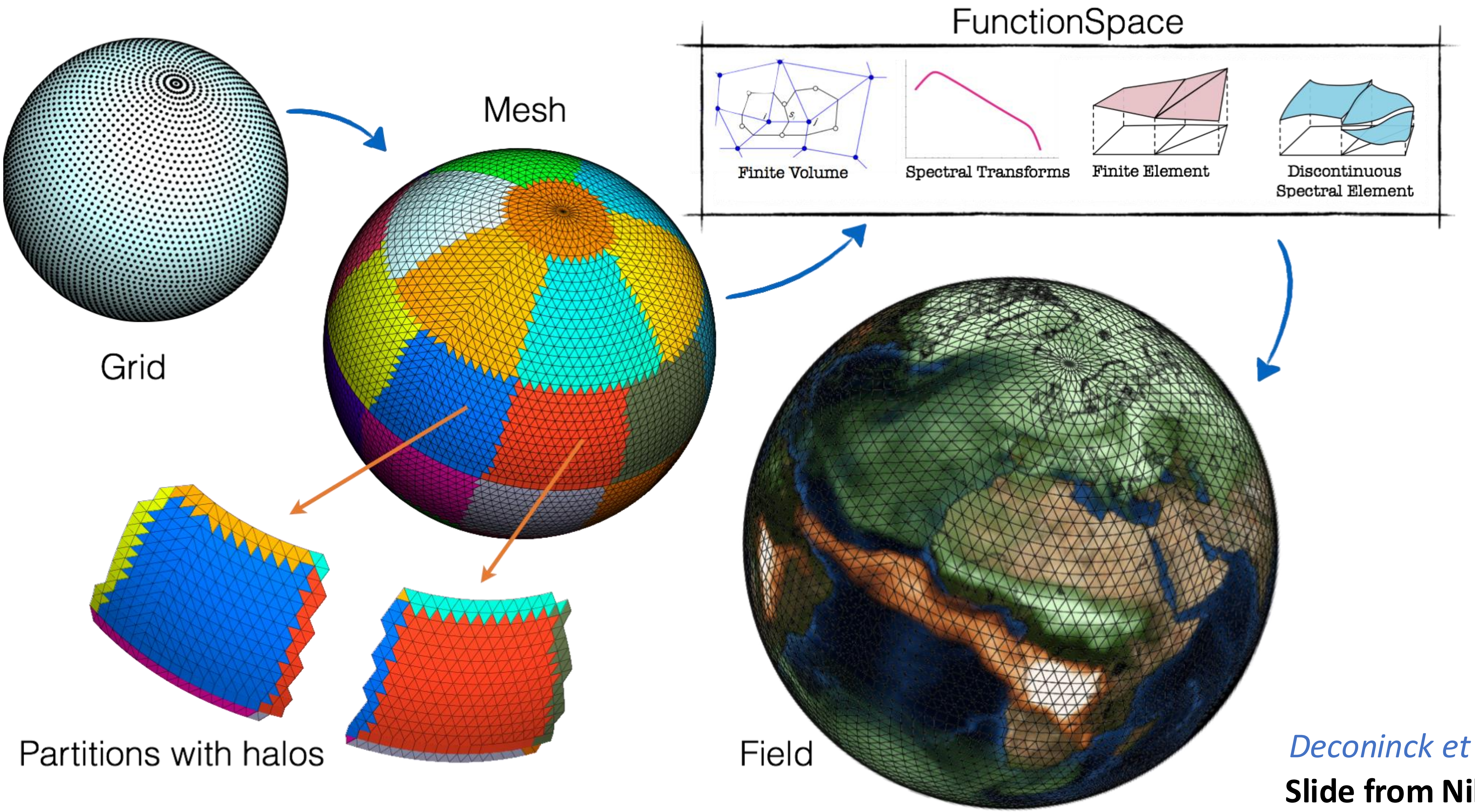
N24 octahedral Gaussian grid

(Wedi et al, 2014, 2015)

Slide from Nils Wedi

Atlas: a library for NWP and climate modelling

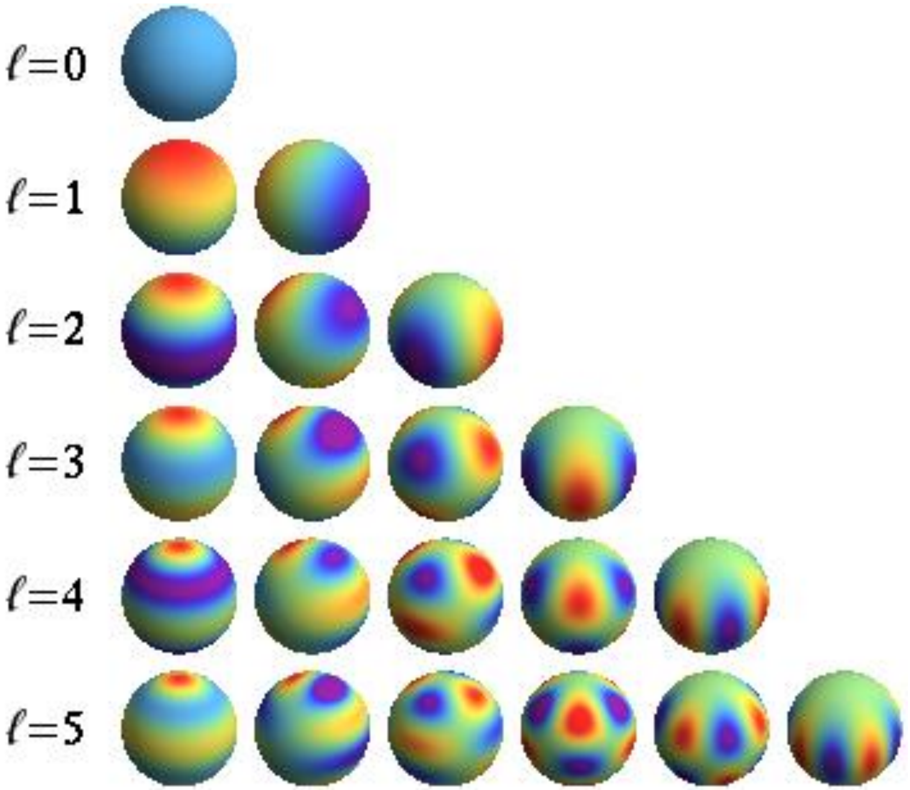
Funded by the European Union

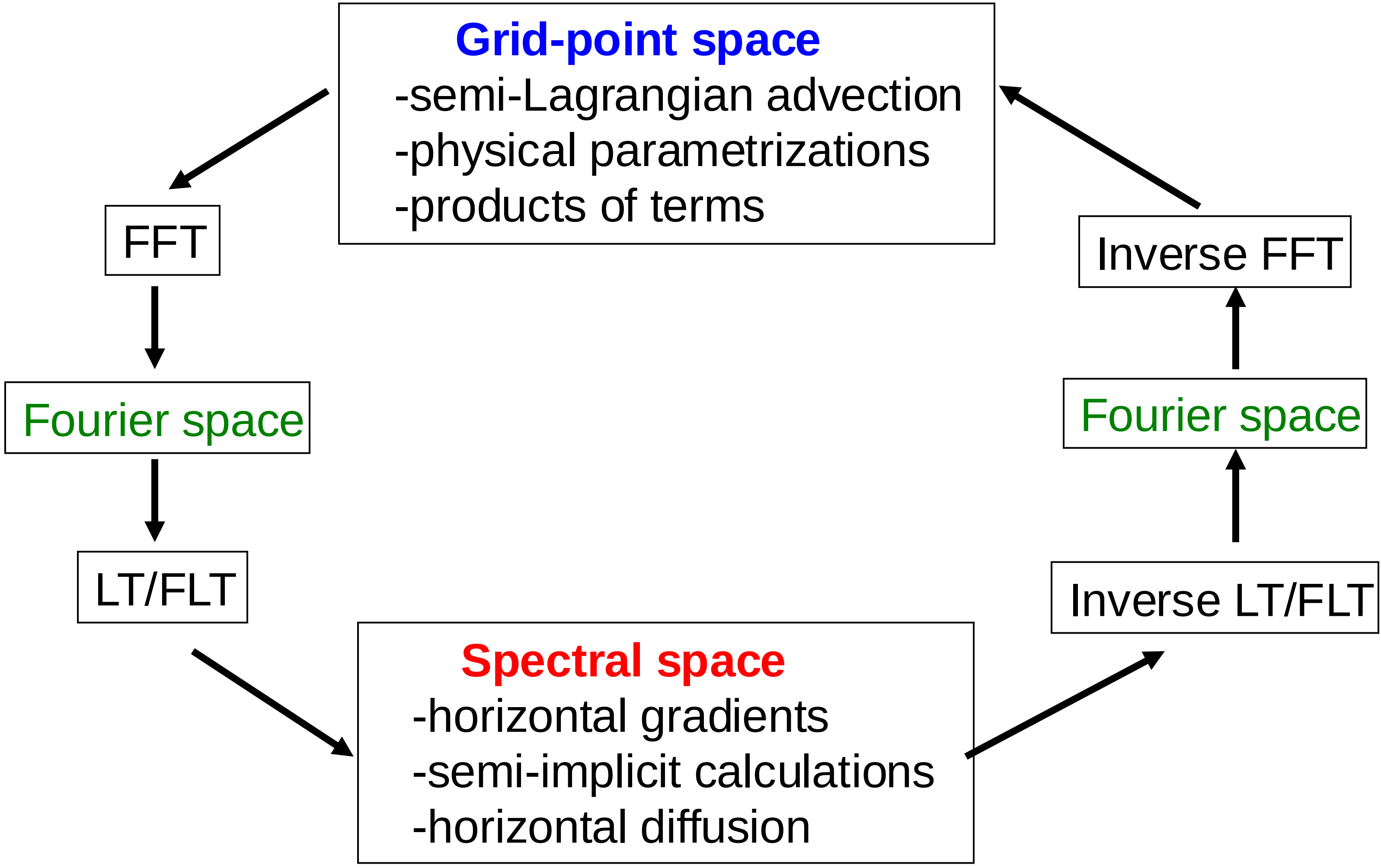


Deconinck et al. 2017

Slide from Nils Wedi

Spectral discretisation in the Integrated Forecast System (IFS)





FFT: Fast Fourier Transform, LT/FLT: Legendre Transform

The equations of motion can also be evaluated for spherical harmonics.

There are plenty of options to discretise… and they are used

Short name	Equation set	Prognostic variables	Horizontal grid	Numerical method	Horizontal staggering
ACME-A	H/NH	$\boldsymbol{u}_\mathrm{h}$, w, ρ_s, $\rho_\mathrm{s}\theta$, Φ, $\rho_\mathrm{s}q_i$	Cubed sphere (Sect. 3.2)	SE	A grid
CSU	NH (unified)	ζ, D, w, p_s, θ_v, q_i	Geodesic (Sect. 3.4)	FV	Z grid
DYNAMICO	H/NH	$\boldsymbol{v}_\mathrm{h}$, $\rho_\mathrm{s}w$, ρ_s, $\rho_\mathrm{s}\theta_\mathrm{v}$, Φ, $\rho_\mathrm{s}q_i$	Geodesic (Sect. 3.4)	FV	C grid
FV^3	NH	$\boldsymbol{u}_\mathrm{h}$, w, ρ_s, $\rho_\mathrm{s}\theta_\mathrm{v}$, Φ, $\rho_\mathrm{s}q_i$	Cubed sphere (Sect. 3.2)	FV	D grid
FVM	NH (D)	ρ_d, $\boldsymbol{u}_\mathrm{h}$, w, θ', q_i	Octahedral (Sect. 3.6)	FV	A grid
GEM	NH	$\boldsymbol{u}_\mathrm{h}$, w, $\dot{\zeta}$, T_v, p, q_i	Yin–Yang (Sect. 3.7)	FD	C grid
ICON	NH (D)	$\boldsymbol{u}_\mathrm{h}$, w, ρ, θ_v, ρq_i	Icosahedral triangular (Sect. 3.3)	FV	C grid
MPAS	NH	$\rho_\mathrm{d}\boldsymbol{u}_\mathrm{h}$, $\rho_\mathrm{d}w$, ρ_d, $\rho_\mathrm{d}\theta_\mathrm{v}$, $\rho_\mathrm{d}q_i$	CCVT (Sect. 3.5)	FV	C grid
NICAM	NH	$\rho\boldsymbol{u}_\mathrm{h}$, ρw, ρ, ρe, ρq_i	Geodesic (Sect. 3.4)	FV	A grid
OLAM	NH (D)	$\rho\boldsymbol{u}_\mathrm{h}$, ρw, ρ, $\rho\theta_\mathrm{il}$, ρq_i	Geodesic (Sect. 3.4)	FV	C grid
Tempest	NH	$\boldsymbol{u}_\mathrm{h}$, w, ρ, $\rho\theta_\mathrm{v}$, ρq_i	Cubed sphere (Sect. 3.2)	SE	A grid

DCMIP2016: a review of non-hydrostatic dynamical core design and intercomparison of participating models, Ullrich et al 2016

IFS dynamical core options at ECMWF

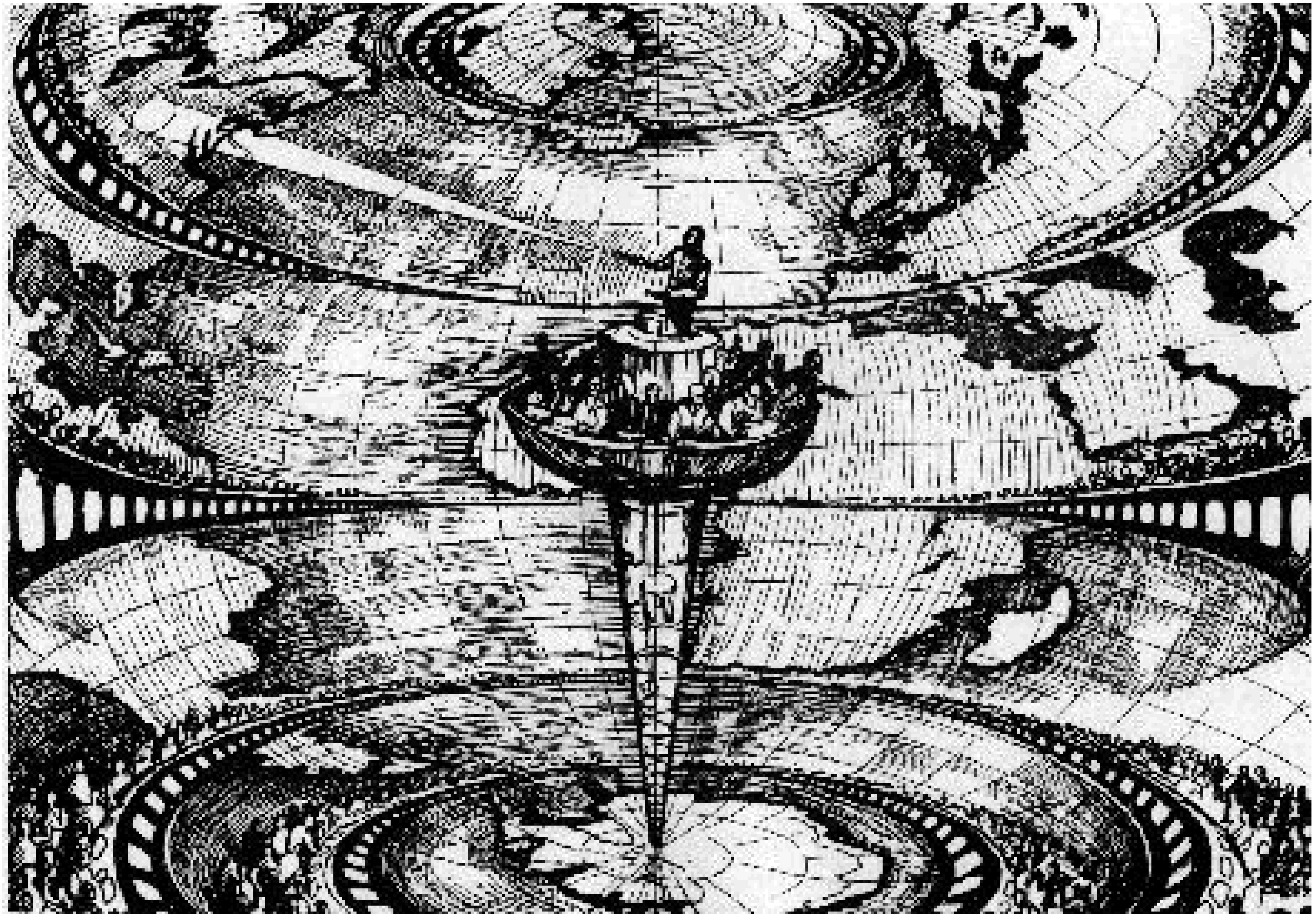
Christian Kuehnlein

| currently operational |

Model aspect	IFS-FVM	IFS-ST	IFS-ST (NH option)
Equation system	fully compressible	hydrostatic primitive	fully compressible
Prognostic variables	ρ_d, u, v, w, θ', φ', r_v, r_l, r_r, r_i, r_s	$\ln p_s$, u, v, T_v, q_v, q_l, q_r, q_i, q_s	$\ln \pi_s$, u, v, d_4, T_v, $\hat{q}$, q_v, q_l, q_r, q_i, q_s
Horizontal coordinates	λ, ϕ (lon–lat)	λ, ϕ (lon–lat)	λ, ϕ (lon–lat)
Vertical coordinate	generalized height	hybrid sigma–pressure	hybrid sigma–pressure
Horizontal discretization	unstructured finite volume (FV)	spectral transform (ST)	spectral transform (ST)
Vertical discretization	structured FD–FV	structured FE	structured FD or FE
Horizontal staggering	co-located	co-located	co-located
Vertical staggering	co-located	co-located	co-located, Lorenz
Horizontal grid	octahedral Gaussian or arbitrary	octahedral Gaussian	octahedral Gaussian
Time stepping scheme	2-TL SI	2-TL constant-coefficient SI	2-TL constant-coefficient SI with ICI
Advection	conservative FV Eulerian	non-conservative SL	non-conservative SL

Slide from Nils Wedi

Richardson's forecast factory, 1922



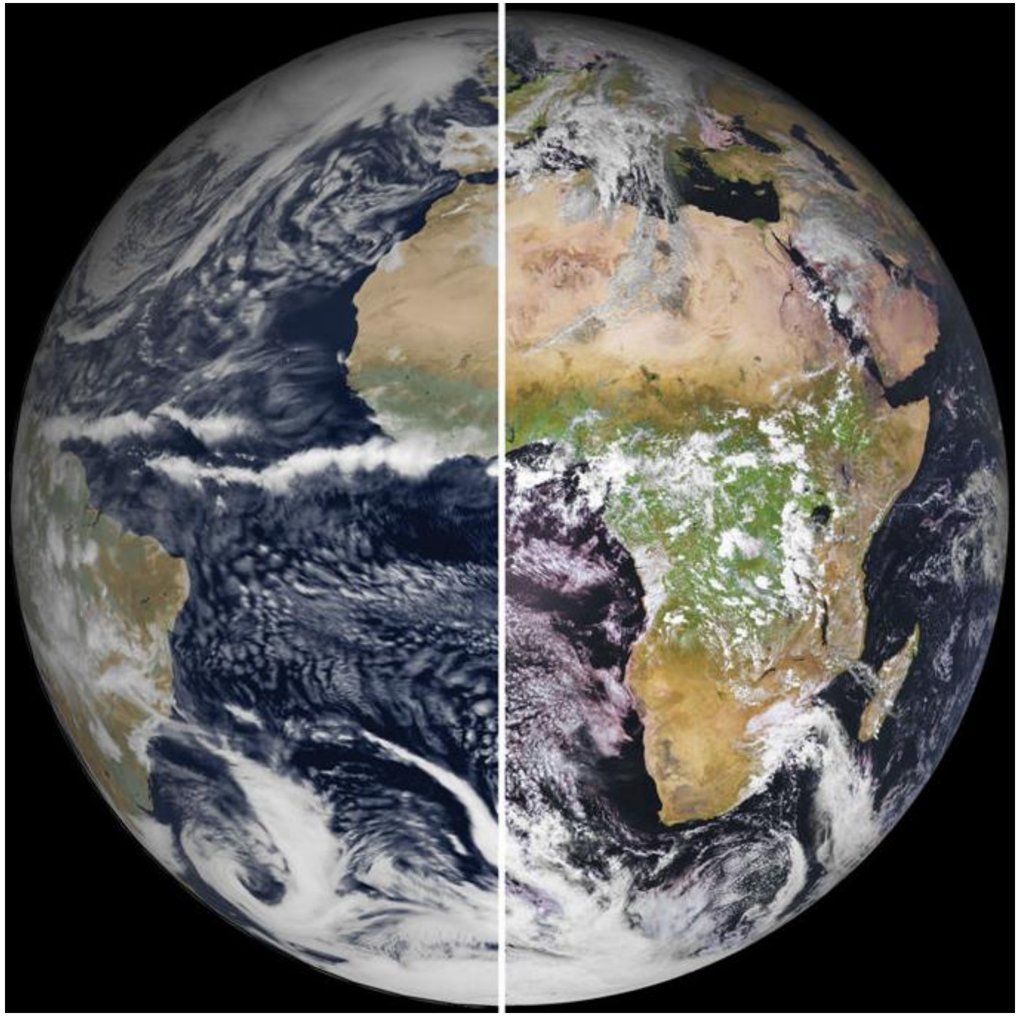
Sketch by A. Lannerback (© Dagens Nyheter, Stockholm)
Found at http://mathsci.ucd.ie

So let's just discretise the equations and all problems are solved…?

Slide from Nils Wedi

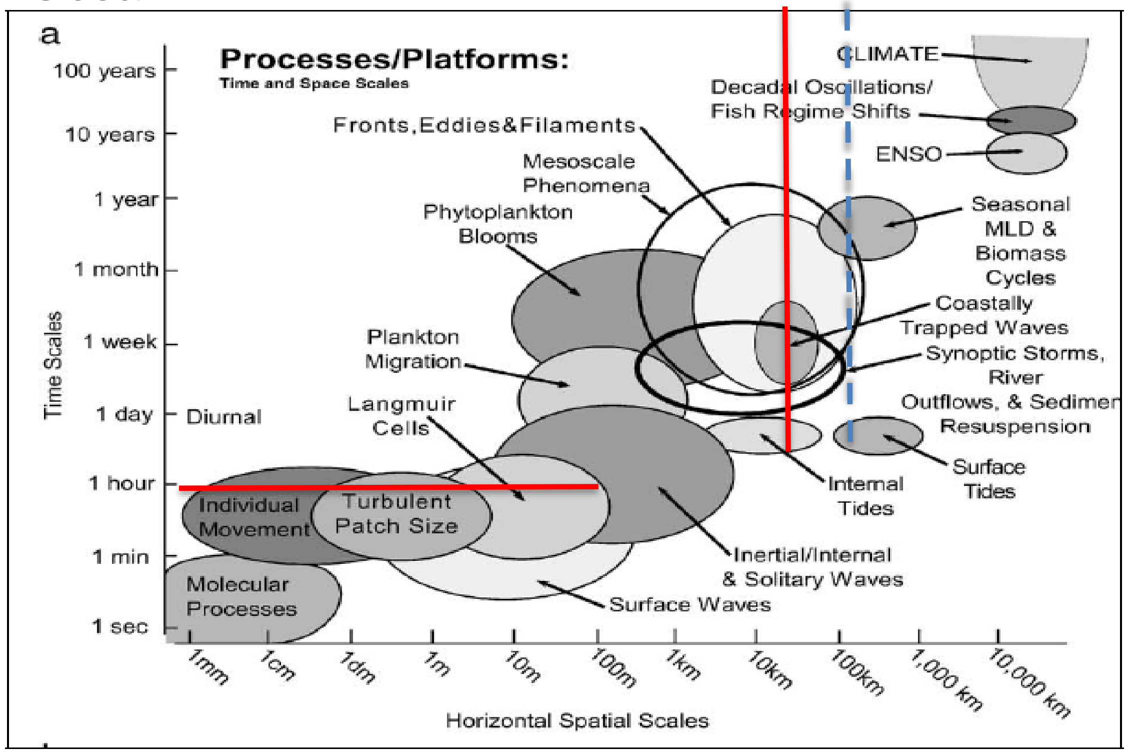
Why is it difficult to predict the weather?

- The Earth is huge, resolution is limited and we cannot represent all important processes within model simulations
- We do not know the exact initial conditions
- The Earth System shows “chaotic” dynamics which makes it difficult to predict the future based on equations
- All Earth System components (atmosphere, ocean, land surface, cloud physics,…) are connected in a non-trivial way
- Some of the processes involved are not well understood



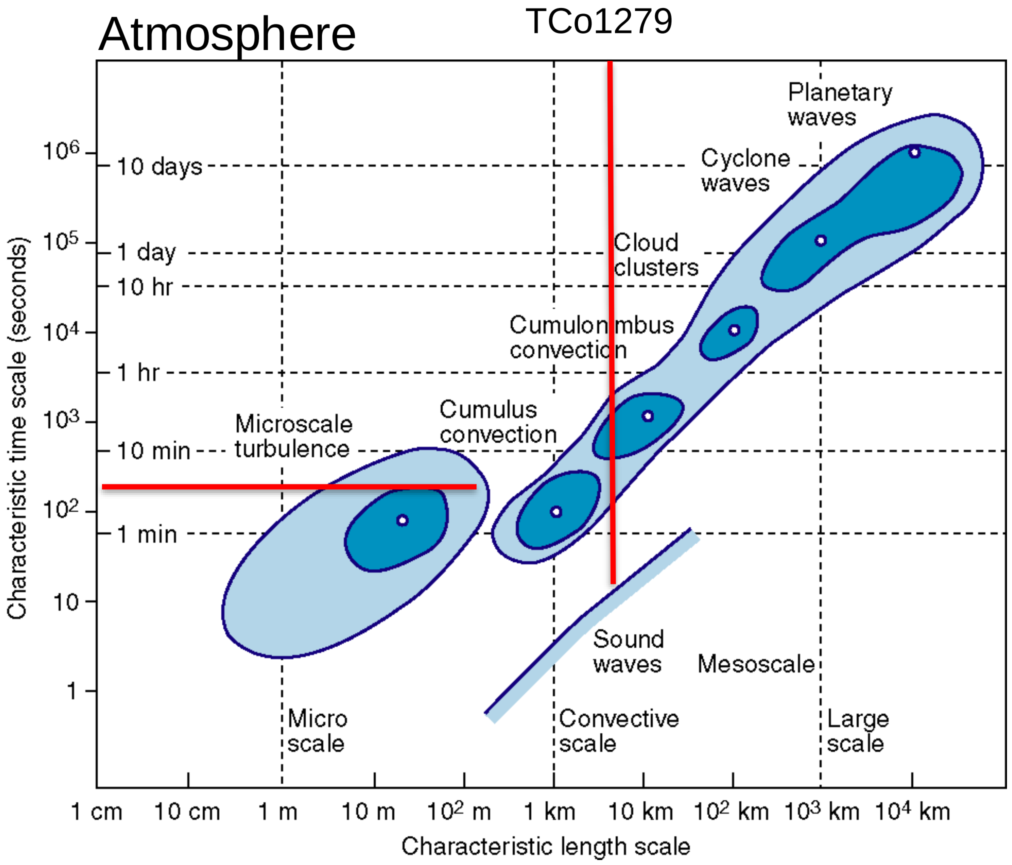
The Earth system as a multi-scale problem

Ocean



From Dickey (2003)

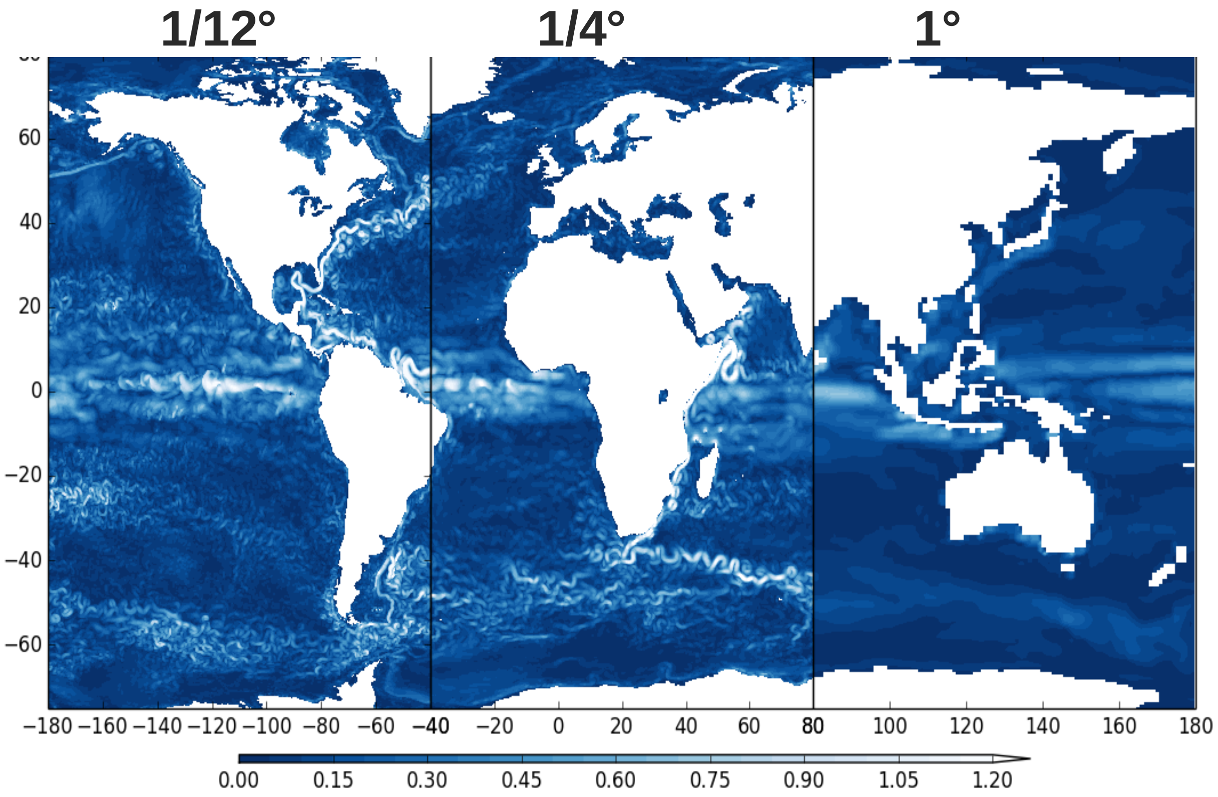
Atmosphere



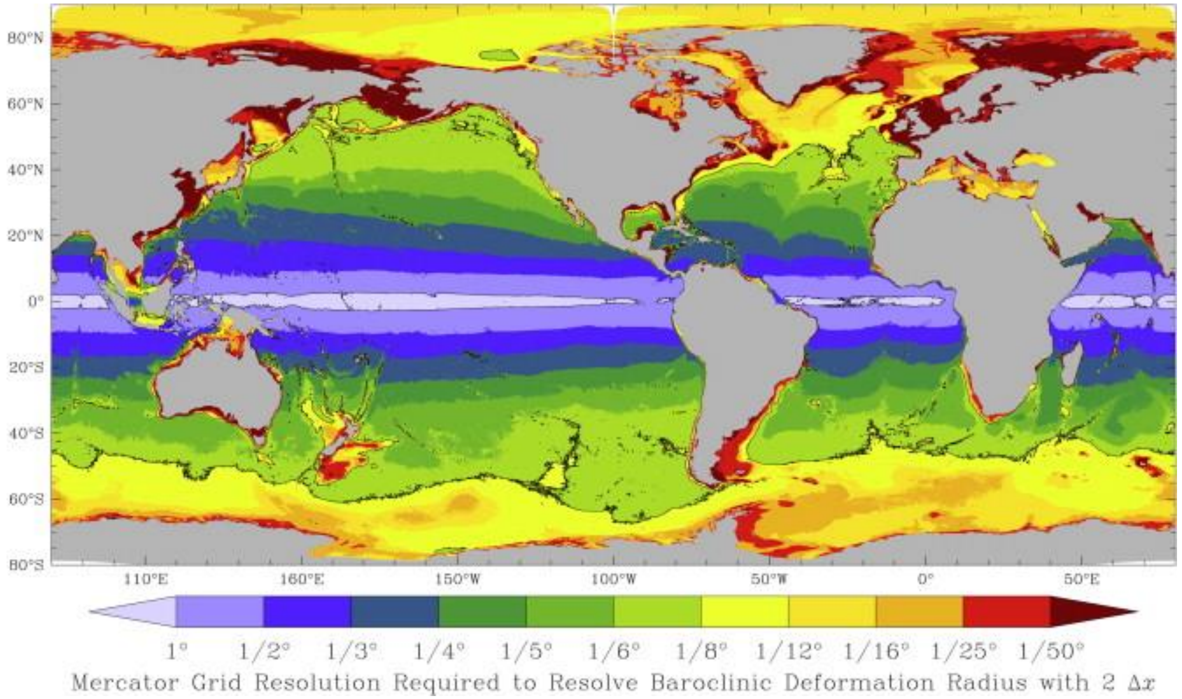
Range of fast and slow waves …

Slide from Nils Wedi

Ocean model - resolution



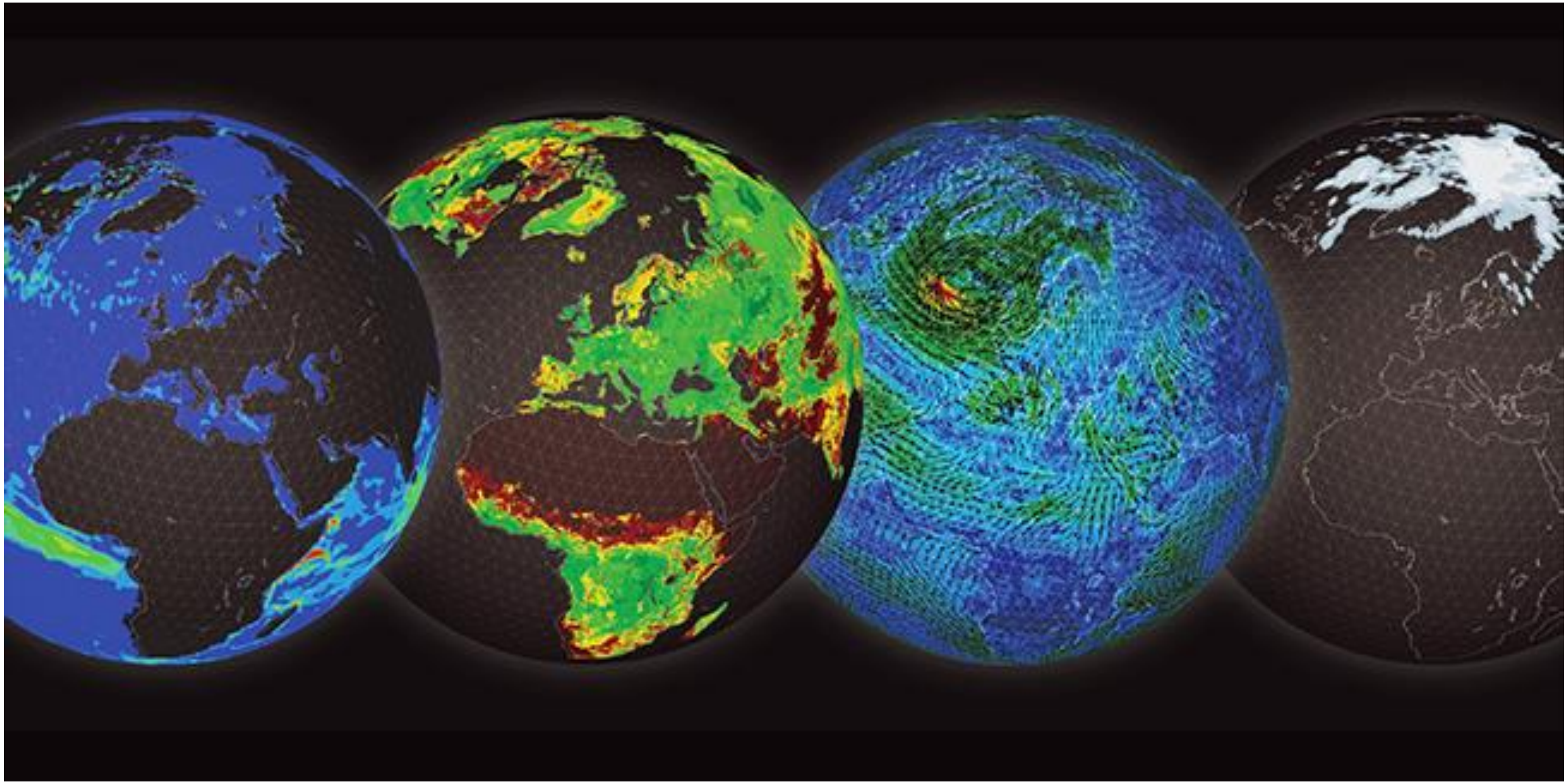
Hewitt et al. (2017)



Hallberg (2013)

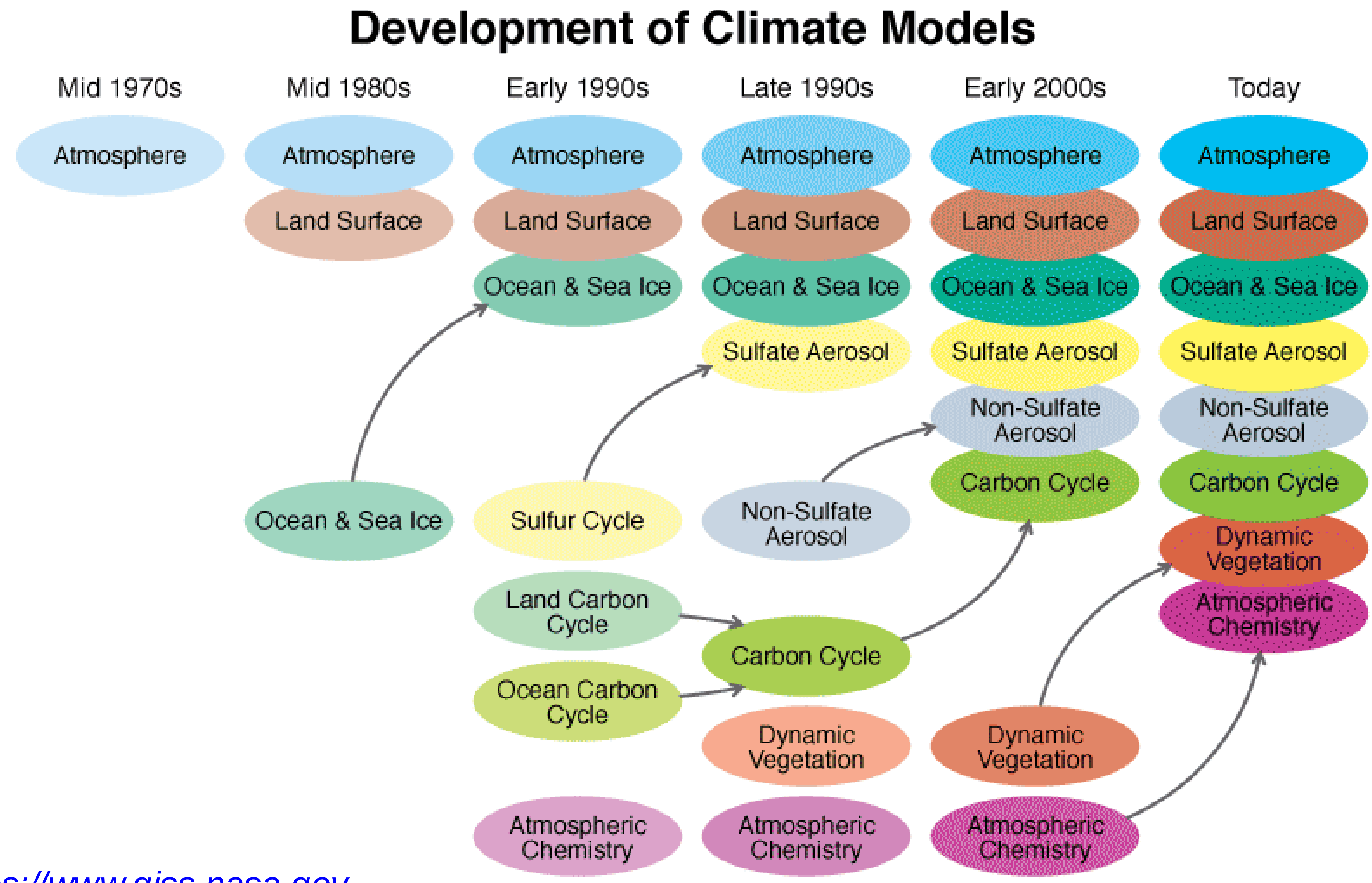
Slide from Nils Wedi

Ocean – Land – Atmosphere – Sea ice



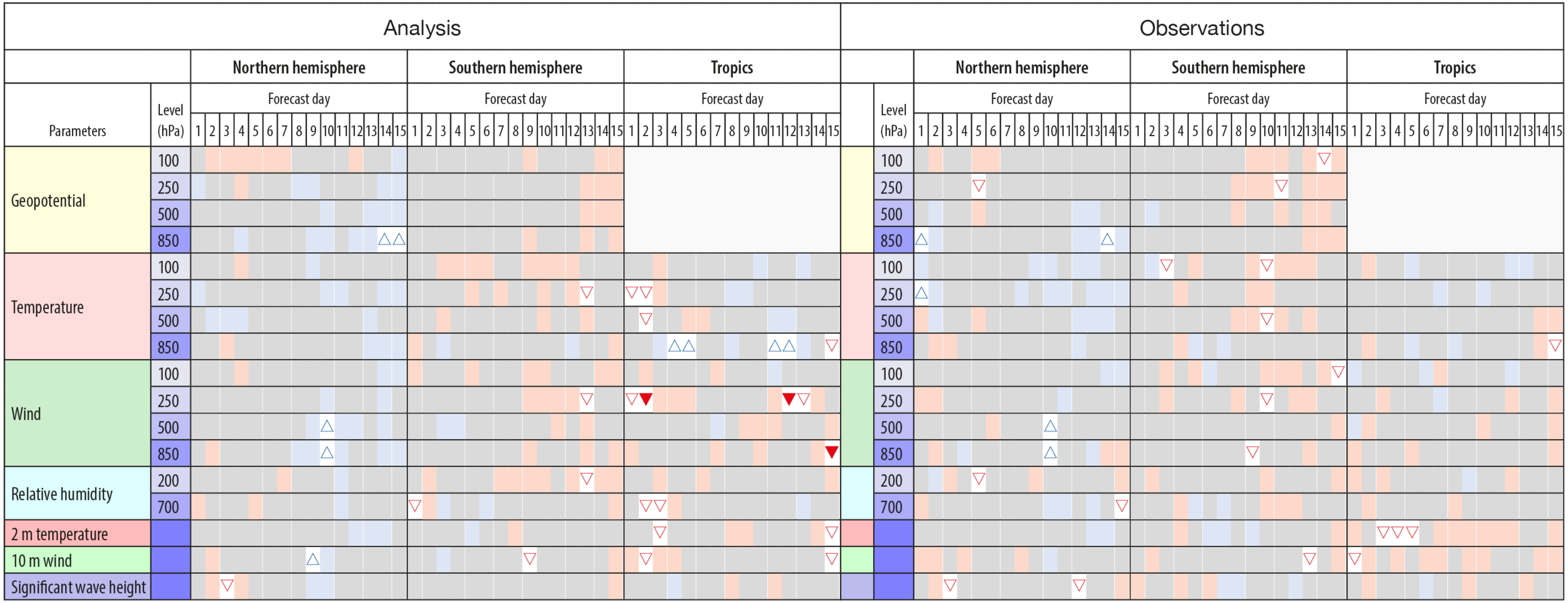
Slide from Nils Wedi

Earth System model complexity



Source: https://www.giss.nasa.gov

The Earth system as a coupled system



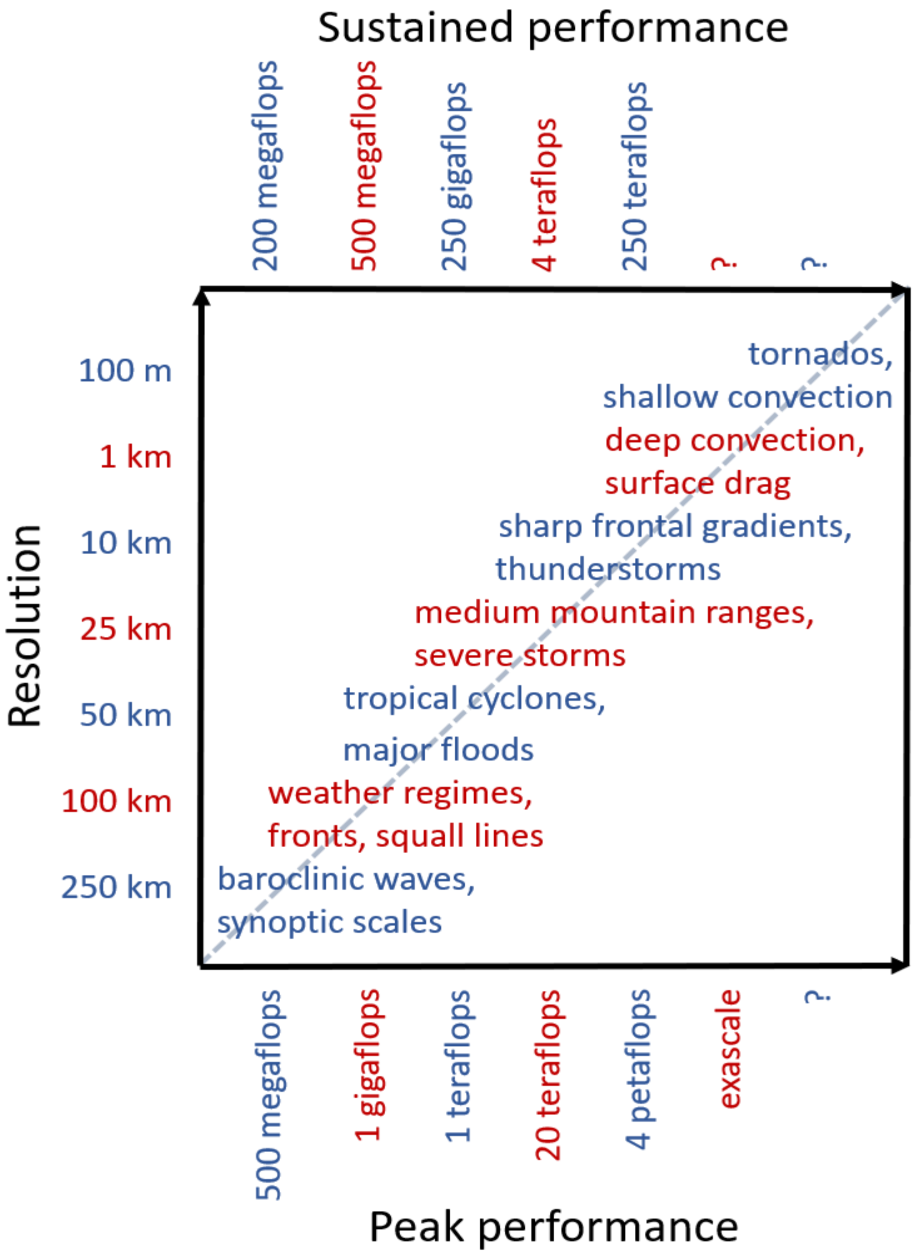
Symbol legend: for a given forecast step...

- ▲ SP better than DP statistically significant with 99.7% confidence
- △ SP better than DP statistically significant with 95% confidence
- SP better than DP statistically significant with 68% confidence
- no significant difference between DP and SP
- SP worse than DP statistically significant with 68% confidence
- ▽ SP worse than DP statistically significant with 95% confidence
- ▼ SP worse than DP statistically significant with 99.7% confidence

Dueben et al. ECMWF Newsletter 2018

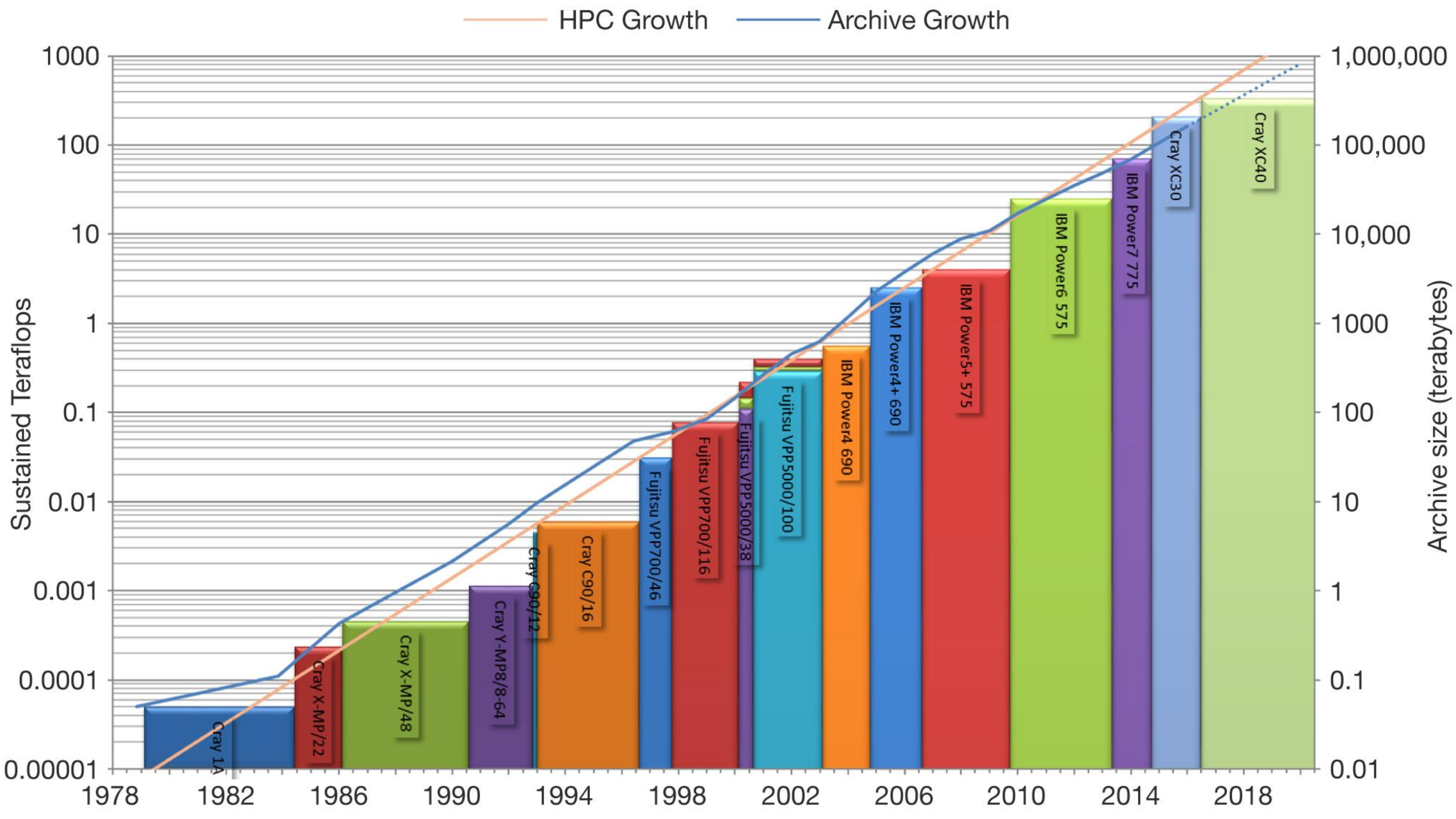
Beyond the grid…

- Not all processes can be discretised on a given grid
- Sub-grid-scale processes need to be parametrised including very important processes of the Earth system such as clouds, boundary layer turbulence, gravity wave drag, ocean eddies, land/snow/ice processes…



Adjusted from Neumann et al. Phil. Trans. A 2019

HPC and HPDA for weather and climate modelling



Bauer et al. ECMWF SAC paper 2019

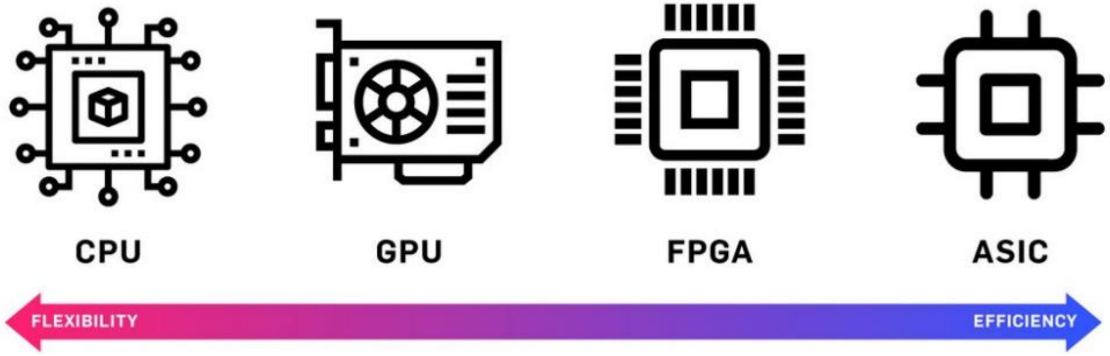
Current challenges in high performance computing?

- Individual processors will not be faster
  → Parallelisation / power consumption / hardware faults

- Hardware is heterogeneous
  → CPUs / GPUs / FPGAs / ASICs

- Machine learning has strong impact on hardware development
  → High floprate at low precision

- I/O is becoming a nightmare and the optimisation of data movement will be the key



Source: venturebeat.com

Energy-aware computing

- All 51 ENS members consume about 300KWh, approximately the same as a single (~5km) global 10-day forecast
- The energy consumption of *one ENS member* is equivalent to leaving the Kettle on for 2 hours !

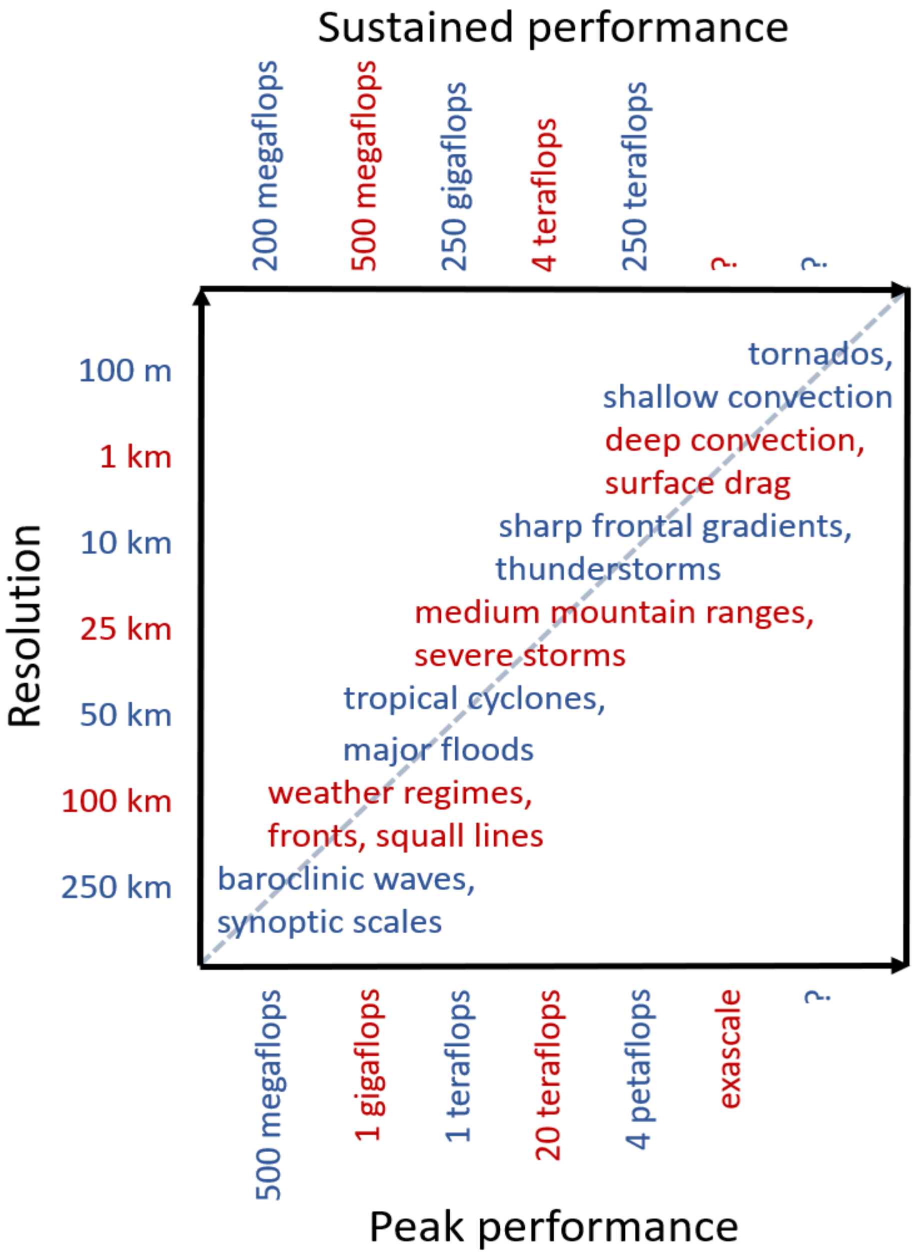
http://ukbusinessblog.co.uk

Time-to-Solution vs.
Energy-to-Solution

Slide from Nils Wedi

Beyond the grid…

- Not all processes can be discretised on a given grid
- Sub-grid-scale processes need to be parametrised including very important processes of the Earth system such as clouds, boundary layer turbulence, gravity wave drag, ocean eddies, land/snow/ice processes…



Adjusted from Neumann et al. Phil. Trans. A 2019

But progress in km-scale modelling is tough…

Compute power?
9 km → 1 km → Factor 9^3 = 729 compute power

Waiting for Moore's law.
→ 2^9 = 512 → Let's wait for 18 years?

Data and storage?
9km: 6,599,680 points x 137 levels x 10 variables
→ 9 billion points → > 0.5 TB

1.5km: 256,800,000 points x 137 levels x 10 variables
→ 352 billion points → > 20 TB

Uff…

ECMWF

TOP500 LIST - JUNE 2023

R_{max} and R_{peak} values are in PFlop/s. For more details about other fields, check the TOP500 description.

R_{peak} values are calculated using the advertised clock rate of the CPU. For the efficiency of the systems you should take into account the Turbo CPU clock rate where it applies.

← 1-100 | 101-200 | 201-300 | 301-400 | 401-500 →

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,699,904	1,194.00	1,679.82	22,703
2	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
3	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,220,288	309.10	428.70	6,016
4	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, Atos EuroHPC/CINECA Italy	1,824,768	238.70	304.47	7,404
5	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148.60	200.79	10,096

Destination Earth to the rescue…

comment

A digital twin of Earth for the green transition

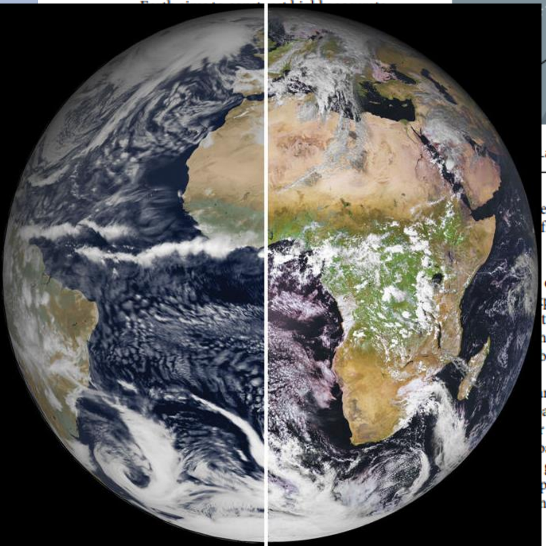
For its green transition, the EU plans to fund the development of digital twins of Earth. For these twins to be more than big data atlases, they must create a qualitatively new Earth system simulation and observation capability using a methodological framework responsible for exceptional advances in numerical weather prediction.

Peter Bauer, Bjorn Stevens and Wilco Hazeleger

The European Union (EU) intends to become climate neutral by 2050, and the set of policies designed to bring about this green transition — the European Green Deal — was announced in December 2019 (ref. [1]). Accompanied by €1 trillion of planned investment, Green Deal policies aim to help the world's second-largest economy sustainably produce energy, develop carbon-neutral fuels and advance circular products in energy-intensive industrial sectors with zero waste and zero pollution.

A key element of the Green Deal is its dependence on the 'digital transformation' — an openly accessible and interoperable European dataspace as a central hub for informed decision making. The EU identified two landmark actions to support the necessary information systems: GreenData4All[2] and Destination Earth[3]. Whereas GreenData4All will develop the European approach to discover, manage and exploit geospatial information, Destination

Layerace / Freepik



e coordinated development
fic disciplines.

of Earth is an information
poses users to a digital
the state and temporal
he Earth system constrained
servations and the laws of

re familiar with a plethora of
ased monitoring tools that
impact on the environment,
ased simulation models
grasp the causes of change
ptions for future adaptation
actions. The ongoing step

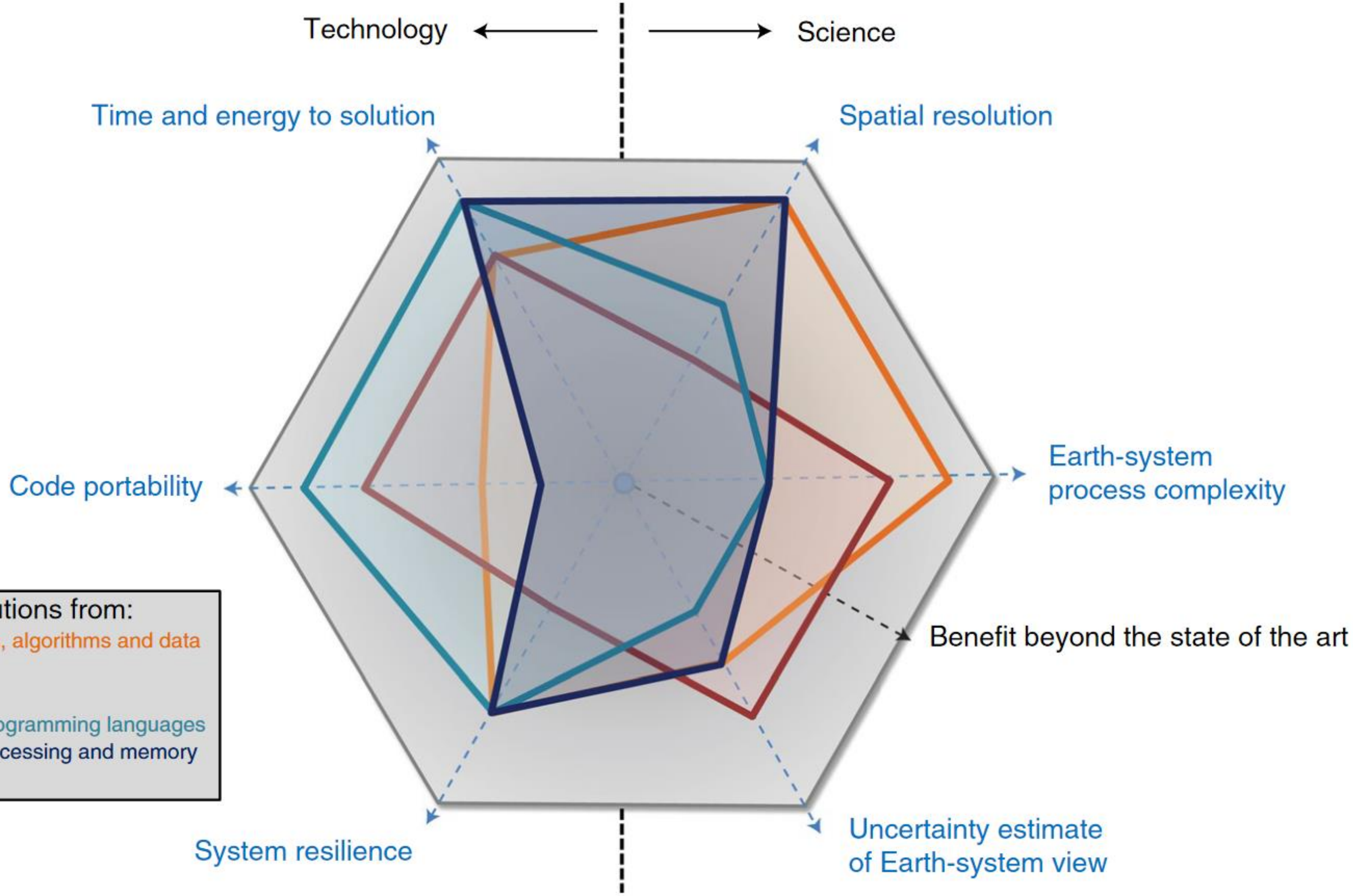
PERSPECTIVE
https://doi.org/10.1038/s43588-021-00023-0
nature computational science

Check for updates

The digital revolution of Earth-system science

Peter Bauer[1], Peter D. Dueben[1], Torsten Hoefler[2], Tiago Quintino[3], Thomas C. Schulthess[4] and Nils P. Wedi[1]

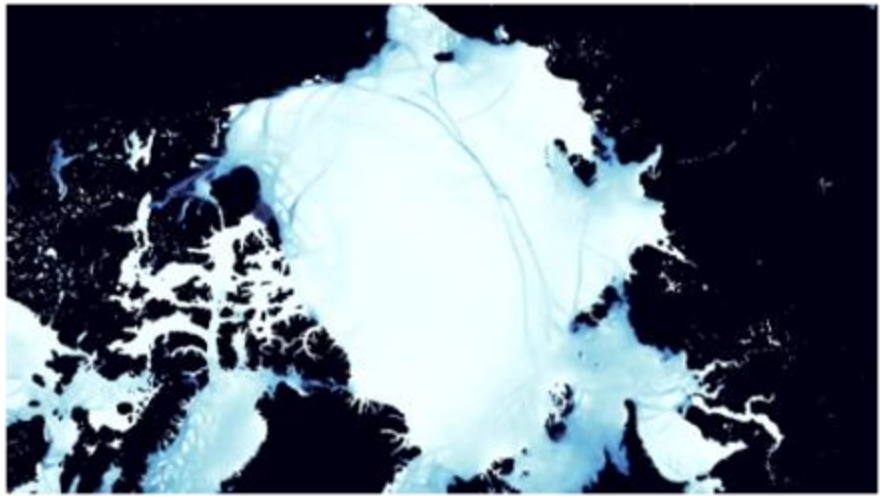
Computational science is crucial for delivering reliable weather and climate predictions. However, despite decades of high-performance computing experience, there is serious concern about the sustainability of this application in the post-Moore/Dennard era. Here, we discuss the present limitations in the field and propose the design of a novel infrastructure that is scalable and more adaptable to future, yet unknown computing architectures.



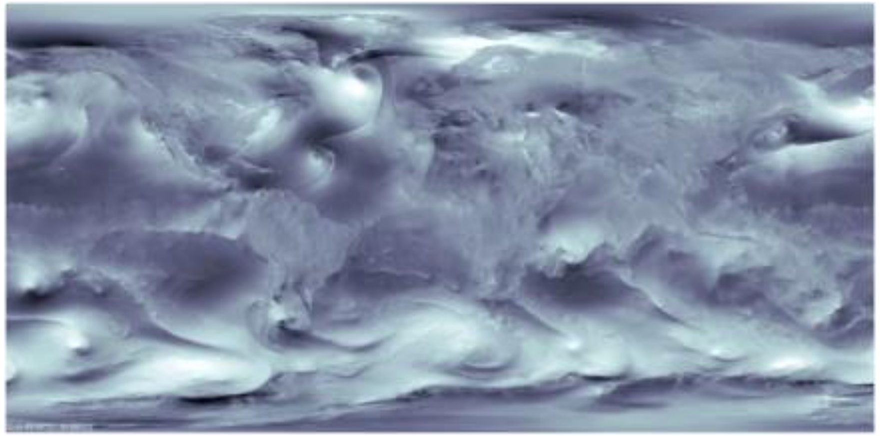
Individual contributions from:

- Numerical methods, algorithms and data structures
- Machine learning
- Domain-specific programming languages
- Heterogeneous processing and memory architectures

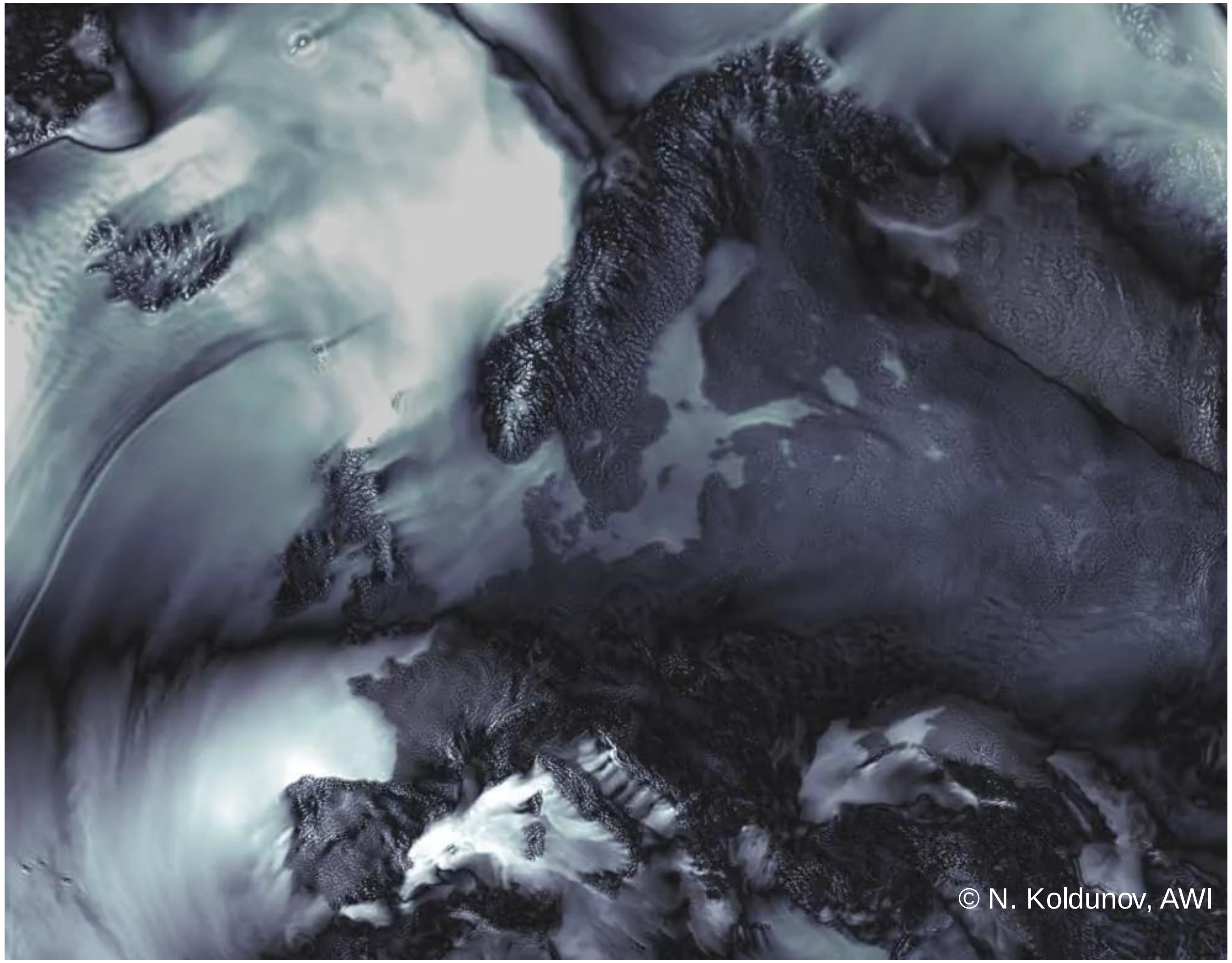
The digital revolution to allow for km-scale models



More realistic at local scale



More realistic at global scale



Better results via a coupled model system

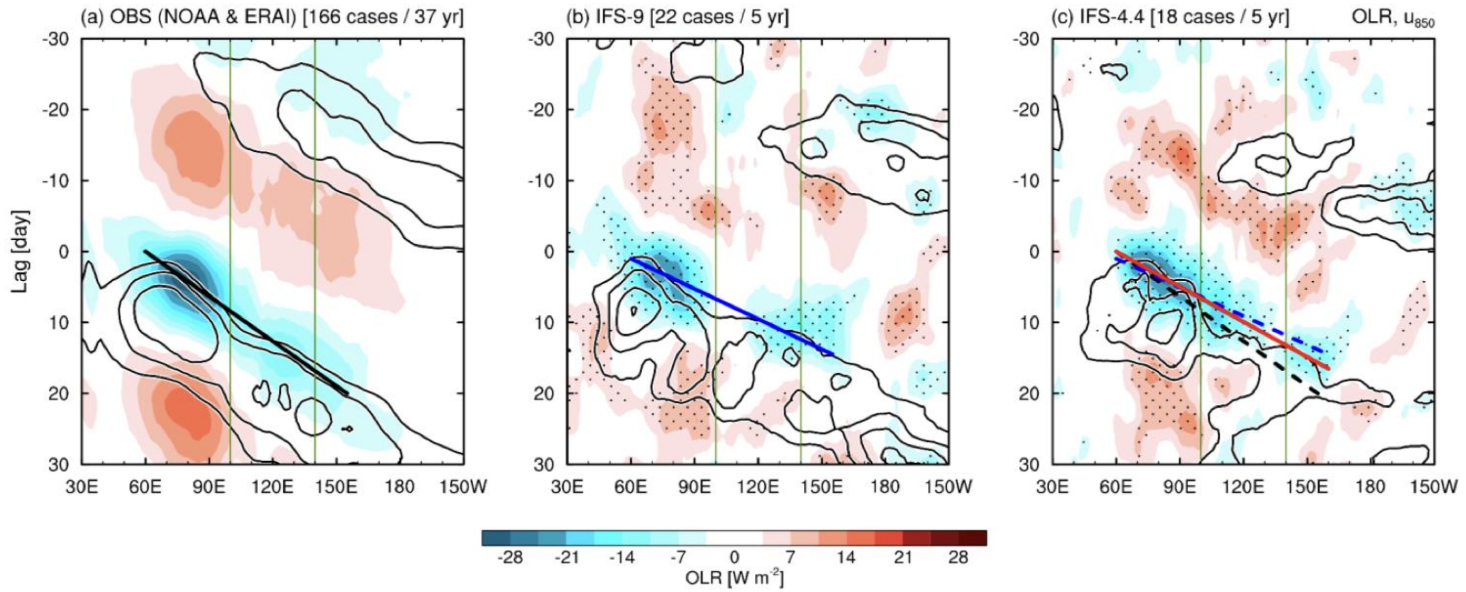
Global km-scale models improve realism of simulations significantly and are now becoming available.

ECMWF

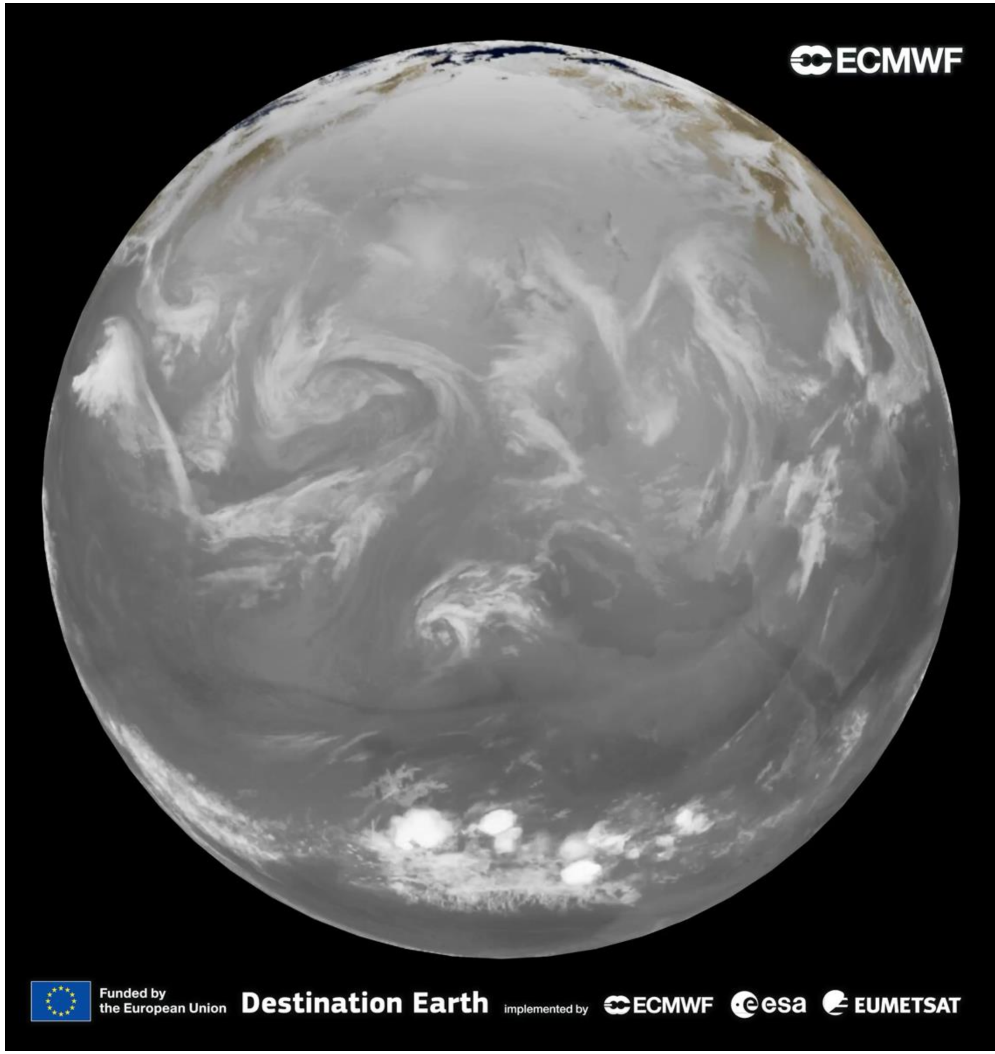
Destination Earth | EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

nextGEMS

The digital revolution to allow for km-scale models



Propagation characteristics of the Madden-Julian Oscillation and composition from (a) observations, (b) IFS 9km simulation and (c) IFS 4.4km simulation with FESOM

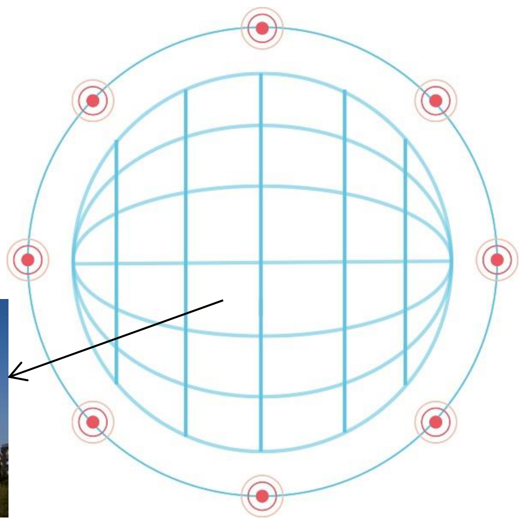


Andreas Mueller and Philippe Lopez

Rackow et al. GMDD 2024

Current Systems

Earth System
models & observations



Limited resolutions

Small-scall processes not represented

Separation of earth system & impact sector models

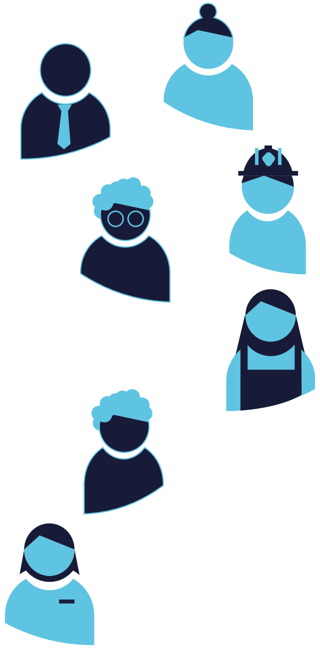
Impact sectors

Limited ways to change experiment design or output

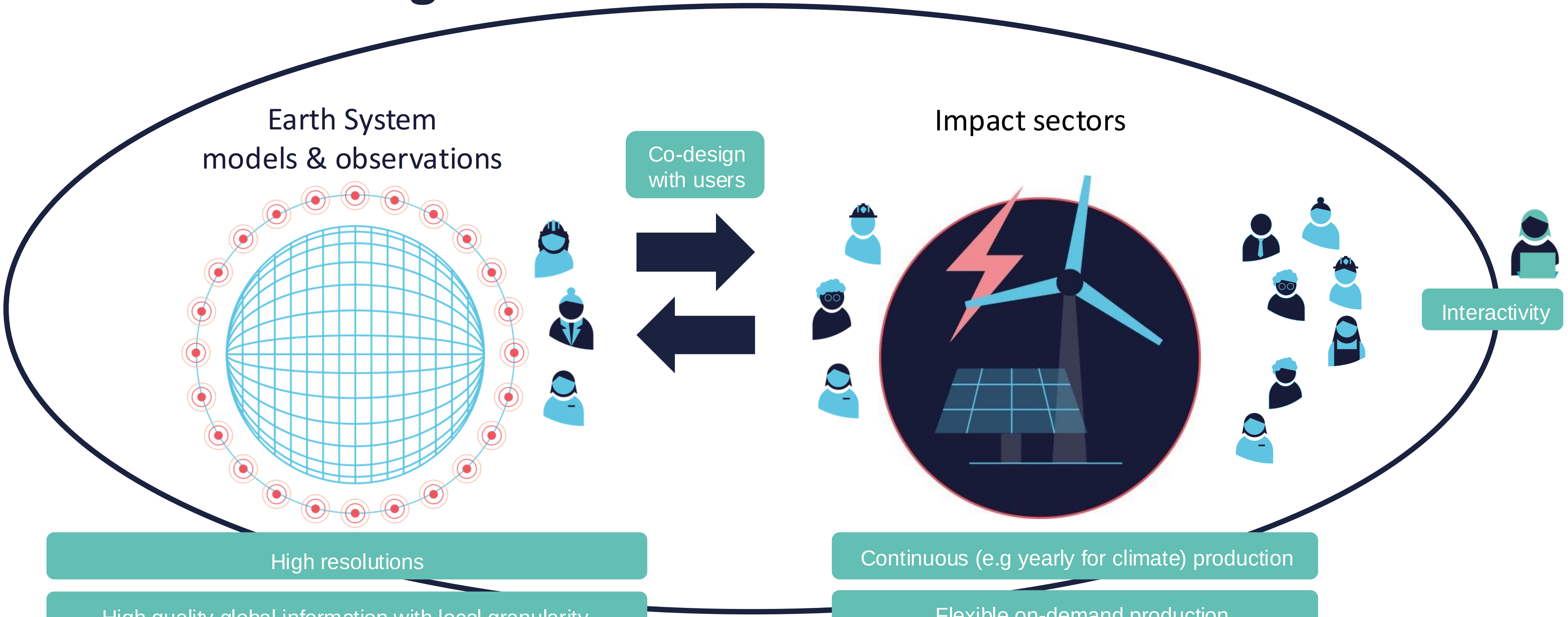
Limited versatility to access and interact with the data

Pre-defined configurations & simulation run times (every 7 years for climate!)

Users



DestinE builds Digital Twins of the Earth



High resolutions

High quality global information with local granularity

Bringing impact sector models WITHIN the same workflow

Continuous (e.g yearly for climate) production

Flexible on-demand production

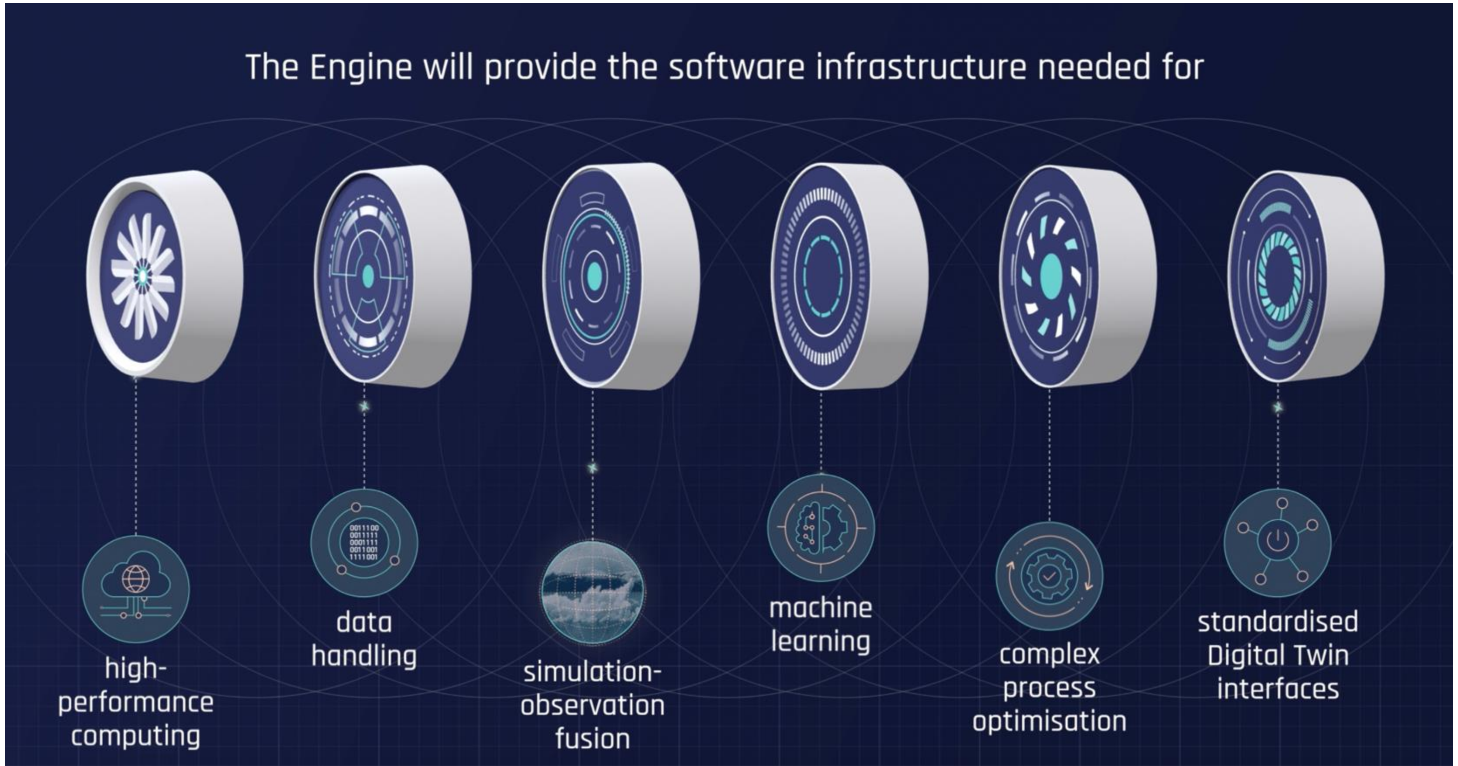
Data streaming and tailored data access

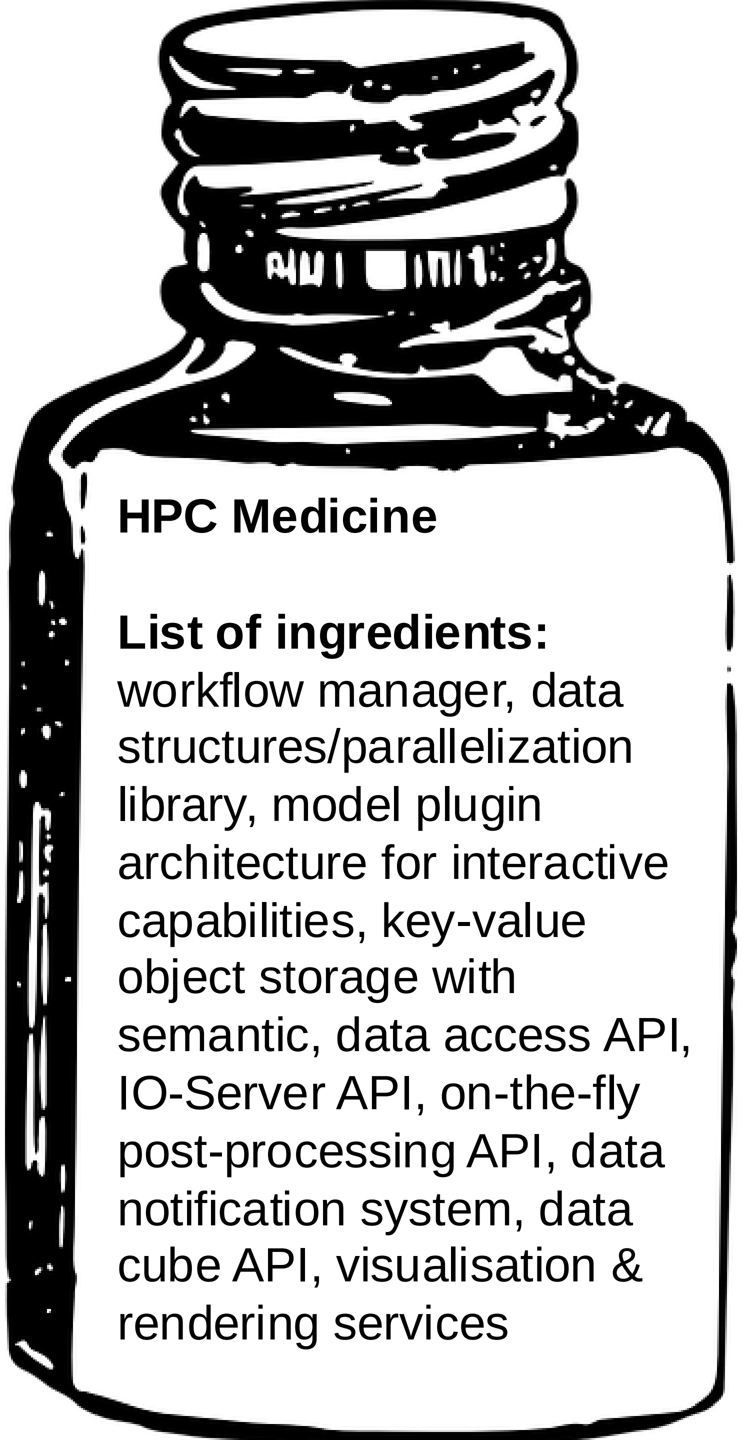
DestinE's Digital Twin Engine

Framework for Digital Twin Workflows

- High Performance Computing adaptation / Digital Twin optimisation
- IO and data workflows
- Software management, controlling workflows, cloud environments
- Visualization

A Game Engine type framework but for Earth Systems…





HPC Medicine

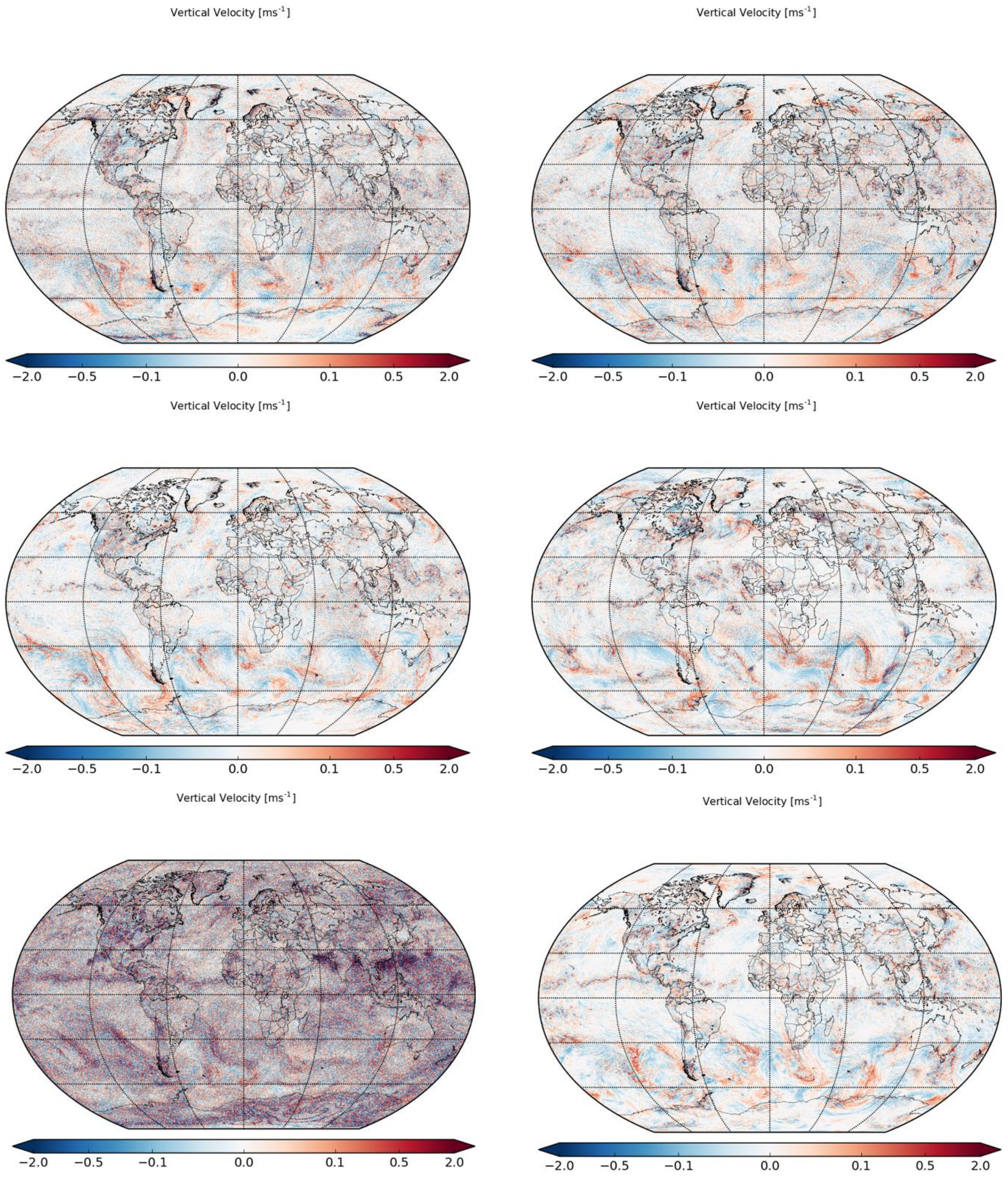
List of ingredients:
workflow manager, data structures/parallelization library, model plugin architecture for interactive capabilities, key-value object storage with semantic, data access API, IO-Server API, on-the-fly post-processing API, data notification system, data cube API, visualisation & rendering services

Global storm resolving models

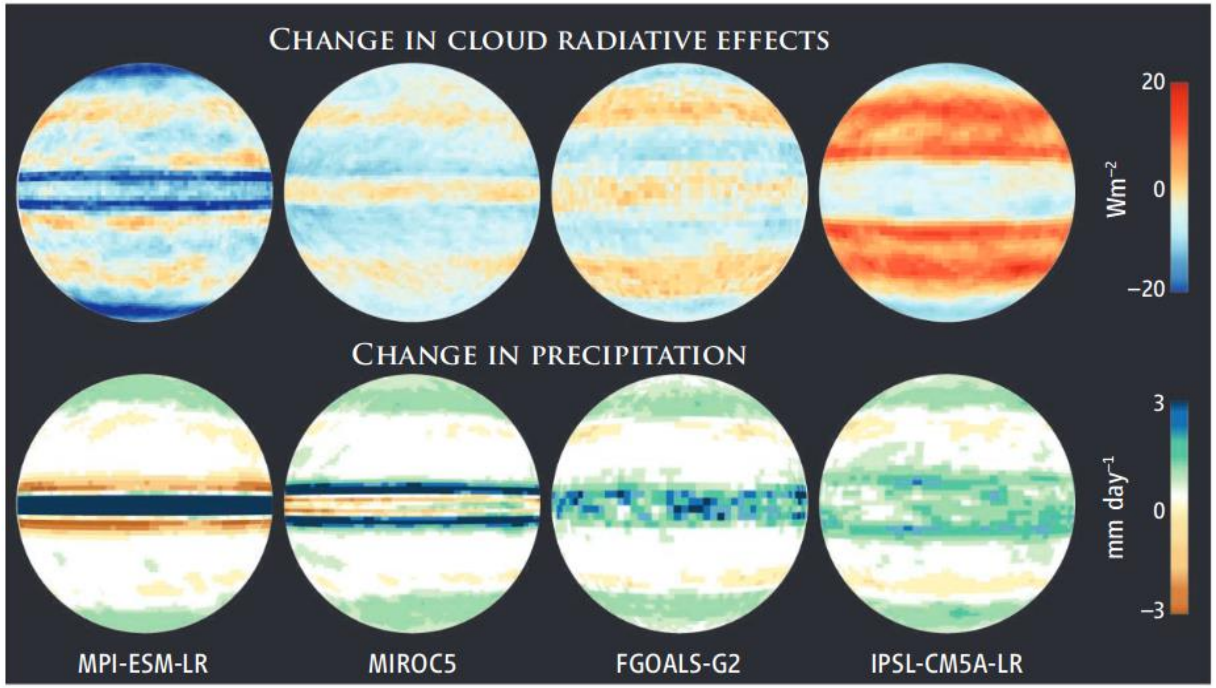
Big steps toward operational use of global storm resolving simulations

- Month-long integration of a number of models at < 5 km grid-spacing as part of DYAMOND
- Season-long integrations of the IFS model at 1.45 km grid-spacing on Summit as part of INCITE
- Year-long coupled ICON integration with 5 km grid-spacing
- 1024-member ensemble data assimilation with 3.5-km grid-spacing with NICAM
- NextGems and DestinE coming
- …

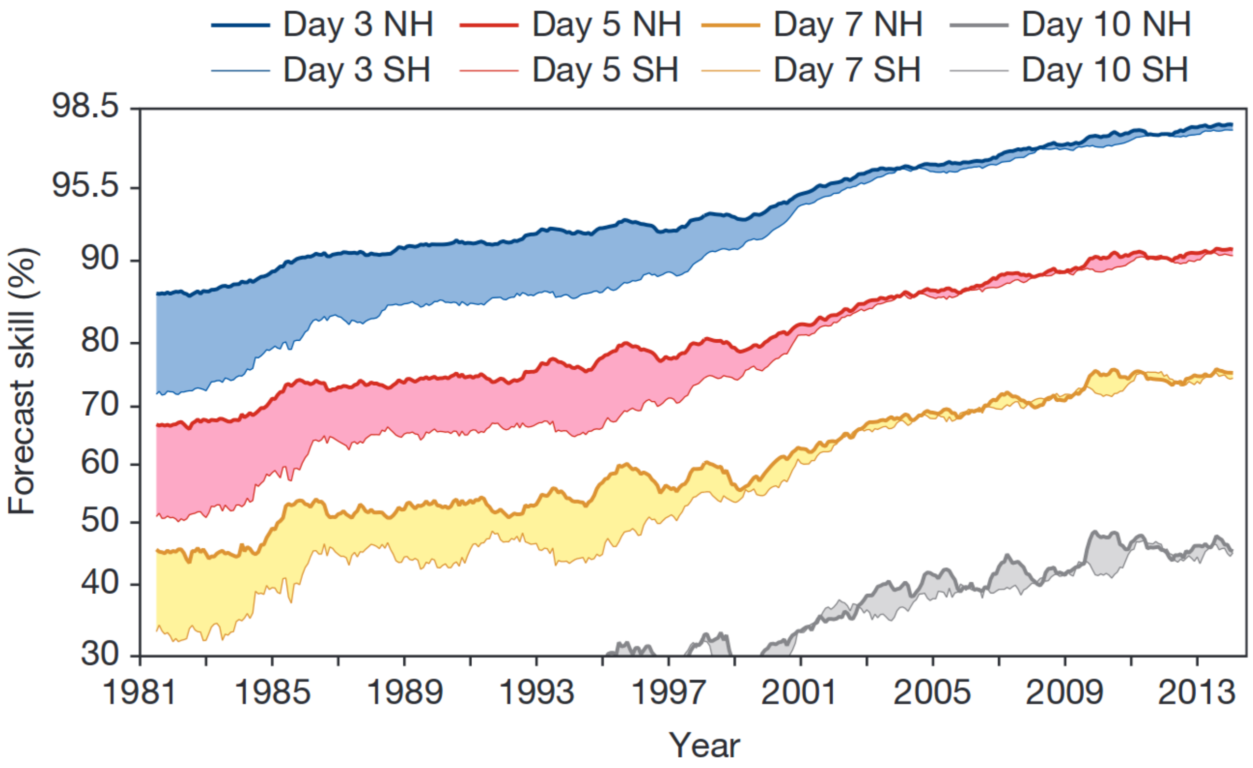
But rather a digital family than digital twins?
Figures by Roland Schrödner and Thibaut Dauhut



Are our current models up for the challenge?

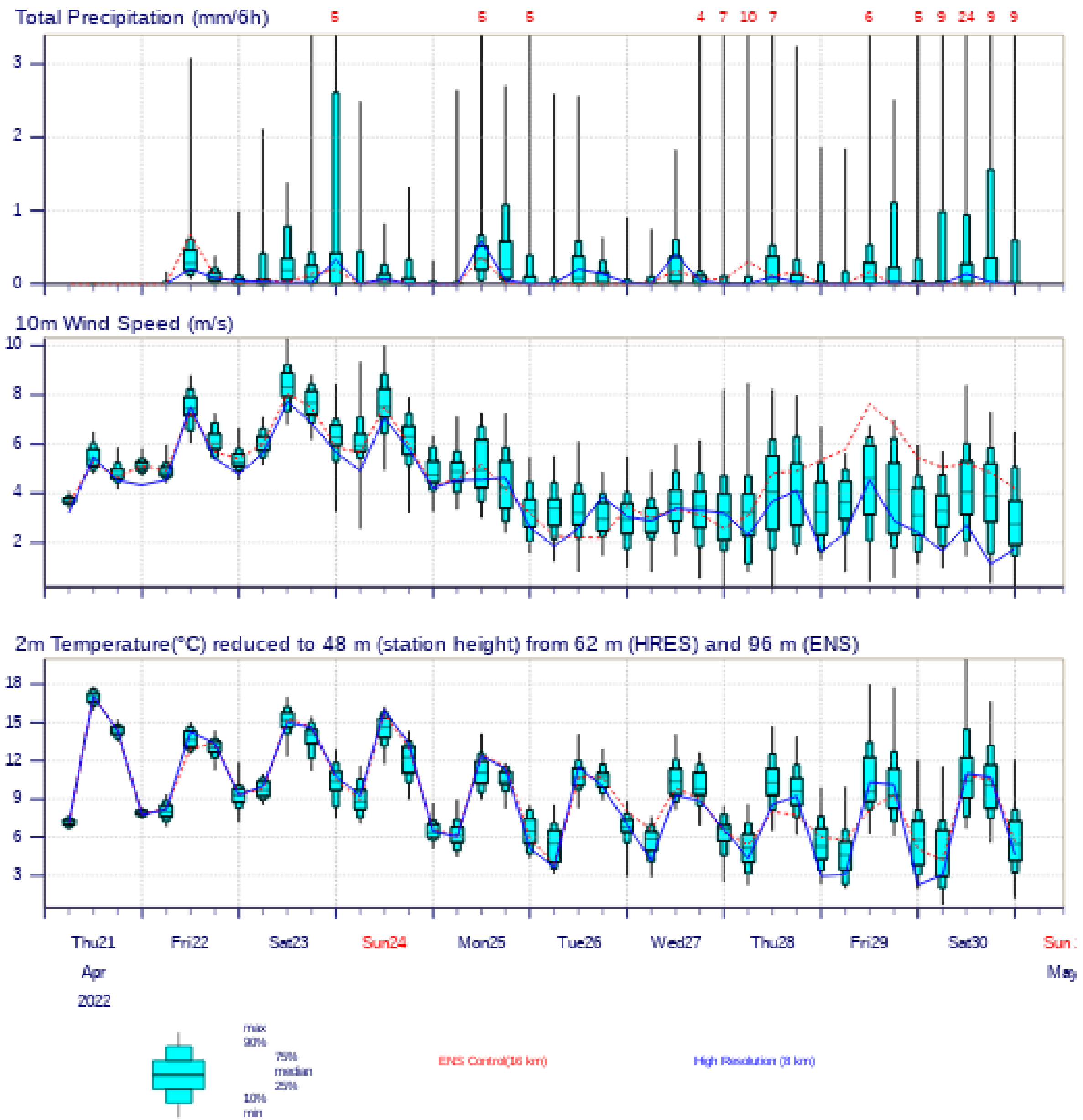


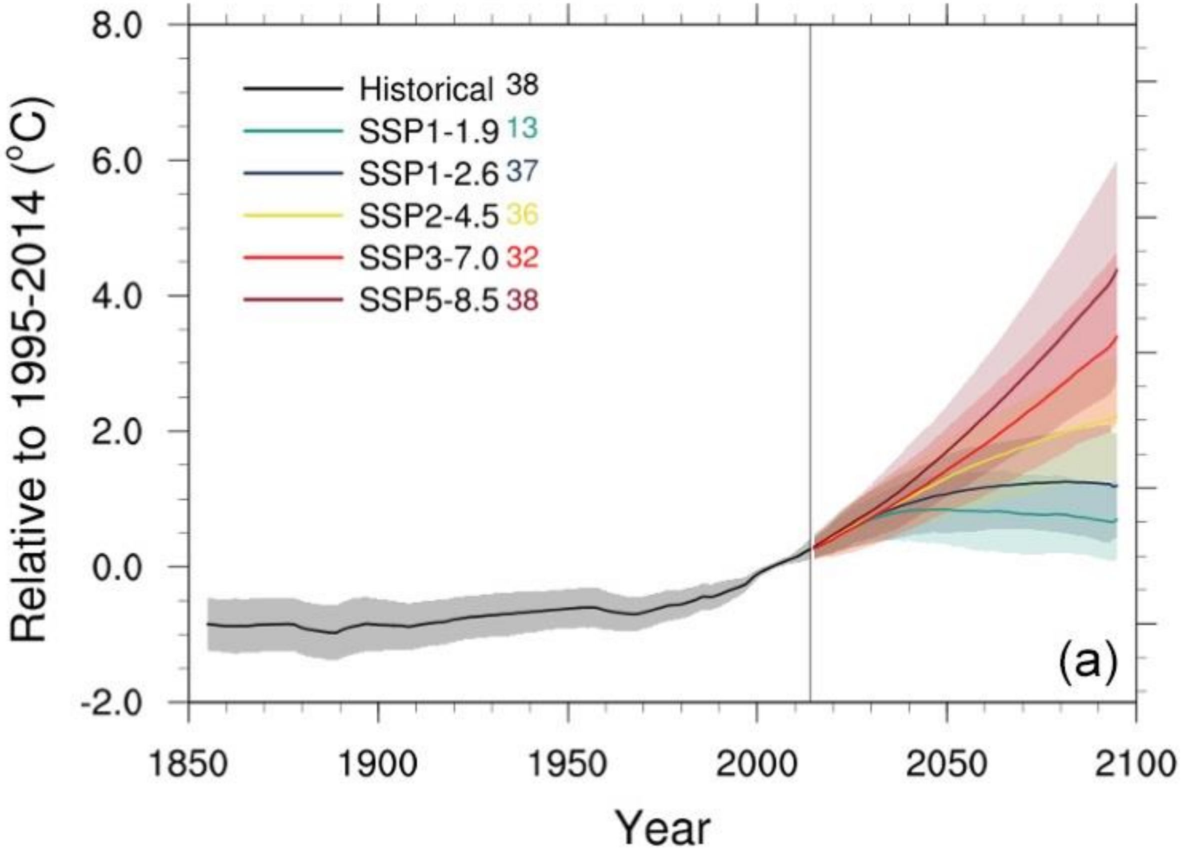
Wide variation. The response patterns of clouds and precipitation to warming vary dramatically depending on the climate model, even in the simplest model configuration. Shown are changes in the radiative effects of clouds and in precipitation accompanying a uniform warming (4°C) predicted by four models from Phase 5 of the Coupled Model Intercomparison Project (CMIP5) for a water planet with prescribed surface temperatures.



Stevens and Bony, Science, 2013.

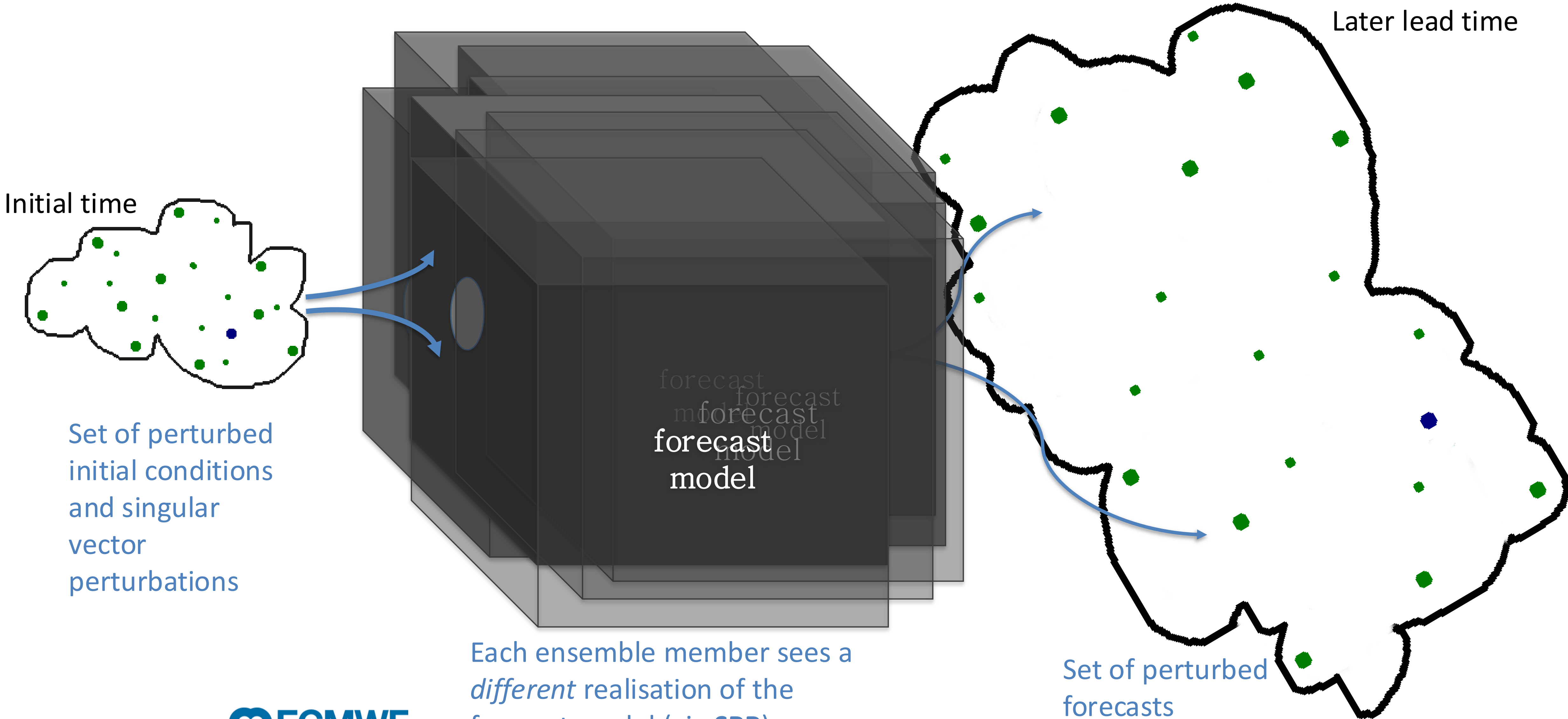
A story of uncertainties





Tebaldi et al. Earth System Dynamics 2021

Sources of uncertainty: accounting for model uncertainty



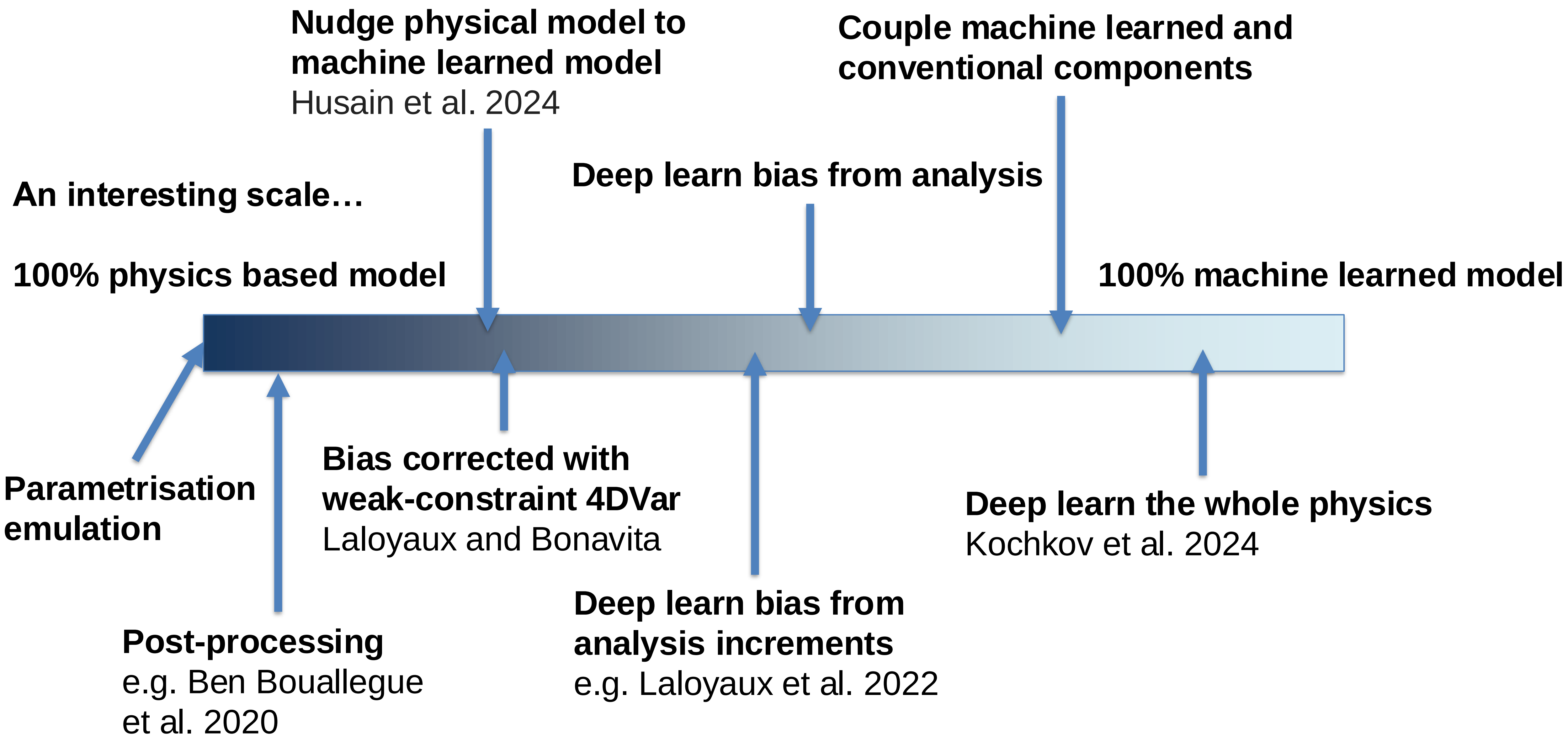
Set of perturbed initial conditions and singular vector perturbations

Each ensemble member sees a *different* realisation of the forecast model (via **SPP**)

Set of perturbed forecasts

Slide from Sarah-Jane Lock

What about hybrid machine-learned physics-based models?

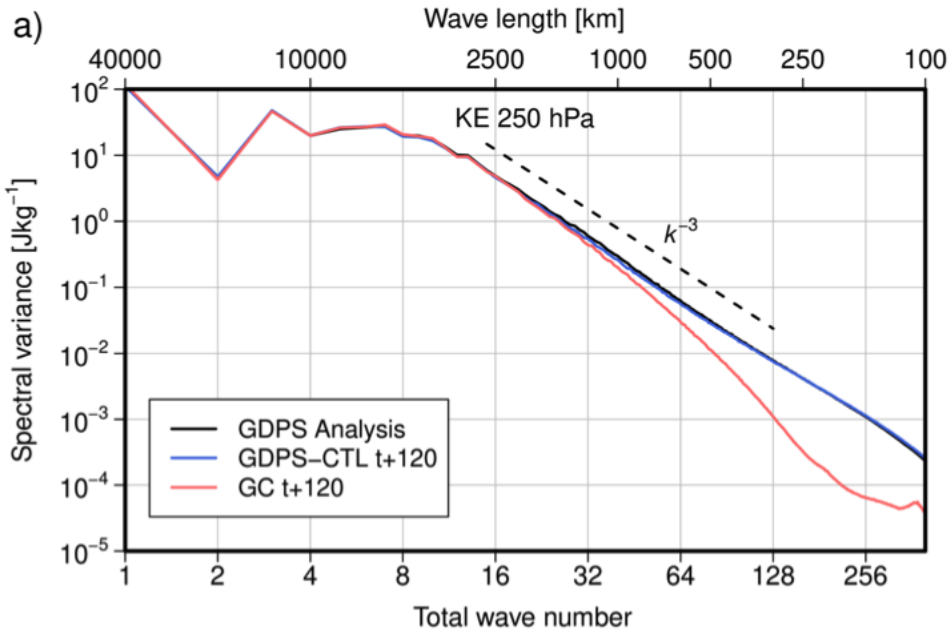


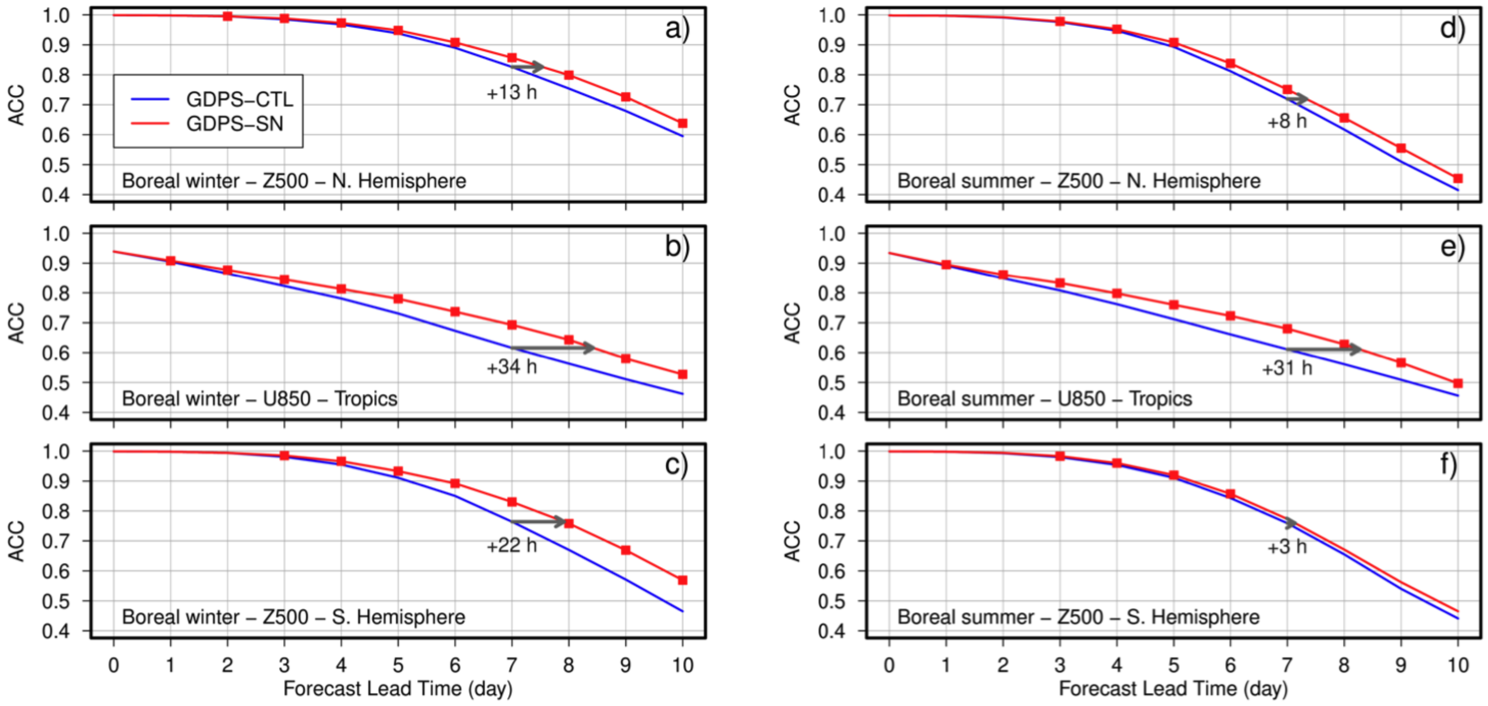
What is the best way to combine machine learning and physical models?

One of the general assumptions of the quiet revolution and physical modelling:
The large scales of the model simulations are well resolved and therefore correct.
The small scales of the model simulations are not well resolved and therefore incorrect.
→ Higher resolution leads to better predictions

However… Machine learned models are coarse, fail to represent small scales, and are still competitive.

→ Get best of both worlds by nudging large scales of machine learned models to the physical models.

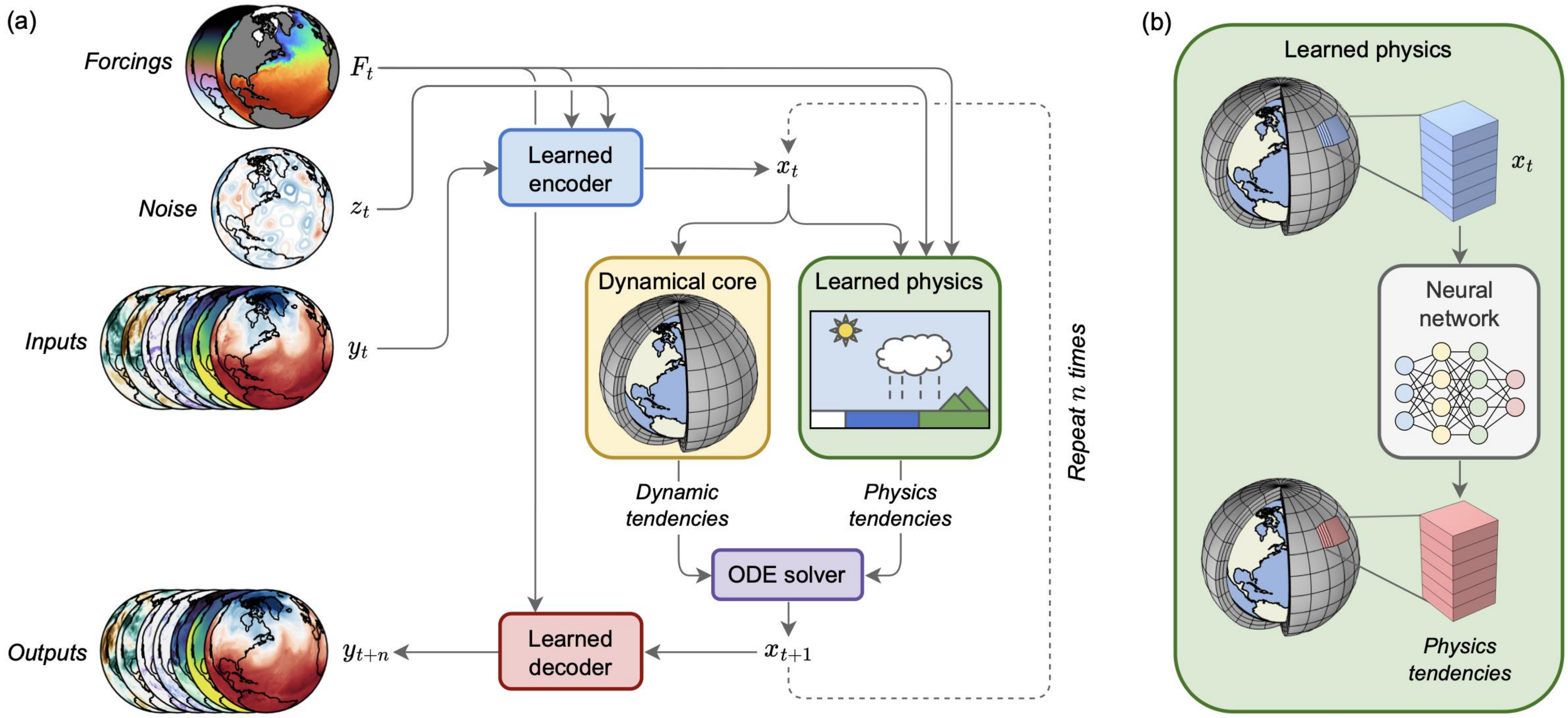




Husain et al. *arXiv:2407.06100* (2024).

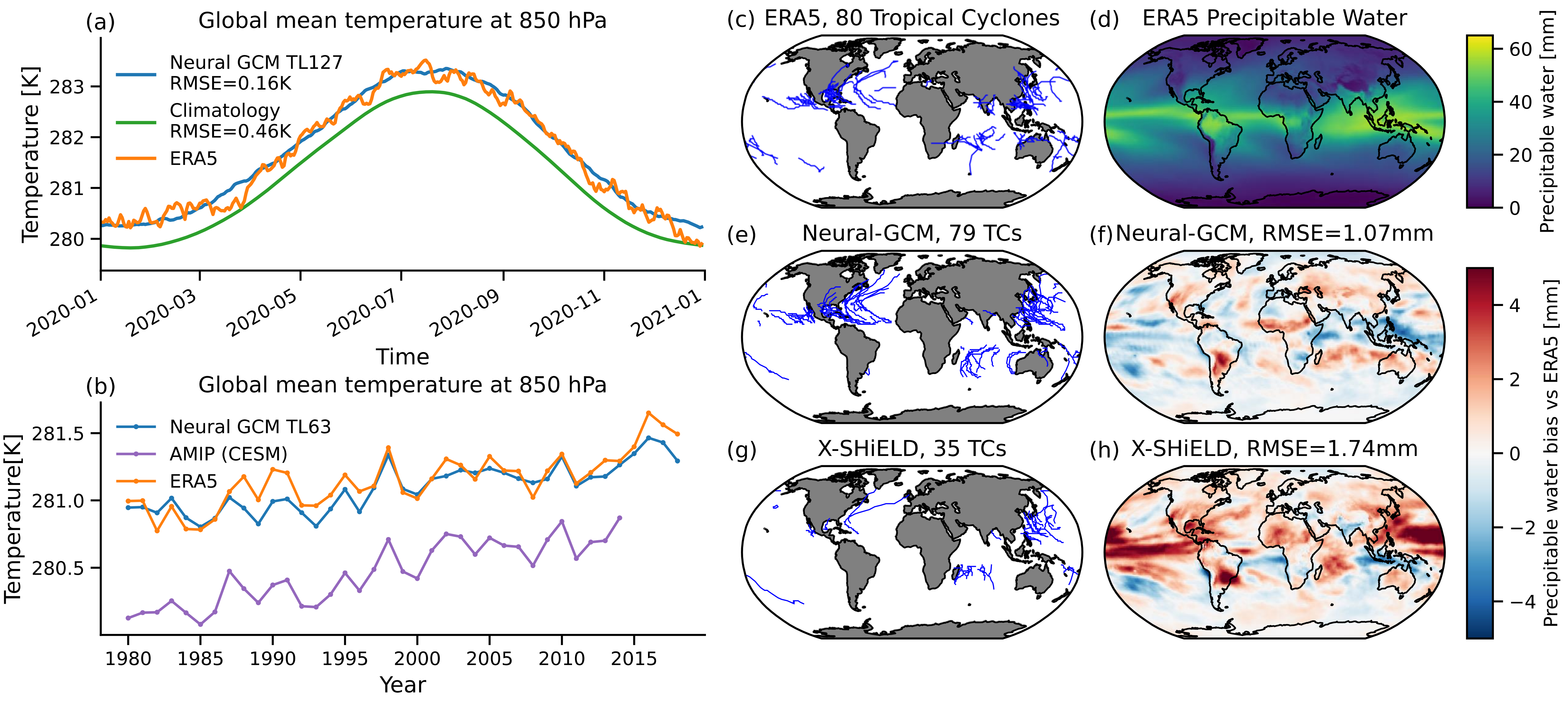
What is the best way to combine machine learning and physical models?

What about hybrid models? – see NeuralGCM from Google



Kochkov et al., *Nature* **632**, 1060–1066 (2024)

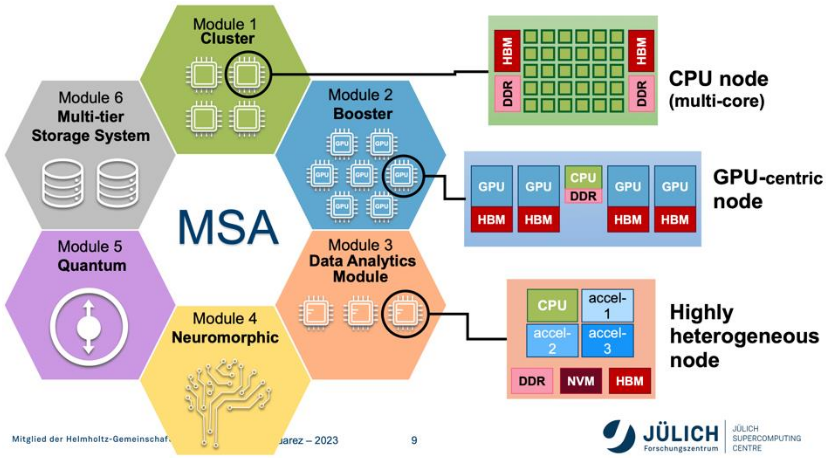
What is the best way to combine machine learning and physical models?



Machine learned models can now also do AMIP simulations.

Kochkov et al., *Nature* **632**, 1060–1066 (2024)

Change of gear in Earth system modelling



Slide from Estela Suarez

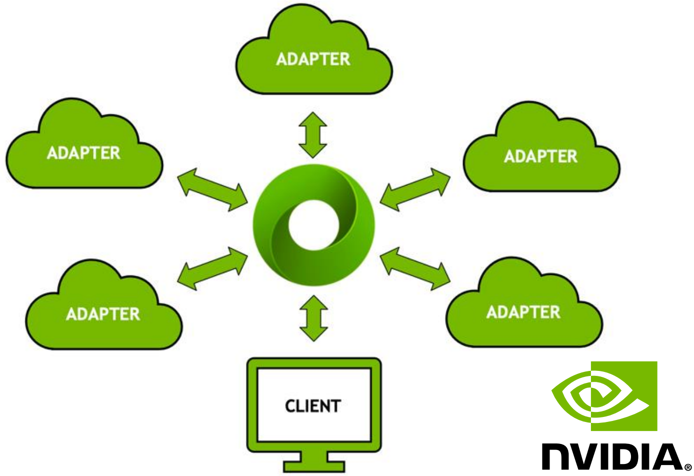
Workmode of 2010:

- Earth system models consist of 1,000,000 lines of Fortran Code
- Code is shared via tarballs, data is stored locally
- Models run on CPUs and Moore's law is still working


Tim Palmer's A380 comparison

Workmode of 2020:

- A team of software developers is needed to use heterogeneous hardware
- Models start to run on GPUs, Moore's law is dying
- Data is stored locally but meta information is available online
- Online code repositories are used to control quality and share model code



Workmode of 2030:

- Machine learning models of 10,000 lines of Python code compete with conventional models
- There will hundreds of machine learning applications using a couple of Foundation models
- HPC is federated
- Data is federated

Summary

- Equations → Discrete Models → Supercomputing
- Numerical models can act as a virtual laboratory for weather and climate
- Numerical models are not perfect and need to be evaluated critically with quantified uncertainties

We need your help to build the models of the future!