CANARI

The science and logistics of the CANARI large ensemble: big science to investigate future weather in your backyard!

Bryan Lawrence

(and a cast of thousands, some of whom I'll introduce as we go)

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

What is CANARI?

CANARI is a 5-year research programme dedicated to understanding how climate variability and change in the Arctic-North Atlantic region will impact on the UK, with a focus on extreme weather and rapid, disruptive change.

We'll get to "What is a large ensemble"!

Outline

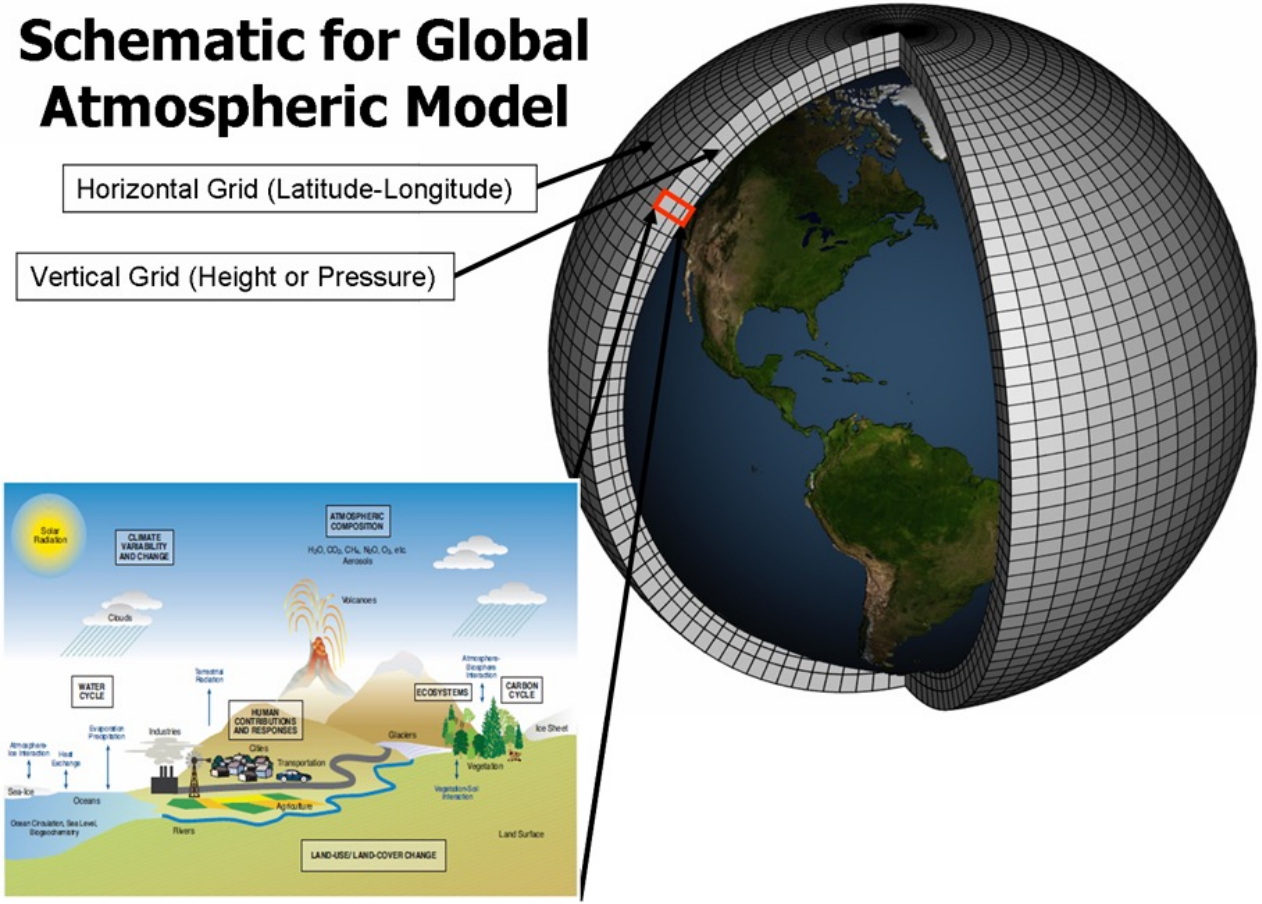
- Part A:
  - A modelling primer
  - Uncertainty and large ensembles
- Part B: CANARI
  - The large ensemble, design and logistics
  - Some early results.

Summary

- Going beyond CANARI to other relevant NCAS programmes.

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)*!



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Modelling 101

State Variables:

u,*v*,*w* — wind

Π —non-dimensional pressure

Θ — Potential temperature

Coordinates:

r, ϕ, λ — radial position, lat, lon

Things that can cause change:

D/*Dt*— time derivative following motion

S — Forcing (External and parametrised)

Objective is given knowledge of the external forcing *S* and the state (*u*, *v*, *w*, Π, Θ) at time *t* to advance knowledge of the state to time $t + \Delta t$, where Δt is the **timestep**.

Newton's second law

$$\frac{D_r u}{Dt} - \frac{uv\tan\phi}{r} - 2\Omega\sin\phi v + \frac{c_{pd}\theta}{r\cos\phi}\frac{\partial\Pi}{\partial\lambda} = -\left(\frac{uw}{r} + 2\Omega\cos\phi w\right) + S^u$$

$$\frac{D_r v}{Dt} + \frac{u^2\tan\phi}{r} + 2\Omega\sin\phi u + \frac{c_{pd}\theta}{r}\frac{\partial\Pi}{\partial\phi} = -\left(\frac{vw}{r}\right) + S^v$$

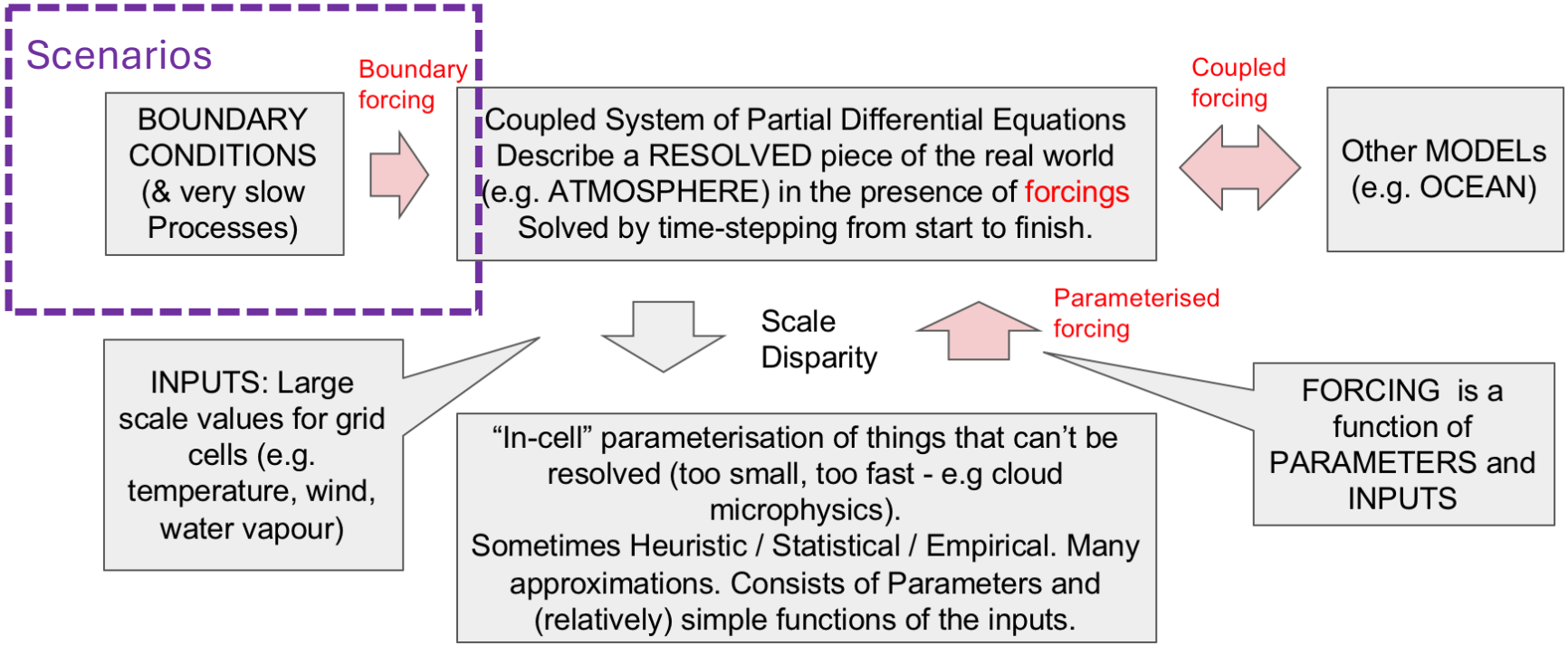
$$\frac{D_r w}{Dt} + c_{pd}\theta\frac{\partial\Pi}{\partial r} + \frac{\partial\Pi}{\partial r} = \left(\frac{u^2+v^2}{r}\right) + 2\Omega\cos\phi u + S^w$$

mass continuity

$$\frac{D_r}{Dt}\left(\rho_d r^2\cos\phi\right) + \rho_d r^2\cos\phi\left[\frac{\partial}{\partial\lambda}\left(\frac{u}{r\cos\phi}\right) + \frac{\partial}{\partial\phi}\left(\frac{v}{r}\right) + \frac{\partial w}{\partial r}\right] = 0$$

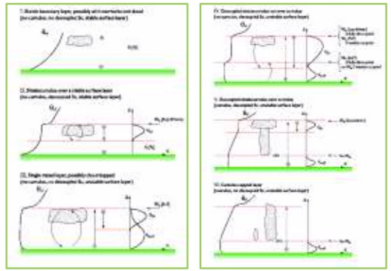
We run climate models with thousands of grid points, with tens to hundreds of variables, solving these equations using timesteps of a few minutes, for durations of decades – you do the math!

One slide intro to climate modelling

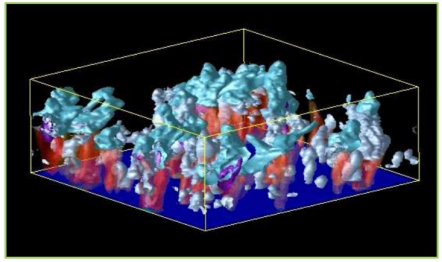


Many of the forcing terms come from parameterisations

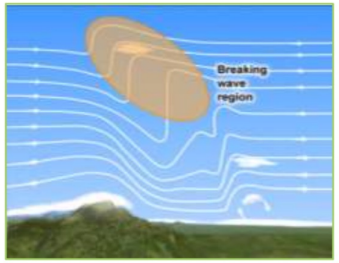
Boundary layer turbulence and mixing



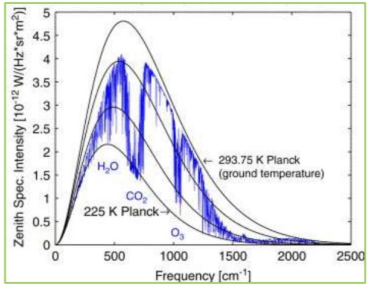
Cumulus convection



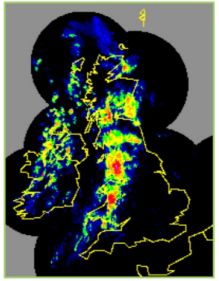
Effects of mountains



Radiation



Precipitation



Clouds and microphysics



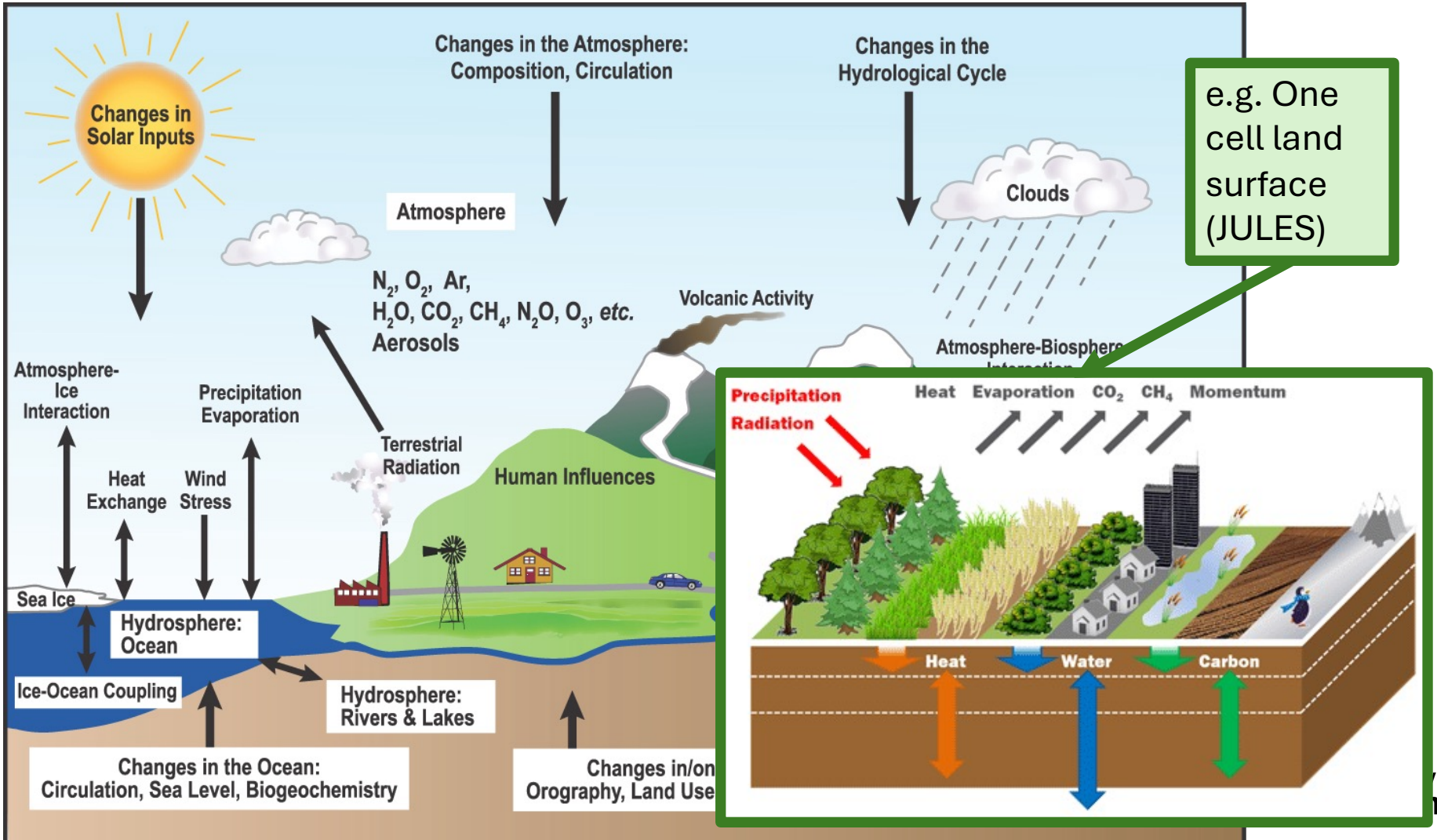
Atmospheric composition



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

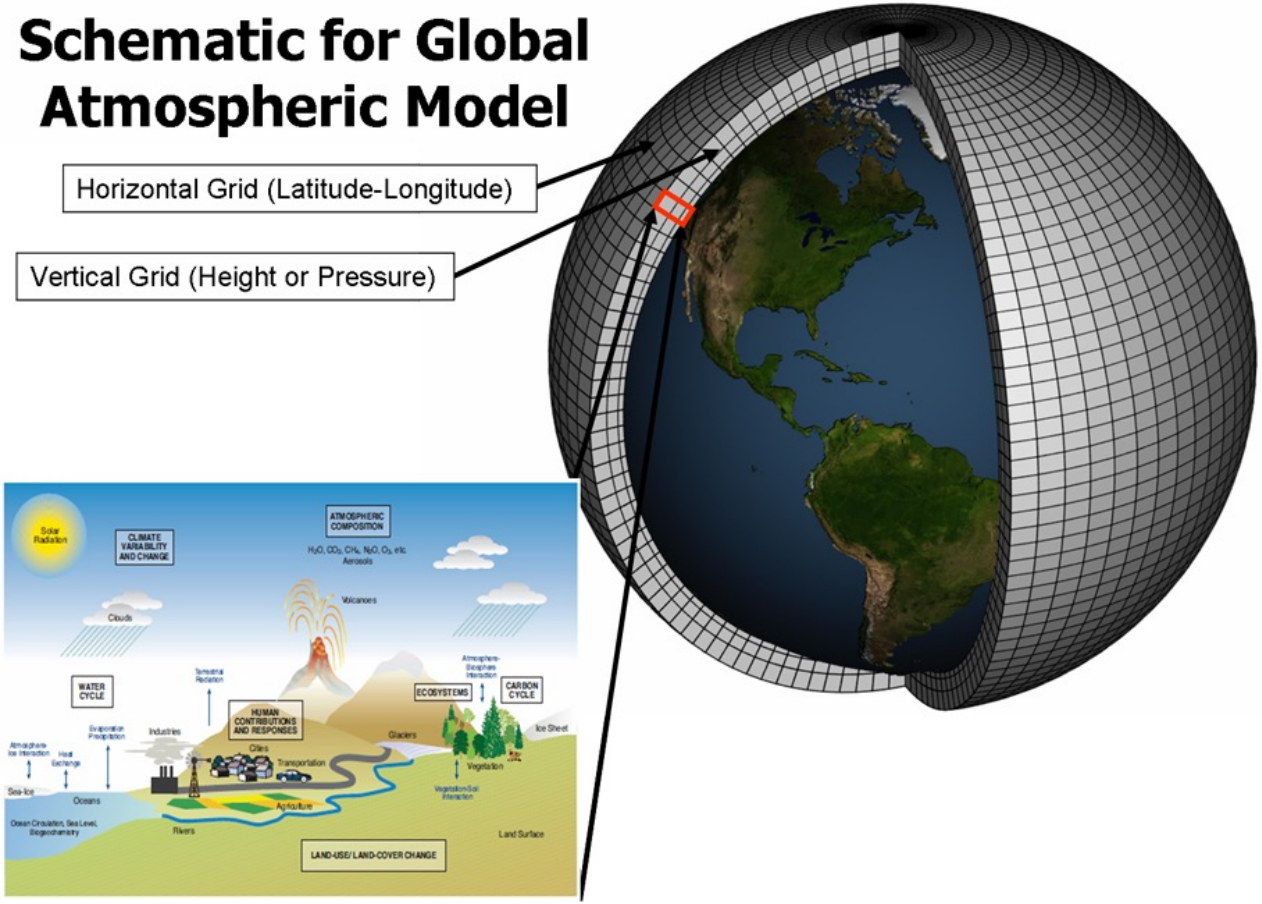
Slide Images from Slingo, 2013

University of Reading


Changes in Solar Inputs
Changes in the Atmosphere: Composition, Circulation
Changes in the Hydrological Cycle
Atmosphere
Clouds
N_2, O_2, Ar,
H_2O, CO_2, CH_4, N_2O, O_3, *etc.*
Aerosols
Volcanic Activity
Atmosphere-Biosphere
Atmosphere-Ice Interaction
Precipitation Evaporation
Heat Exchange
Wind Stress
Terrestrial Radiation
Human Influences
Sea Ice
Hydrosphere: Ocean
Ice-Ocean Coupling
Hydrosphere: Rivers & Lakes
Changes in the Ocean: Circulation, Sea Level, Biogeochemistry
Changes in/on Orography, Land Use
e.g. One cell land surface (JULES)
Precipitation
Radiation
Heat
Evaporation
CO_2
CH_4
Momentum
Heat
Water
Carbon

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

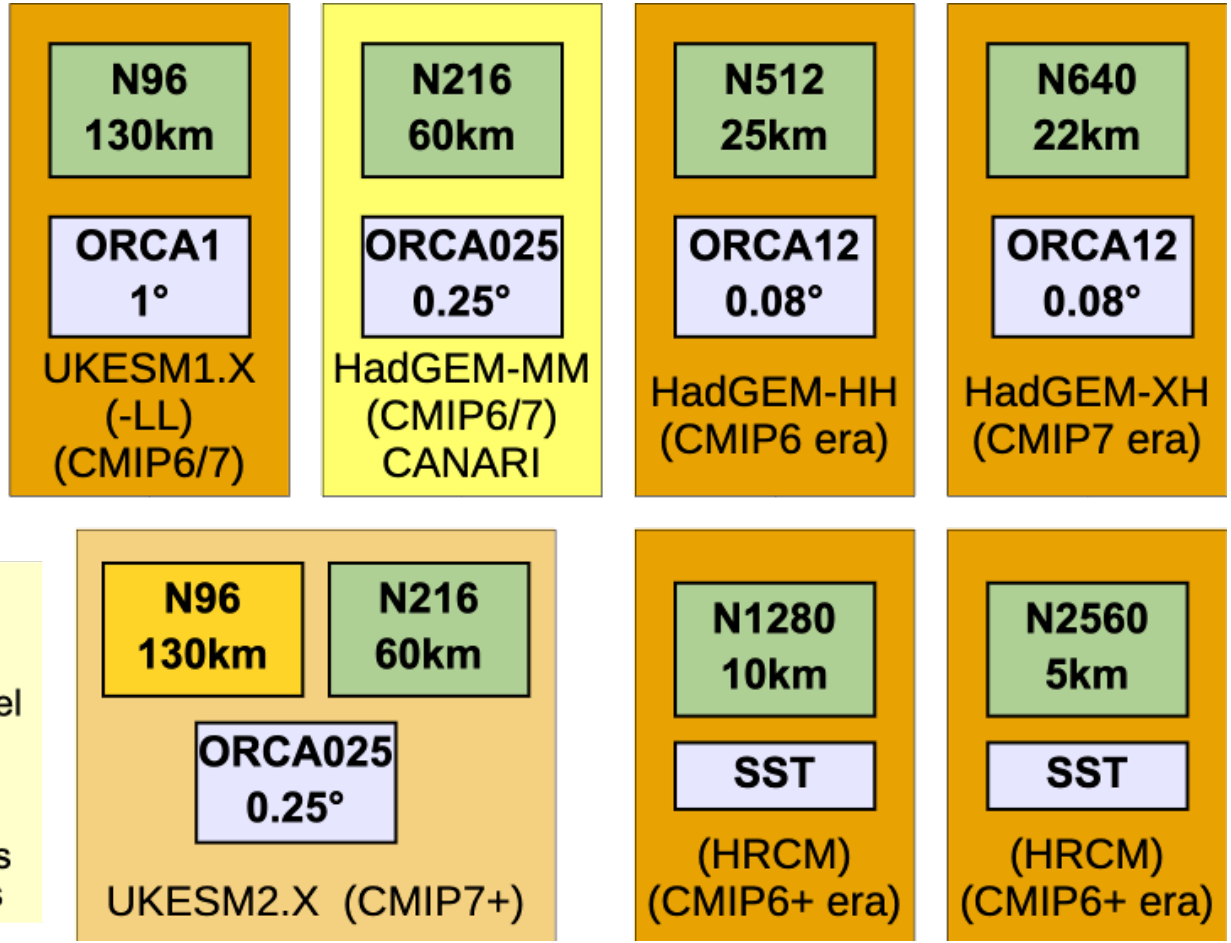
Many points, much state, integrated for years with timestep of *o(minutes)*!



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

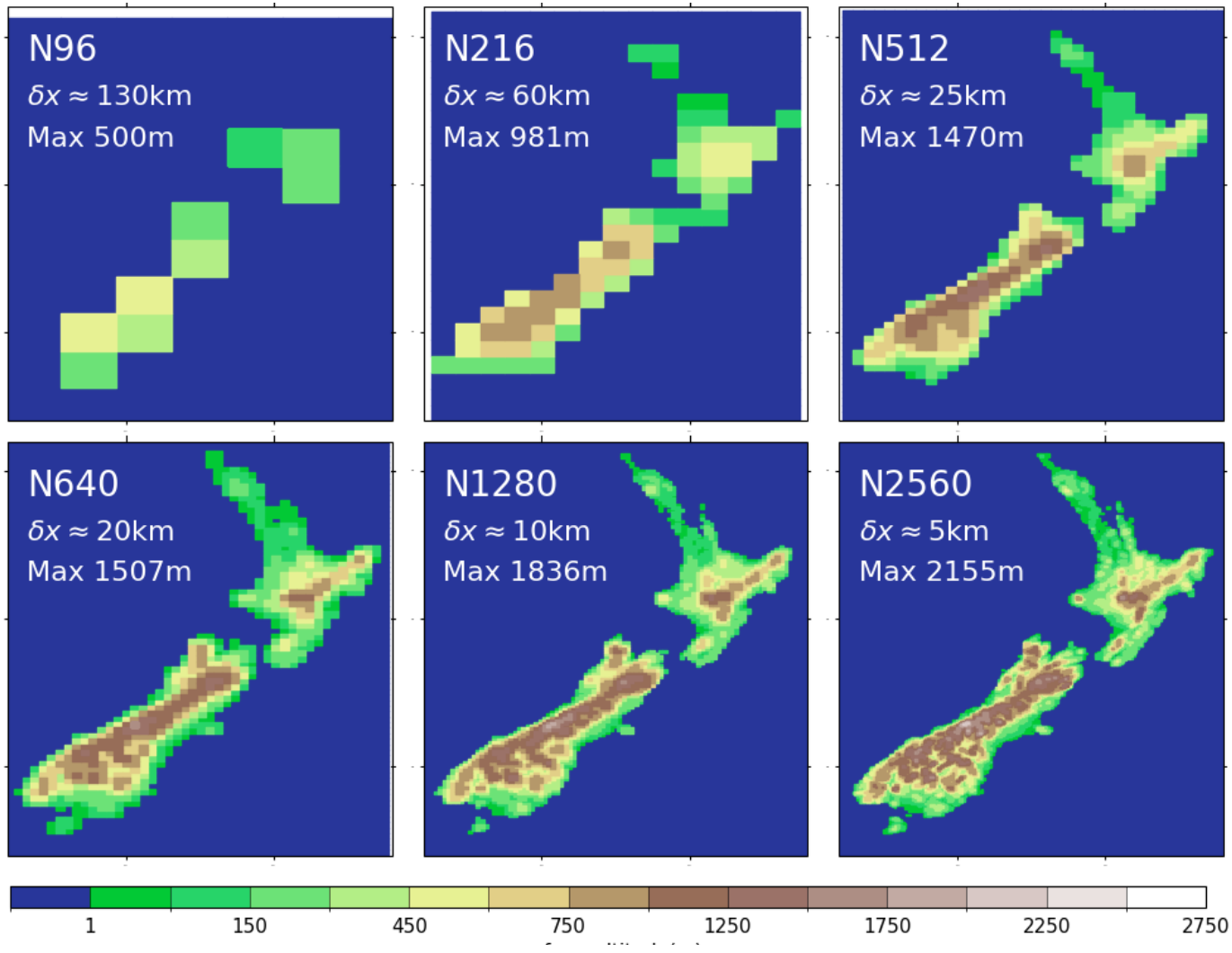
The UM Global Model Collection (2025)



Essentially* the same physics & dynamics used throughout model hierarchy

*Physics evolves with generations

Attempting a visceral feeling for resolution

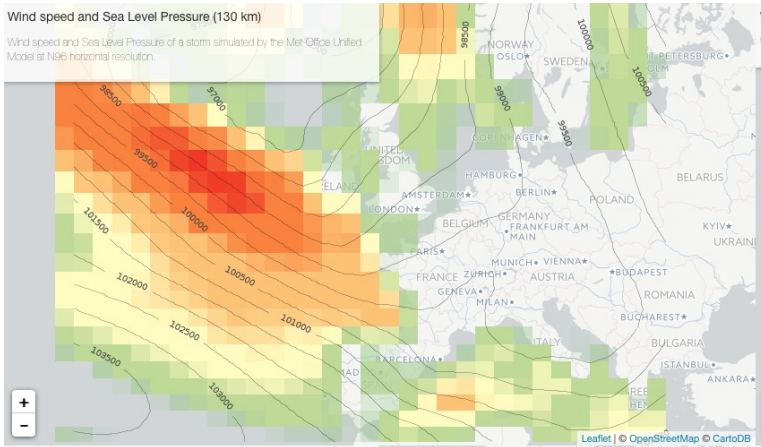


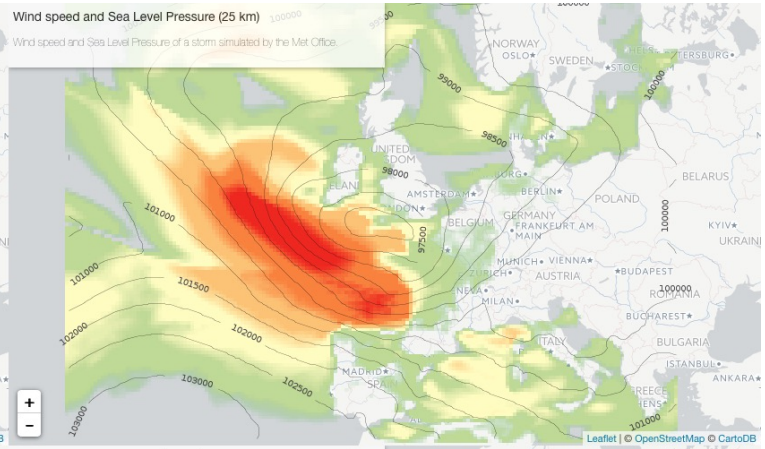
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

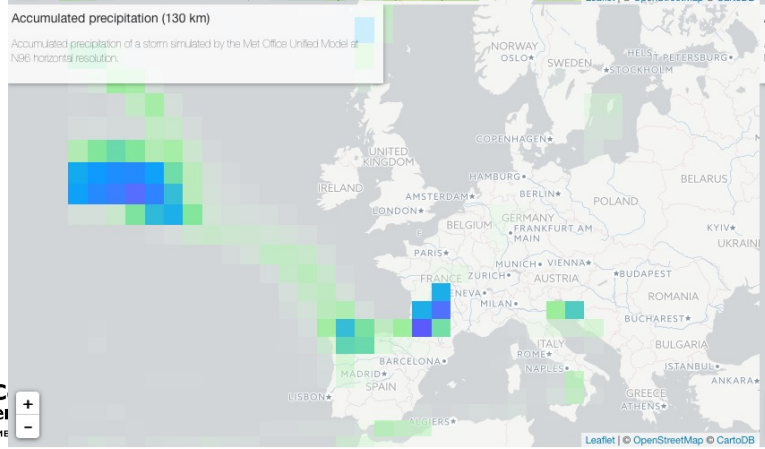
The influence of resolution

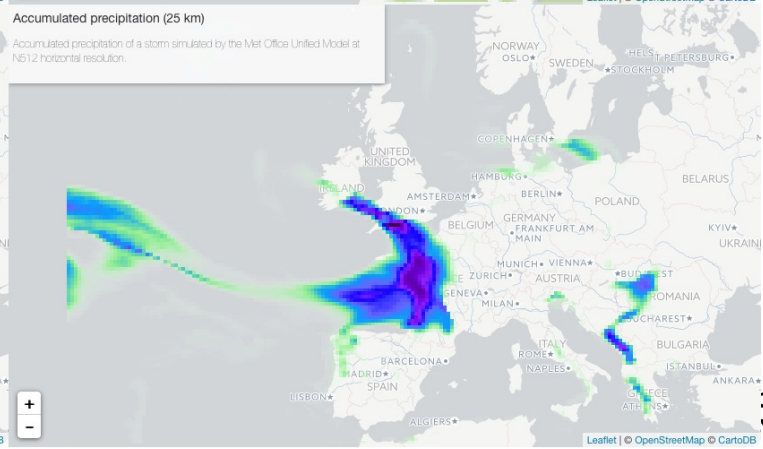


P-L. Vidale
(PI NCAS
HCRCM)









Model Resources

- CANARI simulations run on ARCHER2 at ~2 simulated years/per day (SYPD, more like <1 actual SYPD after queuing).
  - 16 atmosphere nodes
  - 9 ocean nodes
  - 7 IO server nodes (!)
  - Writing 5 GB/node-hour

  i.e using 384 Node Hours per simulated year (NHSY), producing 2 TB per model year (but we don't keep it all)
- The model has little more scaling, so we can't go faster, so 150 years is doomed to take at least 6 months … at best … and we can't keep all 40 simulations in the machine at the same time (otherwise we'd have quarter of the machine for half a year, we'd not be popular!)
- cf N1280 (no ocean) 7200 NHSY, about four times slower (and using a lot more nodes – 150!!).

archer2

5,860 nodes
128 core/node
(Physically in Edinburgh.)

JASMIN

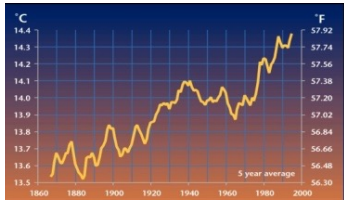
Lotus2
55,000 cores
~ 60 PB Disk
(Physically near Oxford)

Give me more compute?

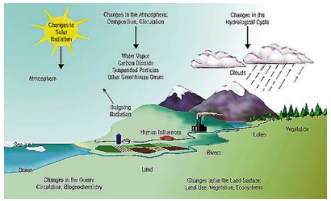
WE can't have it all!

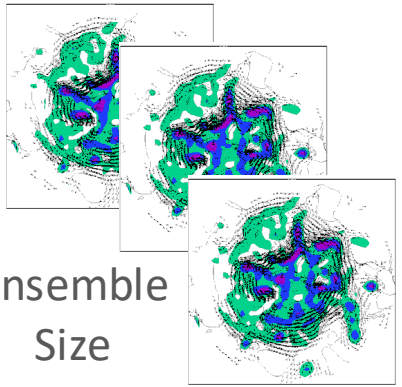
WE need to spend compute on the key aspects of the problem.

Which are?



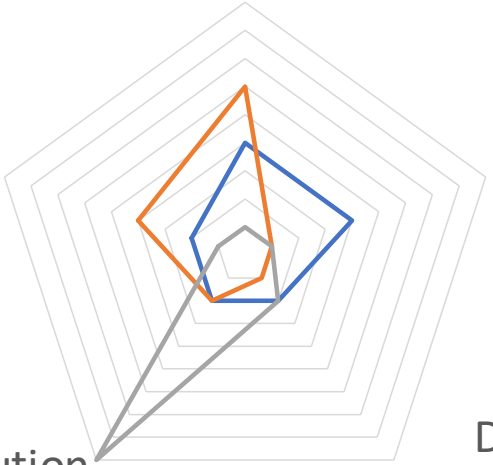
Duration



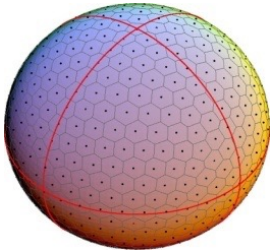


Ensemble Size

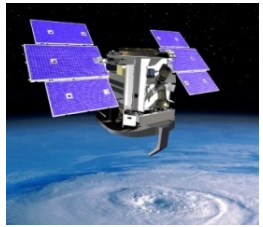
Complexity



Resolution



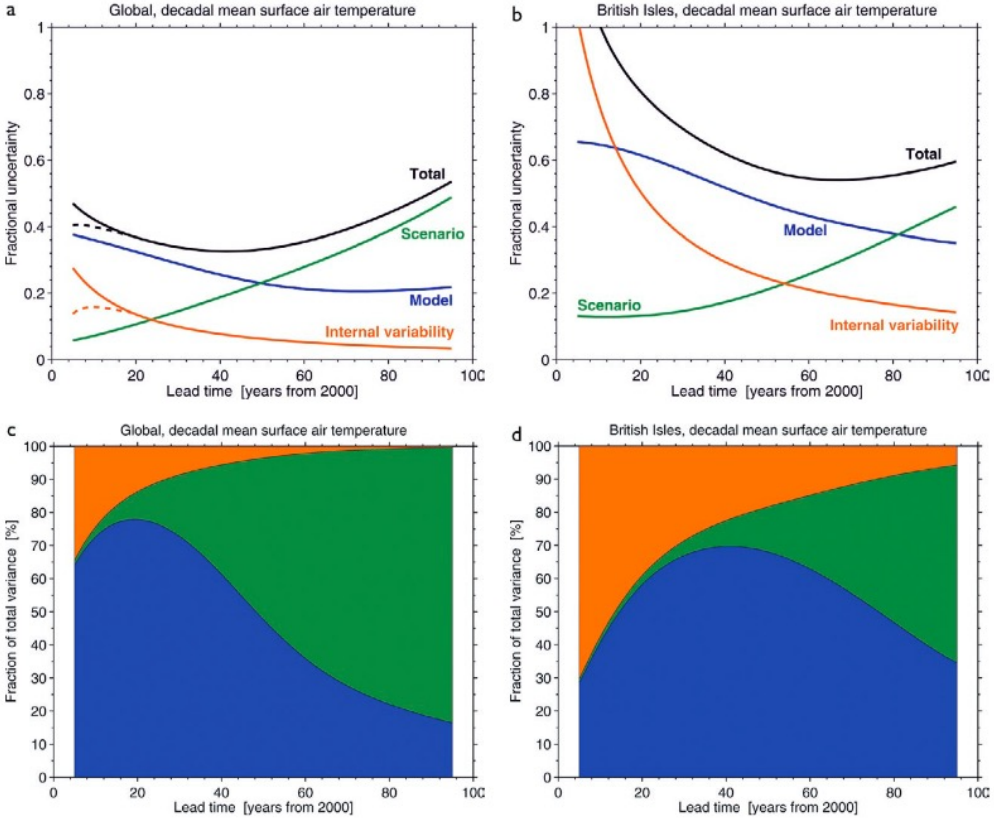
Data Assimilation



Variability Scenarios Extremes

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Uncertainty



But

- we (humans) don't experience "Decadal Mean Surface Temperature", we experience events;
- the ecosystem experiences both extreme events and integrated changes.
- we believe that most of the uncertainty in extremes in the next few decades arises from a combination of **seen and unseen** climate states and model uncertainty.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

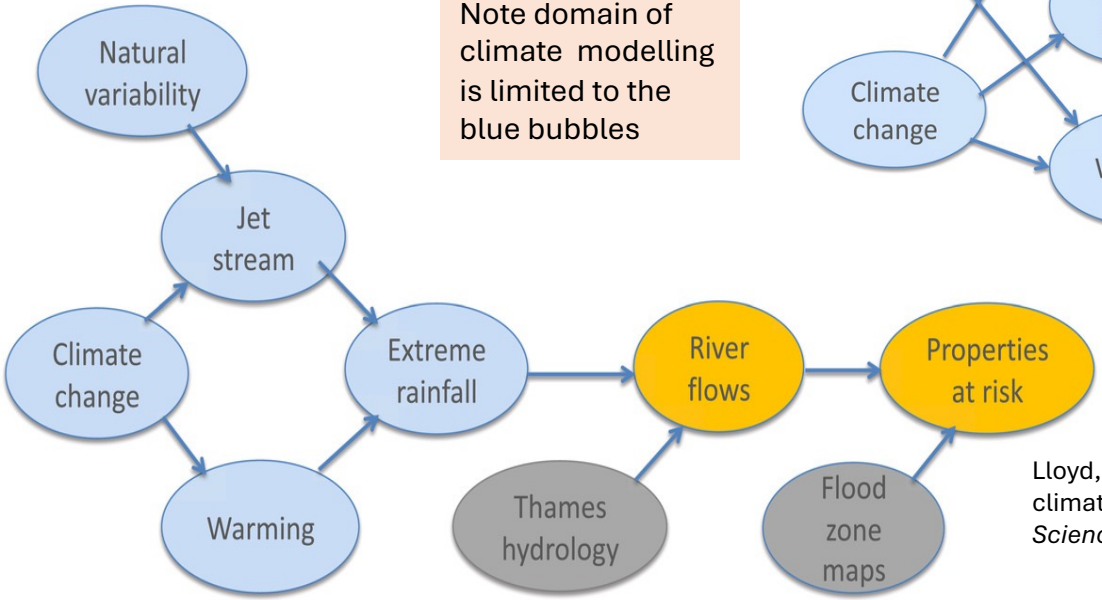
Hawkins, E., & Sutton, R. (2009). The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1107. https://doi.org/10.1175/2009BAMS2607.1

University of Reading

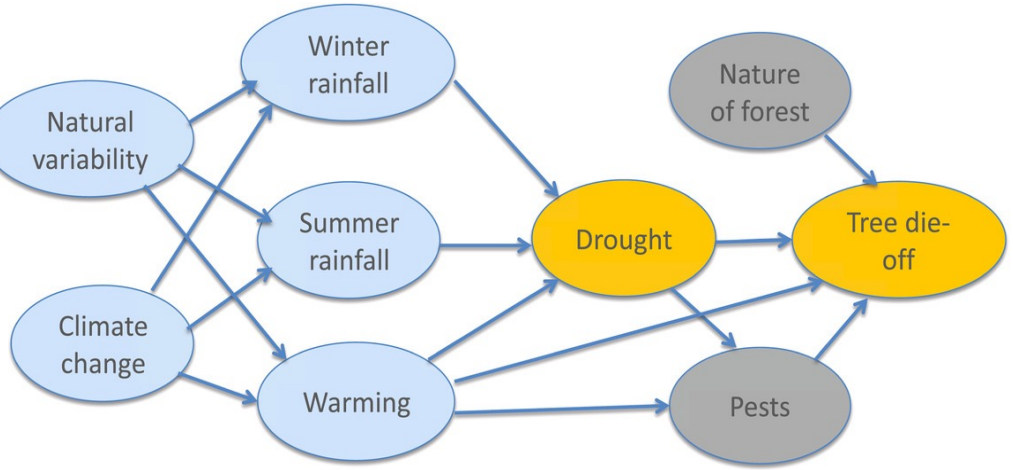
Storylines – going beyond statistics & climate modelling

Two examples:

Thames Valley Flooding



Note domain of climate modelling is limited to the blue bubbles



Tree Die-off

Lloyd, E. A., & Shepherd, T. G. (2020). Environmental catastrophes, climate change, and attribution. *Annals of the New York Academy of Sciences*, *1469*(1), 105–124. https://doi.org/10.1111/nyas.14308

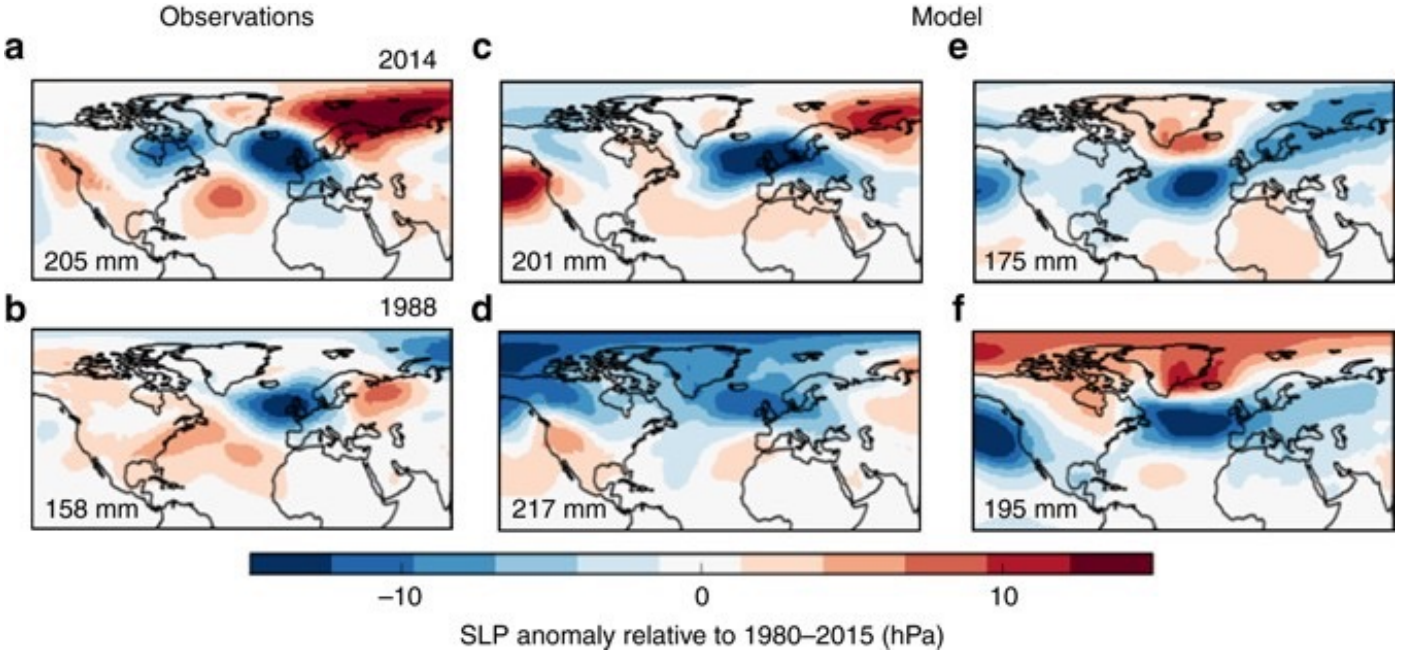
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

The seen and unseen climate

Exploiting global ensemble prediction systems to investigate seasons and extremes that COULD have occurred.

Interesting question as to importance of UPSCALE effects. Could these have been seen in RCM simulations?



Observed and not yet realised climate states. [a, b]: The sea level pressure anomaly fields (in hPa, relative to the January mean sea level pressure field) of the two observed Januarys with highest rainfall totals, 2014 and 1988, using the ERA data set. [c-f]: The sea level pressure anomalies of four extreme simulated Januarys, one of which [d] presents a potential new record rainfall scenario

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S., & Belcher, S. E. (2017). High risk of unprecedented UK rainfall in the current climate. *Nature Communications*, *8*(1), 107. https://doi.org/10.1038/s41467-017-00275-3

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

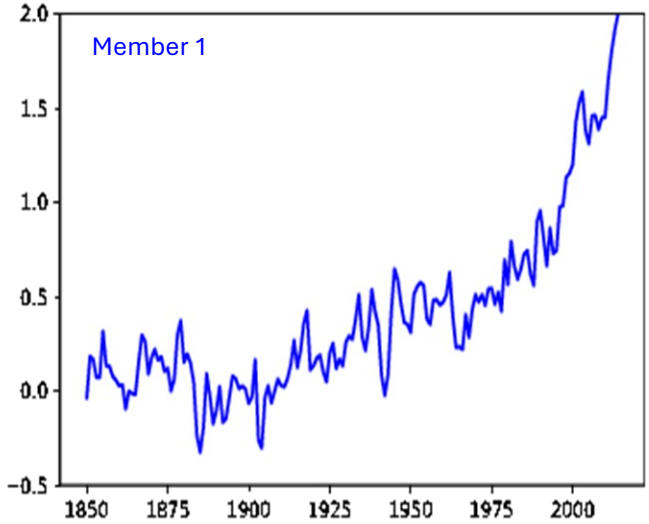
What is a large ensemble?

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

Global mean temperature from a 50-member CanESM5 ensemble



Andrea Dittus



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

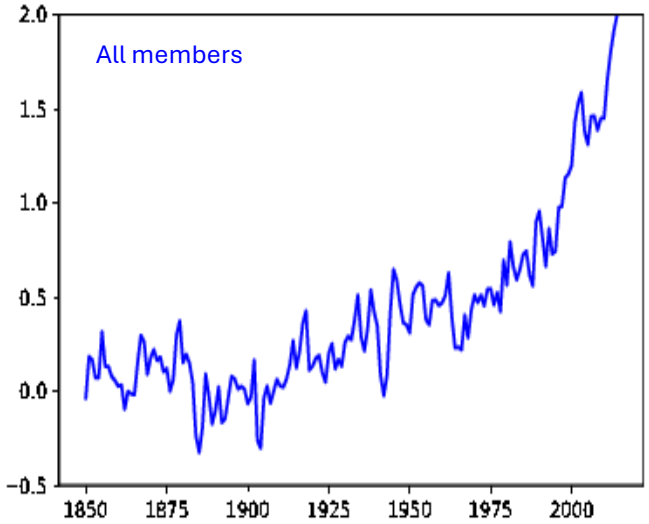
What is a large ensemble?

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

Global mean temperature from a 50-member CanESM5 ensemble



Andrea Dittus

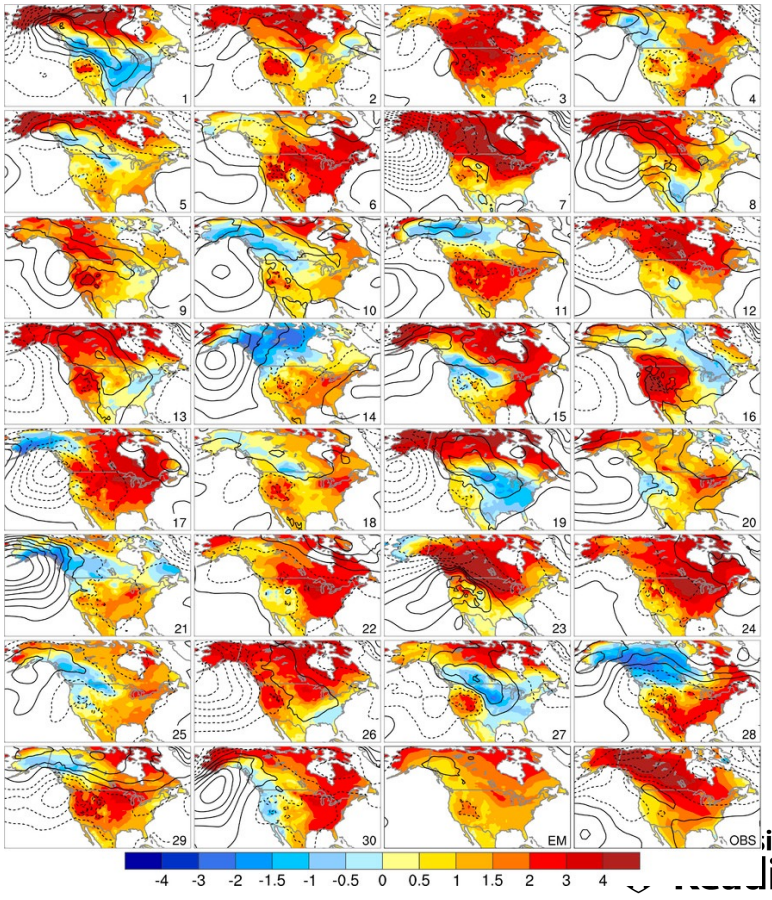


What is a large ensemble?

- 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*



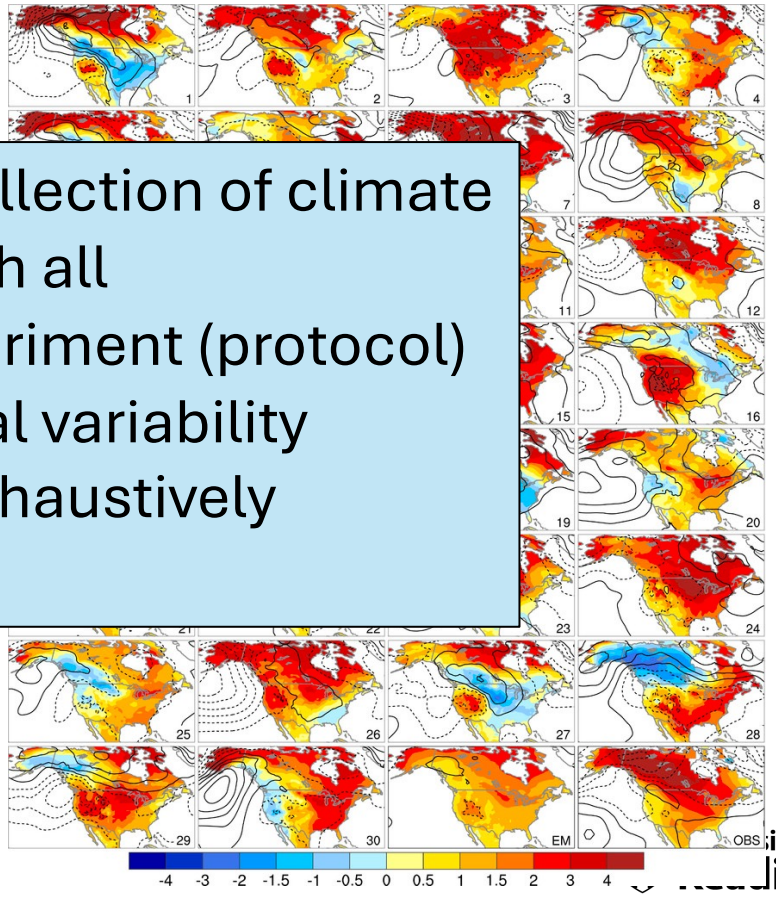
What is a large ensemble?

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



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

A NERC NC Multi-Centre Science Programme



- CANARI PI: Len Shaffrey (NCAS)
- Apr 2022 – Mar 2027
- Further info: canari.ac.uk

Aim: Advance understanding of the impacts on the UK arising from climate variability and change in the Arctic-North Atlantic region, with a focus on extreme weather and the potential for rapid and disruptive change.

- Inland flooding
- High temperatures and health
- Drought and water supply
- Extreme winds and wind damage
- Changes in UK Shelf Seas





Reinhard Schiemann (Large Ensemble)



Ben Harvey (UK Regional)



WP1 Simulations

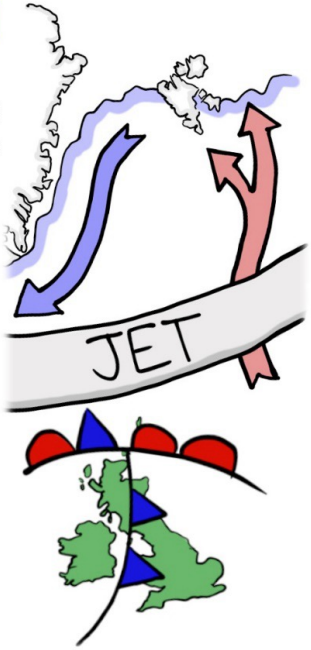
- N Atl-Arctic Ocean-Ice
- N640 High Resolution
- N216 Large Ensemble
- 2.2km European
- 1.5km Sea Shelf

Processes

WP4 Arctic Change
Sea ice, ocean mixing and transport, clouds stratosphere, mid-latitude pathways

WP3 Large-scale Flow and Weather
Eddy-mean interact-ions, large-scale drivers, s2d forecasts

WP2 UK Impacts
Flooding, extreme winds, heatwaves, droughts, and sea shelves



WP5 Knowledge Integration

Process Understanding
combined with
User Knowledge

Key uncertainties in climate projections to inform co-created **Storylines**

Collaborators
- Met Office
- Universities
- Other LTSM2 projects

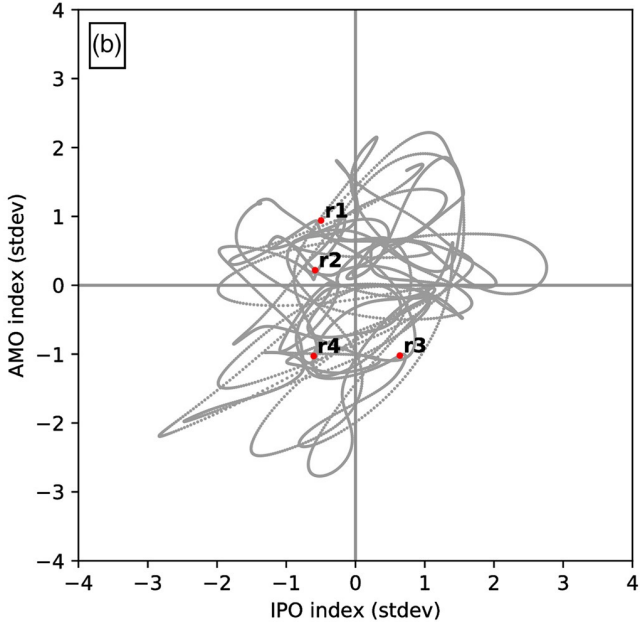
Stakeholders
- Policy-makers
- International programmes
- Business
- Media/Public

Global HadGEM3 model
Regional HadGEM3 model
Regional Ocean Model

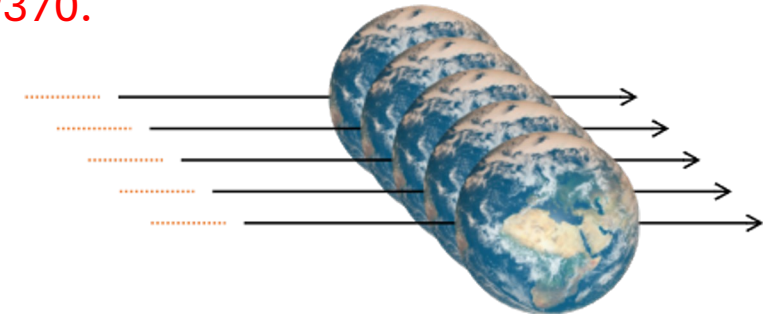
A Suite of Community Simulations
- Produced on archer2
- Stored at JASMIN
- Available for wider community use

Initialising and running a large ensemble





1. (CMIP6: 500-year picontrol experiment)
2. Run 8 “macro” historical experiments from differing initial ocean states for 1850-1950.
3. Make 5 perturbations of the 1950 atmospheric state.
4. Continue 40 historical simulations from 1950 to 2014, and
5. Continue 40 future simulations onward using SSP370.



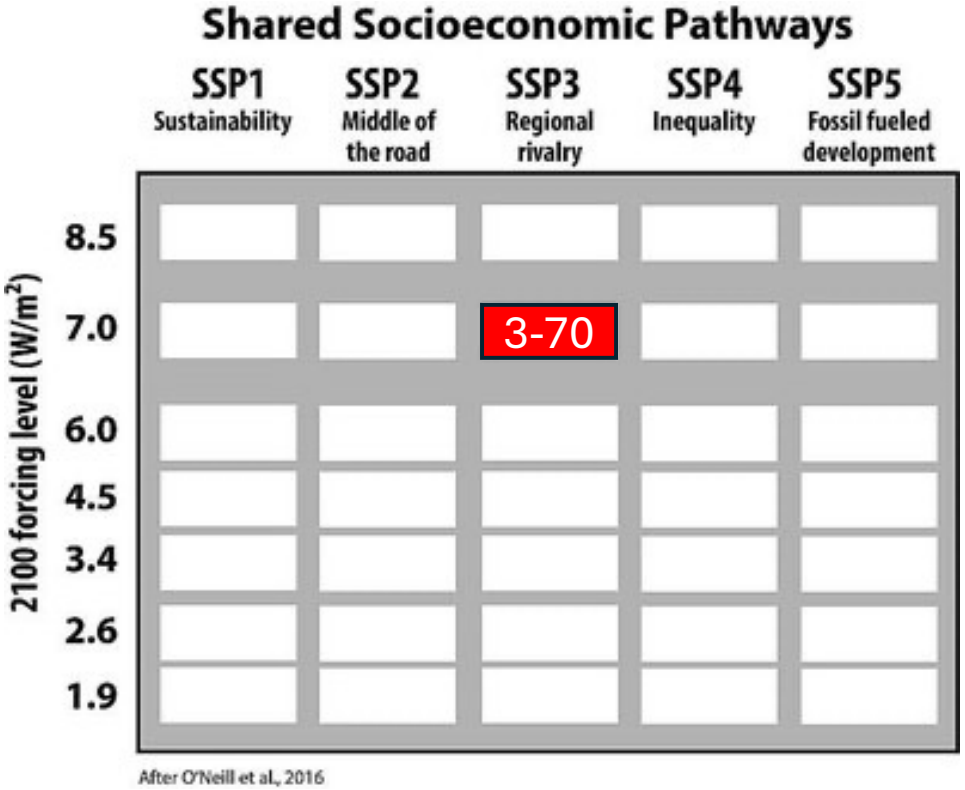
HadGEM3.1-MM piControl: Atlantic Multidecadal Oscillation (AMO) and the Interdecadal Pacific Oscillation (IPO).

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Scenarios

SSP3-7.0 (Shared Socioeconomic Pathway 3, 7.0 W/m^2) represents a high-emission future scenario:

- Characterized by regional rivalry and slow economic growth.
- High fossil fuel reliance and weak climate policies.
- Projected warming of 2.8–4.6°C by 2100.



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

CANARI Regional Model: Motivation



Convection-permitting regional climate models (CPMs/RCMs) add value over coarser resolution GCMs:

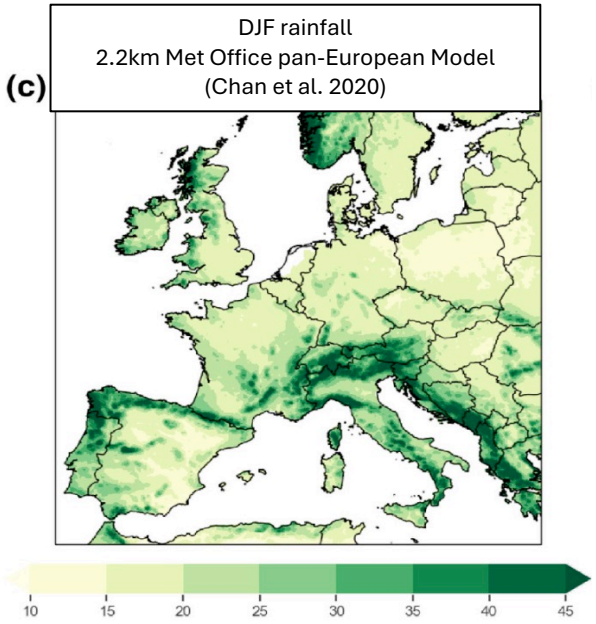
- Better representation of hourly rainfall distributions (Berthou et al, 2020)
- Better representation of mesoscale wind features associated with extratropical cyclones (Manning et al, 2022)

They also exhibit modified future projections:

- Enhanced intensification of rainfall extremes (Chan et al. 2020)
- Enhanced intensification of wind extremes (Manning et al, 2022)

However, these studies are based on short (~10 year) time slice experiments and have limited outputs (due to cost of RCM)

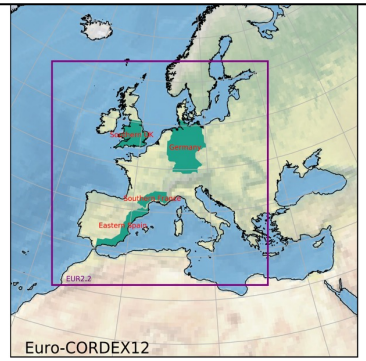
➢ There is a need to better sample present-day and future extreme events as simulated by CPMs



CANARI Regional Model: Experimental Design

- Identify the most extreme UK storms in the CANARI Large Ensemble
- Downscale with the Met Office 2.2km pan-EU regional model (atmosphere only, within a 12km nested domain)
- Case study approach allows for extensive model output for detailed study of atmospheric dynamics and impact modelling
- Targets: ~50 present-day and ~50 future storms for each of
    - Extreme winds
    - Extreme precipitation associated with
        - Extratropical cyclones (winter frontal rainfall)
        - Slow-moving cut-off lows (summer convective rainfall)

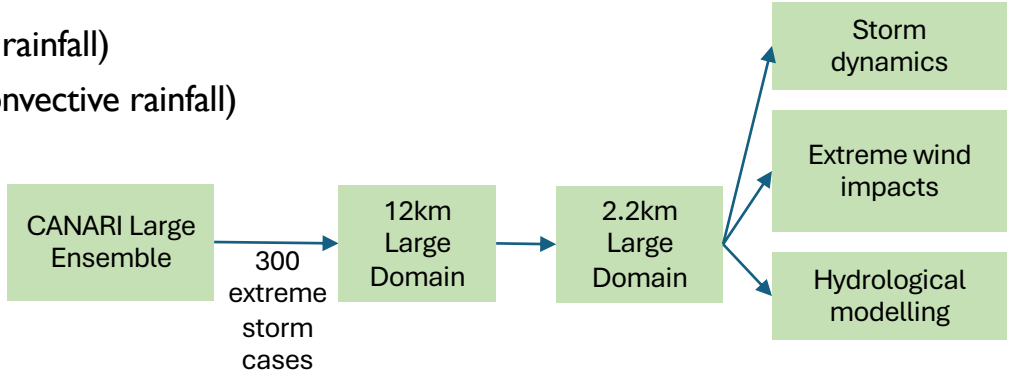
Domains for nested 12km and 2.2km models (Chan et al. 2020)



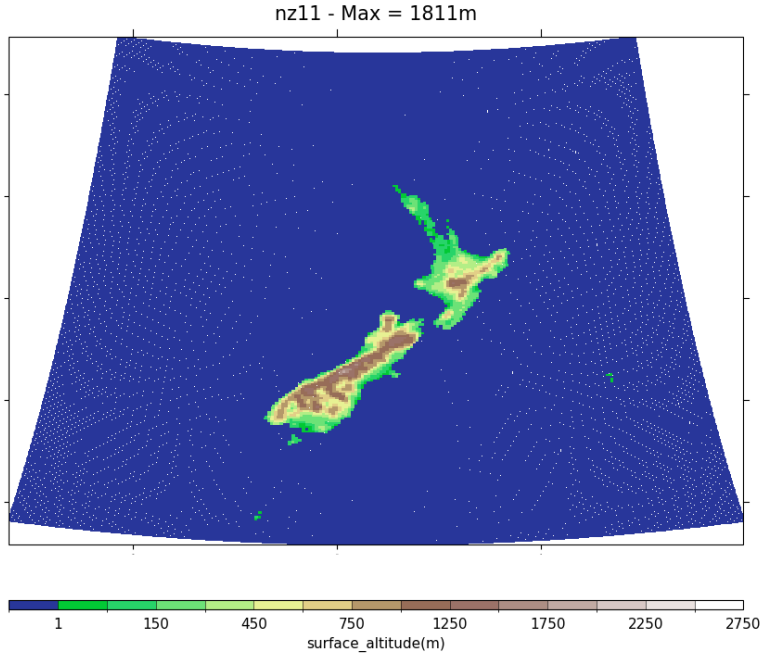


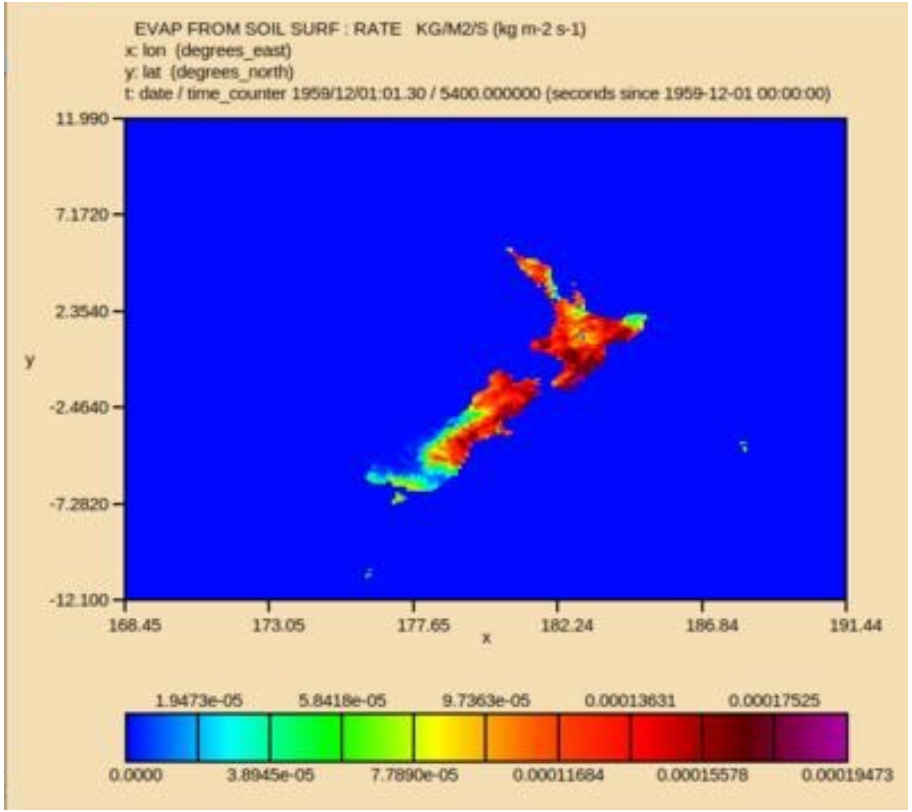


Oscar Martínez-Alvarado



Beyond Europe

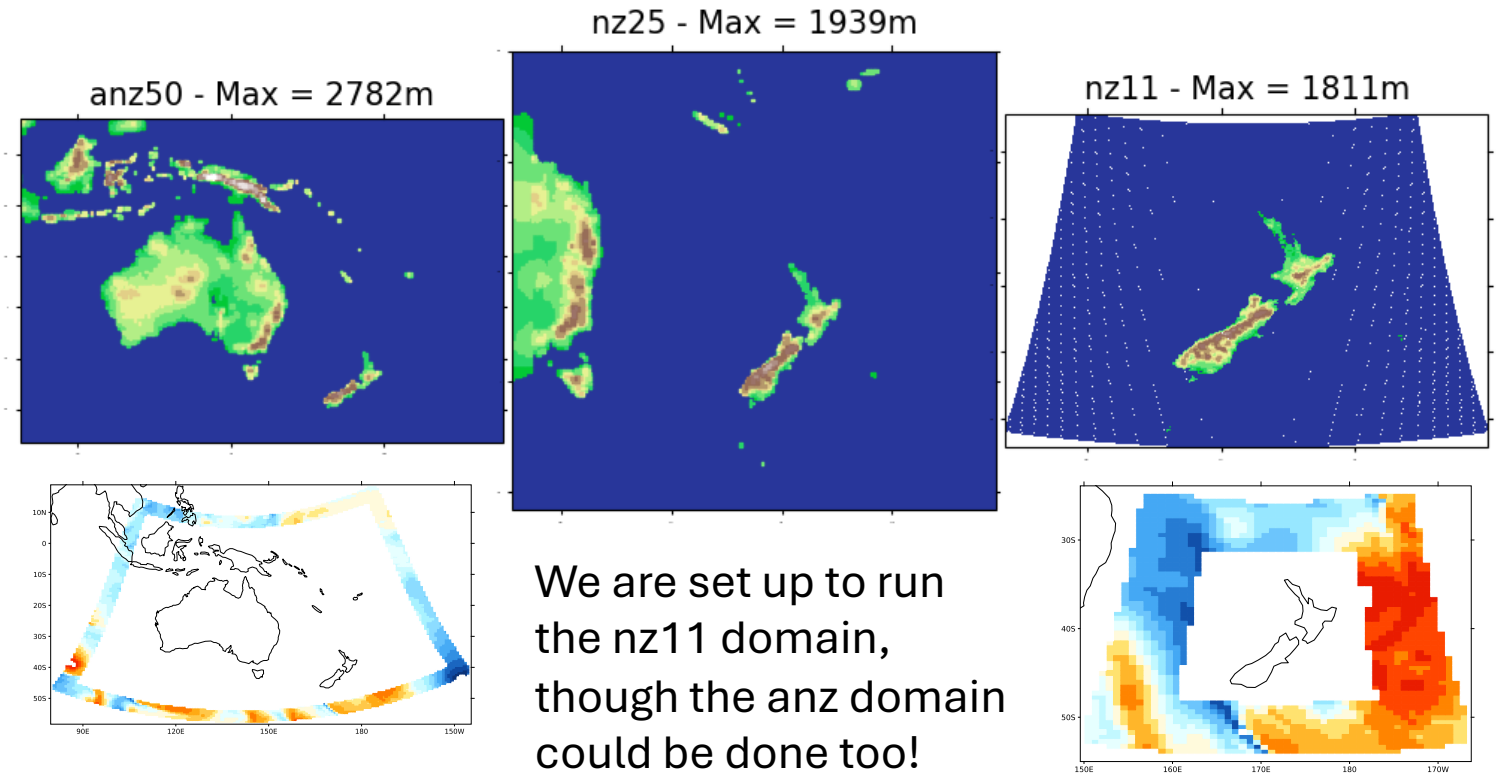




Timestep forcing + Monthly start dumps for all 6000 years!

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

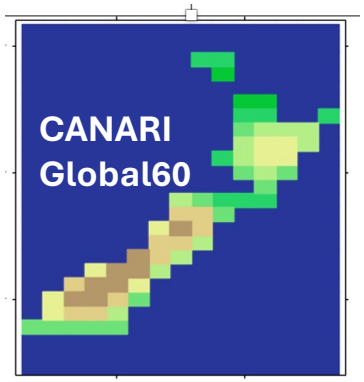
University of Reading

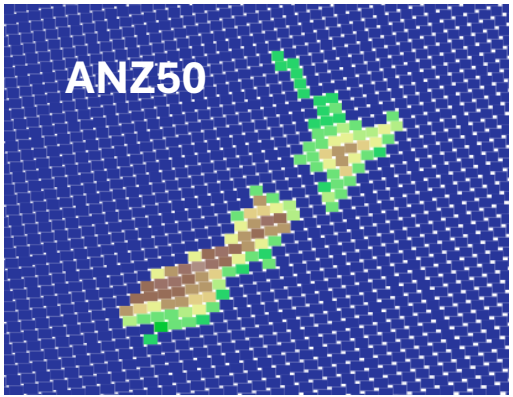


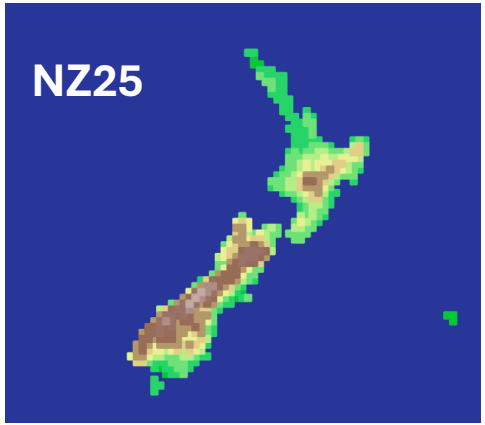
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

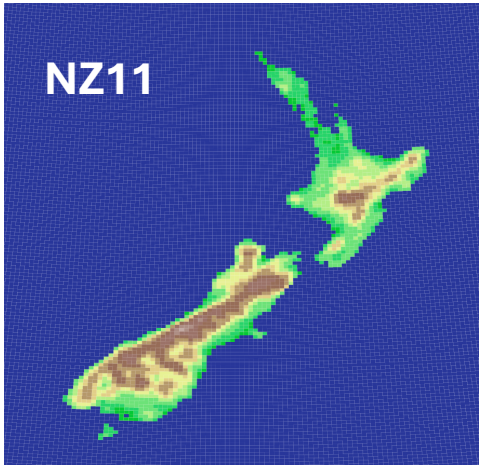
University of Reading

Current view of NZ on these domains



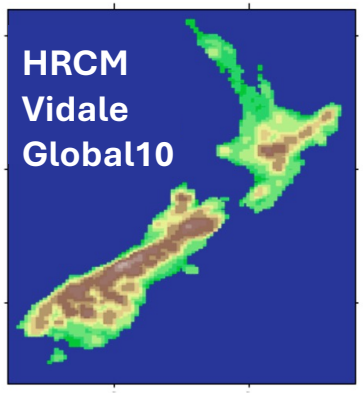






(ANZ50 and NZ25 existing WAH – Rosier et al)

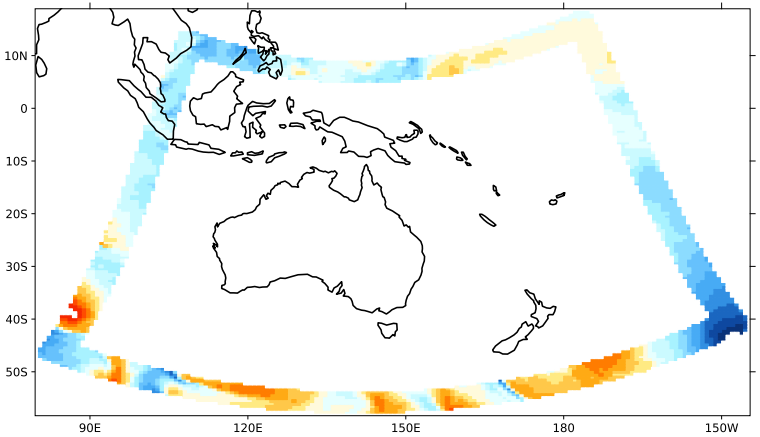
Entirely feasible to nest even higher resolution (2/5 km?) inside ANZ or NZ domains!

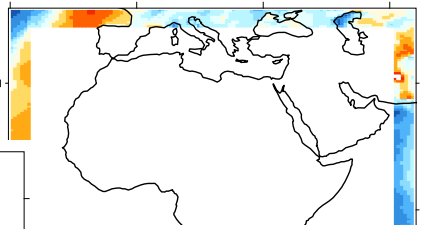


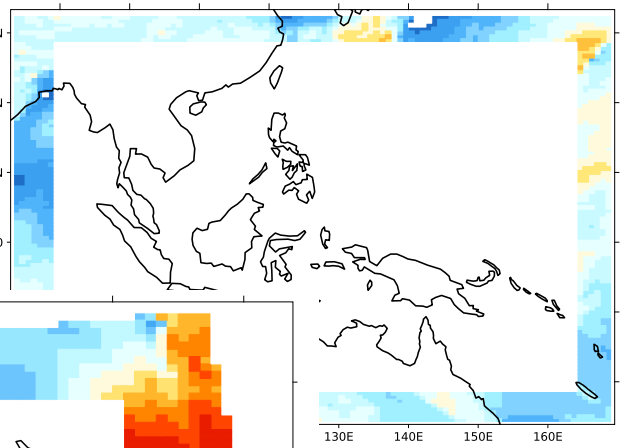
(Recall Global10 uses 150 nodes, cf 2 nodes for NZ11 and they run at roughly the same speed in SYPD.)

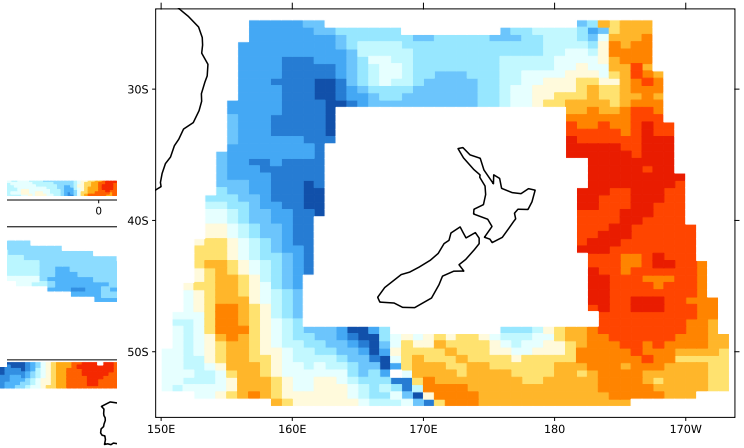
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

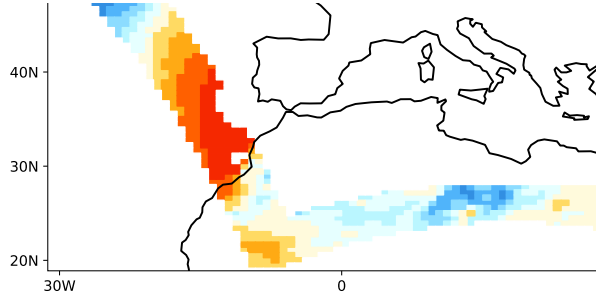
Many regional frames (Not all shown here)

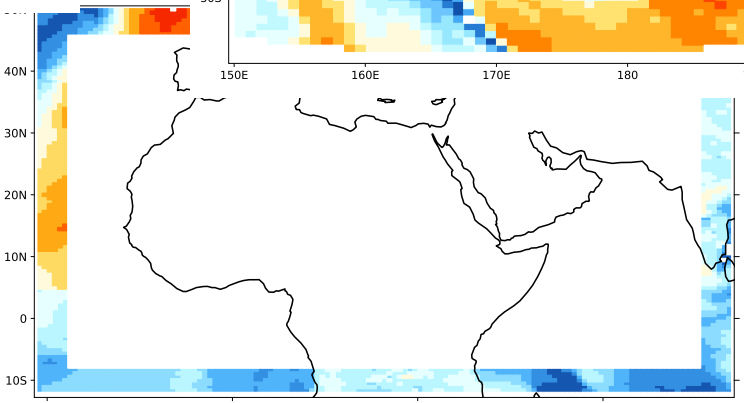


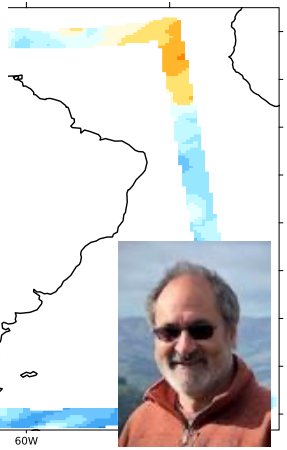










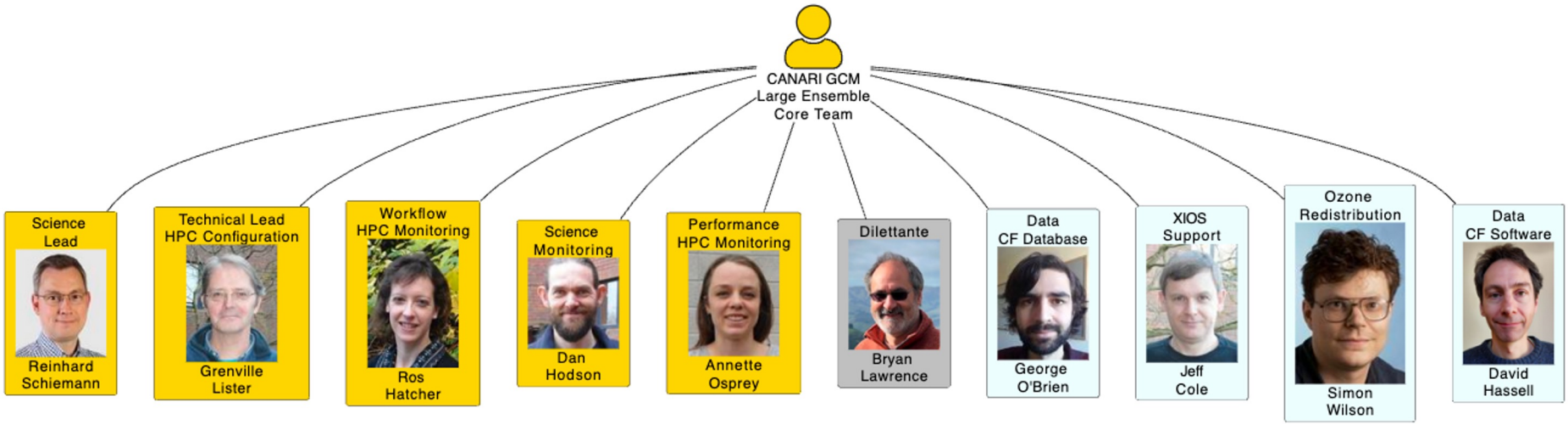


National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Timestep output for driving variables

University of Reading

CANARI:
Large team put the model and workflow into place

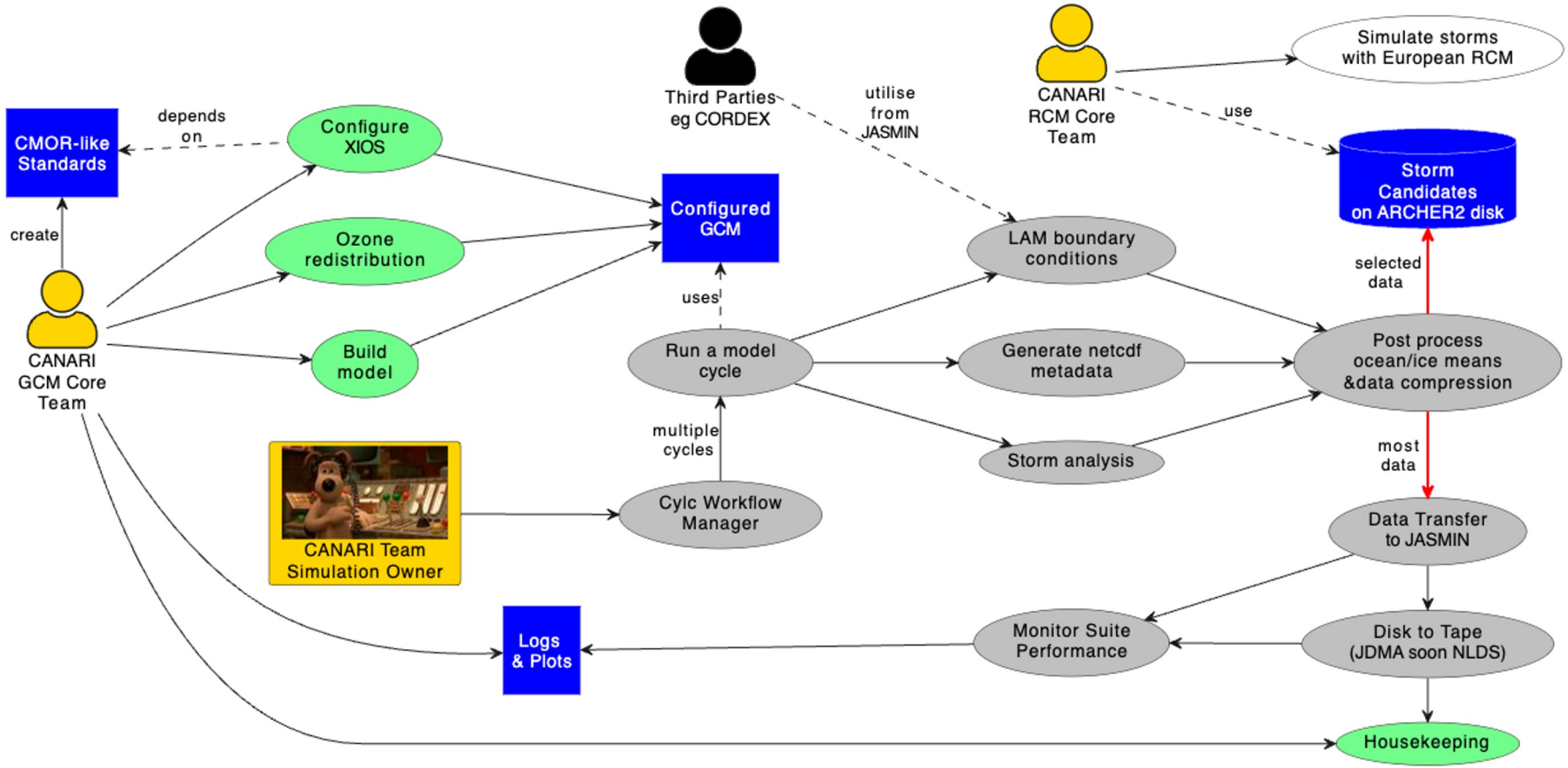


Every one of whom had/has a role to play in setting up and/or sustaining the large-ensemble.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

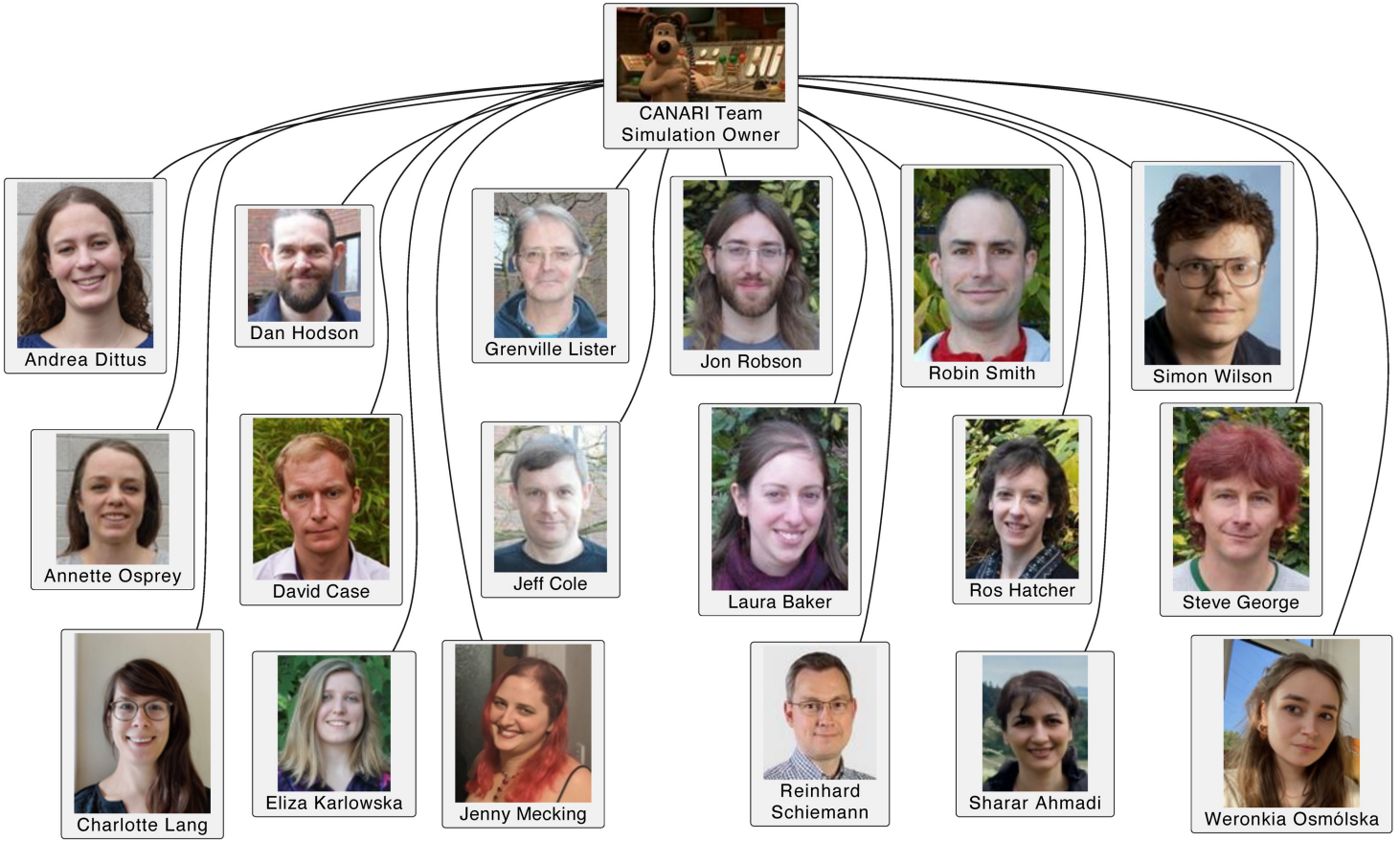
University of Reading

CANARI Activities



Each simulation broken into (typically) three month cycles.

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


Grenville Lister

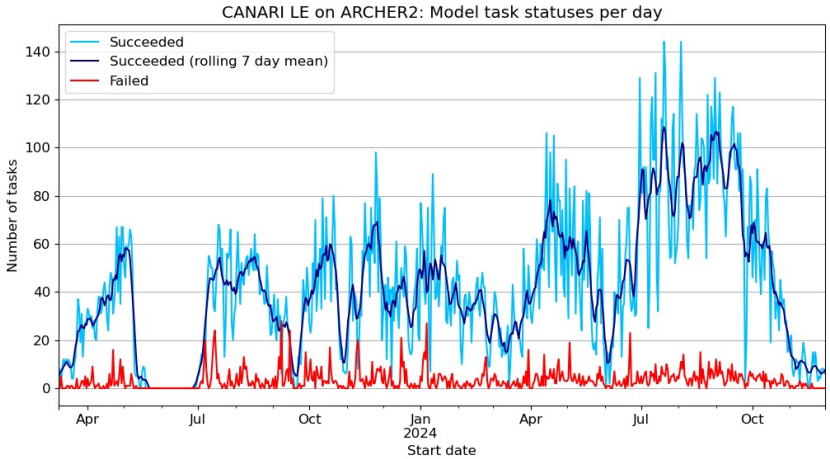

Ros Hatcher

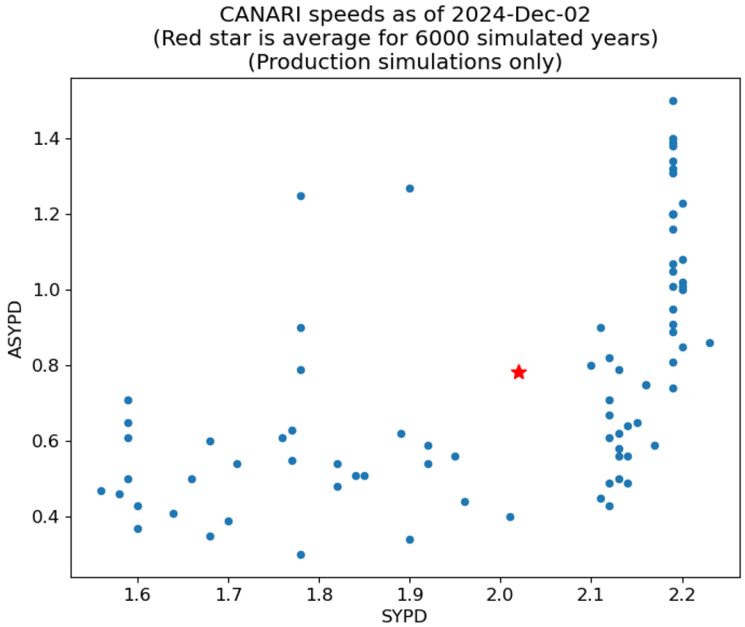

Annette Osprey

Each dot is one simulation cycle. One cycle is 3 months

There were 24,000 cycles!
Most happily resubmit …

but many do not





NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Large Ensemble Data

Simulations

- 40 Historical
- 40 Future
- = 40 x 150 = 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

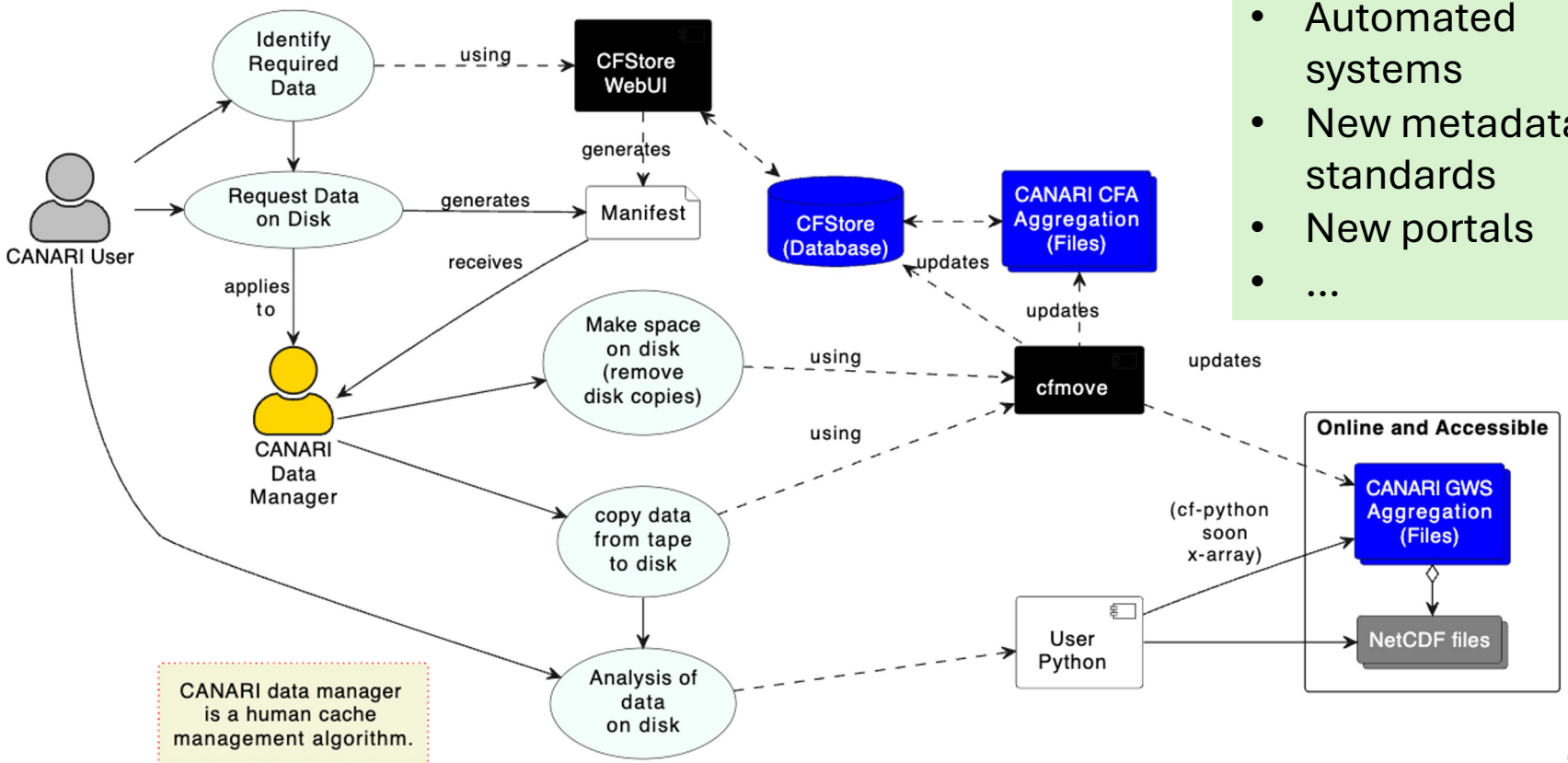
Volumes (compressed): ~5.5 PB

- 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”

A raft of innovations for the data management alone!



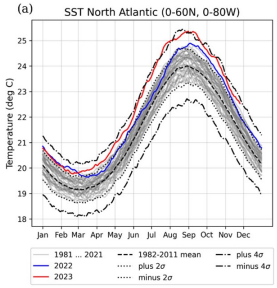
- Automated systems
- New metadata standards
- New portals
- …

Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

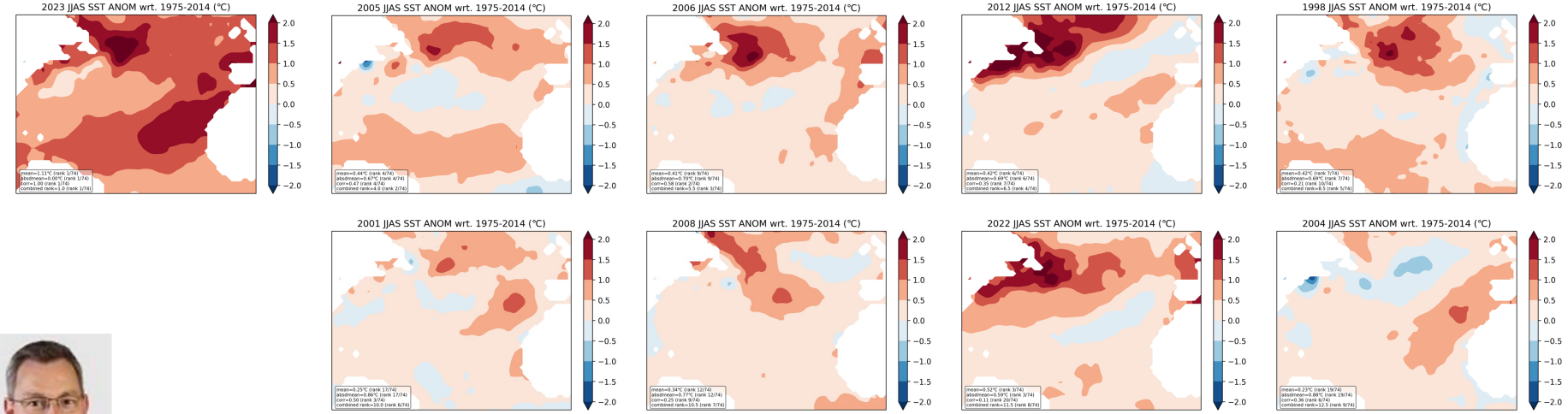
University of Reading

2023 Extreme North Atlantic SSTs: Analogues in observations (HadISST)

- Analogues defined in terms of both mean SST anomaly and pattern correlation



Kuhlbrodt et al. 2024



'best' 8 from 73 observed years (1950-2022)



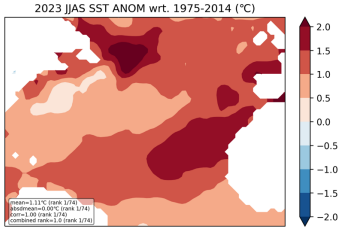
2023 is exceptional in the observational record, there are no close analogues

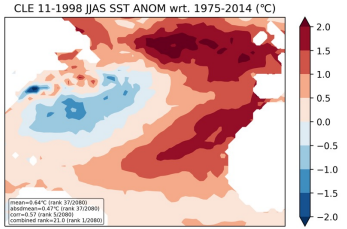
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

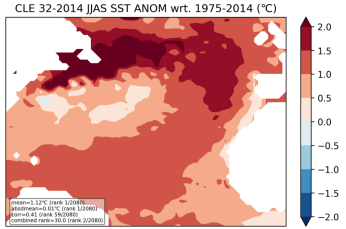
University of Reading

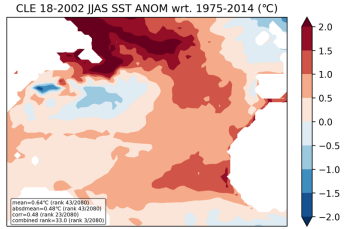
2023 Extreme North Atlantic SSTs: Analogues in the CANARI Large Ensemble

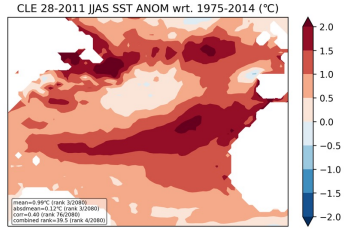
- Analogues defined in terms of both mean SST anomaly and pattern correlation

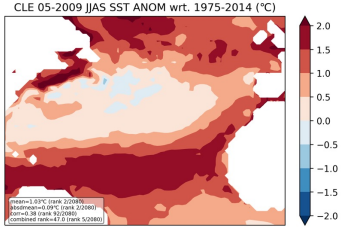


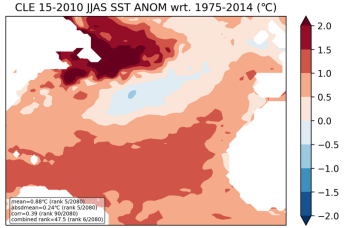


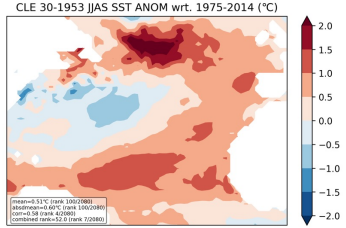


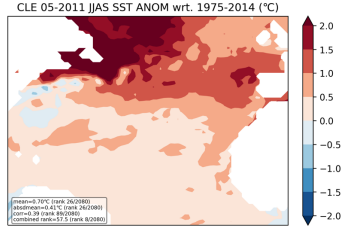












‘best’ 8 from 2080 simulated years (1950-2014)

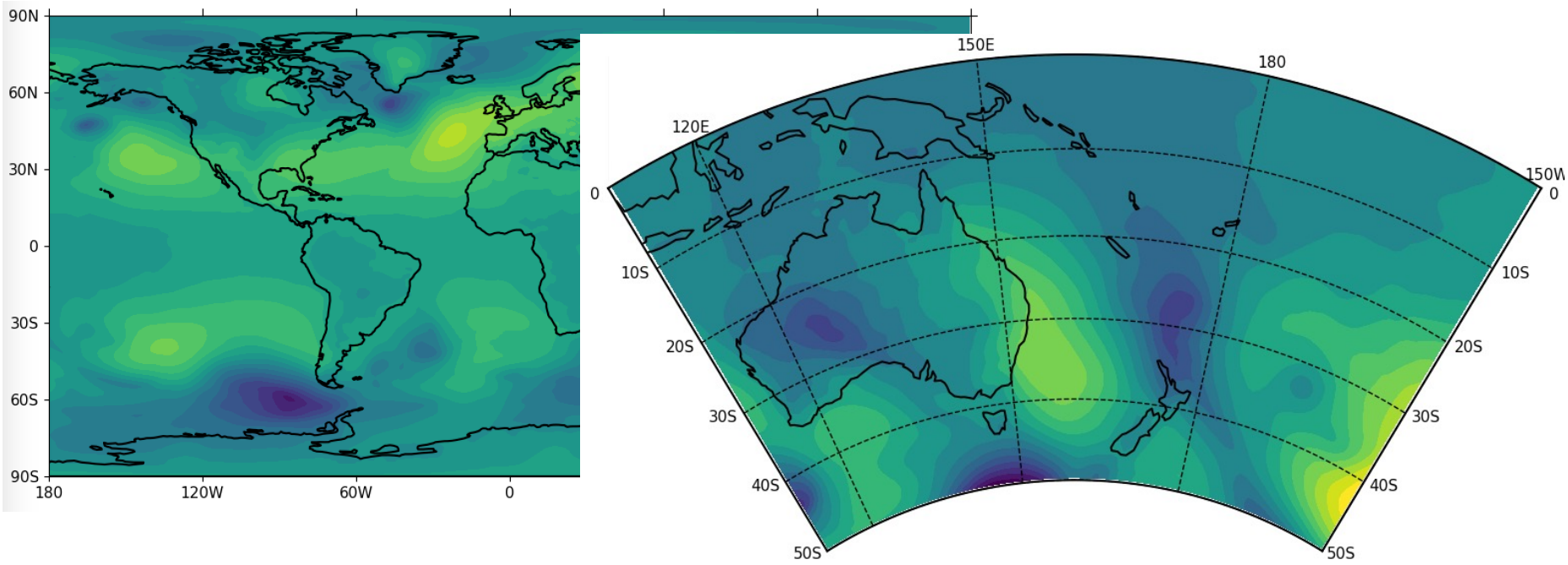


A small number of events in the historical CANARI LE look similar to 2023

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

6000 years of >800 variables
data at temporal resolution from hours to months
(depending on the variable)



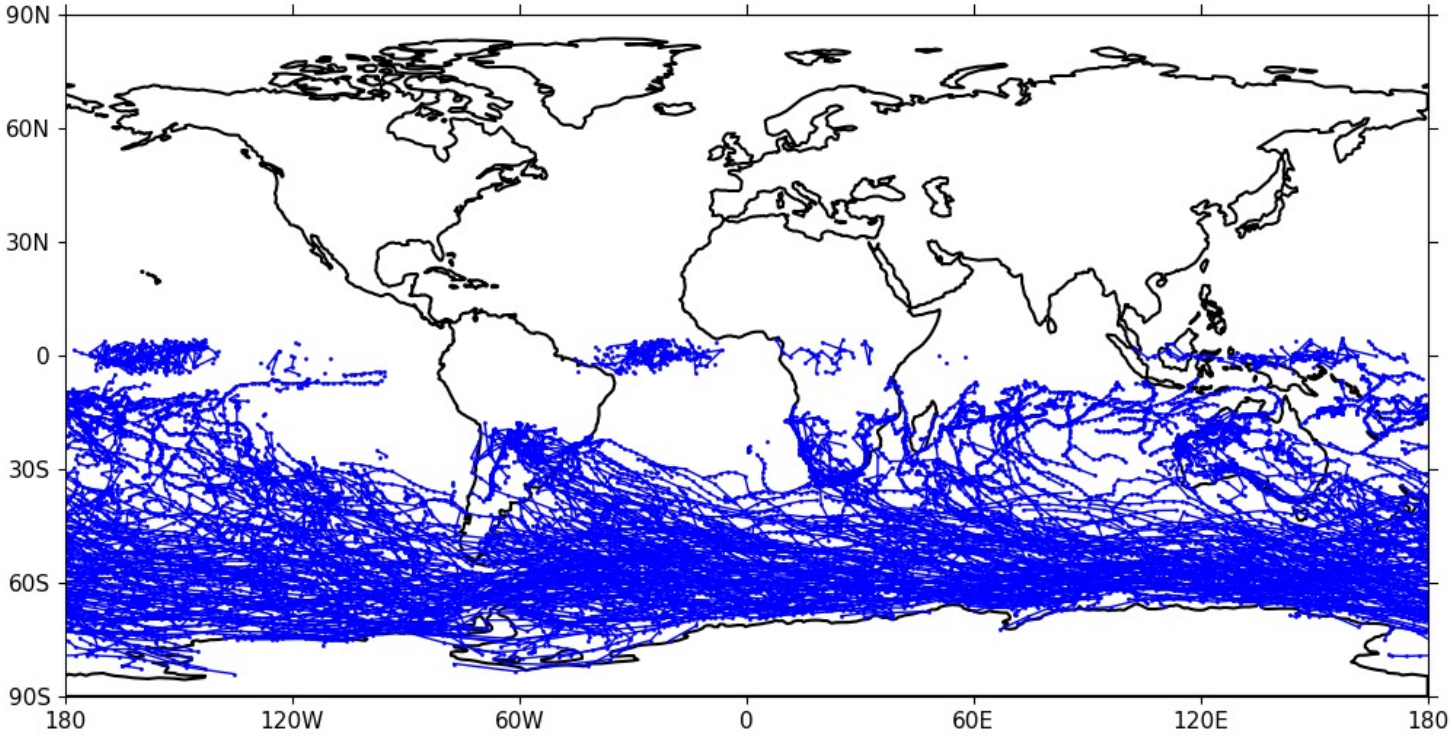
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Playing with the data

University of Reading

The entire large ensemble has been “tracked”.

Storms which meet a certain threshold are identified and their tracks and properties such as ”associated precipitation” are recorded.



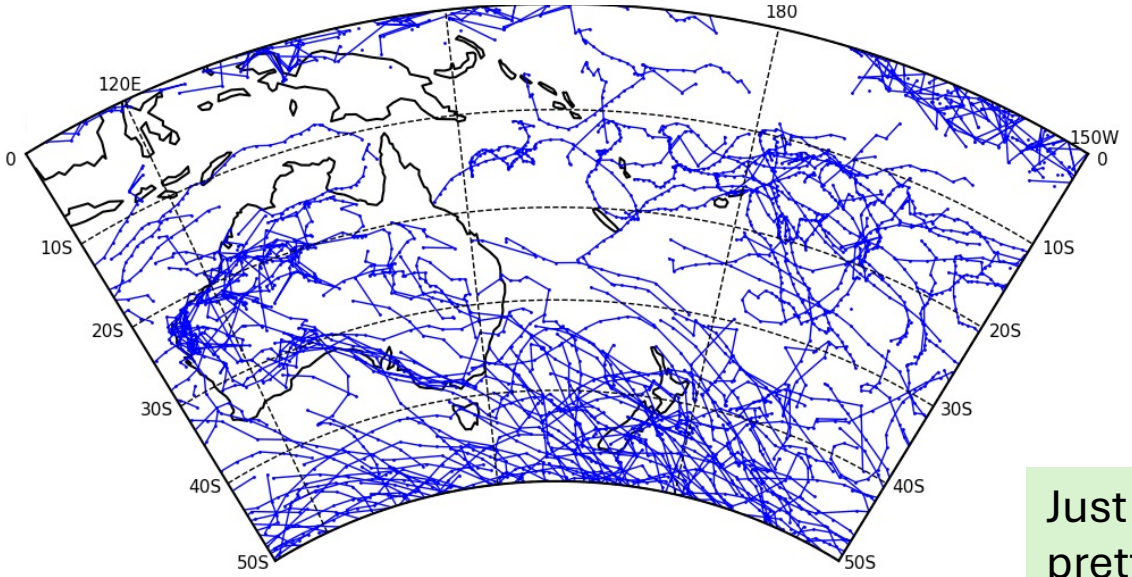


Kevin Hodges

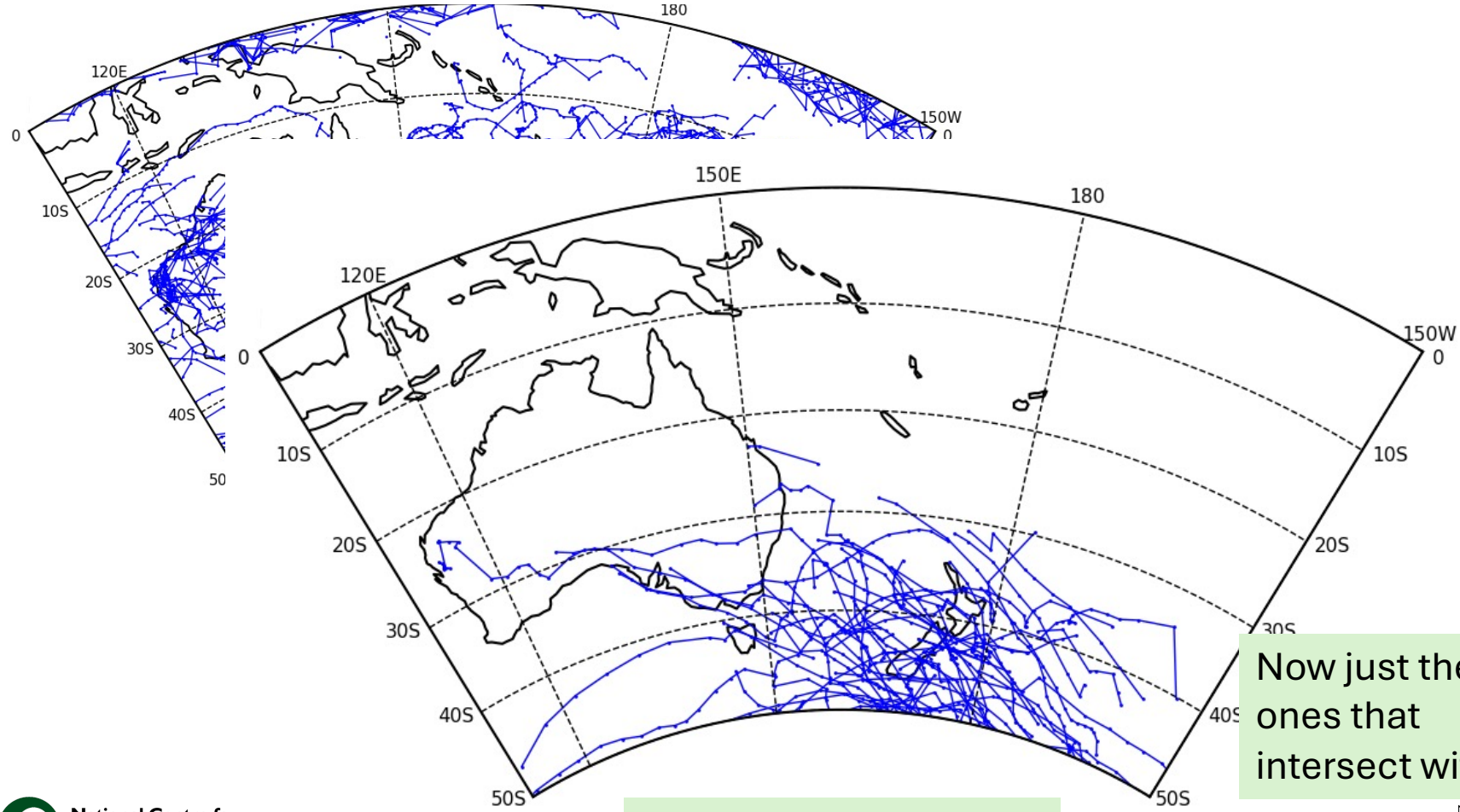
(SH storms for one 6-month period from one simulation)

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading



Just the tracks in a pretty arbitrary region, for one season



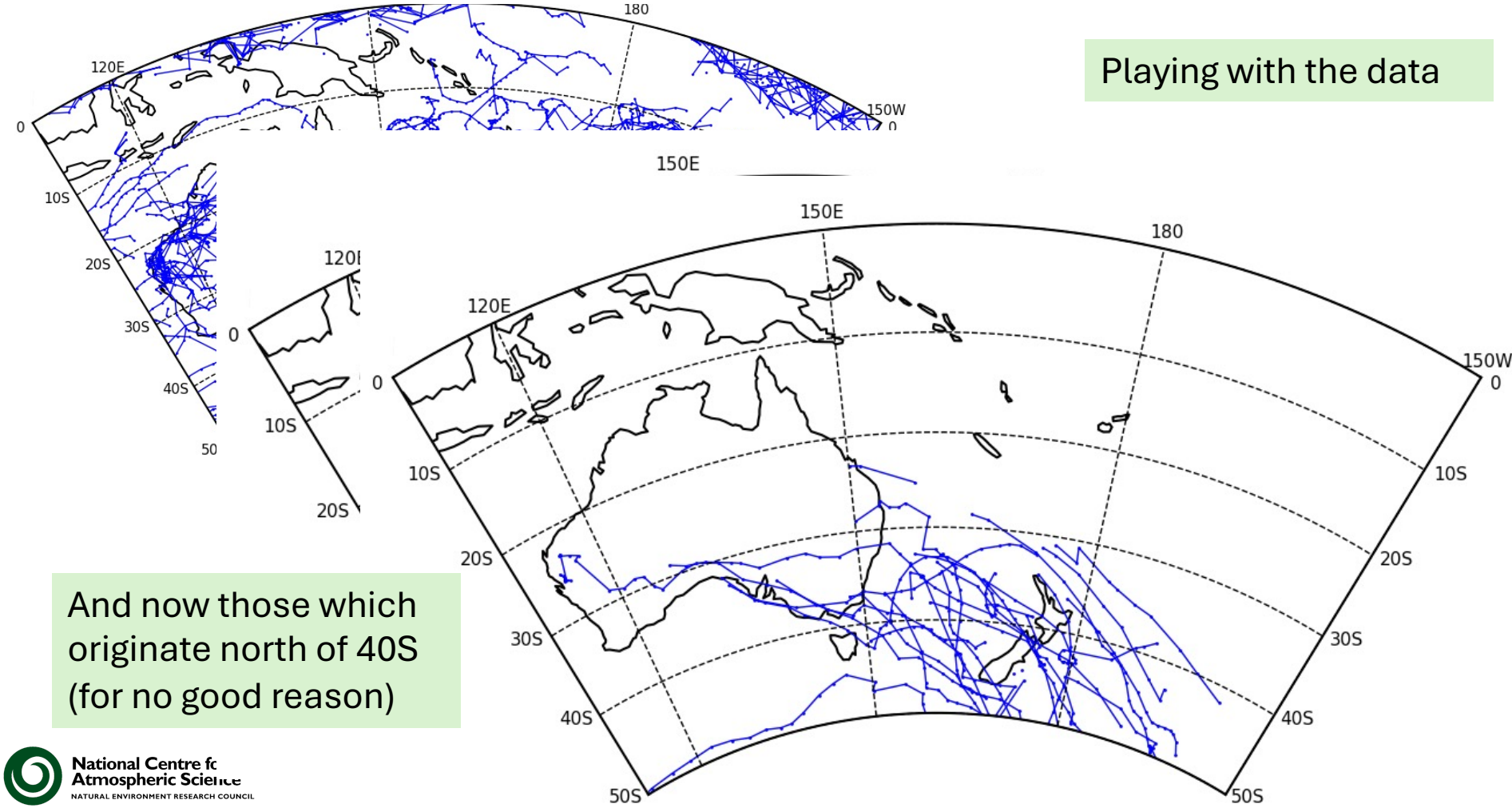
Playing with the data

Now just the ones that intersect with NZ

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Playing with the data



And now those which originate north of 40S (for no good reason)

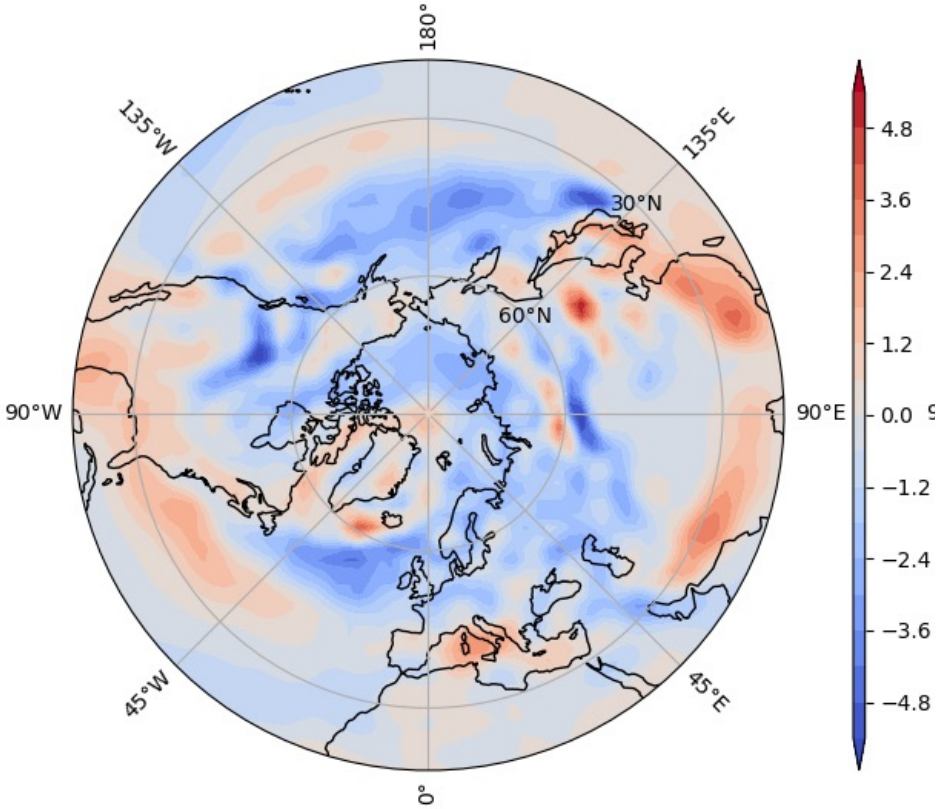
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

NH Track density bias

Track density bias during September–April in the CANARI LE against ERA5 (CANARI LE – ERA5).

The units of the plot are number of cyclones per month per unit area (5^{o} spherical cap).

There is a lot to understand (& AFIK no one has looked at the southern hemisphere) …



Warning: Preliminary version





National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Mid-latitude storm track trends







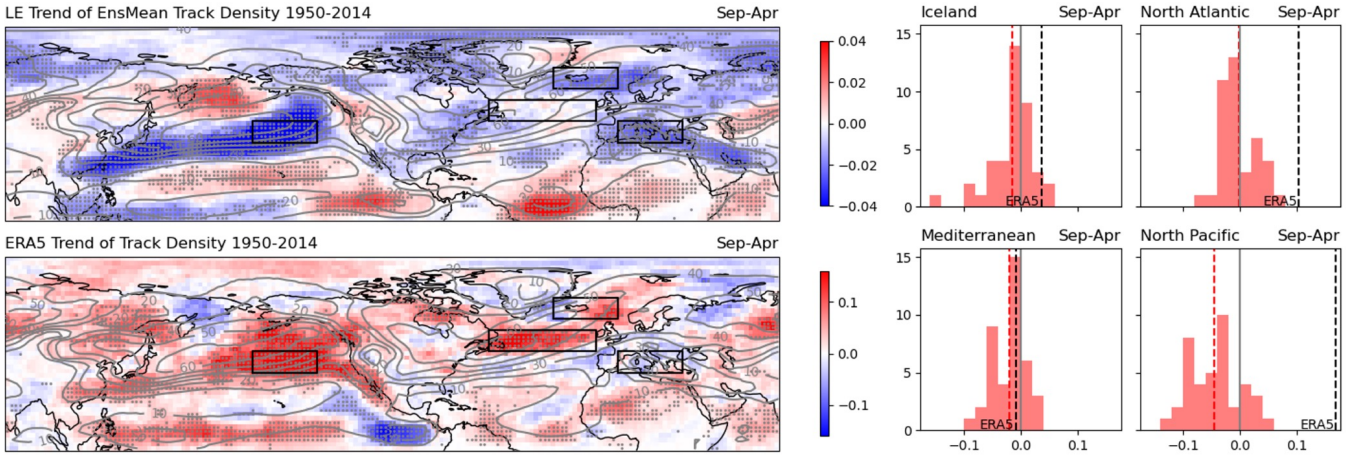


Figure 1. Linear trend (shading; units: tracks passing with 5 degrees per winter per year) and climatology (contours; units: tracks passing within 5 degrees per winter) of 1950-2014 DJF (a) ensemble mean track density and (b) ERA5 track density. Note the different colorbars. Stippling indicates trend p-values < 0.05. Panels (c) to (f) show histograms of the trend in track density for individual members, and ERA5 (black dashed), averaged over the boxes indicated.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

Europe Storm Identification

Extreme windstorms are being identified using a storm severity index (SSI) based on instantaneous 10m winds (U)

$$SSI = \sum_{\substack{UK\ land \\ U_i > U_{ref}}} \left(U_i - U_{ref}\right)^3 A_i$$

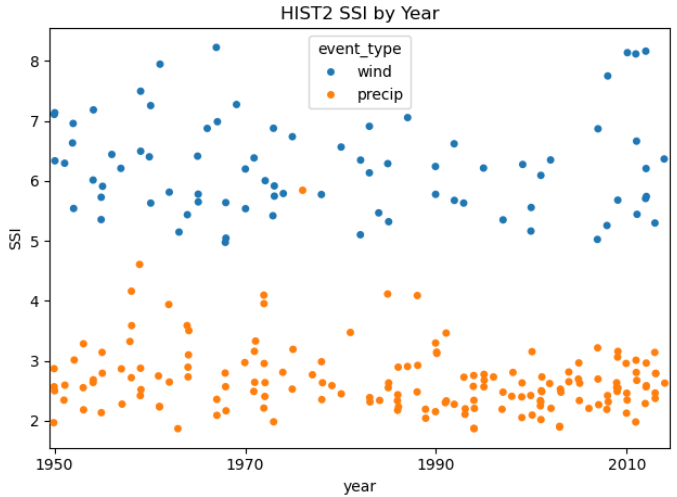
Extreme precipitation storms are being identified using an extreme precipitation index (EPI) based on the 5-day accumulated precipitation (P5d):

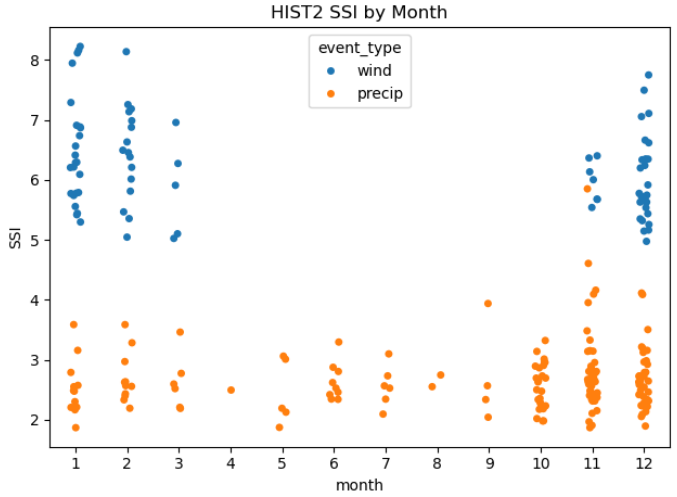
$$EPI = \sum_{\substack{UK\ land \\ P_{5d,i} > P_{98}}} \left(P_{5d,i} - P_{98}\right) A_i$$

(These results from the historical simulations.)









National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

SSI European seasonal variability – CANARI LE v ERA5

- Extreme windstorms are being identified using a Storm Severity Index (SSI) based on exceedance above 25 m/s and area affected.
- All land points for which windspeed exceeds 25 m/s contribute to the SSI.

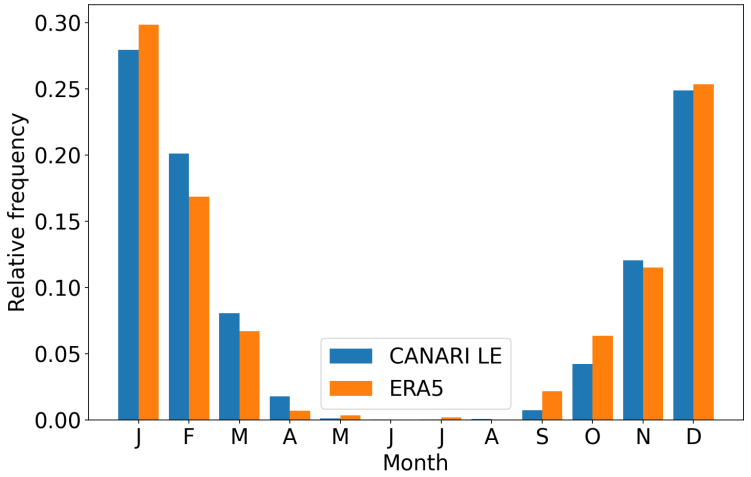
$$SSI(t) = \sum_{\substack{UK\ land \\ u_i(t)>25\ m/s}} U(t)_M A(t)_i$$

$U_M(t)$ is the maximum wind speed at a time t.





SSI > $2x10^{16}$ m^5/s^3



- The seasonality of UK extreme windstorms in ERA5 is represented well by the CANARI LE.
- This is true throughout the year, except perhaps in late spring and summer in which ERA5 exhibits a very small fraction of storms (Is the SSI threshold value too low?).

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

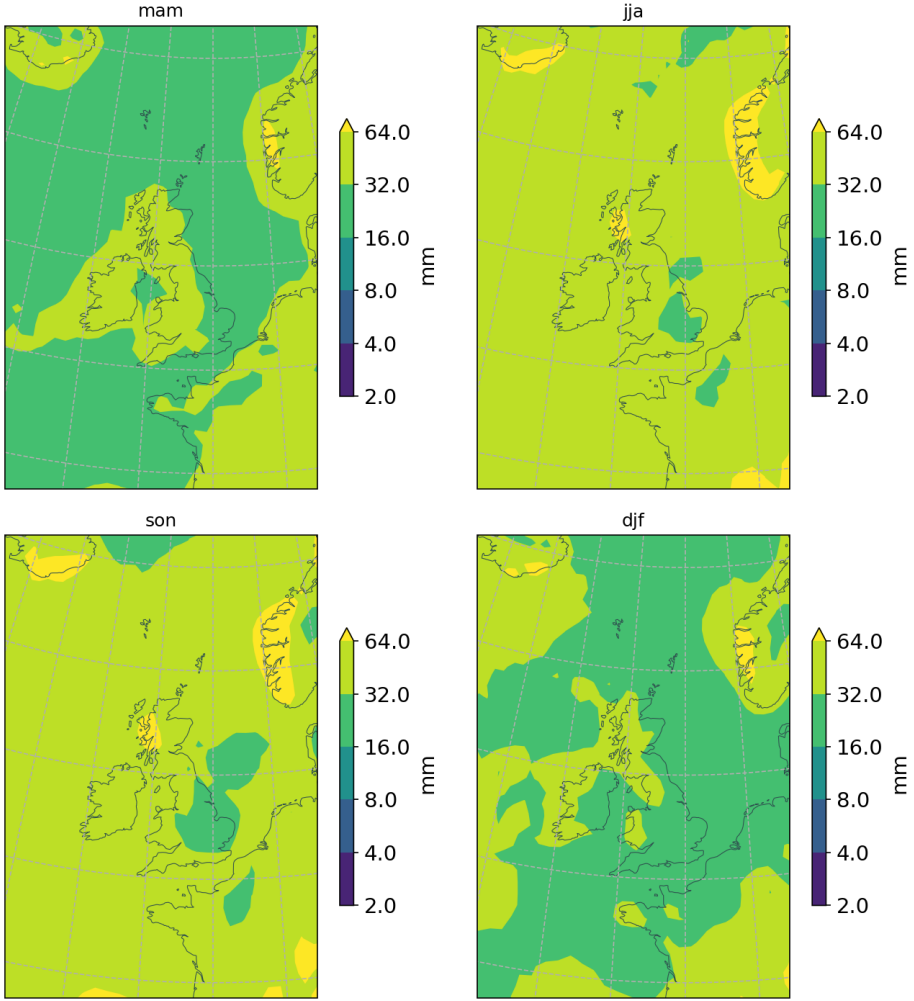
University of Reading

Extreme precipitation

$$EPI = \sum_{\substack{UK\ land \\ P_{5d,i} > P_{98}}} (P_{5d,i} - P_{98}) A_i$$

- P_{5d} 98th percentile seasonal cycle derived from ERA5 hourly precipitation.





National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

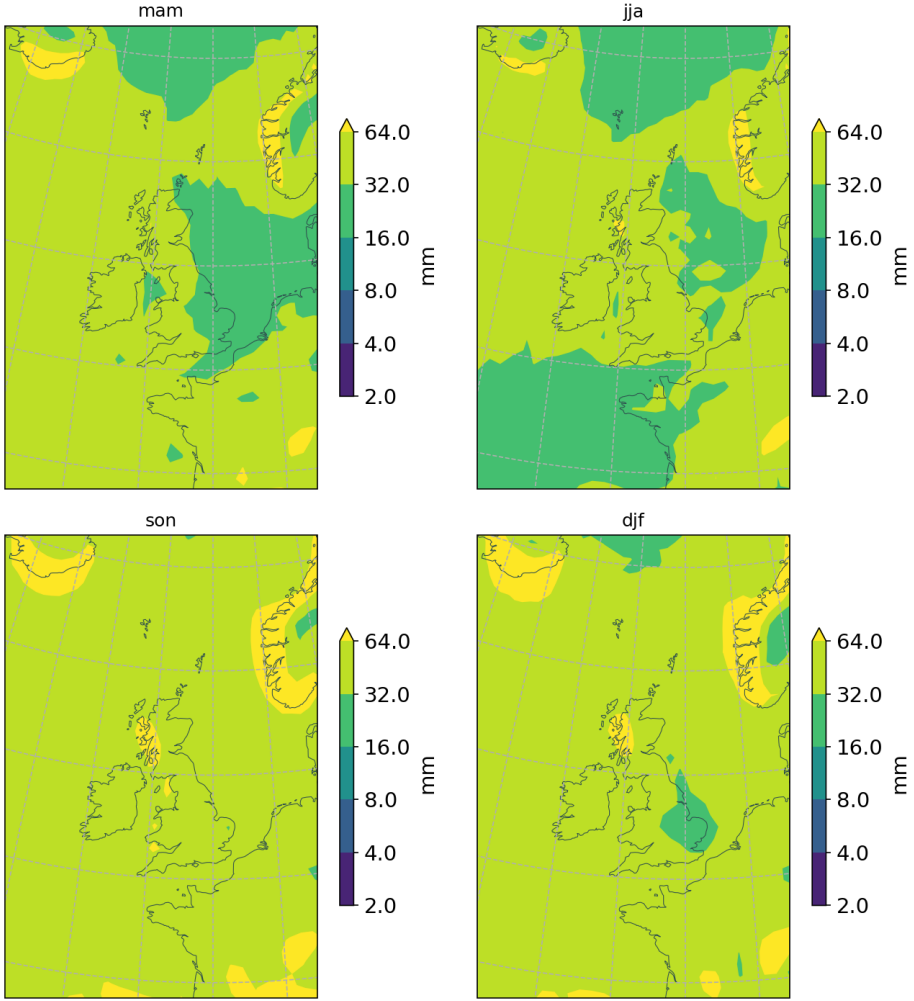
Extreme precipitation

$$EPI = \sum_{\substack{UK\ land \\ P_{5d,i} > P_{98}}} \left(P_{5d,i} - P_{98}\right) A_i$$

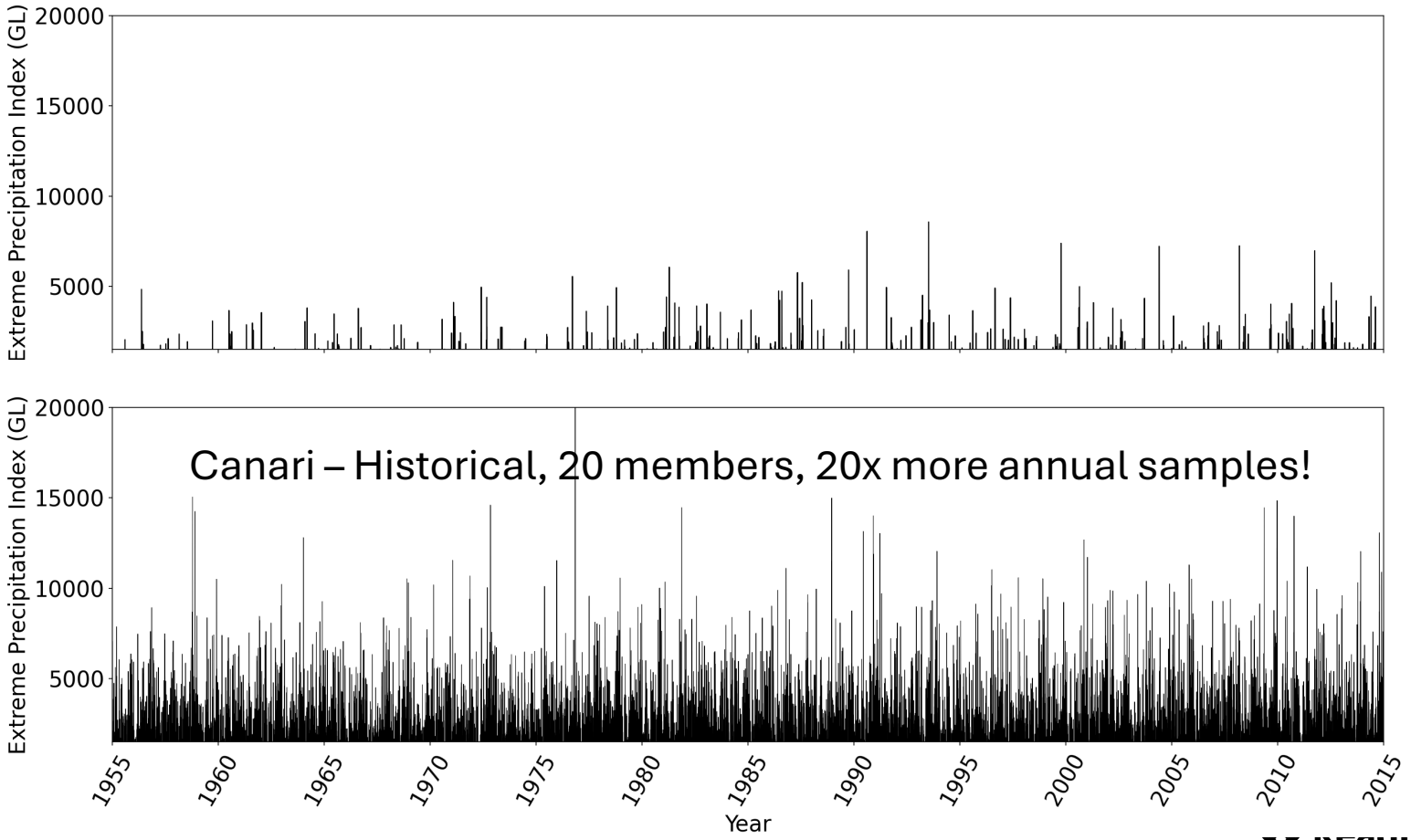
- P_{5d} 98th percentile seasonal cycle derived from CANARI LE member 11 hourly precipitation.
- CANARI LE exhibits a wet bias with respect to ERA5 throughout the year, but noticeable over winter (DJF).



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL



EPI interannual variability – ERA5





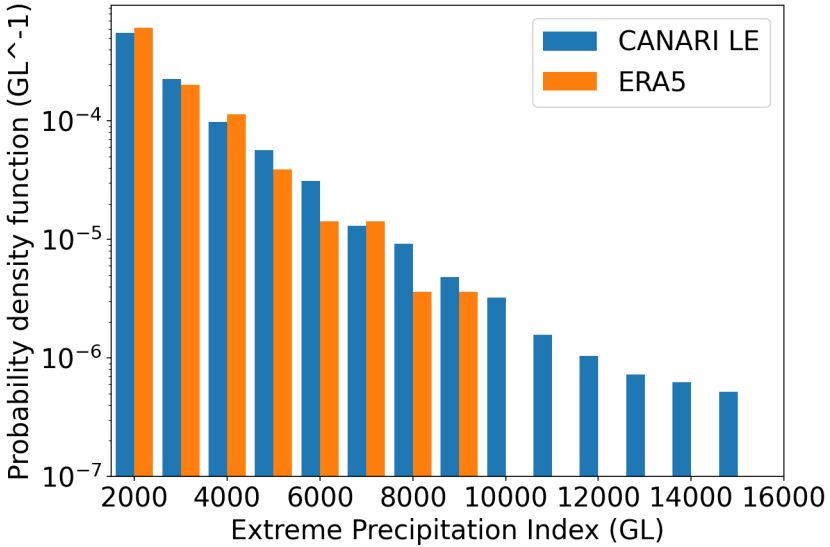
Time se

National Ce
Atmospher
NATURAL ENVIRONMENT RESEARCH COUNCIL

Reading

EPI probability distribution – CANARI LE v ERA5

- Empirical probability density functions derived from the EPI corresponding to the CANARI LE and ERA5.
- Both PDFs decay exponentially at a similar rate.
- The PDFs appear consistent with each other; the CANARI LE **provides an estimate of PDF at the tail of the distribution.**





The climate we have had, would appear to have been a lot drier than the climate we could have had!

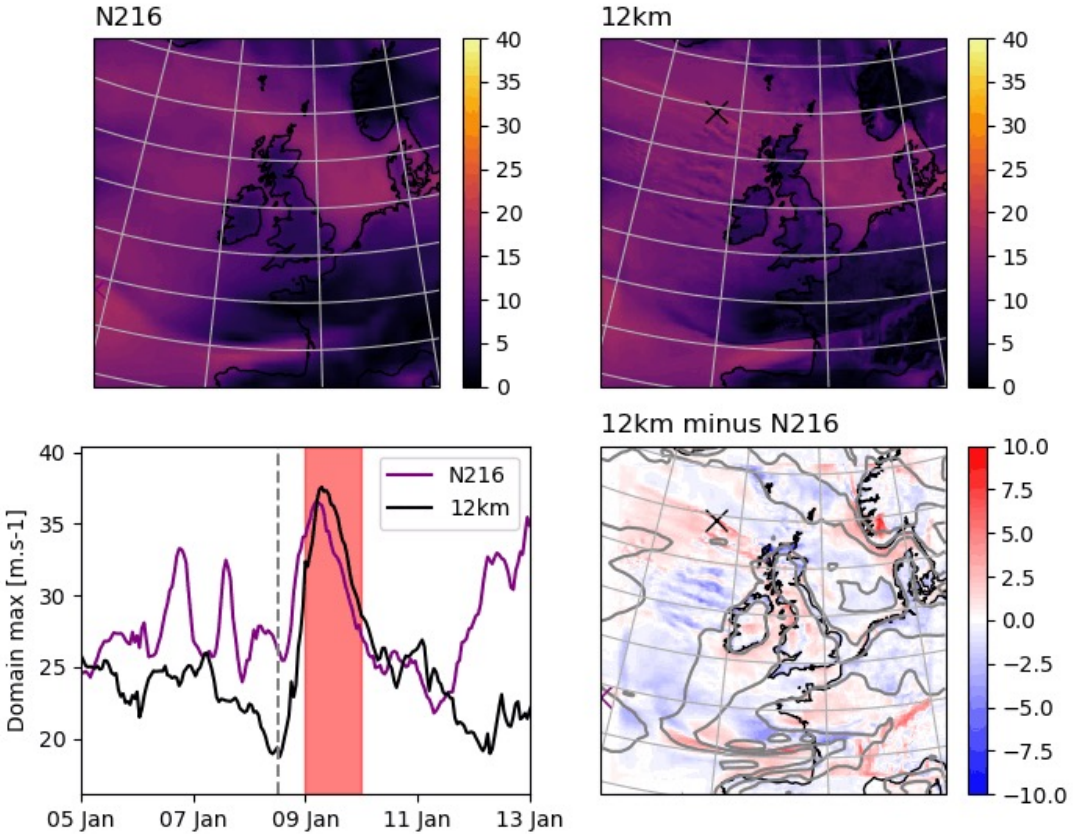
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

UK Illustration

- 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





CANARI Large Ensemble: Documentation paper

- Paper in preparation to
  - document the Large Ensemble and data access
  - evaluate key aspects of the historical simulations
  - author count currently at 40



The CANARI HadGEM3 Large Ensemble: Design and evaluation of historical simulations

Reinhard K. H. Schiemann[1], Grenville Lister[1], Rosalyn Hatcher[1], Dan Hodson[1], Bryan Lawrence[1], Len Shaffrey[1], Andrea Dittus[1], Jon Robson[1], Ben Harvey[1], Oscar Martínez-Alvarado[1], Steve Woolnough[1], Andrew Turner[1], Simon Wilson[1], Jeff Cole[1], Sharar Ahmadi[1], Laura Baker[1], David Case[1], Steve George[1], Eliza Karlowska[1], Charlotte Lang[1], Annette Osprey[1], Robin Smith[1], Jake Aylmer[2], David Schröder[2], Niamh O'Callaghan[2], Weronika Osmolska[3], Scott Osprey[4], Yevgeny Aksenov[5], Adam T. Blaker[5], Andrew Coward[5], Jenny Mecking[5], Stephanie Rynders[5], Bablu Sinha[5], Chris Wilson[6], Martin Andrews[7], Steven Hardiman[7], John Rostron[7], David Sexton[7], Hua Lu[8], Tony Phillips[8], and Jonny Williams[9]

[1]National Centre for Atmospheric Science, Department of Meteorology, University of Reading, United Kingdom
[2]Department of Meteorology, University of Reading, United Kingdom
[3]National Centre for Atmospheric Science, University of Leeds, United Kingdom
[4]National Centre for Atmospheric Science, University of Oxford, United Kingdom
[5]Marine Systems Modelling, National Oceanography Centre, Southampton, SO14 3ZH, United Kingdom
[6]National Oceanography Centre, Liverpool, United Kingdom
[7]Met Office Hadley Centre, Exeter, United Kingdom
[8]British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, United Kingdom
[9]National Institute of Water and Atmospheric Research, Private Bag 99940, Newmarket, Auckland 1149, New Zealand

Correspondence: Reinhard Schiemann (r.k.schiemann@reading.ac.uk)

Summary

- The interplay between different types of uncertainty requires different types of models.
- The CANARI large ensemble, just finished, is the highest resolution high quality model ensemble available (AFIK)
    - Builds from a well characterized CMIP6 model (HadGEM3.1-MM)
    - 40 simulations from 1950 to 2100 = 6000 years of data
    - The ensemble has been designed for downscaling to look at extremes, and we're doing that in the UK, and "to support others" (hence my visit).
    - We're only now starting to touch the data beyond "routine monitoring" and finding lots of things (individually and via community hackathons).

There is more

- NCAS High Resolution Climate Programme (PI: Pier-Luigi Vidale) has several ensemble members which have run some current and future N1280 (global 10km) for 30 years apiece!
- There are global storm-resolving projects spinning up
    - see WCRP Digital Earths
- Conclusion: There is a lot of data to analyse and much to understand about the influence of domain (global v regional) and resolution on what we might conclude about future climates.
    - Do we get the same storylines from regional downscaling as we would get from global downscales? What of the upscale transfer from high resolution weather?