Climate Science and NCAS CMS: An intro!

Scientific context and some examples of the activities of one of the oldest University based RSE groups in the UK.

Bryan Lawrence

Prof of Weather and Climate Computing

NCAS and the University of Reading

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Outline

- Climate Modelling: the inexorable consumer of computing and producer of data.
  - The role of resolution, complexity and ensembles.
- What supporting (academic) climate modelling in the UK requires
  - An intro to CMS itself and the helpdesk
- Some example activities:
  - Modelling at Scale: CANARI
  - Modelling Futures: Preparing for Grace-Hopper
  - Models and Data: Intercomparison with TWINE-VISION
  - Data: A pure python HDF reader, pyfive

Given knowledge of state at every grid point
at time t,
calculate at every grid point
state at $t + \Delta t$.

Many points, much state, integrated for years with timestep of *o(minutes)*!

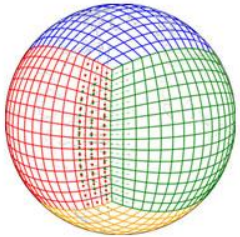
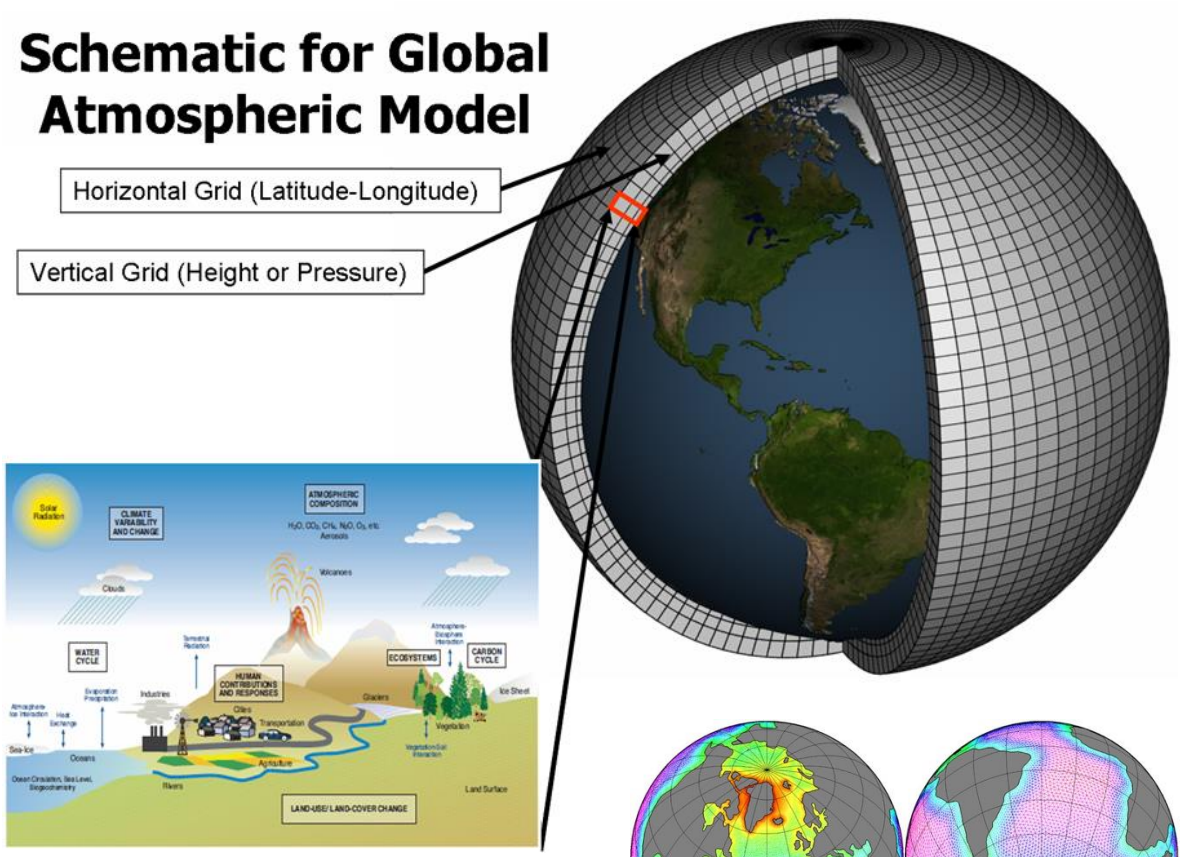


Image credits for those I know:
http://dx.doi.org/10.48550/arXiv.2508.17462
https://www.gfdl.noaa.gov/fv3/fv3-grids/
https://doi.org/10.1016/j.dynatmoce.2017.05.003



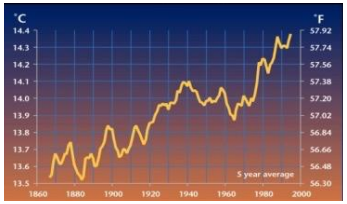
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Give me more compute?

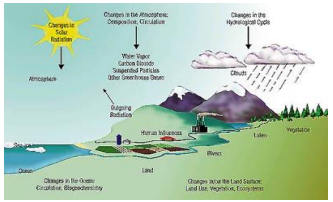
WE can't have it all!

WE need to spend compute on the key aspects of the problem.

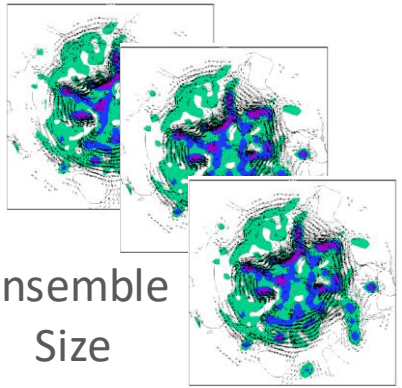
Which are?



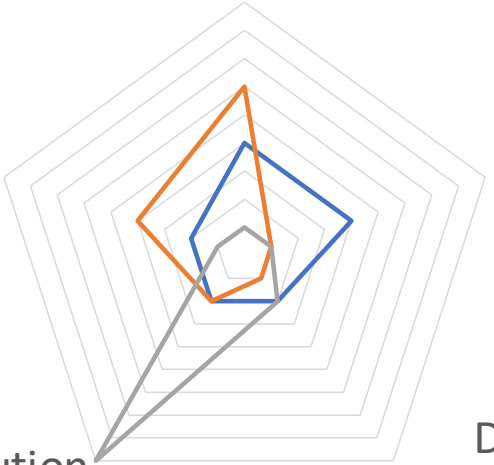
Duration



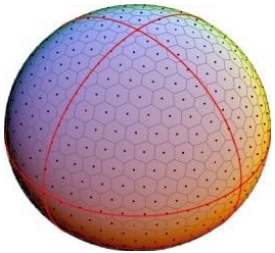
Complexity



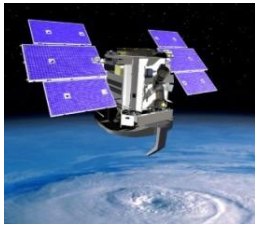
Ensemble Size



Resolution

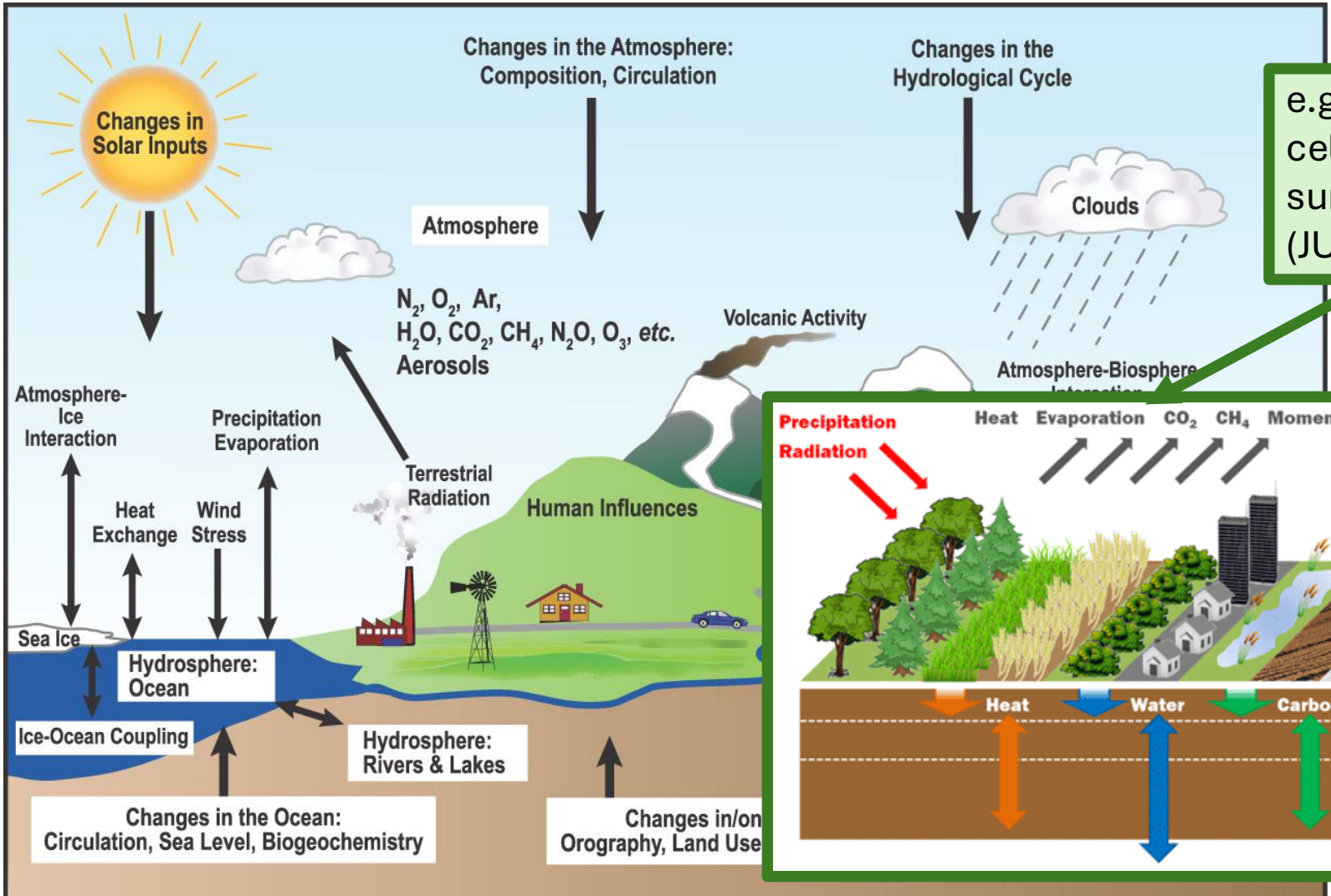


Data Assimilation



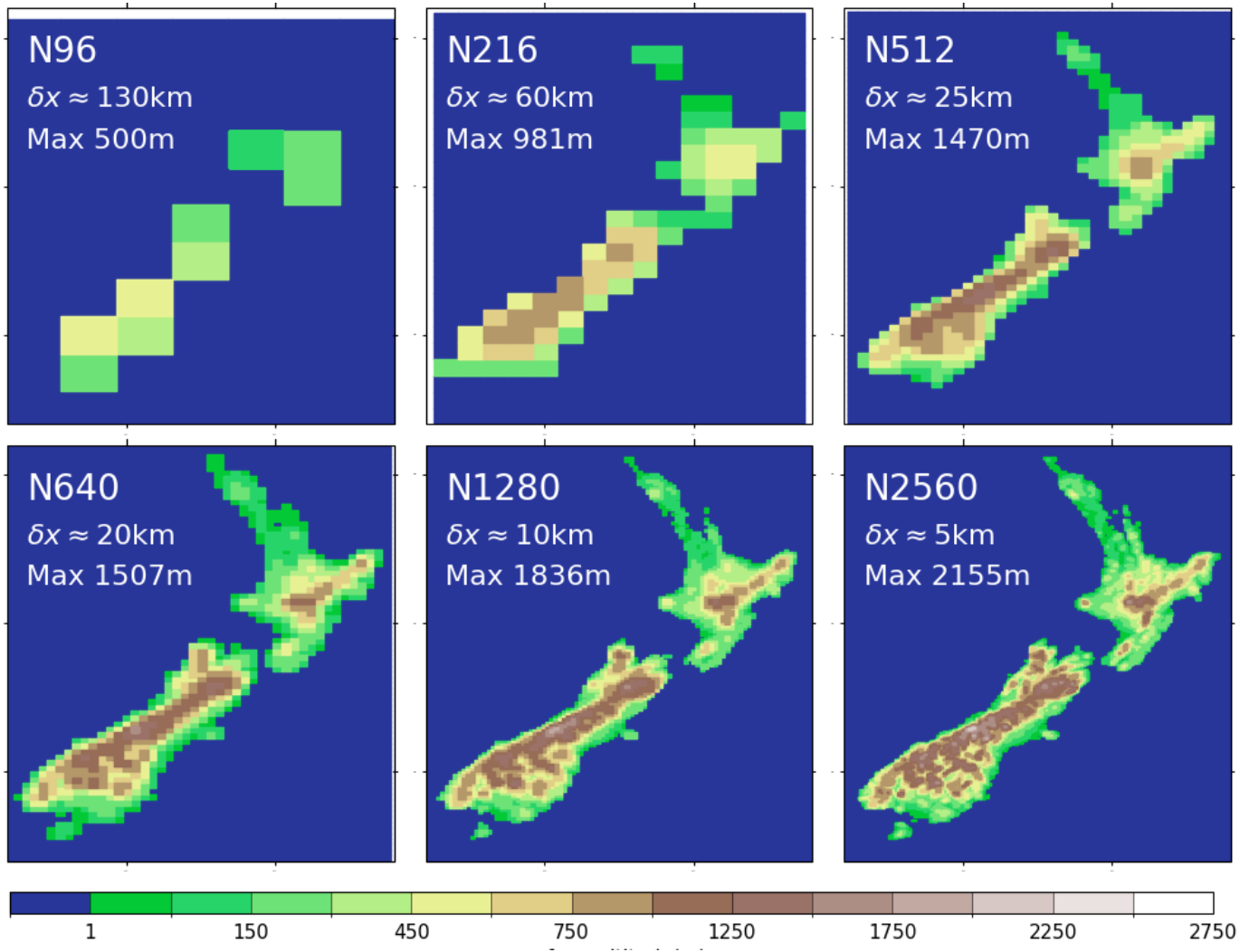
Variability Scenarios Extremes

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL



e.g. One cell land surface (JULES)

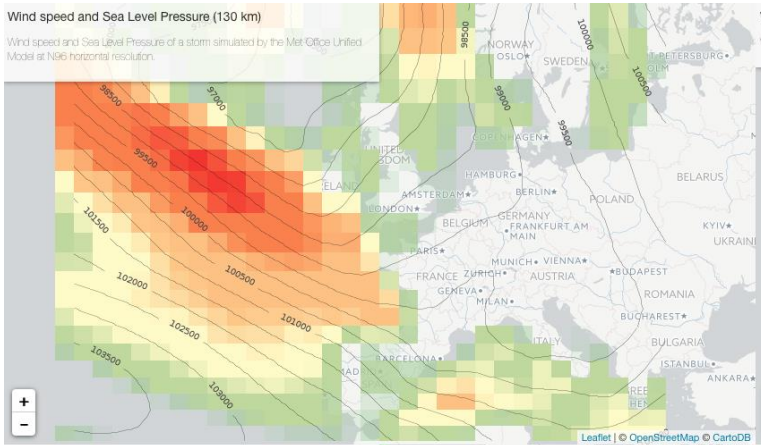
Attempting a visceral feeling for resolution

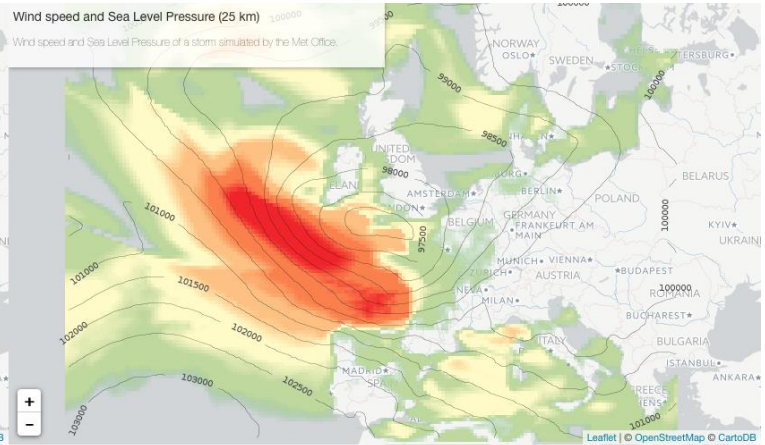


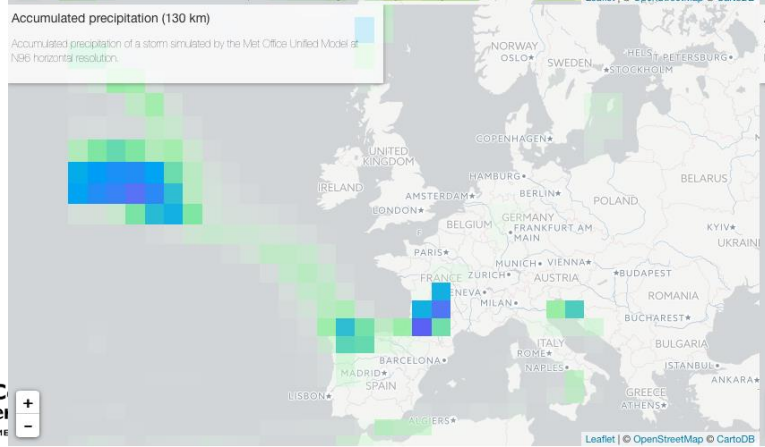
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

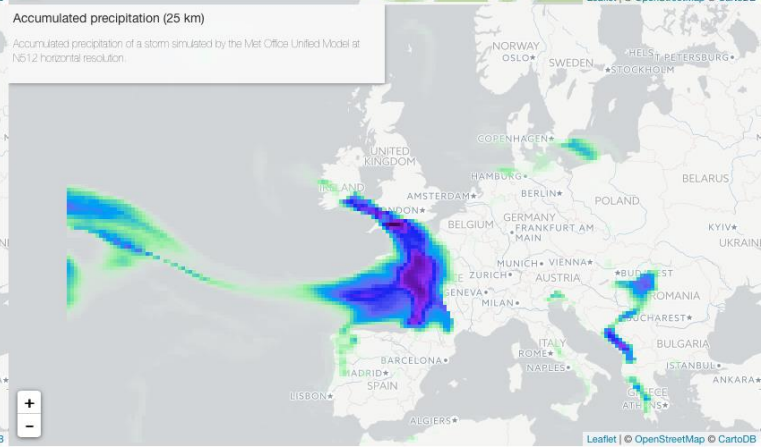
The influence of resolution

P-L. Vidale
(PI NCAS
HCRCM)







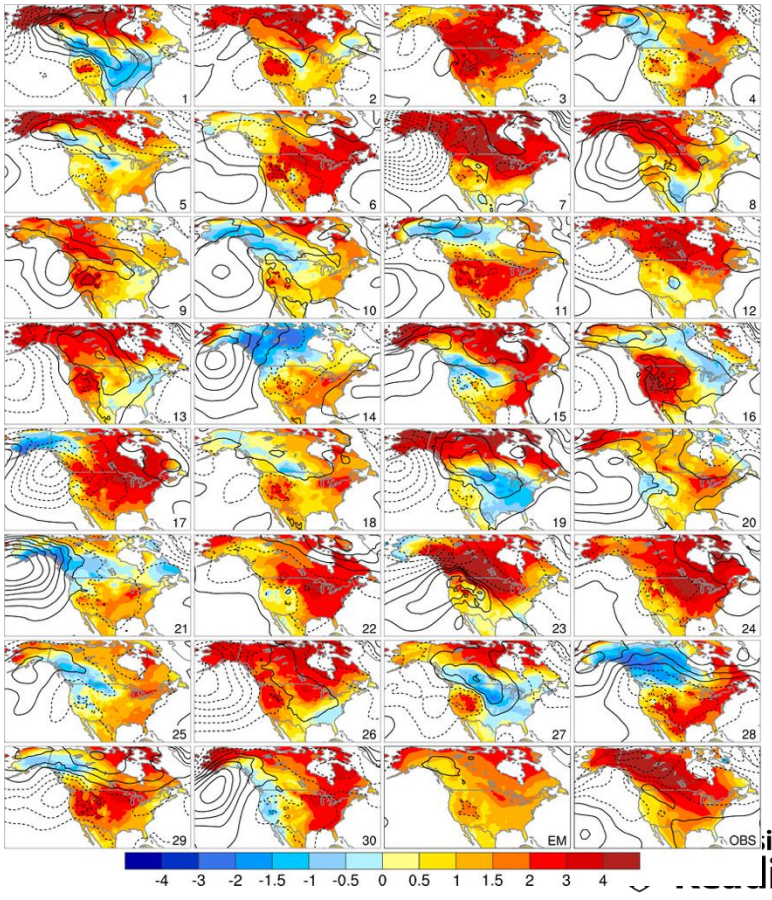


What is a large ensemble?

- A single climate model simulation exhibits variability due to natural internal atmosphere/ocean processes (e.g. ENSO)

- Variability becomes even more important on regional scales:
  - 50-year temperature trends over North America 1963–2012 in 30 historical simulations
  - Warming due to greenhouse gases is clear, but internal variability is also large

From Deser et al., 2016, *JClim*



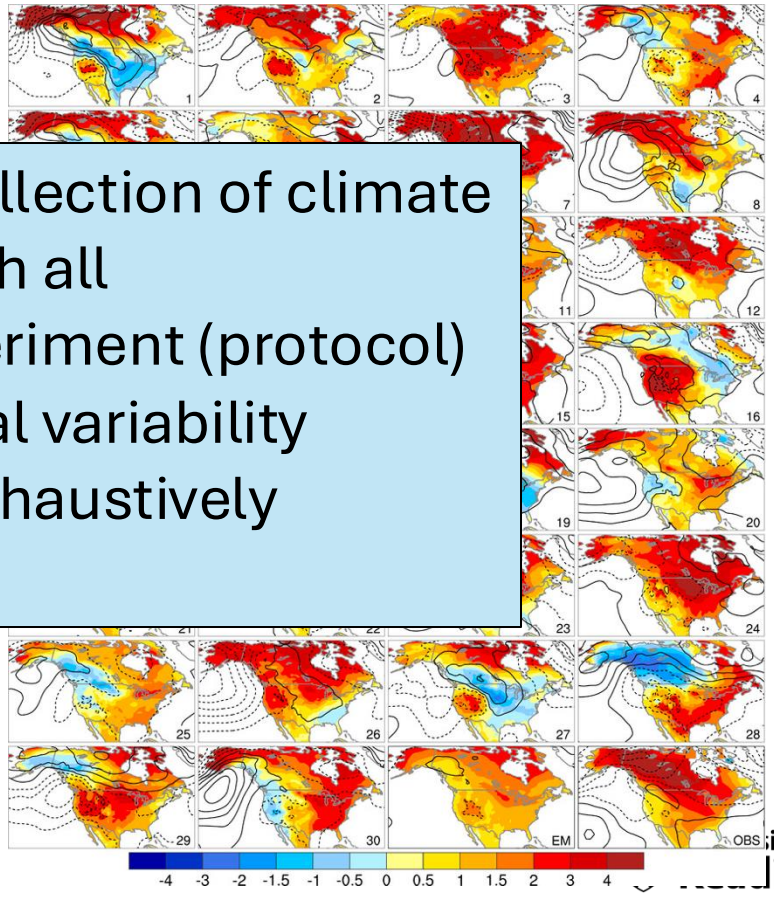
What is a large ensemble?

From Deser et al., 2016, *JClim*

- A single climate model simulation exhibits variability due to natural internal a
  (e.g. ENS
- Variability
  important
  - 50-yea
    North
    histori
  - Warming due to greenhouse gases is clear, but internal variability is also large

A large ensemble is a collection of climate model simulations which all

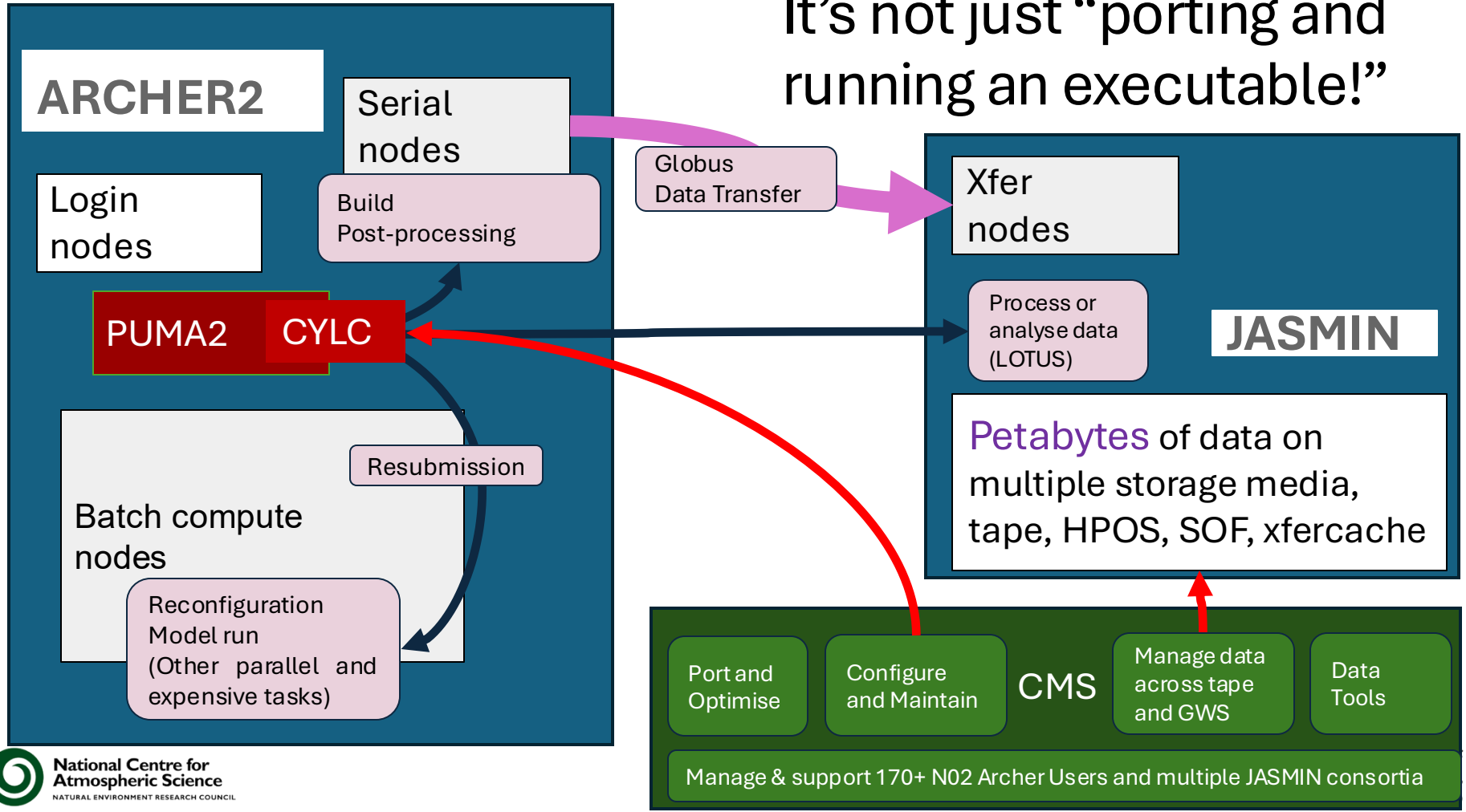
- Follow the same experiment (protocol)
- Aim to sample internal variability independently and exhaustively



Coupled Ocean-Atmosphere Cost!

Resolution	Data	Speed (SYPD)	Nodes	Cost (NHSY)	Typical Use Case
Atmosphere: N96 ~133 km Ocean: O1 ~ 100 km	Simulation: ~ 0.1-> 0.5 TB/D JASMIN : o(PB)s	5-10	5-10	10-50	Coupled centennial climate. CMIP6, CMIP7. Physical climate and Earth System (UKESM) configurations
Atmosphere: N216 ~60 km Ocean: O25 ~ 25 km	Simulation: ~4 TB/D JASMIN : o(PB)s	2	32	384	Large Ensembles (CANARI, 6000 years!).
Atmosphere: N512 ~25 km Ocean: O12 ~ 10 km	Simulation : ~ 2.5 TB/D JASMIN: o(10^2TB)s O(10)PBs	0.6	80	3120	Coupled simulations, selected model intercomparison projects: HIRESMIP, EPOC. Trying to port to Grace-Hopper
Atmosphere: N1280 ~10km Ocean: O12 ~ 10 km	Simulation: ~ 10 TB/D? JASMIN: TBD	0.5	150	7200	Building this now, aiming to use for high resolution climate, extremes, cyclones etc.

It's not just "porting and running an executable!"



NCAS Computational Model Services

CMS October 2025
(*Lawrence*)
Head: *Lister*

- **Data**
  - CF Tools *Bartholomew*
  - CF Tools & Standards Hassell
  - ESMValTool (UKESM) *Predoi*
- **Workflow**
  - PUMA & HPC Workflow *Hatcher*
  - HPC Workflow & Support *A.N.Other*
- **HPC**
  - HPC Specialist *Bilbao*
  - HPC Specialist *Case*
  - HPC Specialist *Osprey*
  - HPC Specialist *Wilson*
- **Applications**
  - AFESP (ECMWF et al) *Ahmadi*
  - Land Surface (JULES) *Livings*
  - Coupled Systems (UKESM) *Stringer*














National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

CMS Help Desk

Supporting “N02 Consortium” on ARCHER2 (and other) HPC platforms.

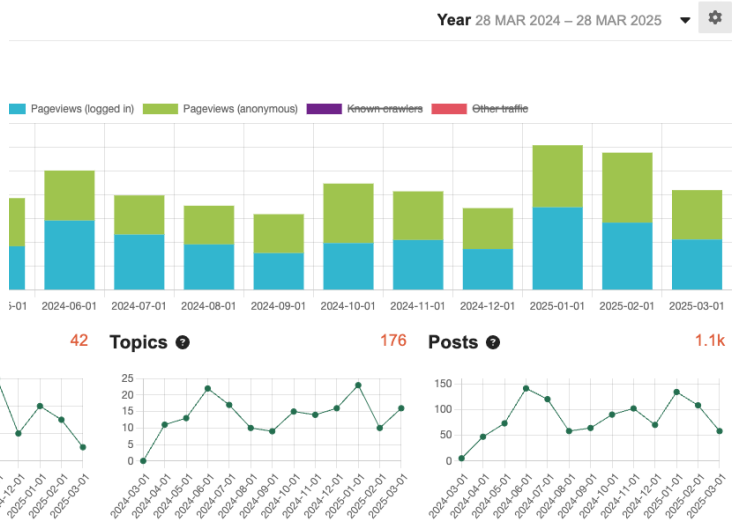
NCAS CMS Modelling Helpdesk Website About

National Centre for Atmospheric Science

Sign Up Log In

categories tags all Latest Hot Categories Docs

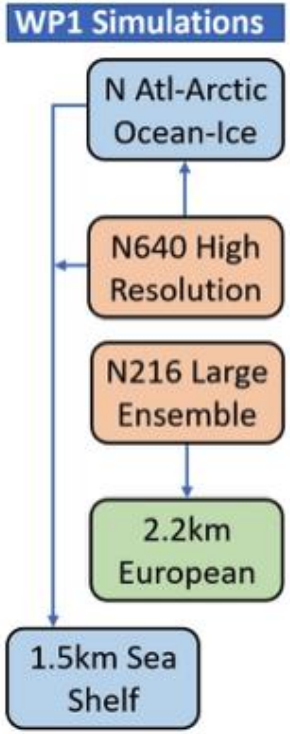
Topic	Replies	Views	Activity
Welcome to the NCAS-CMS Modelling Support Forum Welcome to the CMS Modelling Support Forum. Here you can raise all your modelling support queries (topics) covering Unified Model, NEMO, problems logging into supported platforms like ARCHER2, etc. Before raising a que… read more	0	1.1k	Jun 2021
Validation fail Unified Model	5	9	3d
SVN issues with JULES on Jasmin JULES	5	18	4d
Error activating new UKCA tracers Unified Model Monsoon3	23	50	6d
Merge Conflicts Unified Model ARCHER2	3	6	11d
Nemo postproc error in coupled run	3	8	12d
Validity Time Mismatch Unified Model ARCHER2	3	30	12d
JULES ancillariy suite with ANTS on Jasmin Rose/Cylc and FCM JASMIN	3	7	12d
Porting ENS to ARCHER2 Unified Model ARCHER2	6	19	12d



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

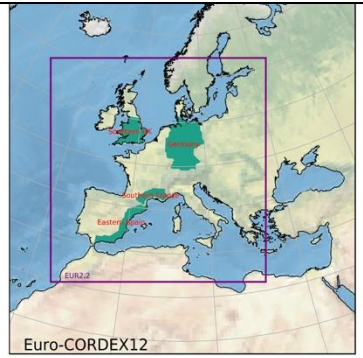
CANARI: Climate Change in the Arctic-North Atlantic Region and Impacts on the UK

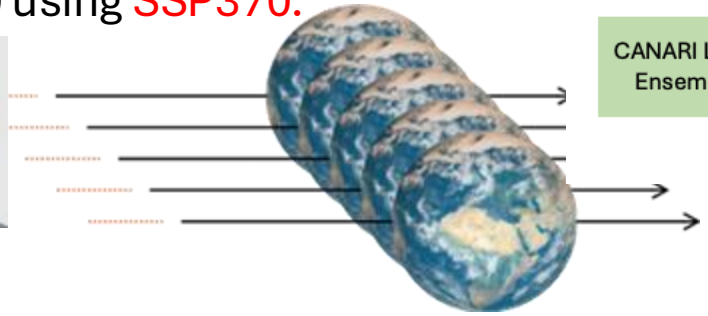
(CMIP6: 500-year picontrol experiment)

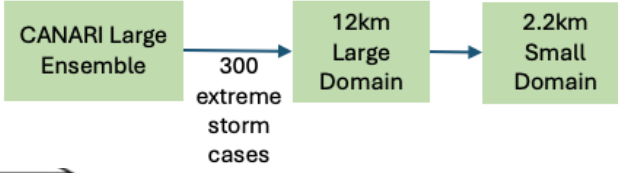


1. Run 8 "macro" historical experiments from differing initial ocean states for 1850-1950.
2. Make 5 perturbations of the 1950 atmospheric state.
3. Continue 40 historical simulations from 1950 to 2014, and
4. Continue 40 future simulations onward to 2100 using SSP370.

Domains for nested 12km and 2.2km models (Chan et al. 2020)



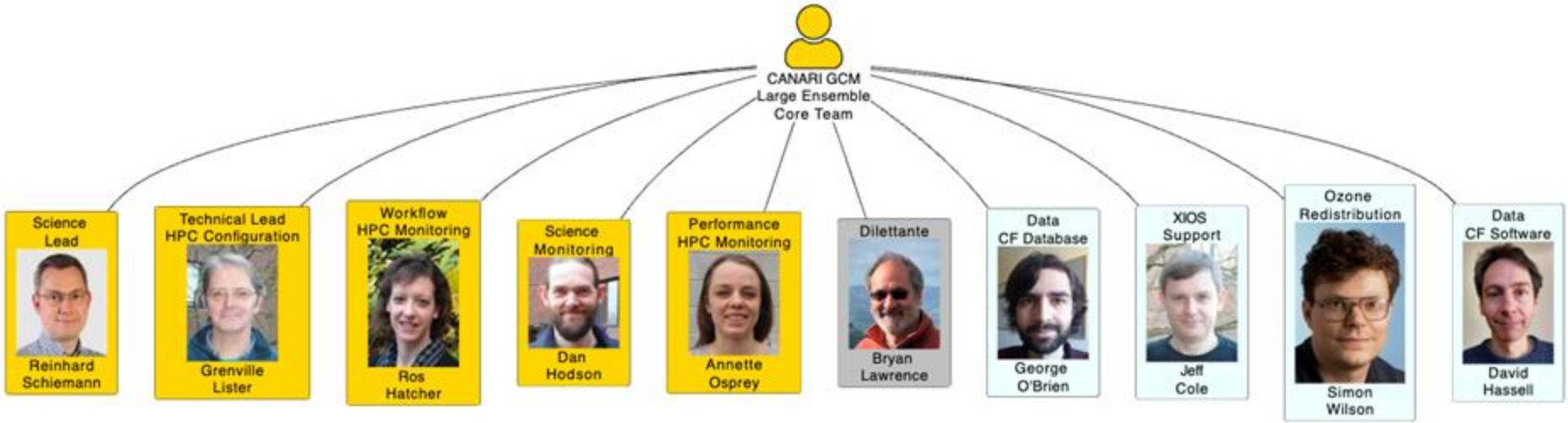




National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

CANARI: 6000 years of simulation
Large team put the model and workflow into place

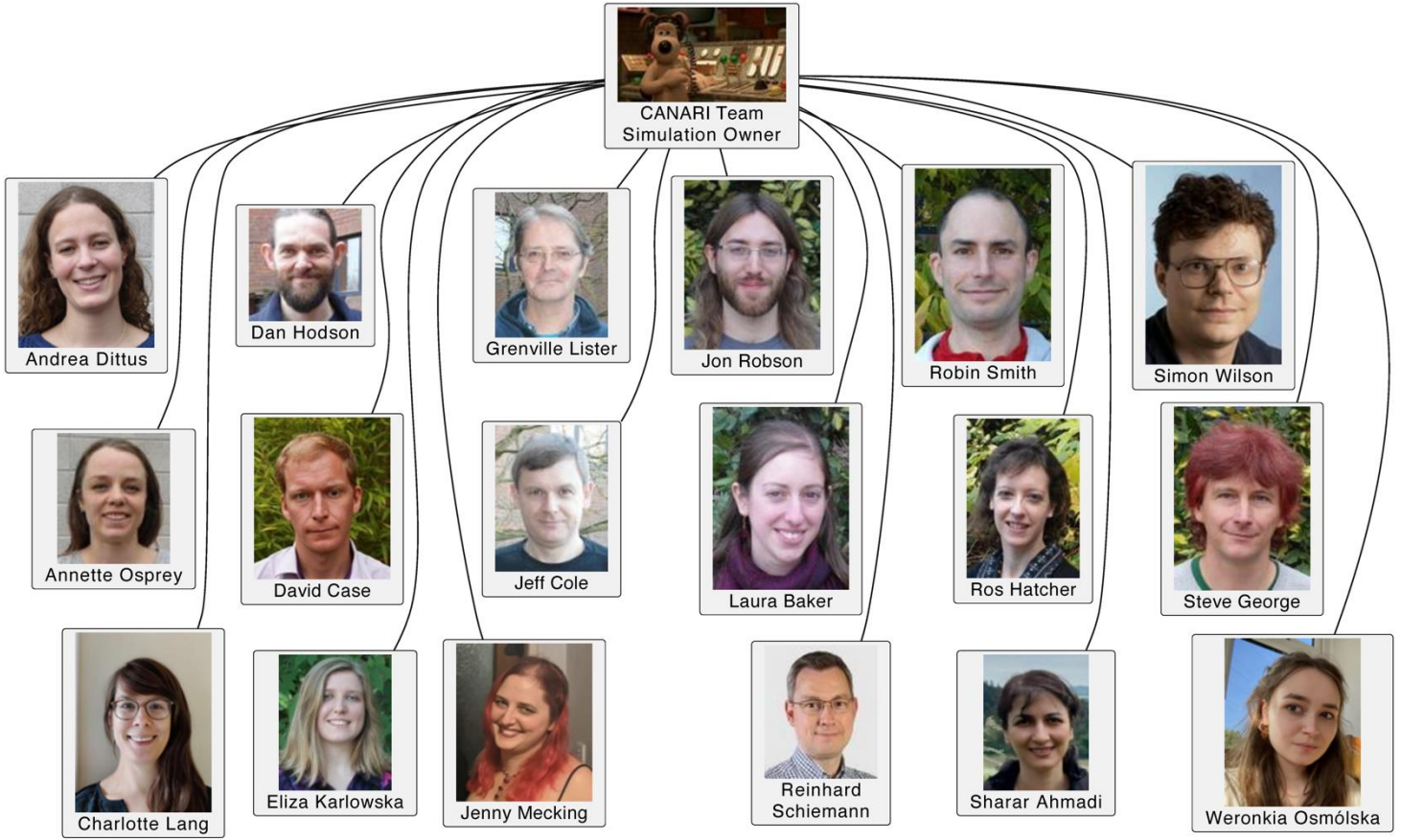


Every one of whom had/has a role to play in setting up and/or sustaining the large-ensemble.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

CANARI: Large team involved in running the model

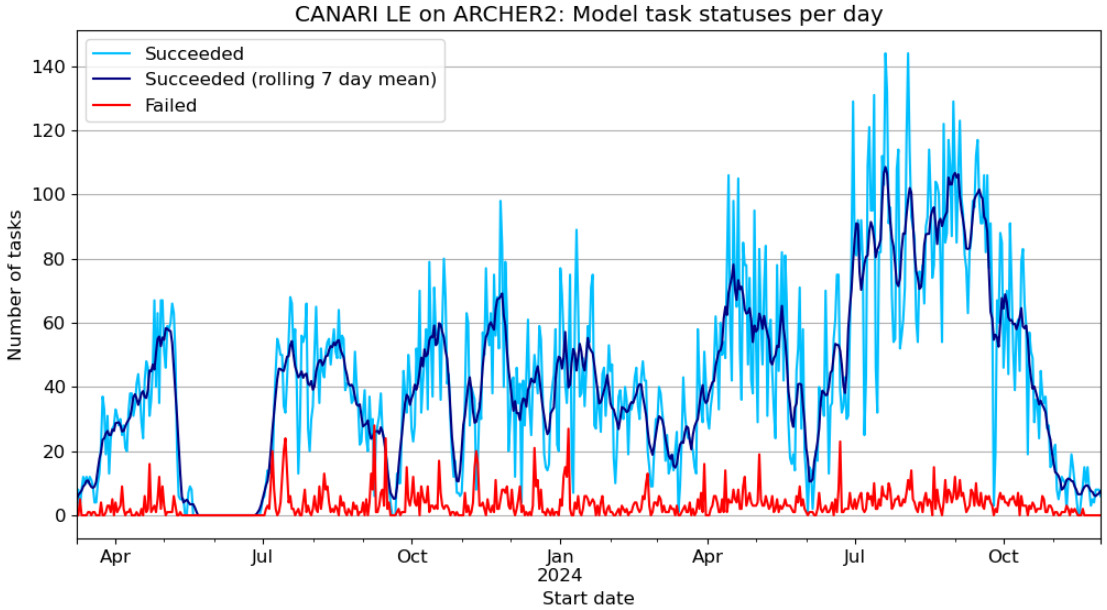


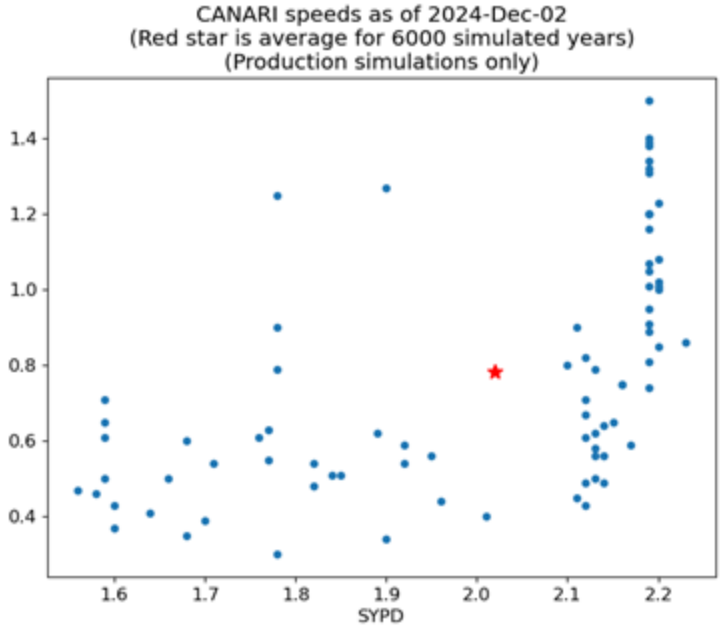
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Reported Average Speeds are Misleading

Each dot is one simulation cycle. One cycle is 3 months
There were 24,000 cycles (6000 x 3-month jobs)!
Most happily resubmit … but many do not





Large Ensemble Data

Simulations

- 40 x 74 years Historical
- 40 x 76 years Future

That is:

- 40 x 150 years (1950-2100), or
- 6000 years of simulation

Output

- 822 ”atomic” datasets per atmospheric simulation.
- UM boundary conditions (“frames”) for several regions.
- Restart dumps monthly plus extra for “specially identified storms”.
- Additional variables to support all CORDEX regions.

Data

Volumes (compressed): ~5.5 PB

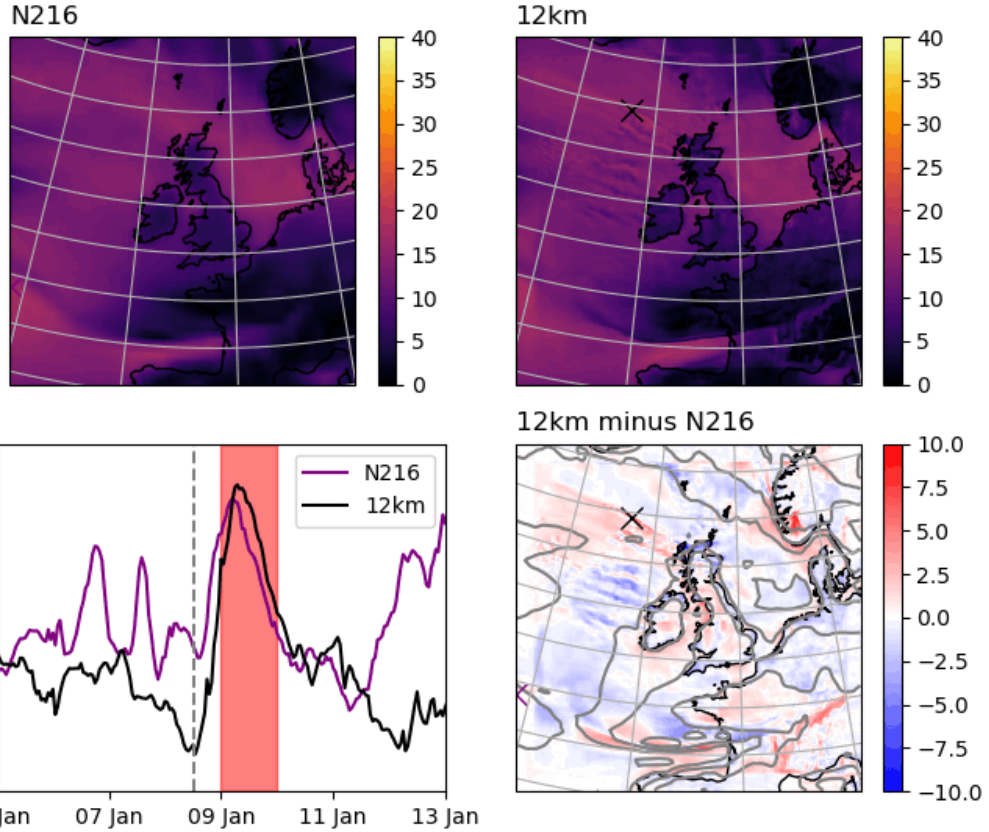
- 2 PB atmos fields
- 0.5 PB atmos restarts
- 0.4 PB ocean fields
- 0.1 PB ocean restarts
- 0.1PB ice fields and restarts
- 2.1 PB of frames etc

4.5 Million Files!

1 PB Disk used as “tape cache”

CANARI also involves nested simulations

- Find the most intense windstorms in the CANARI data, and then re-simulate using nested even higher resolution model.
- The most intense UK windstorm in 2,600 years of LE data simulation
- Cold conveyor belt is more intense and sharper in the 12km RCM than the GCM



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Porting to Grace-Hopper

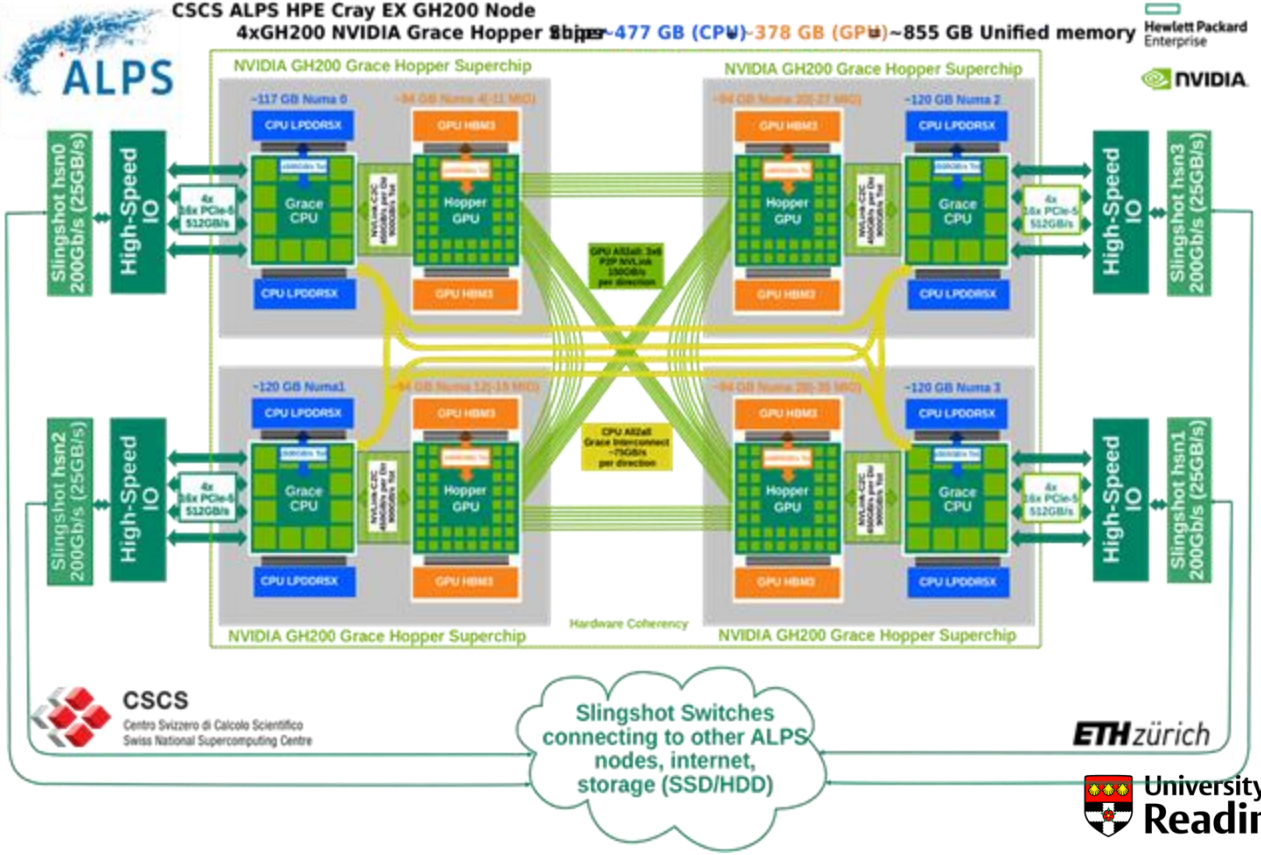




Each node:

- Four “Super-Chips”, each with a Grace CPU and a Hopper GPU.

Can we get a coupled model running on this platform?



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

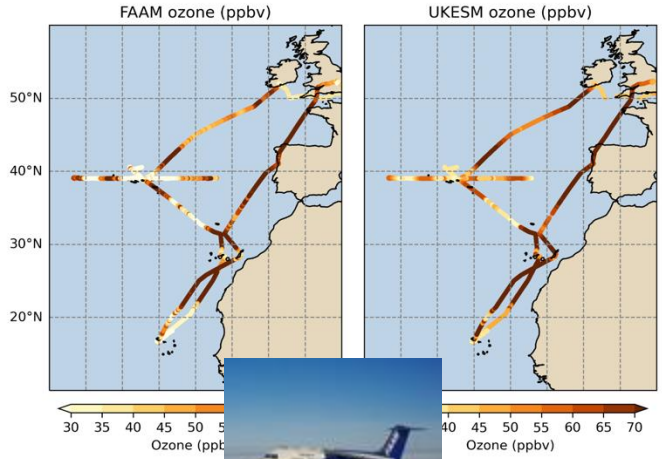
University of Reading

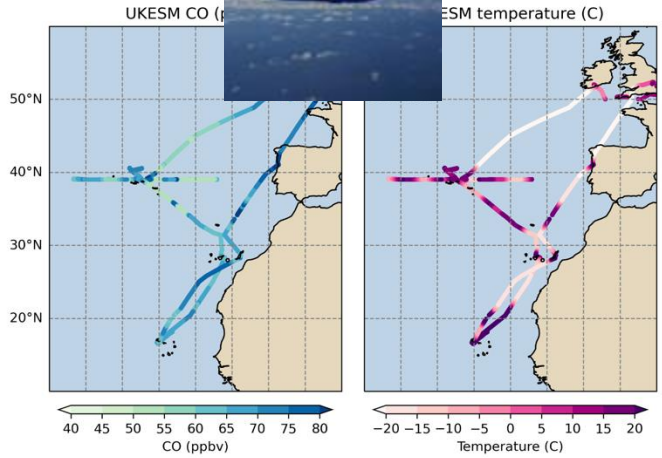
Grace-Hopper: Why, What, and How?

We think N512O12 might fit on this machine and run at around 2SYPD at 384 NHSY if

1. We can get N512 UM atmosphere running on 64 Grace chips of 16 nodes, while
2. The O12 NEMO Ocean is running on the 64 Hopper GPUs, and
3. If we can fit all the required memory on the 16 nodes, and
4. We can make a coupled model with the right version of NEMO, so that we can use
5. a good GPU port of NEMO via PSyclone from NG-ARCH, and
6. We can sort out the XIOS issues on this architecture, and finally:
7. We can find the necessary effort (likely ~ 4 FTE for ~6 months at least).

Much faster and cheaper than CPU version: could consider a large ensemble at this resolution which we could use for training AI systems and …

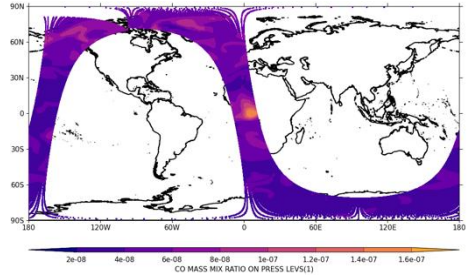


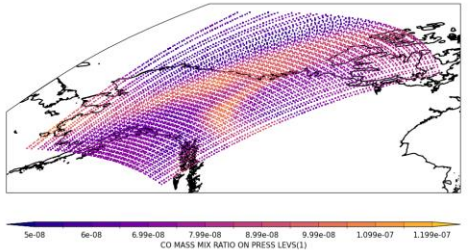


UNIVERSITY OF CAMBRIDGE
Yusuf Hamied Department of Chemistry

VISION: towards seamless integration of Model, Satellite, and In-Situ Observation

- Reduce NERC carbon data emissions by optimising impact and outcomes from the FAAM Airborne Laboratory.
- Improve policy decision making for climate mitigation by providing tools for integrating right-time observational data into Earth System Models.





National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

FAAM

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

UNIVERSITY OF LEEDS

University of Reading

VISION toolkit: interoperable and stable software

- an **accessible** package that is **easy and fast to use**, and is **portable** so can be used by anyone, with *any CF-compliant* model and observational dataset.
- Written in python with **cf-python** and **cf-plot**, data analysis and visualisation libraries developed and maintained within NCAS.
- An important aspect of portability is that the software relies on standardised CF-compliant data from/for the models. We will provide **CF-netCDF usage guidance** to define what this entails.

Climate and Forecasting Conventions

python

cf python

Command Line Interface

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Py[Five]

pyfive.readthedocs.io



Traditionally, accessing HDF5 and NetCDF4 data across networks has been slower than it needs to be for three reasons:

(1) The c-libraries (used by Python and others) which access the data are not thread-safe,

(2) The indexes of chunk locations kept in "b-trees" within the files are potentially distributed through the file, and

(3) The chunking is often inappropriate.

But there are now solutions

(1) We found, and revitalised, a pure python HDF5 reader (pyfive). It is thread-safe, and NCAS-CMS are committed to maintaining it.

(2) The **h5repack** tool can put all the indexes at the front (but either way pyfive can access these efficiently enough).

(3) This can often be fixed at source, otherwise, use **h5repack**!

- `h5netcf` and `xarray` support underway
- already in `cf-python`

No longer a *need* to use zarr or kerchunk to access data on object stores

Integral to another new tool: PyActiveStorage, providing remote reductions on HDF5 data.

Bonus: from a curation point of view, we are no longer dependent on a private US company to read our data in the future!

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Take Home

- Climate Science can use all the available computing, and there are good scientific reasons to continue to grow our use of HPC.
  - (I haven't said anything much about Weather as we don't have time to discuss all the AI implications.)
  - (For climate science, the implications of AI are not as earth shattering as they are for weather. That's the topic of another talk.)
- But this leads to problems:
  - Our models need porting an optimizing.
  - The workflows are complex and need support.
  - Big teams are necessary to do big science. (But there can be, and are in this case, hundreds of downstream users.)
  - The data volumes and tools are challenging. Importance of standards and shared libraries.
  - Even basics like our interfaces to well established data formats need continual maintenance.
- CMS has been around for a couple of decades and continues to support the UK academic weather and climate modelling community.

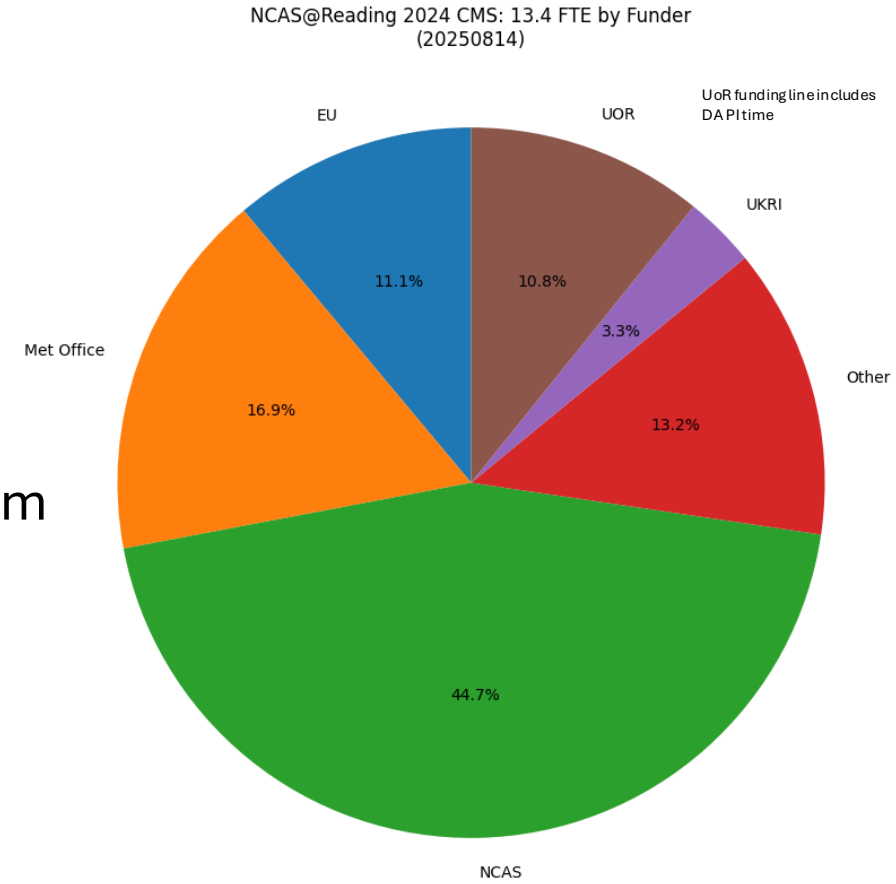
That’s all folks!

CMS is supported by:

- NCAS National Capability
- NCAS Institutional Funding
- The University of Reading

With grant and project funding from

- The European Commission
  - ESIWACE Series
  - IS-ENES Series
  - … more to come
- The Met Office
- UKRI



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading