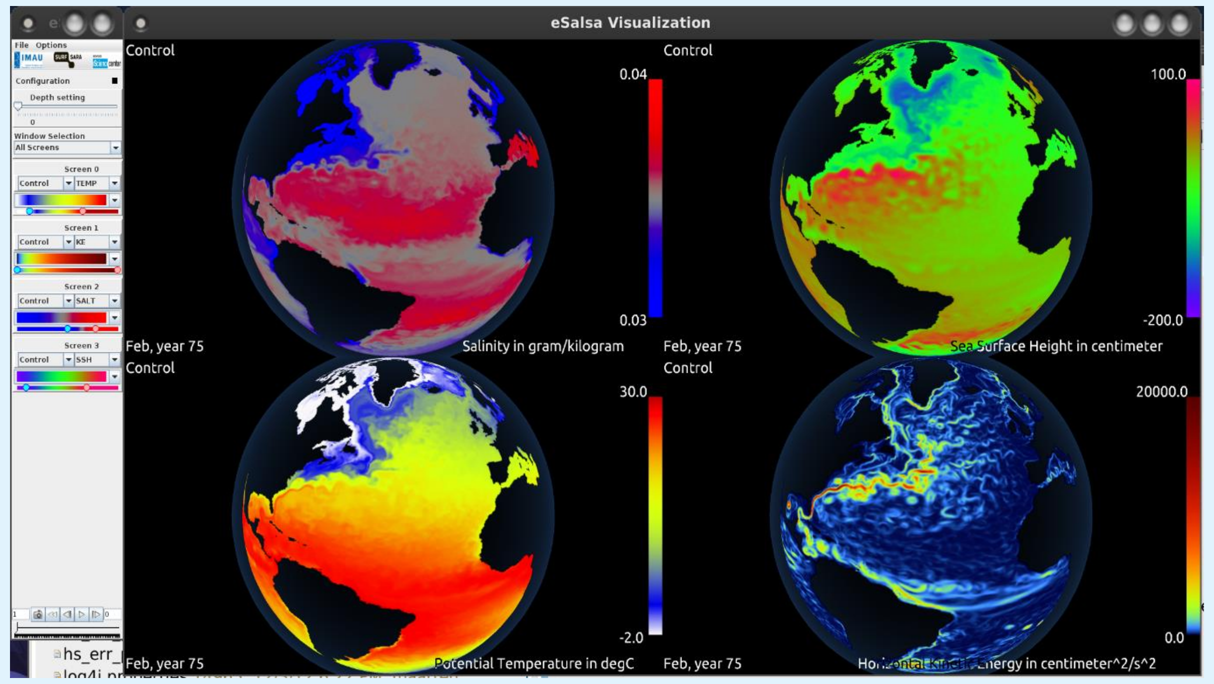
A (c)loud revolution in weather and climate research

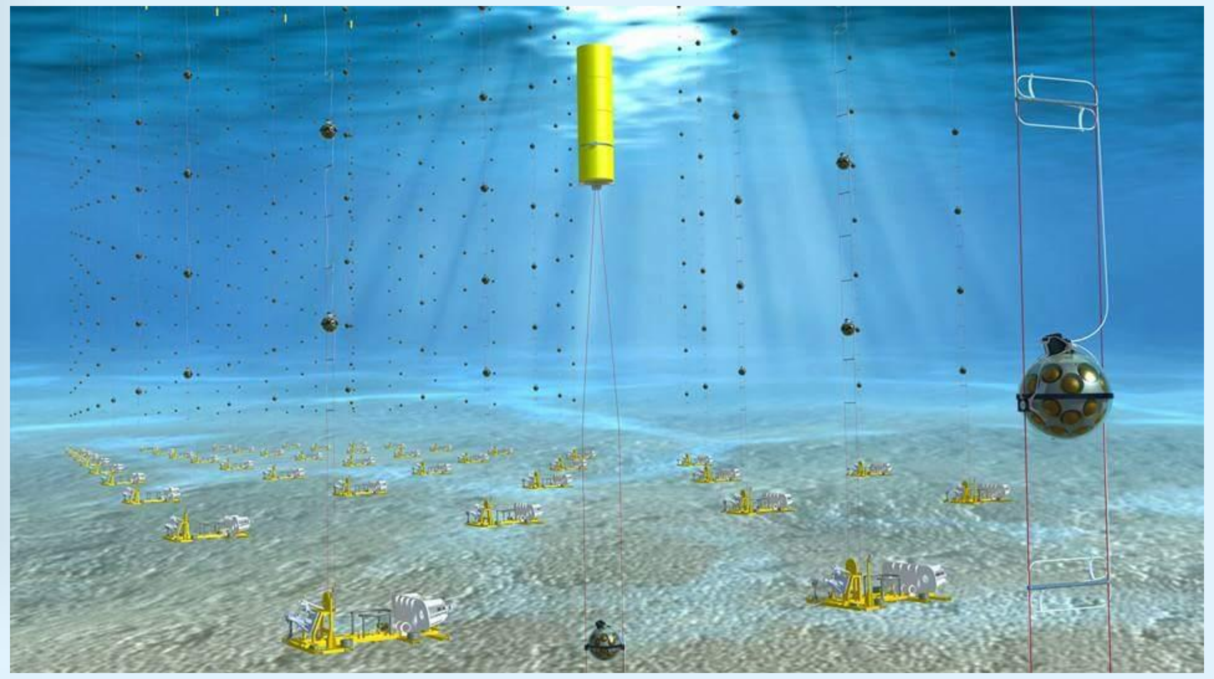
Wilco Hazeleger

Reading, 27/09/2018

Acknowledgement: the team at the Netherlands eScience Center

100+ projects





Humanities & Social Sciences

incl. SMART cities, text analysis, creative technologies

Physics & Beyond

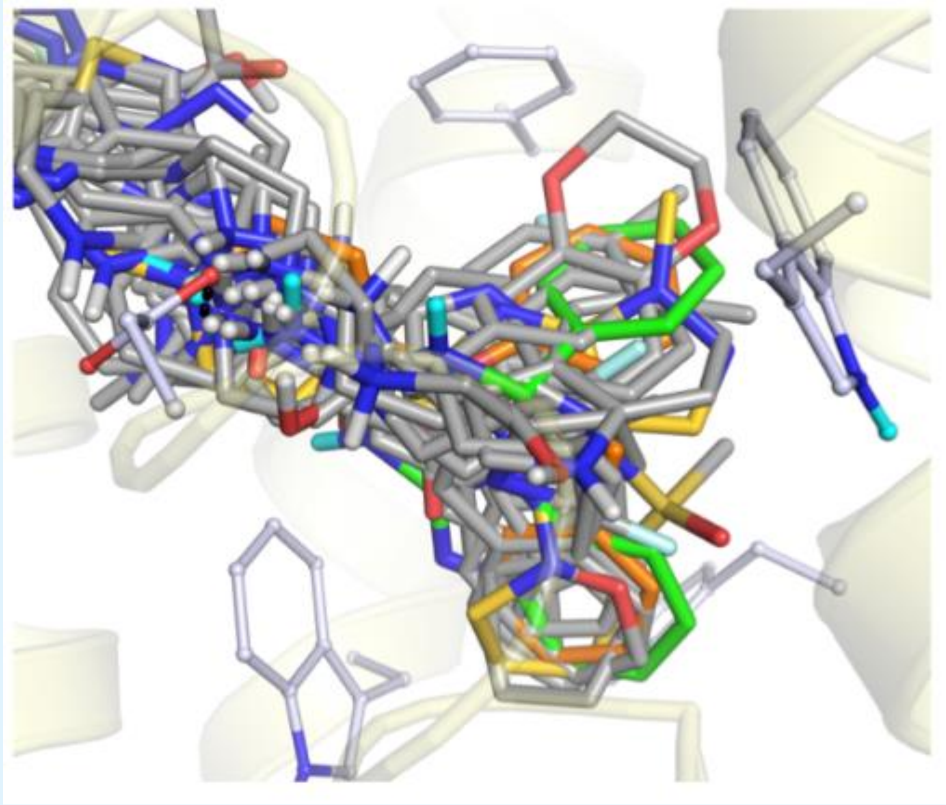
incl. astronomy, high-energy physics, advanced materials

Sustainability & Environment

incl. climate, ecology, energy, logistics, water management

Life Sciences & eHealth

incl. bio-imaging, next generation sequencing, molecules

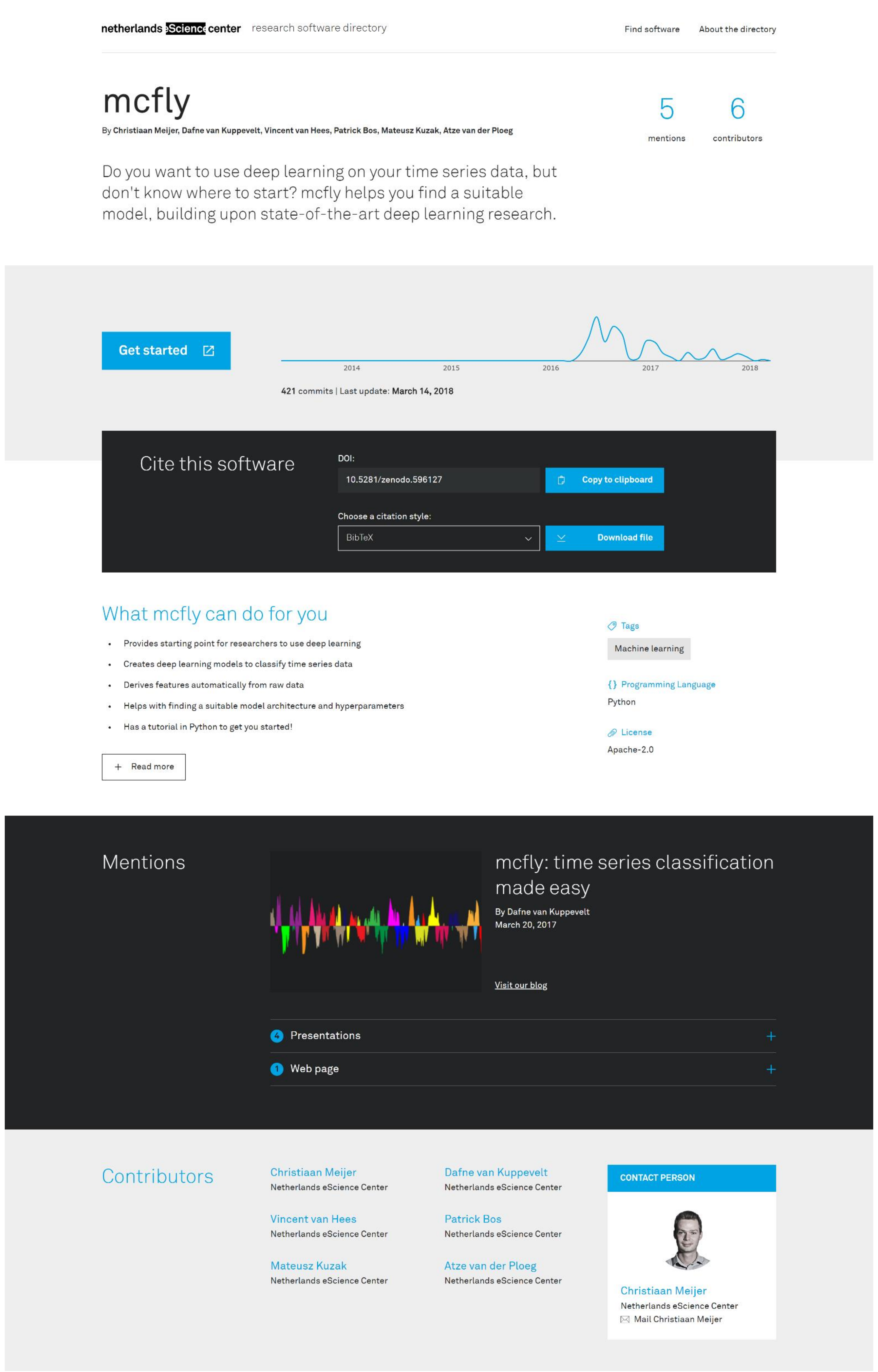


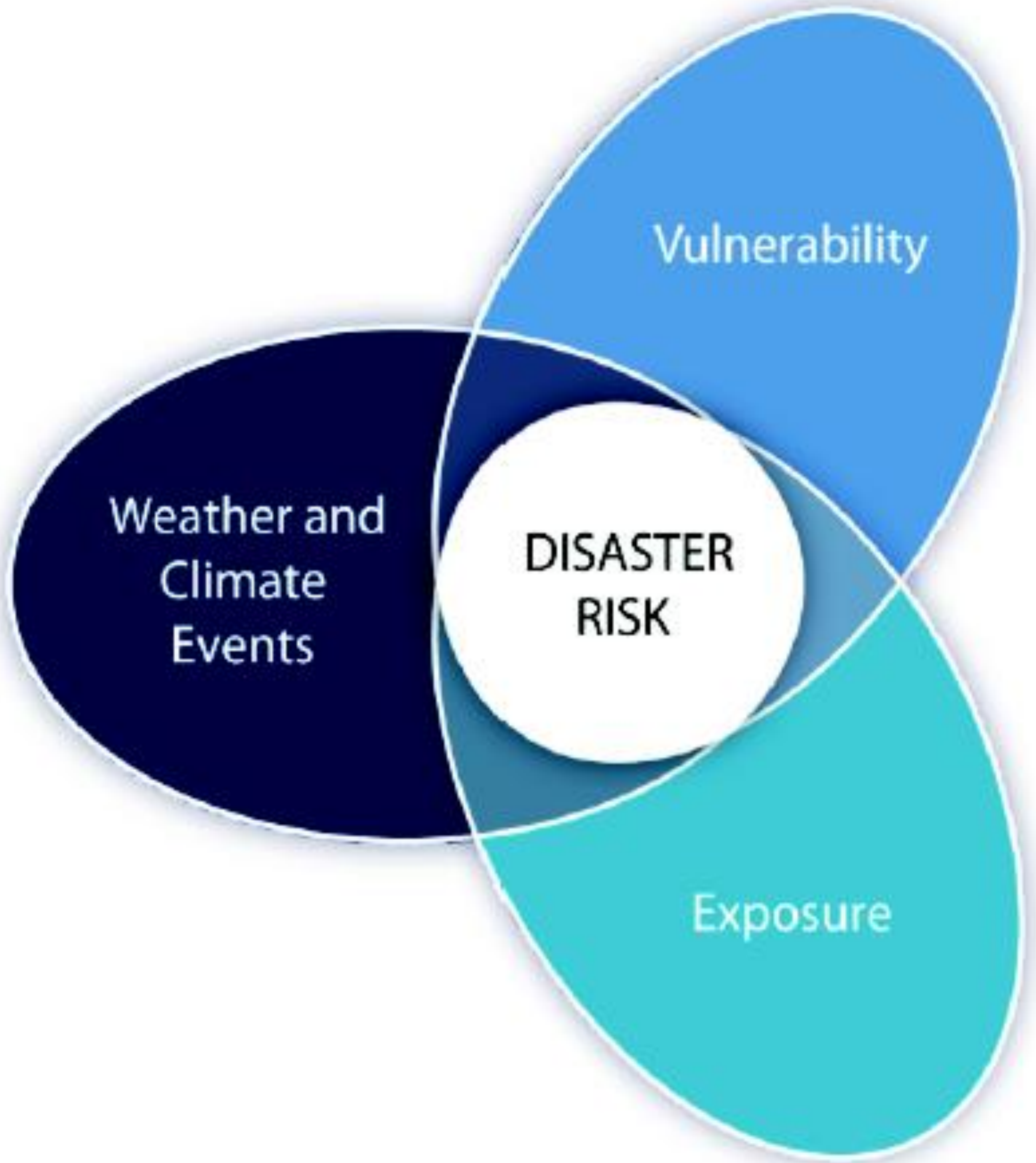
Research Software Directory

FAIR software:

- Finding software
- Making software accessible
- Quickly judge relevance and quality
- Indicating return on investment

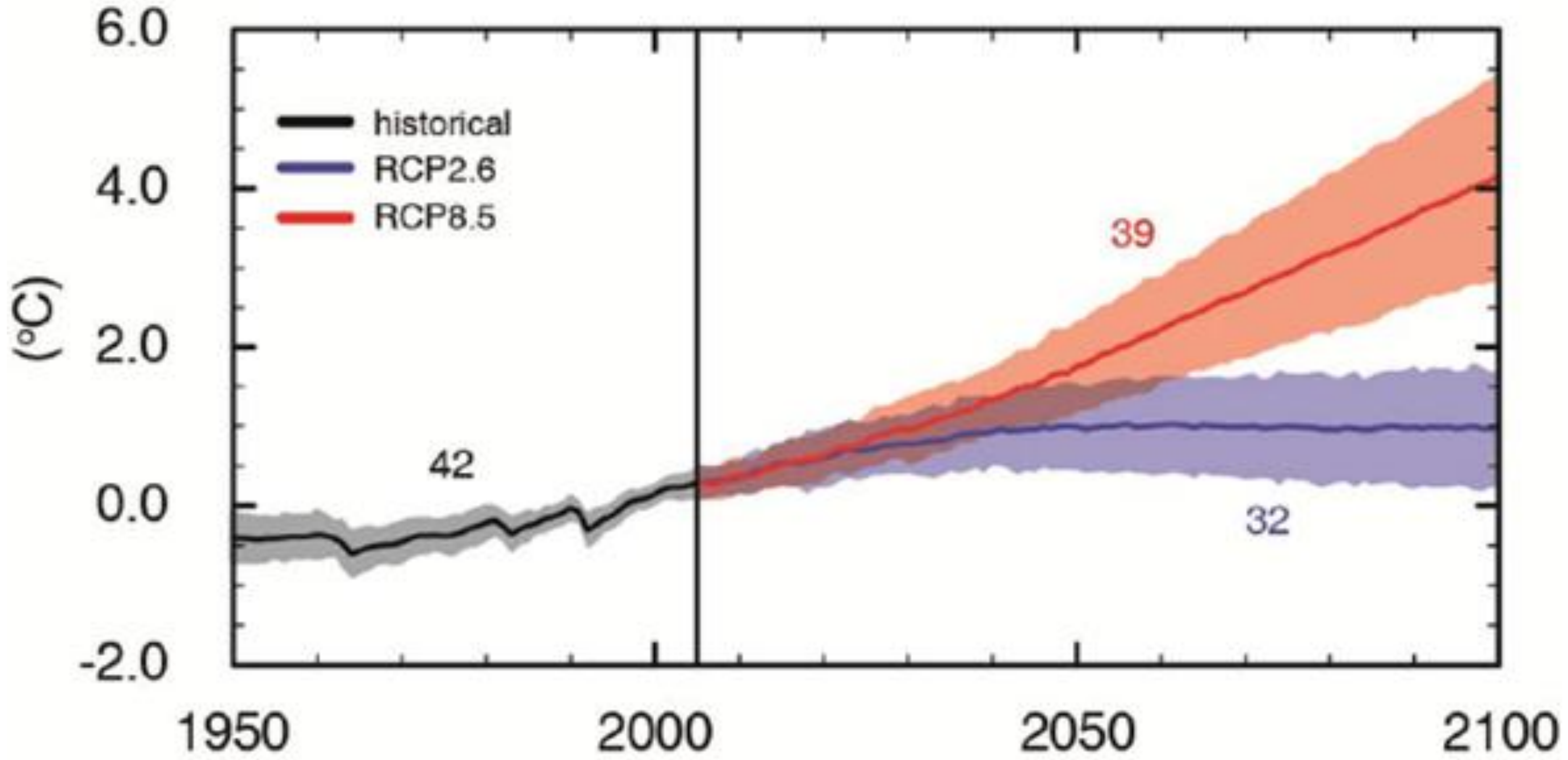
www.research-software.nl



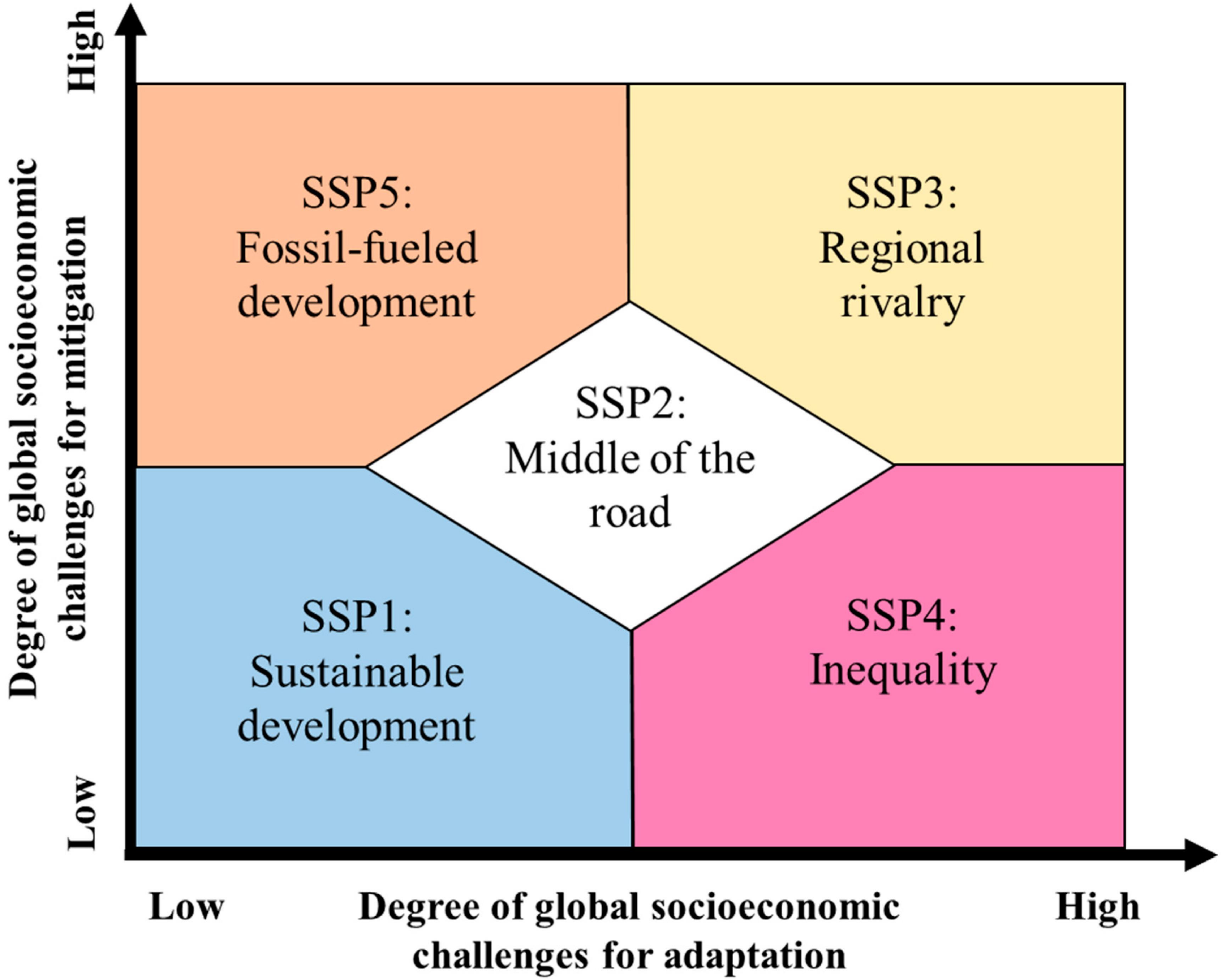

Vulnerability
Weather and Climate Events
DISASTER RISK
Exposure


TOTAL

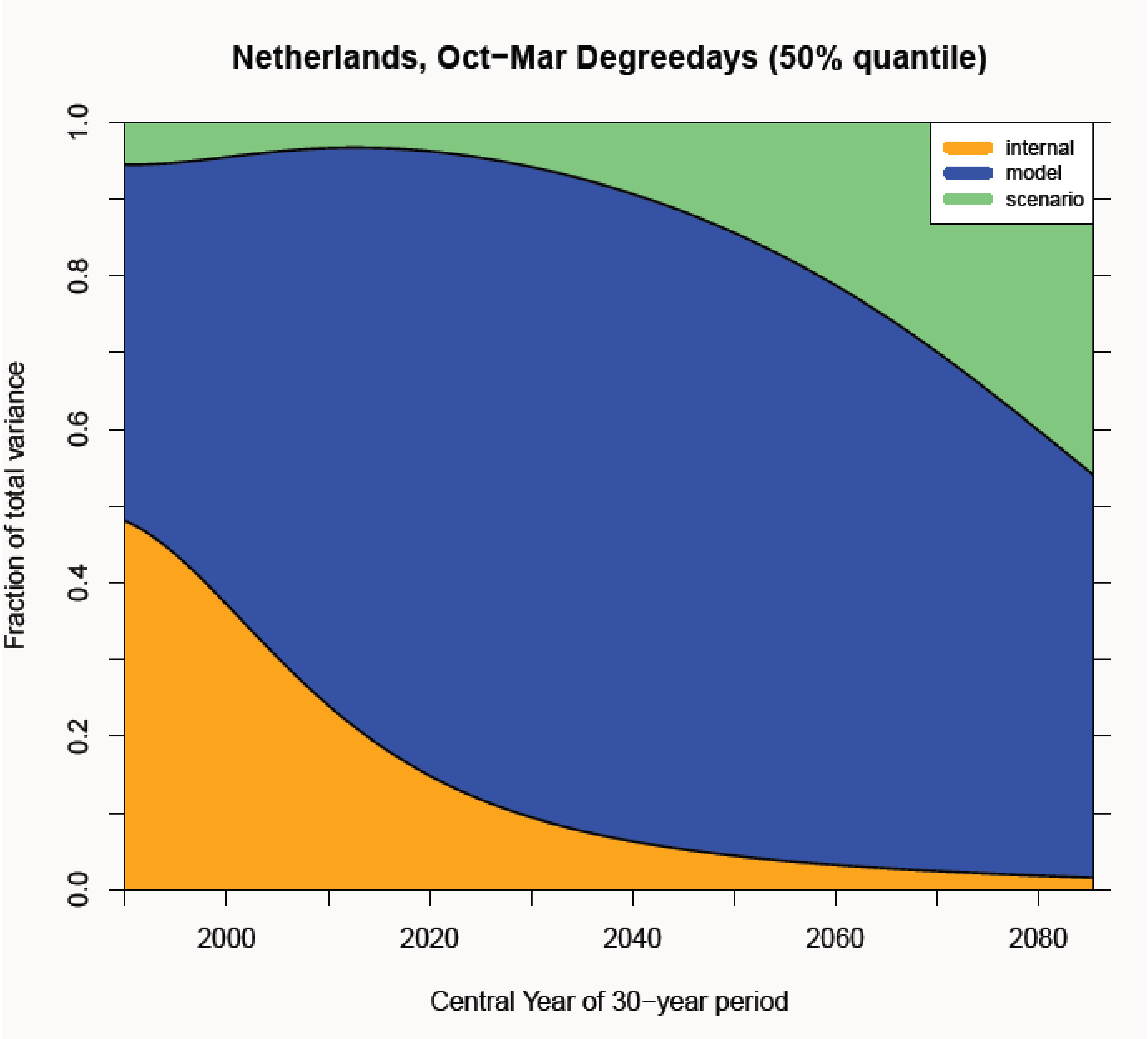
Global climate change (IPCC 5AR)



Socio-economic pathways (e.g. Riahi et al 2017)



Variance in projection of a local relevant climate variable (sum of temperature under 18oC)

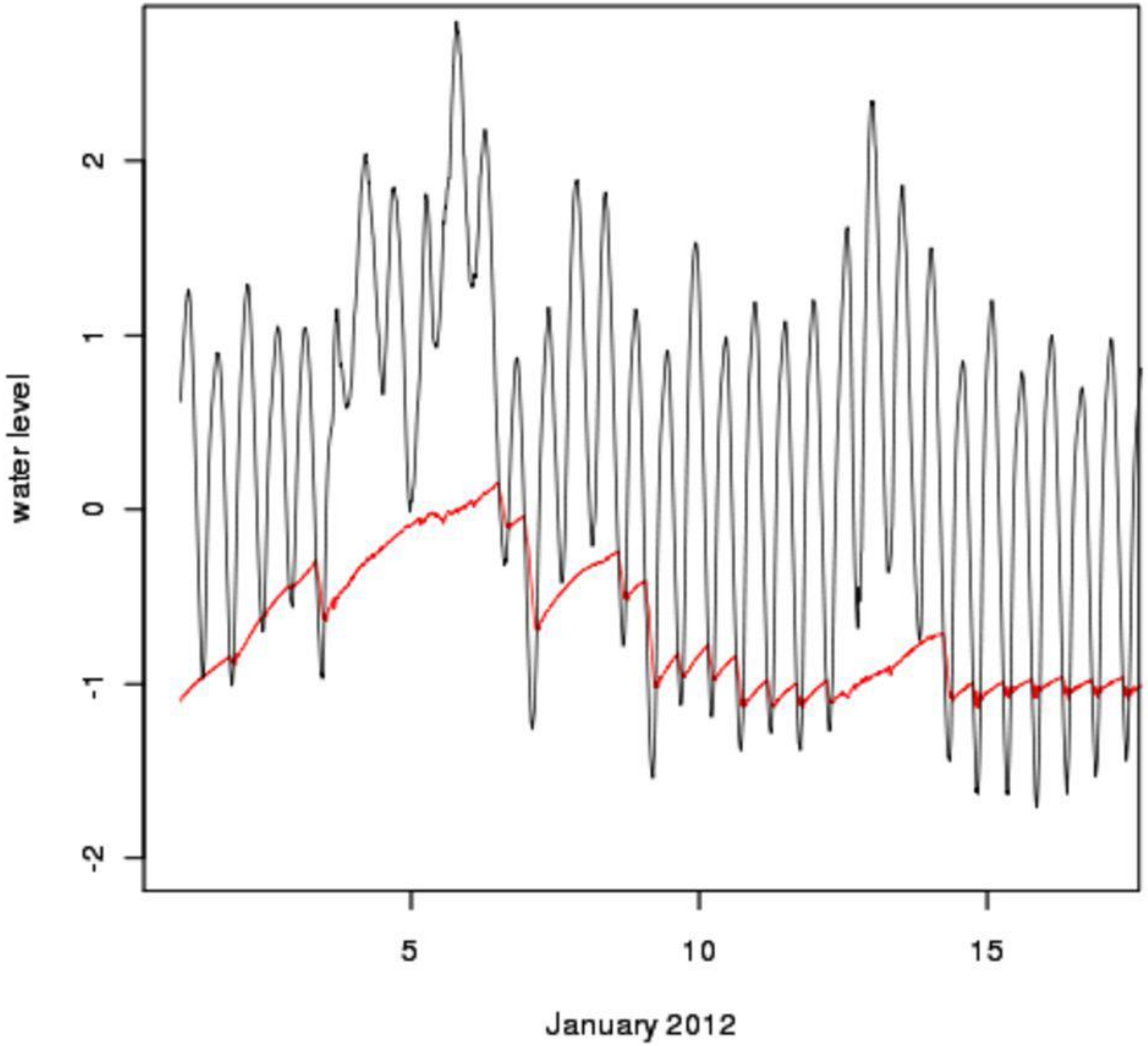


H. de Vries (pers comm)

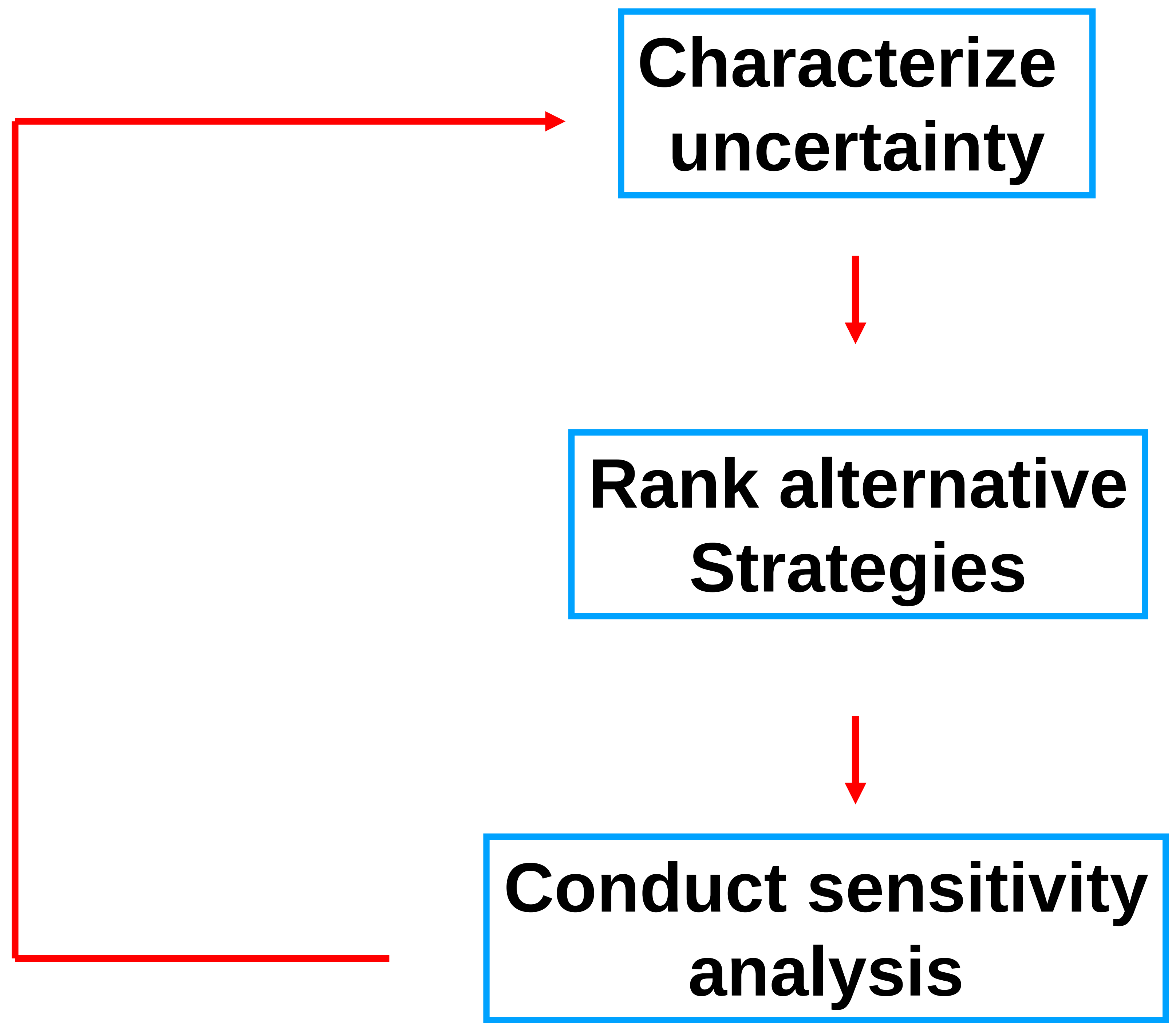
Incident January 2012: evacuation after rising inland water levels (Hazeleger et al 2015)

inland water level

sea level



Predict than act…not suited



Lampert and Groves 2006

nature climate change

PERSPECTIVE

PUBLISHED ONLINE: 28 JANUARY 2015 | DOI: 10.1038/NCLIMATE2450

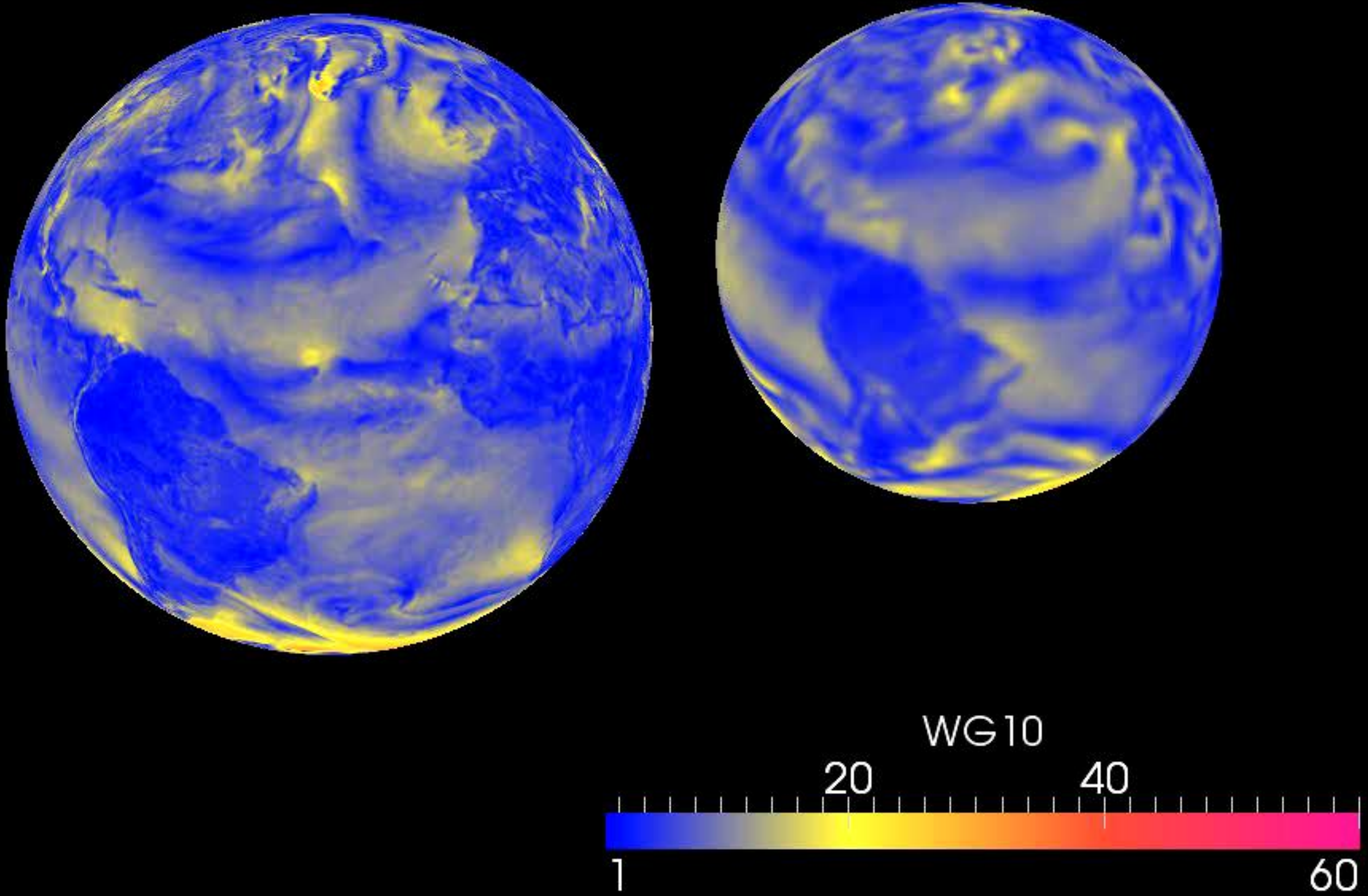
Tales of future weather

W. Hazeleger[1,2,3*], B.J.J.M. van den Hurk[1,4], E. Min[1], G.J. van Oldenborgh[1], A.C. Petersen[4,5], D.A. Stainforth[6,9,10], E. Vasileiadou[4,8] and L.A. Smith[6,7]

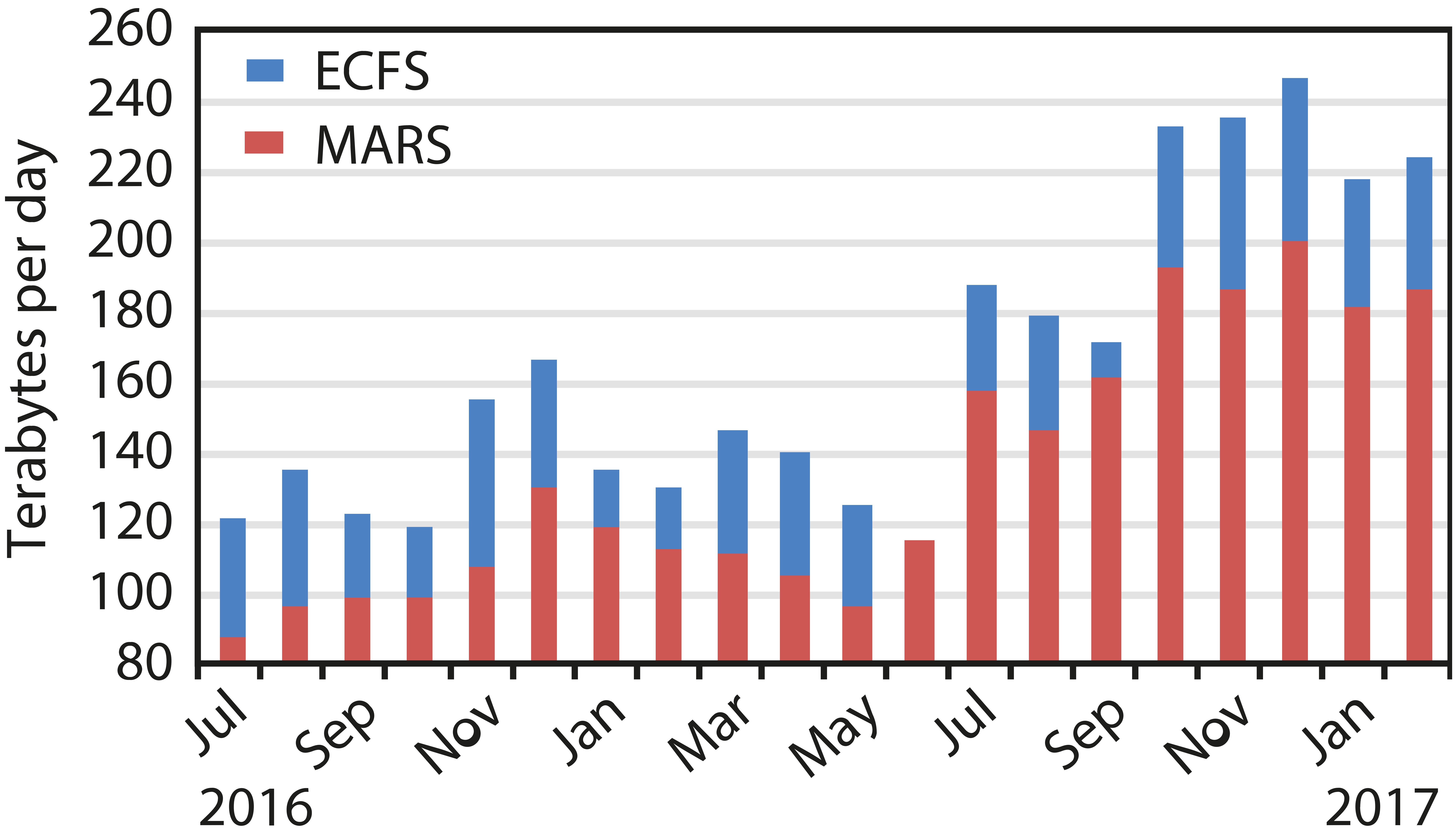
Society is vulnerable to extreme weather events and, by extension, to human impacts on future events. As climate changes weather patterns will change. The search is on for more effective methodologies to aid decision-makers both in mitigation to avoid climate change and in adaptation to changes. The traditional approach uses ensembles of climate model simulations, statistical bias correction, downscaling to the spatial and temporal scales relevant to decision-makers, and then translation into quantities of interest. The veracity of this approach cannot be tested, and it faces in-principle challenges. Alternatively, numerical weather prediction models in a hypothetical climate setting can provide tailored narratives of high-resolution simulations of high-impact weather in a future climate. This 'tales of future weather' approach will aid in the interpretation of lower-resolution simulations. Arguably, it potentially provides complementary, more realistic and more physically consistent pictures of what future weather might look like.

We need an alternative framework to translate the scenarios to the daily lives of users
"Feeding the imagination" not for the sake of forecasting, but preparedness.

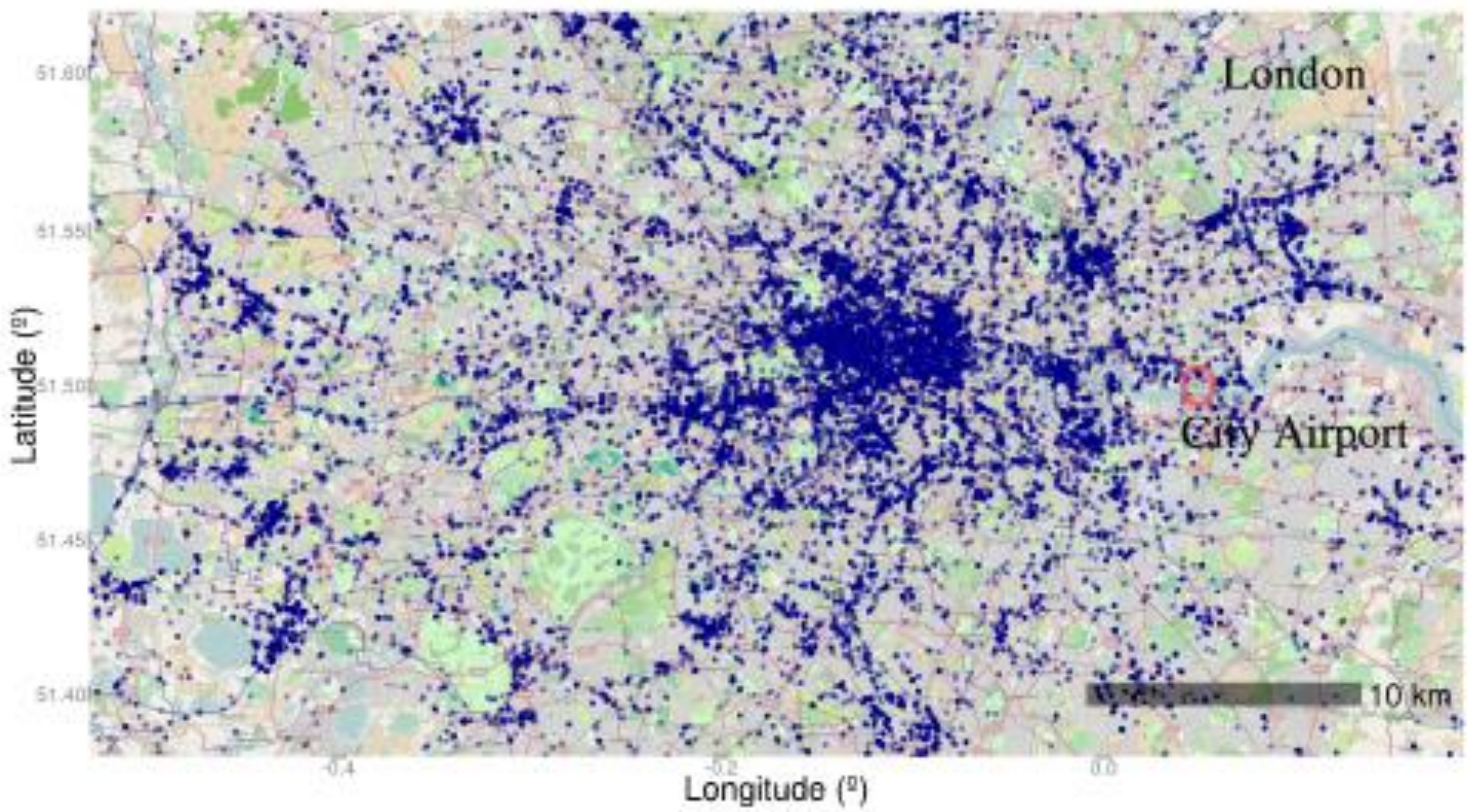
EC EARTH

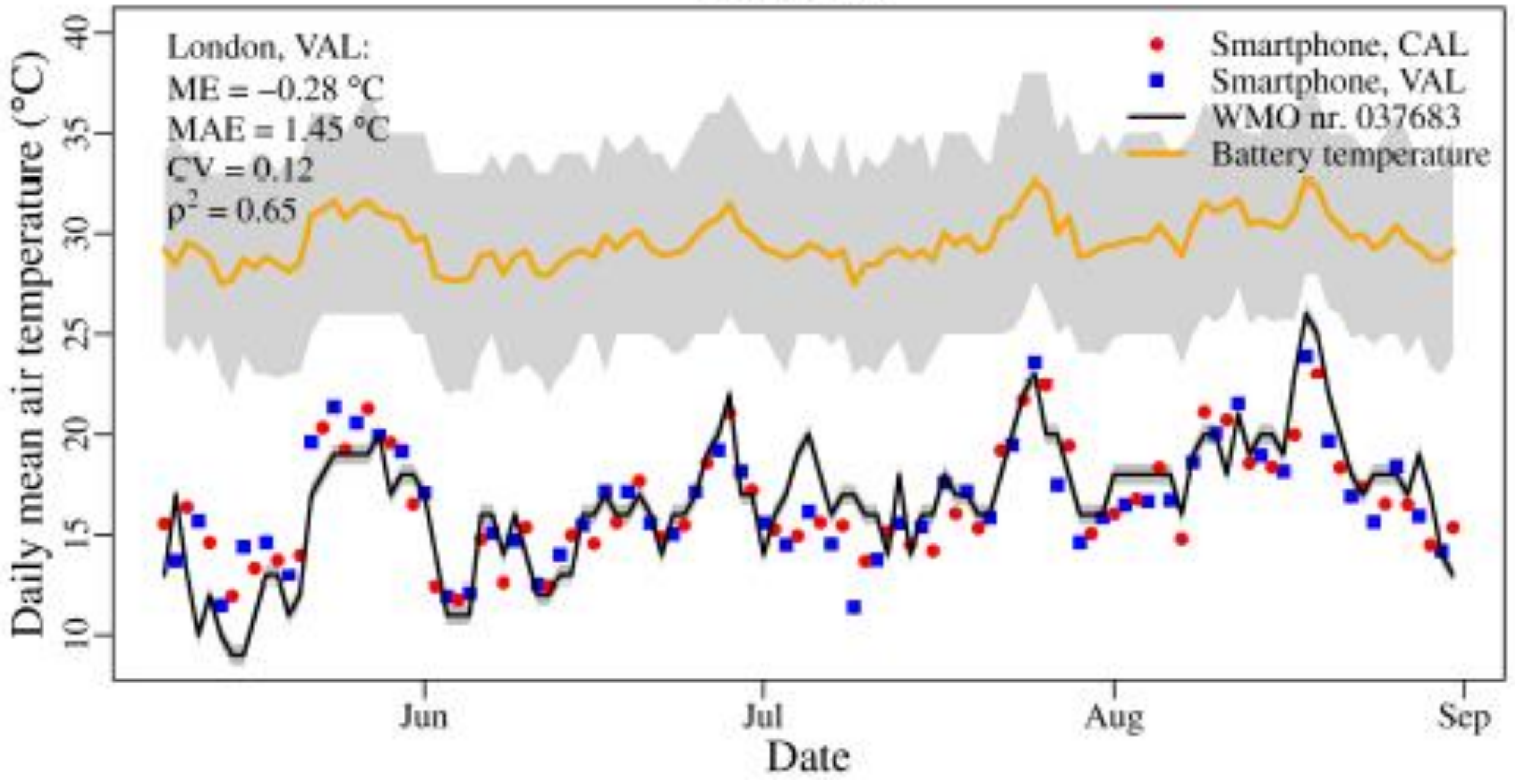


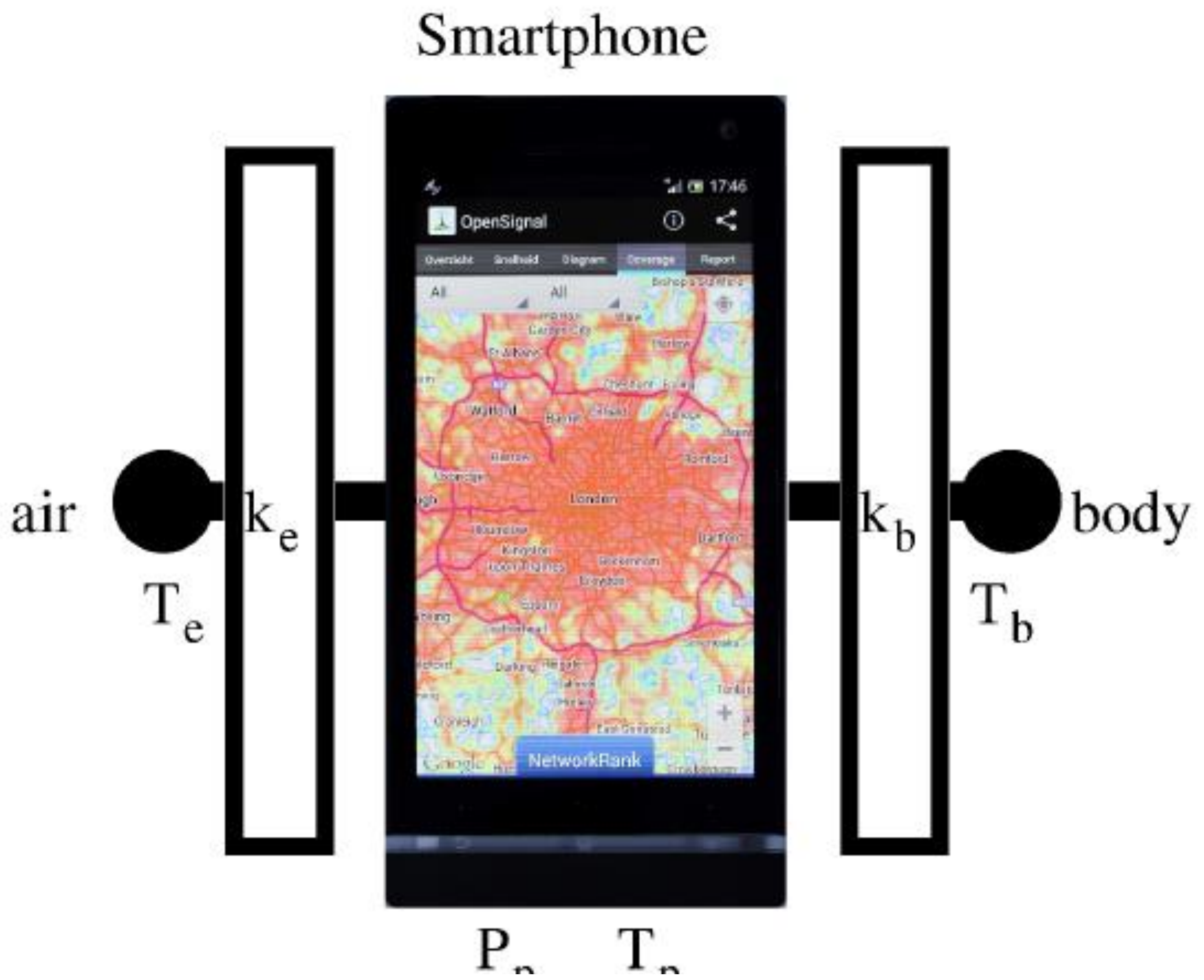
Big science, big data challenge

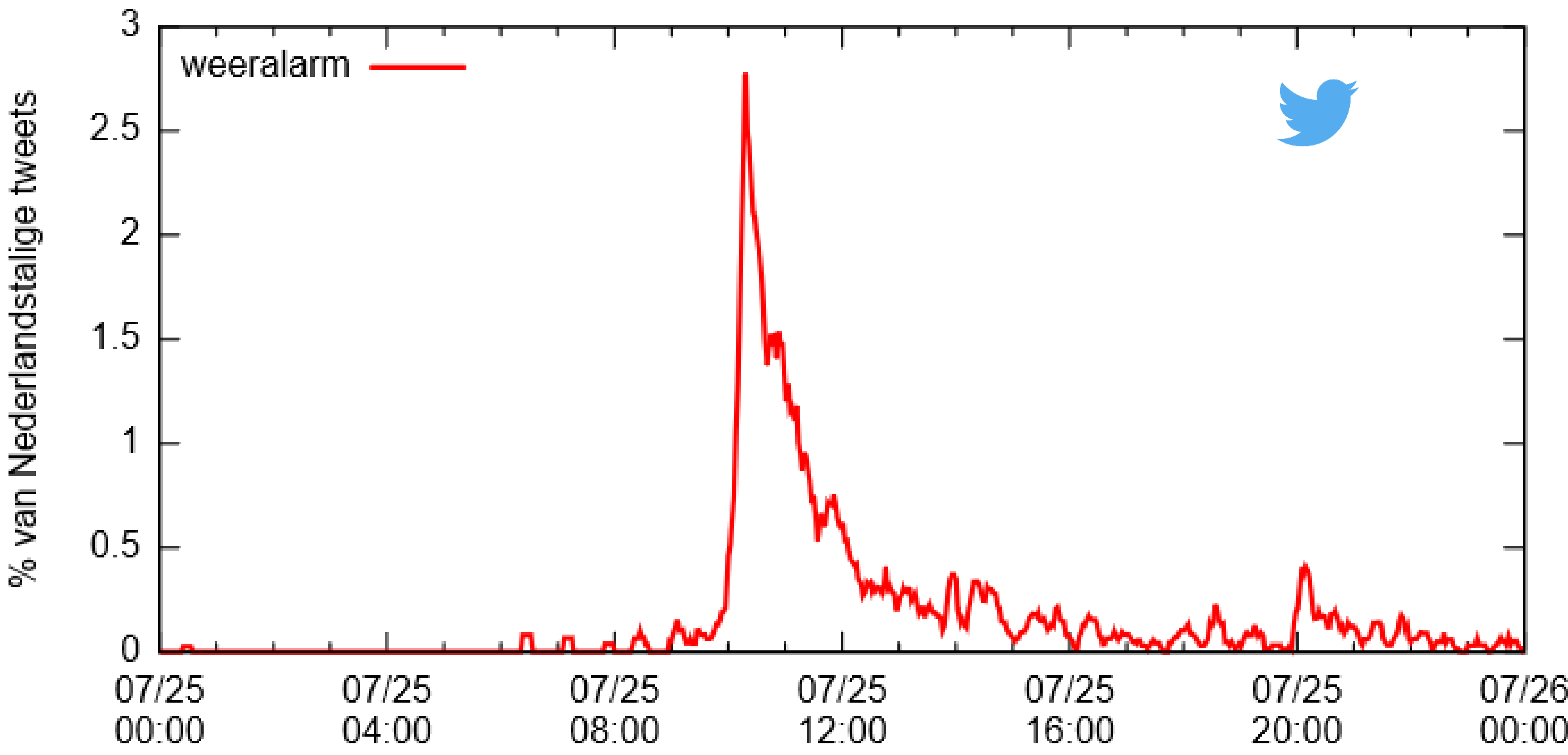

ECFS
MARS
Terabytes per day
260
240
220
200
180
160
140
120
100
80
Jul
Sep
Nov
Jan
Mar
May
Jul
Sep
Nov
Jan
2016
2017

Other big data in meteorology



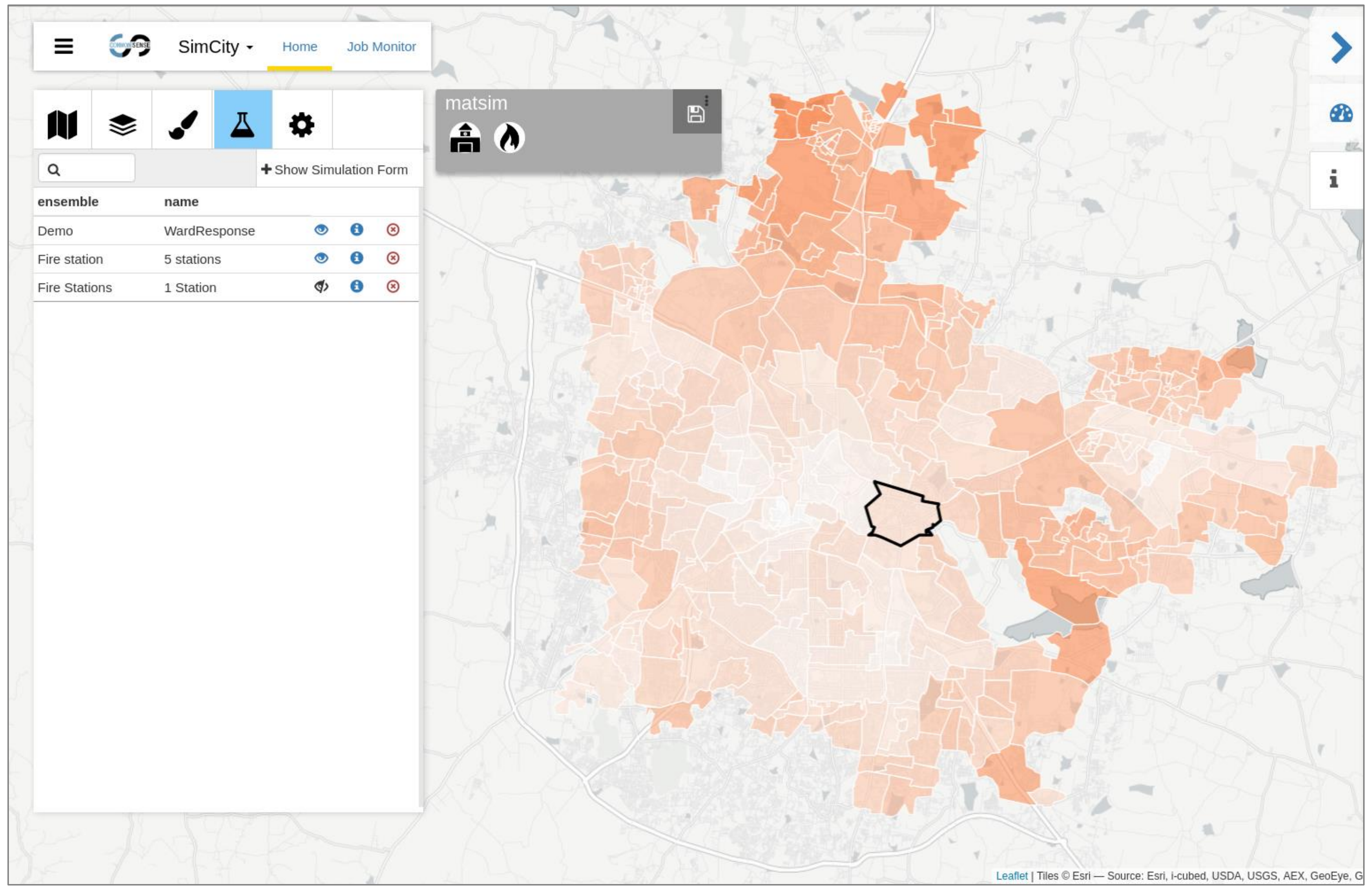




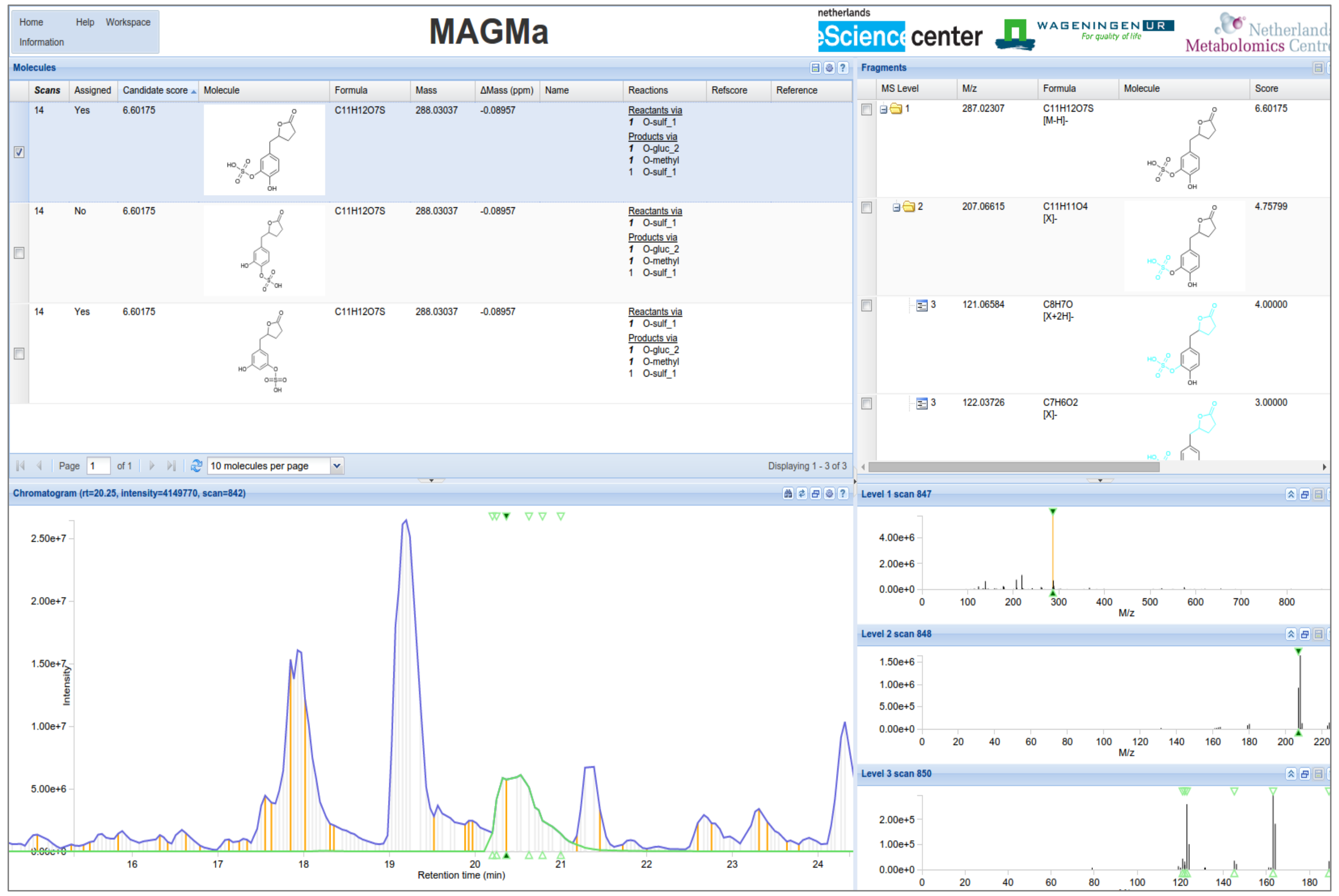


Overeem et al GRL 2013

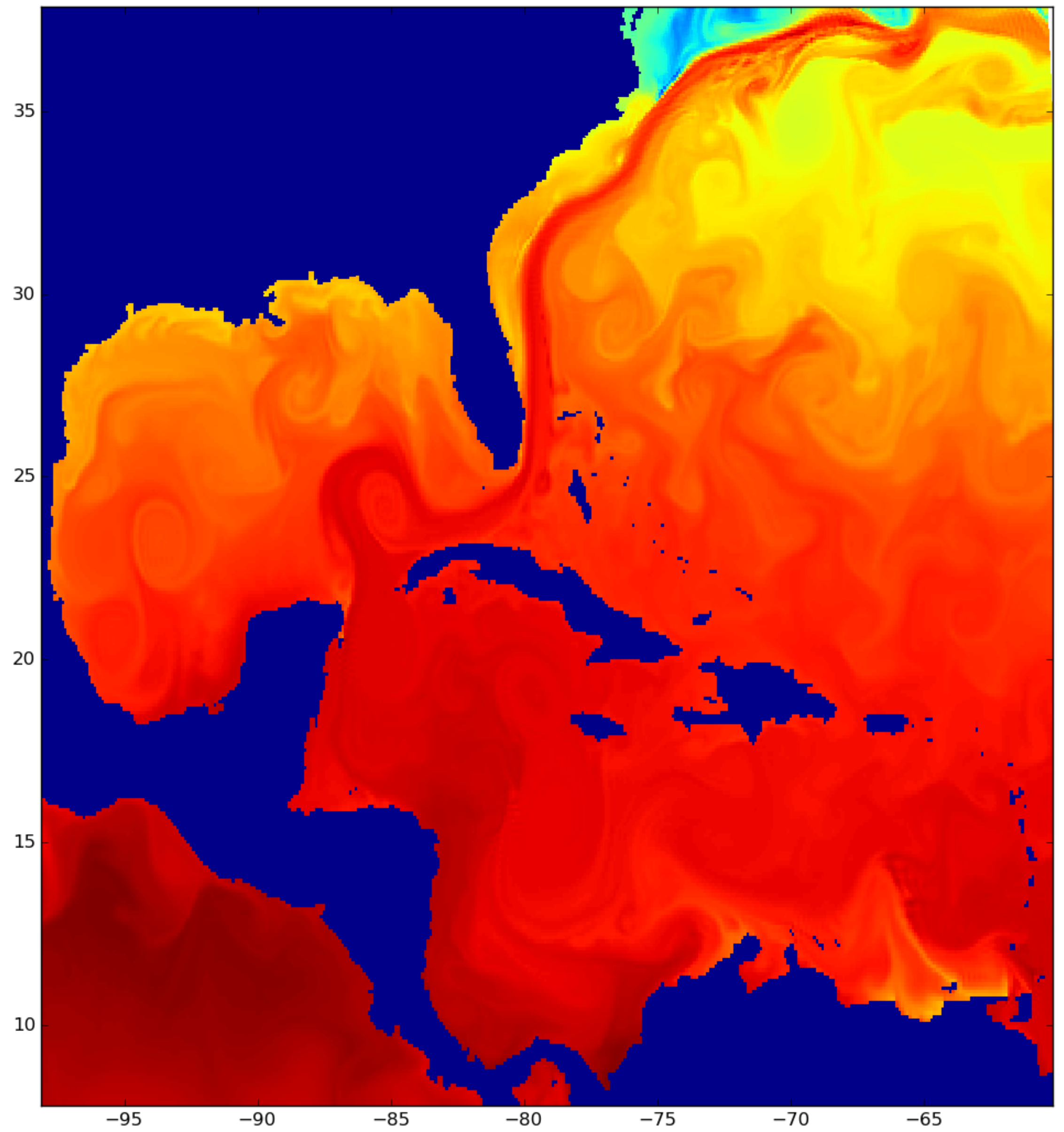
Compute at the touch of a button



SIMCITY



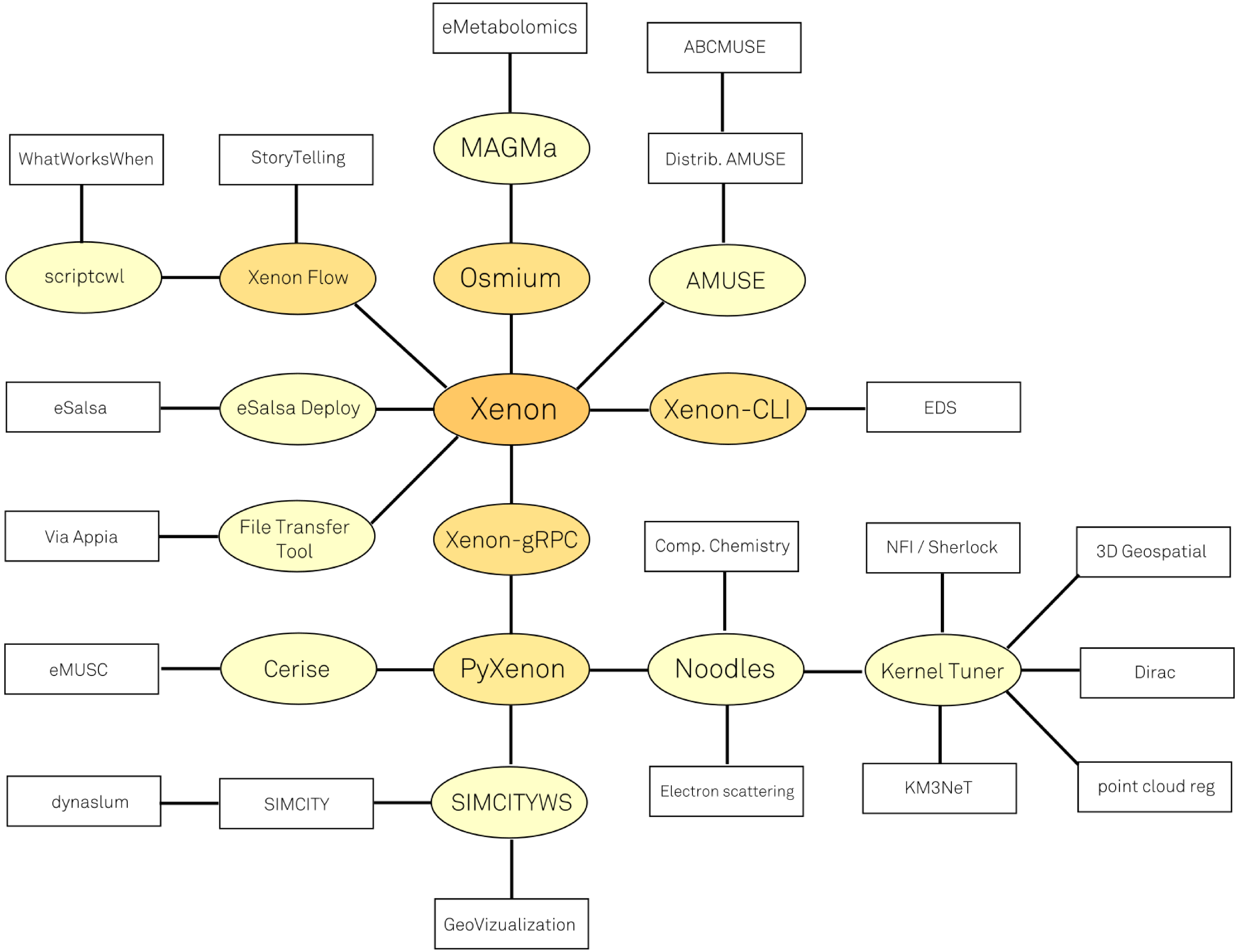
MAGMa



OMUSE

This "deployment and coupling" is a recurring theme in many of our projects: easy access to compute and storage and babysitting applications

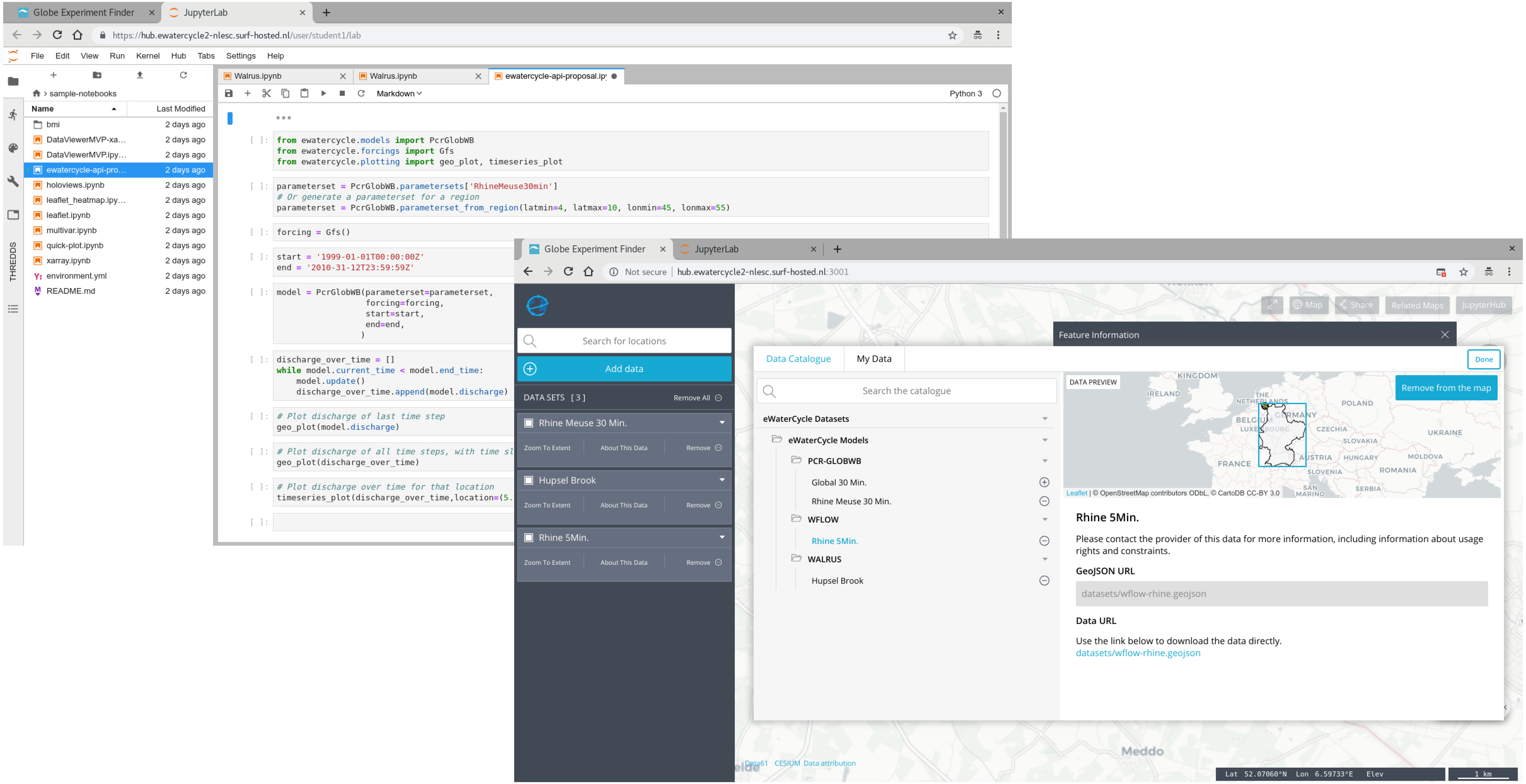
Xenon and friends



Xenon is a software library that provides easy access to compute and storage.

https://github.com/NLeSC/Xenon

Flexible software tools



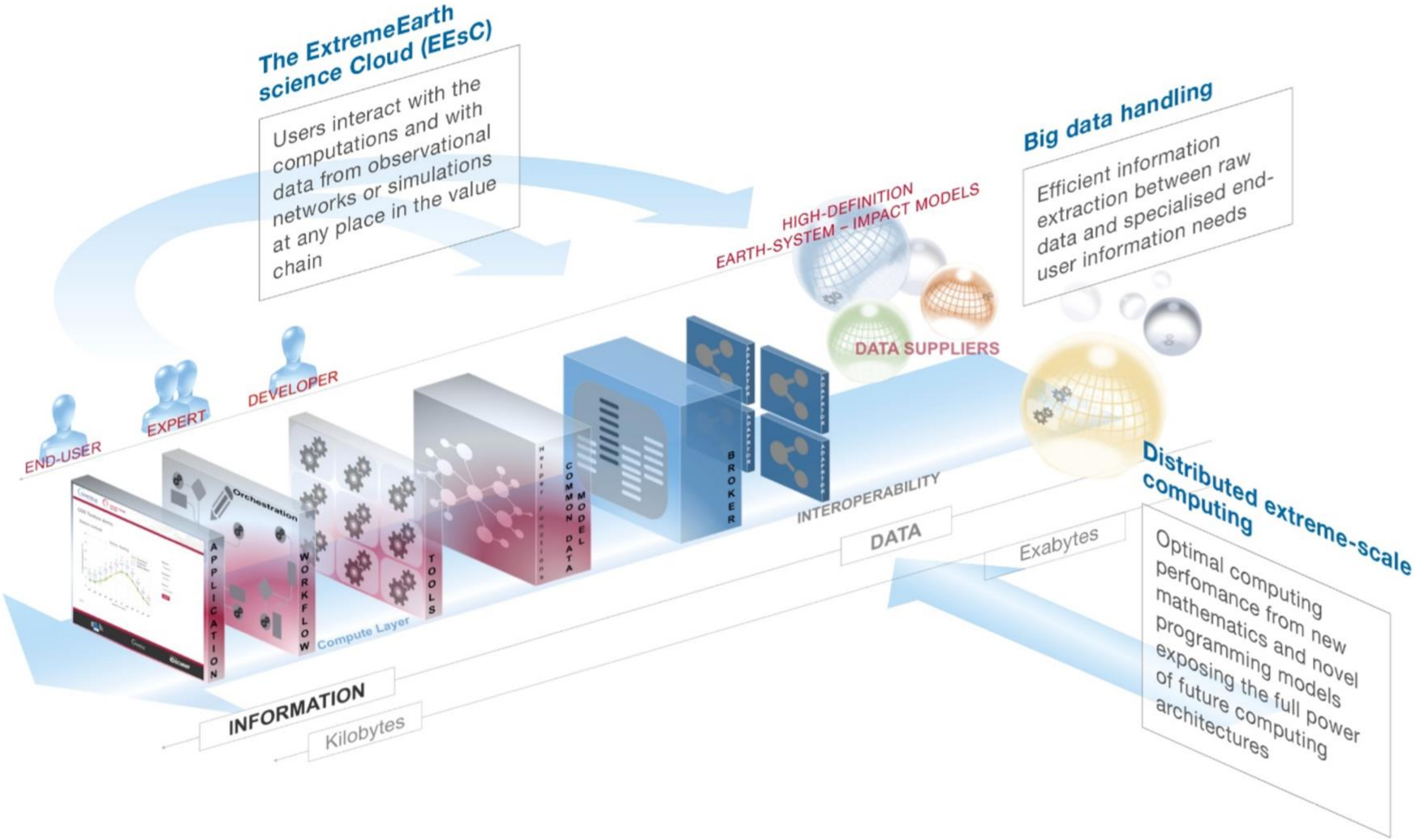
Flexible steering, execution of models and data handling

...

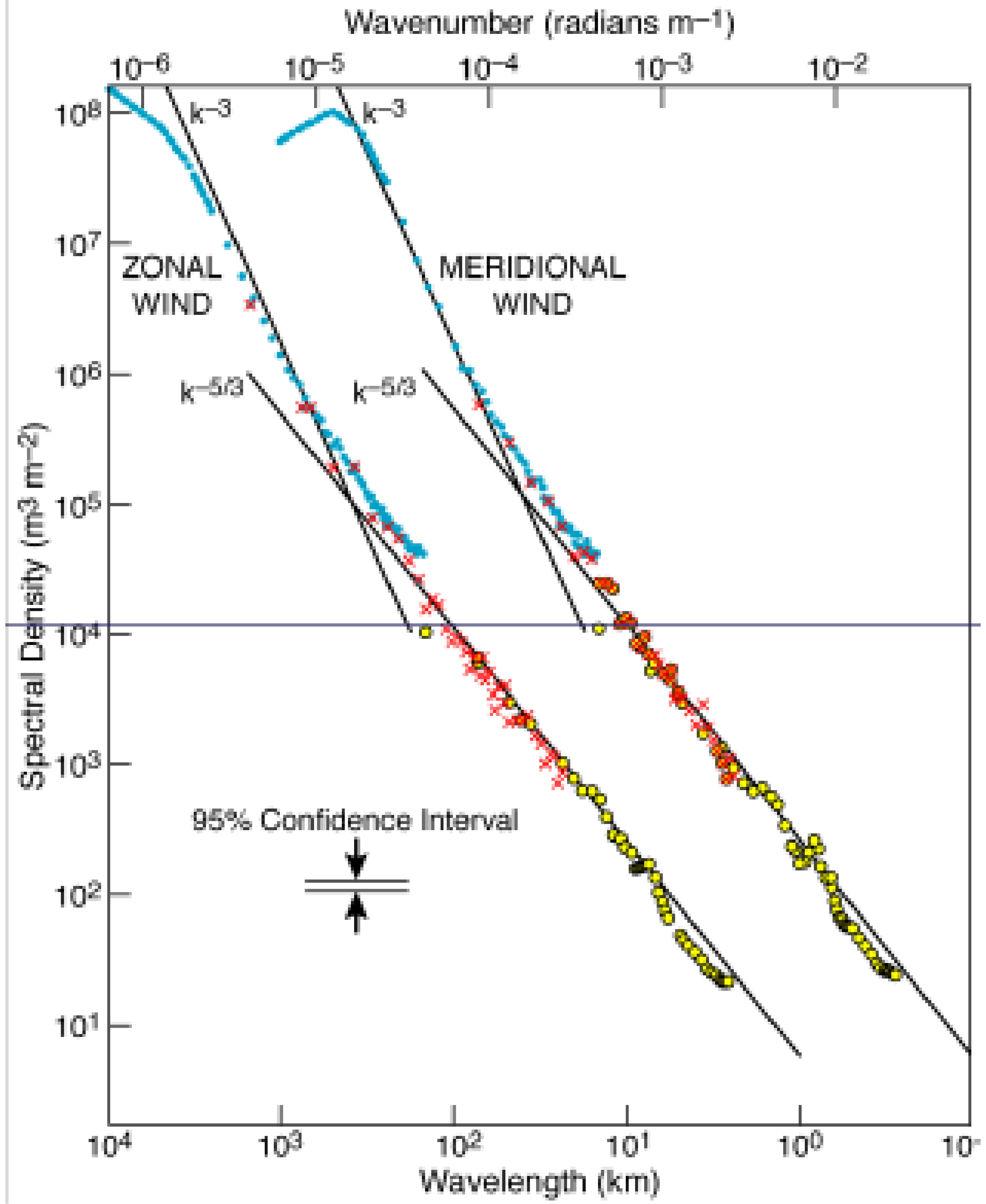
```
[ ]: from ewatercycle.models import PcrGlobWB
     from ewatercycle.forcings import Gfs
     from ewatercycle.plotting import geo_plot, timeseries_plot

[ ]: parameterset = PcrGlobWB.parametersets['RhineMeuse30min']
     # Or generate a parameterset for a region
     parameterset = PcrGlobWB.parameterset_from_region(latmin=4, latmax=10, lonmin=45, lonmax=55)

[ ]: forcing = Gfs()

[ ]: start = '1999-01-01T00:00:00Z'
     end = '2010-31-12T23:59:59Z'

[ ]: model = PcrGlobWB(parameterset=parameterset,
                       forcing=forcing,
                       start=start,
                       end=end,
                      )

[ ]: discharge_over_time = []
     while model.current_time < model.end_time:
         model.update()
         discharge_over_time.append(model.discharge)

[ ]: # Plot discharge of last time step
     geo_plot(model.discharge)
```

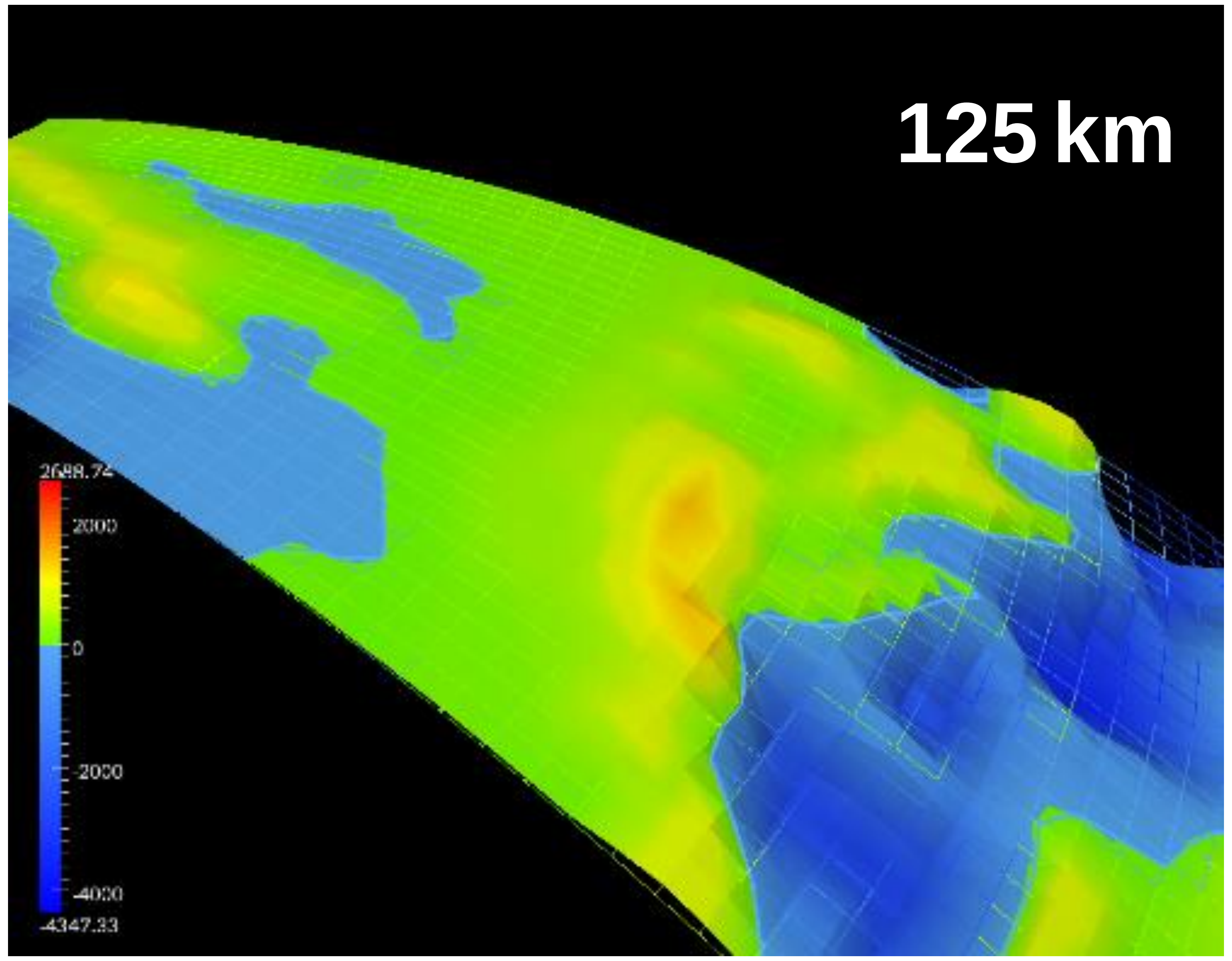
Niels Drost, pers. Comm, NLeSC/TUD/UU/WUR/Deltares eWatercycle II project

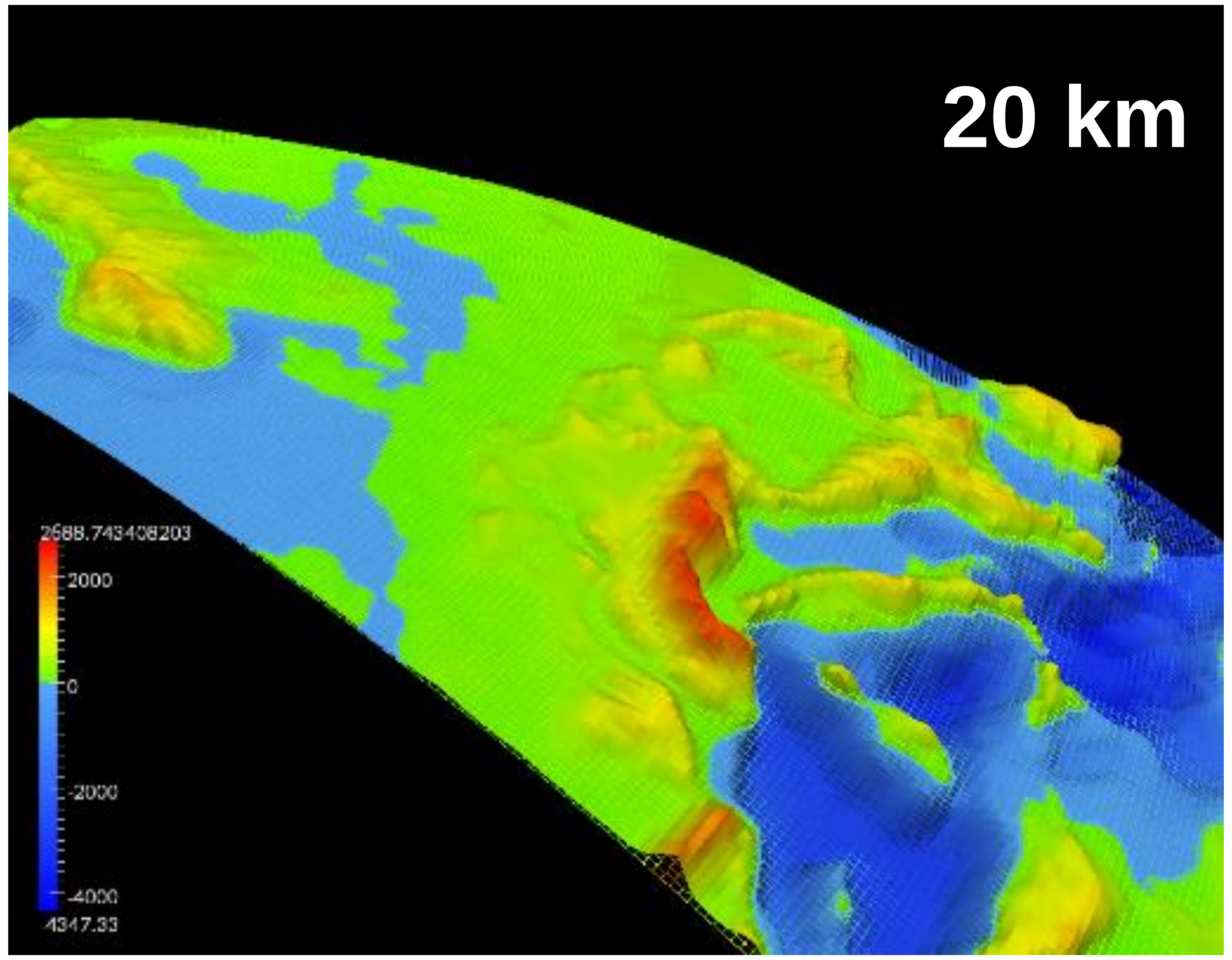
What e-infrastructure does it take?



The need for resolution





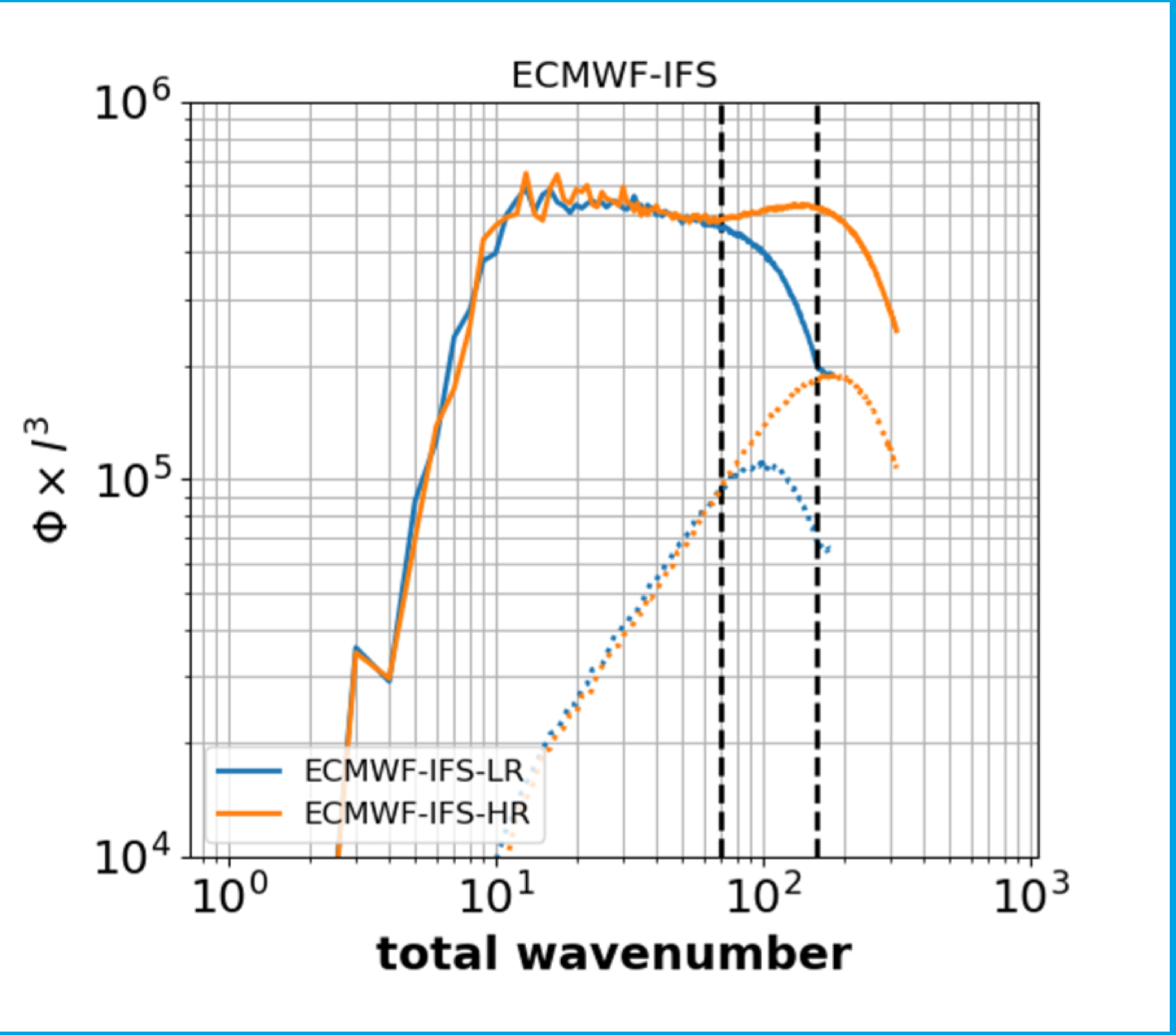


Nastrom & Gage 1985; Hazeleger et al 2012

Kinetic energy spectrum at two grid resolutions

- ECMWF-IFS-LR spectral reduced TCO255 123 290 2.4 ECMWF-IFS-HR spectral reduced TCO511 62.6 125 2.0

- To resolve deep convection, at least factor 10 horizontal resolution (factor 1000 computing) needed



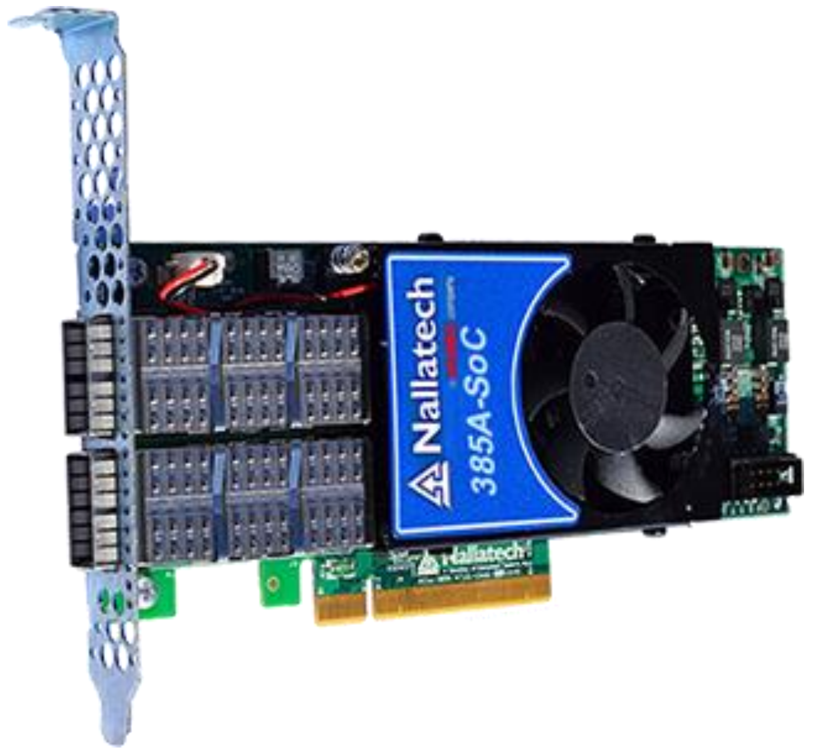
Klaver et al to be submitted

PRIMAVERA

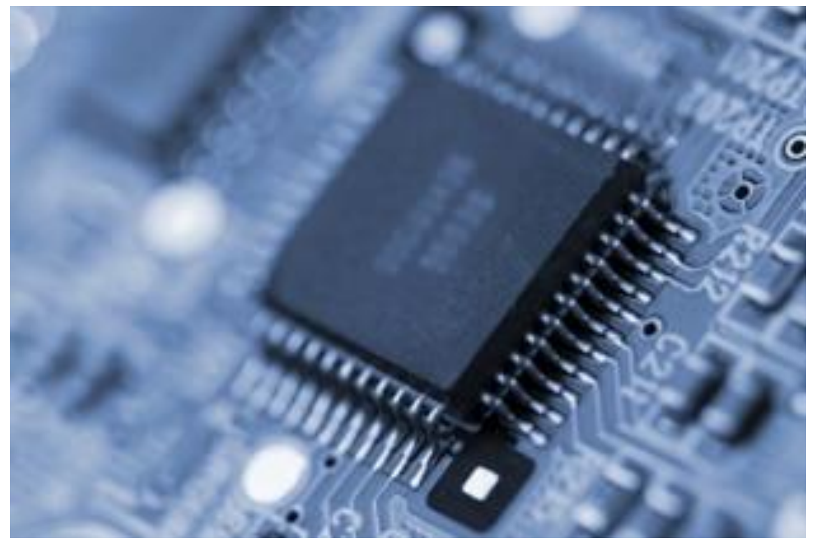
More energy efficient →

CPU

GPU



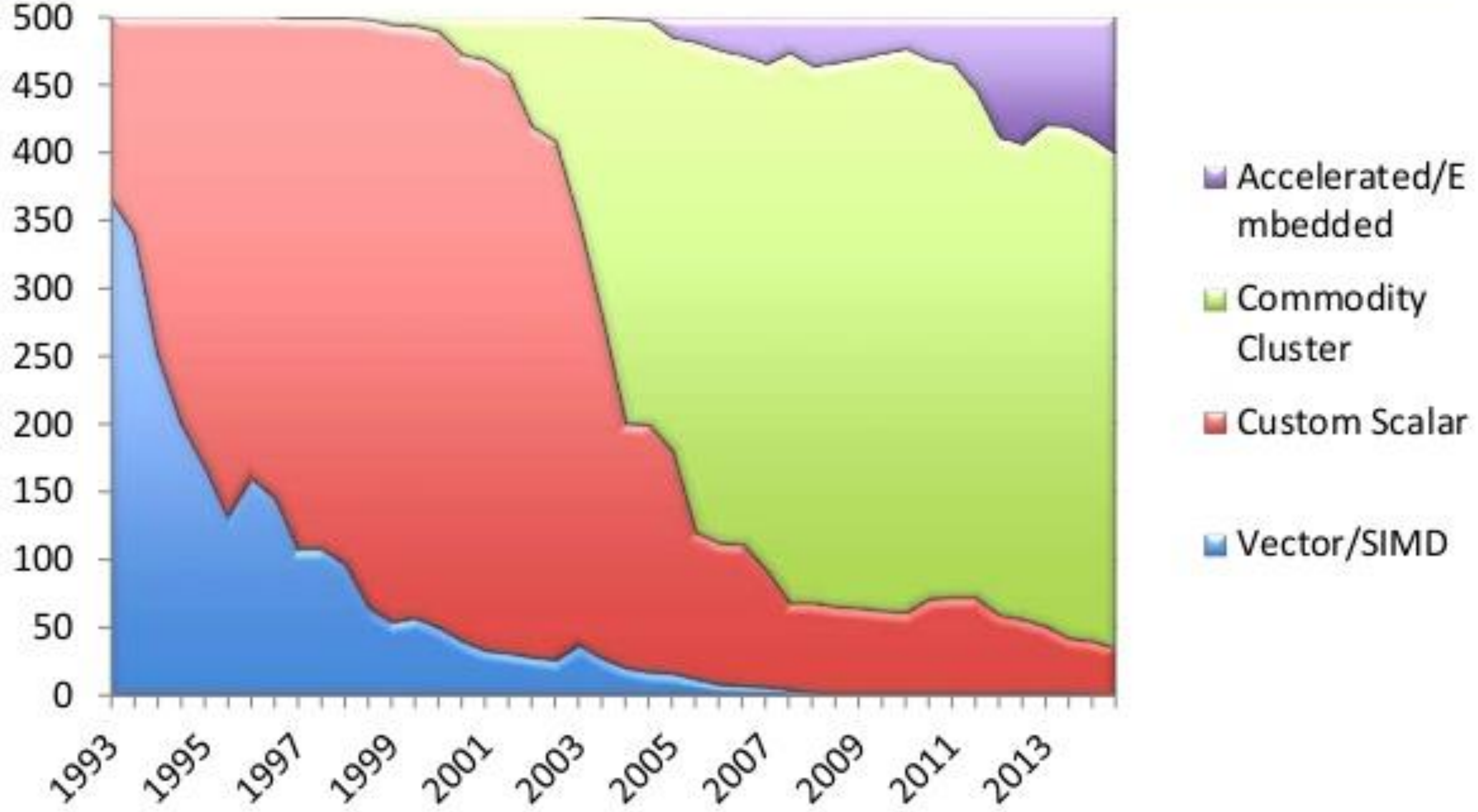
FPGA



ASIC

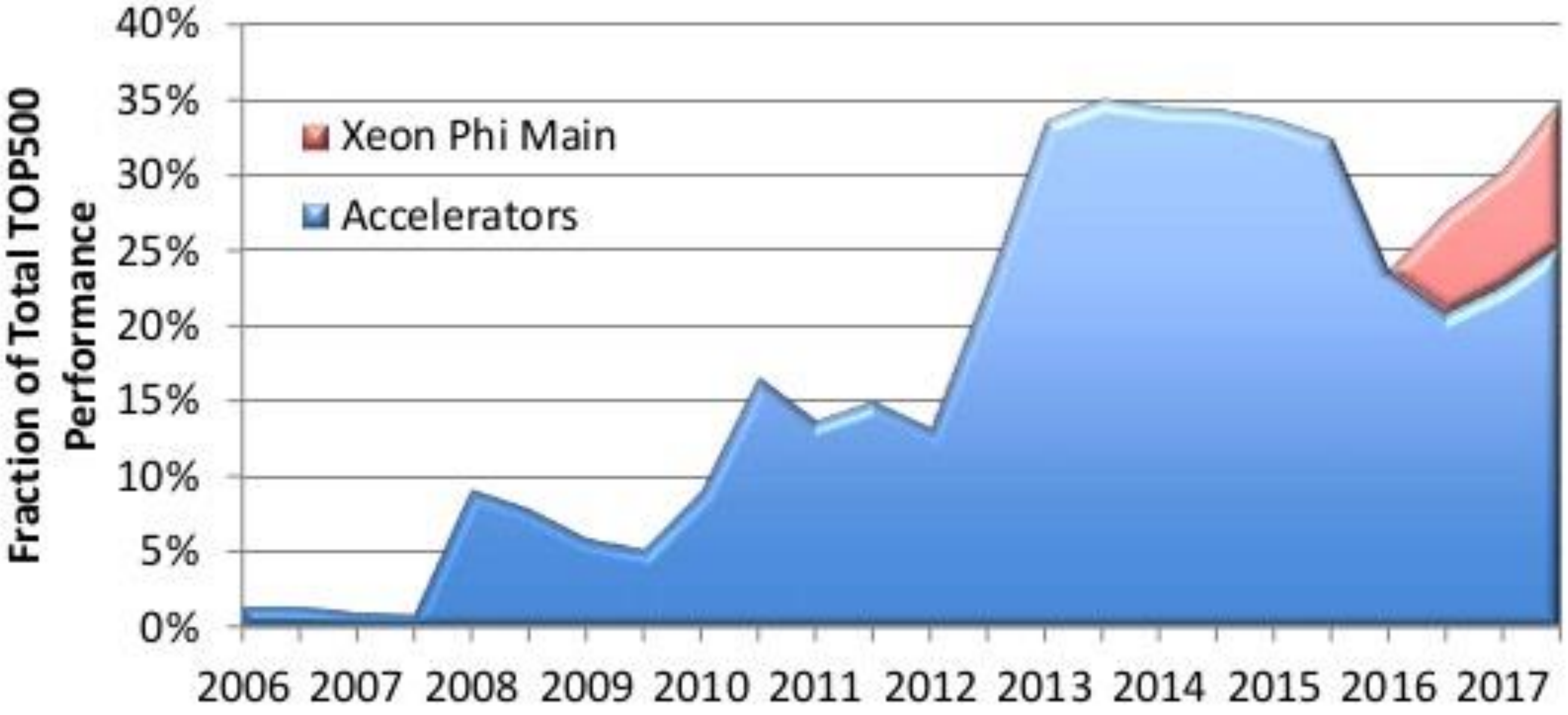
← **Easier to program**

BELL'S LAW



www.top500.orgc

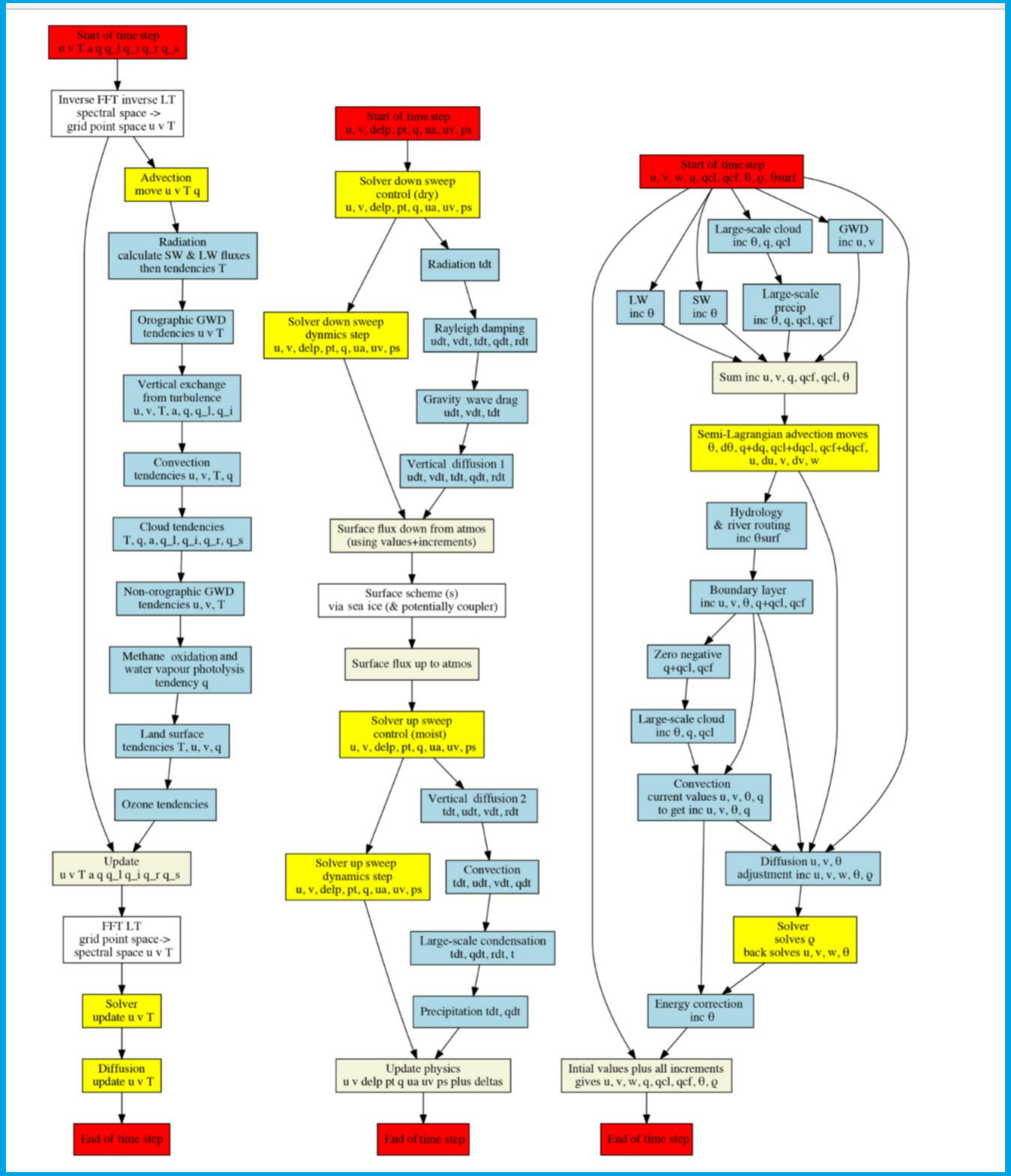
PERFORMANCE SHARE OF ACCELERATORS



www.top500.org

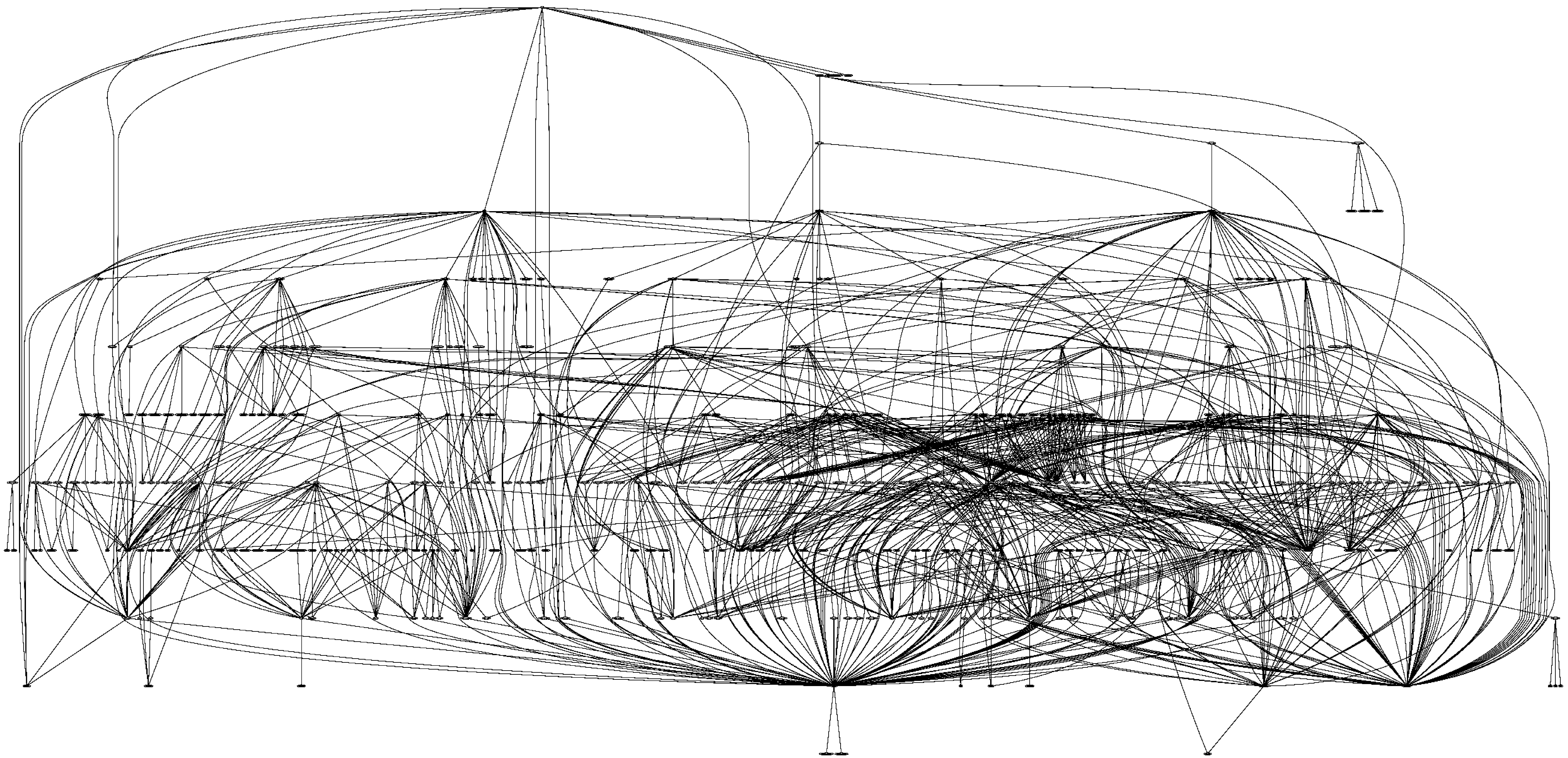
A climate model

- Initialize
- Start loop
  - Dynamics
  - Physics
- Update
- End loop
- I/O

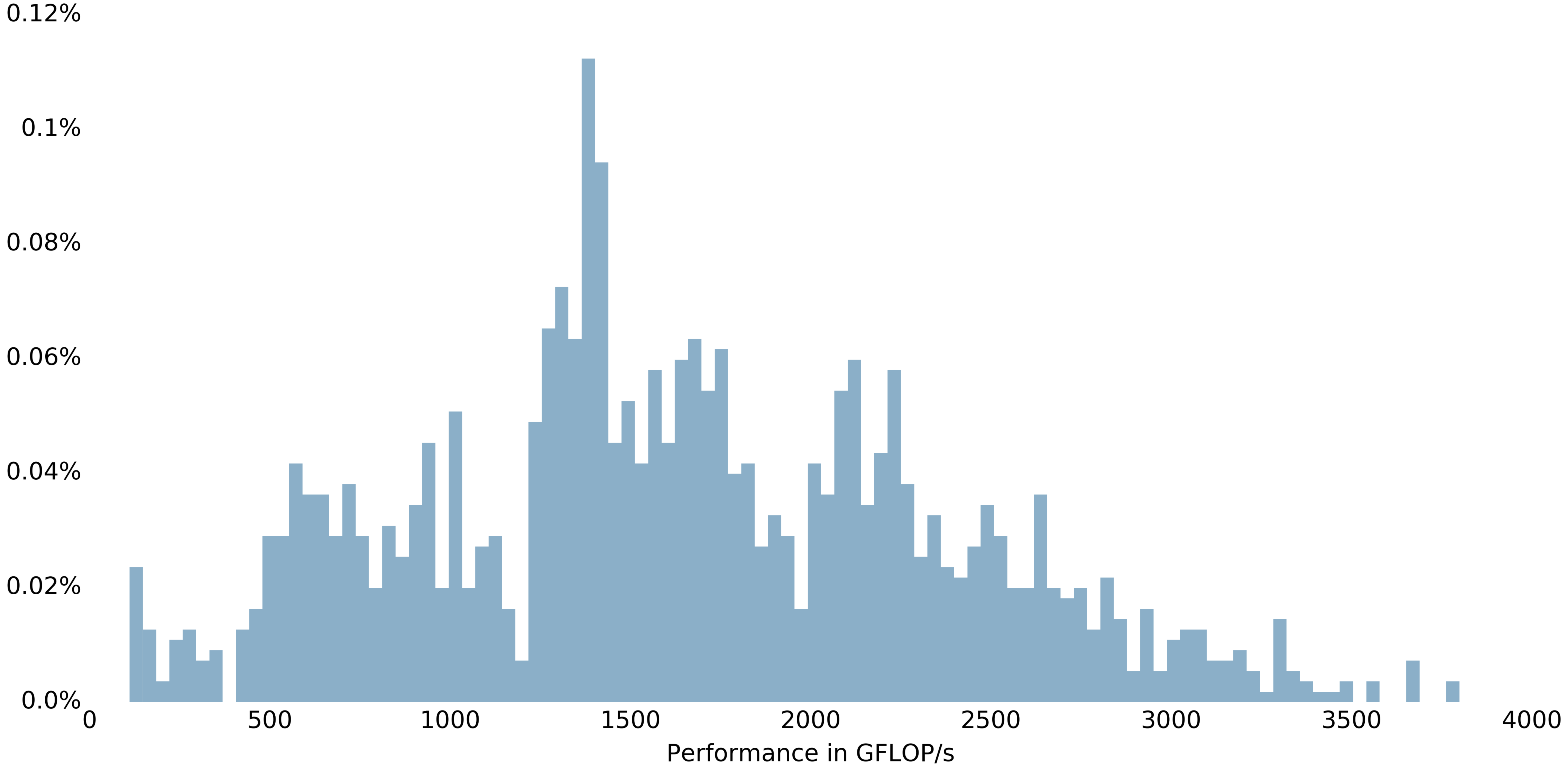


Lawrence et al GMD 2017

Reality check: call graph of ocean GCM POP (courtesy B van Werkhoven)

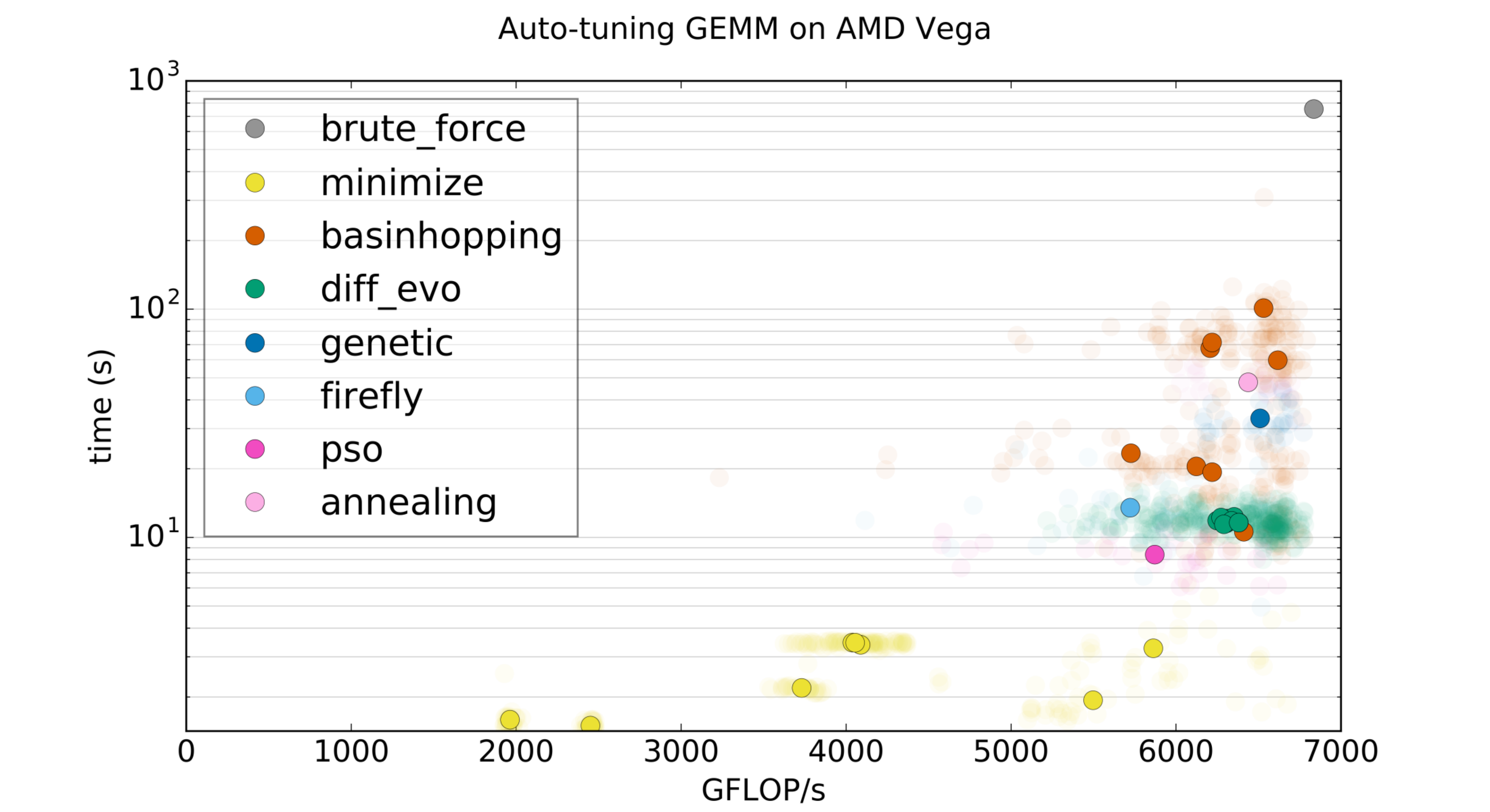


Tuning for performance: 2D Convolution on GTX Titan X (Maxwell; van Werkhoven, FCGS accepted)



Optimized GPU code requires that you get all the details exactly right:

- Mapping of the problem to threads and thread blocks
- Thread block dimensions
- Data layouts in the different memories
- Tiling factors
- Loop unrolling factors
- How to overlap computation and communication
- ...



Problem:

Creates a very large and discontinuous search space

http://benvanwerkhoven.github.io/kernel_tuner/

FGCS, accepted for publication

Project output: AA-Alert exploration radio transient sky

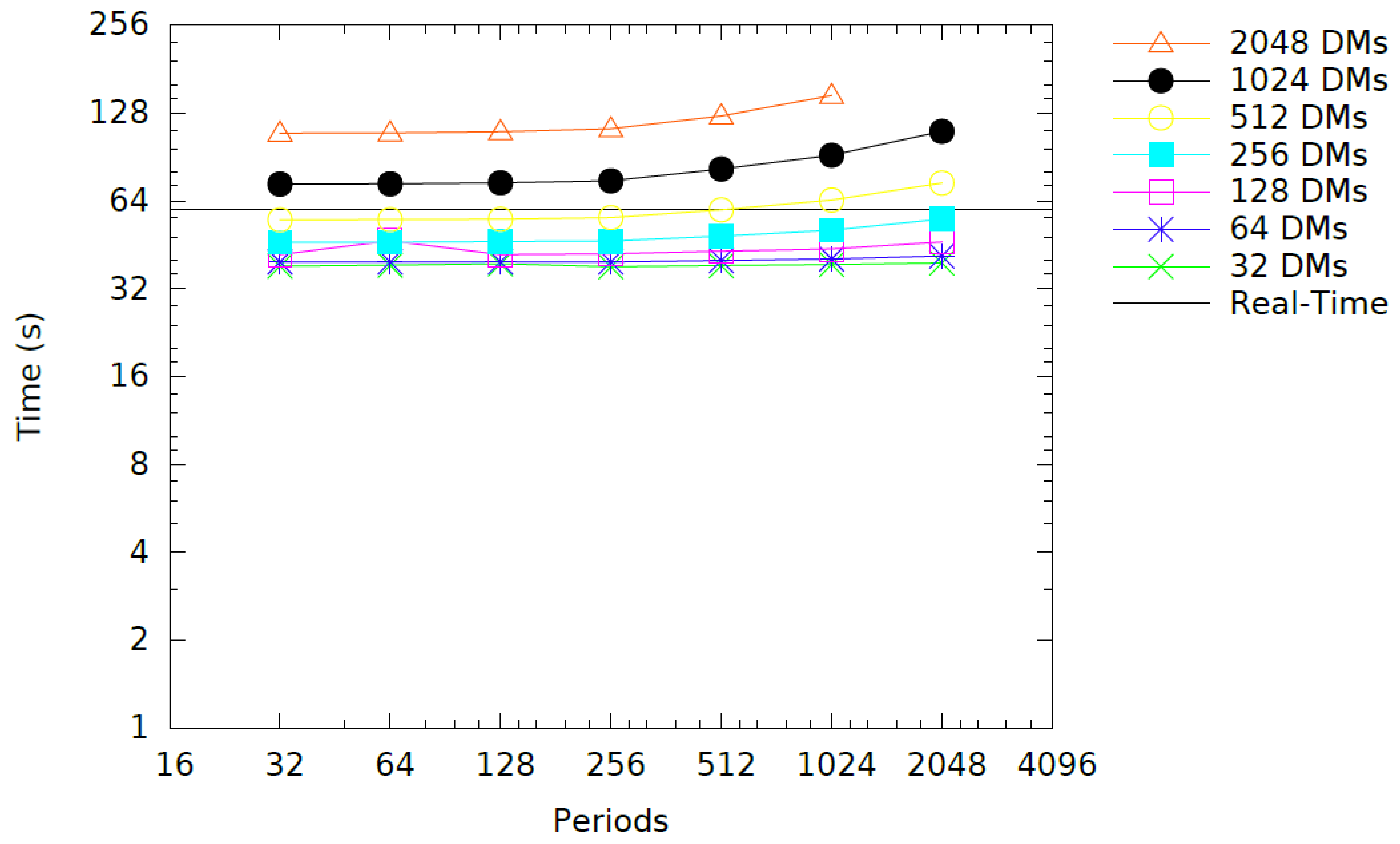
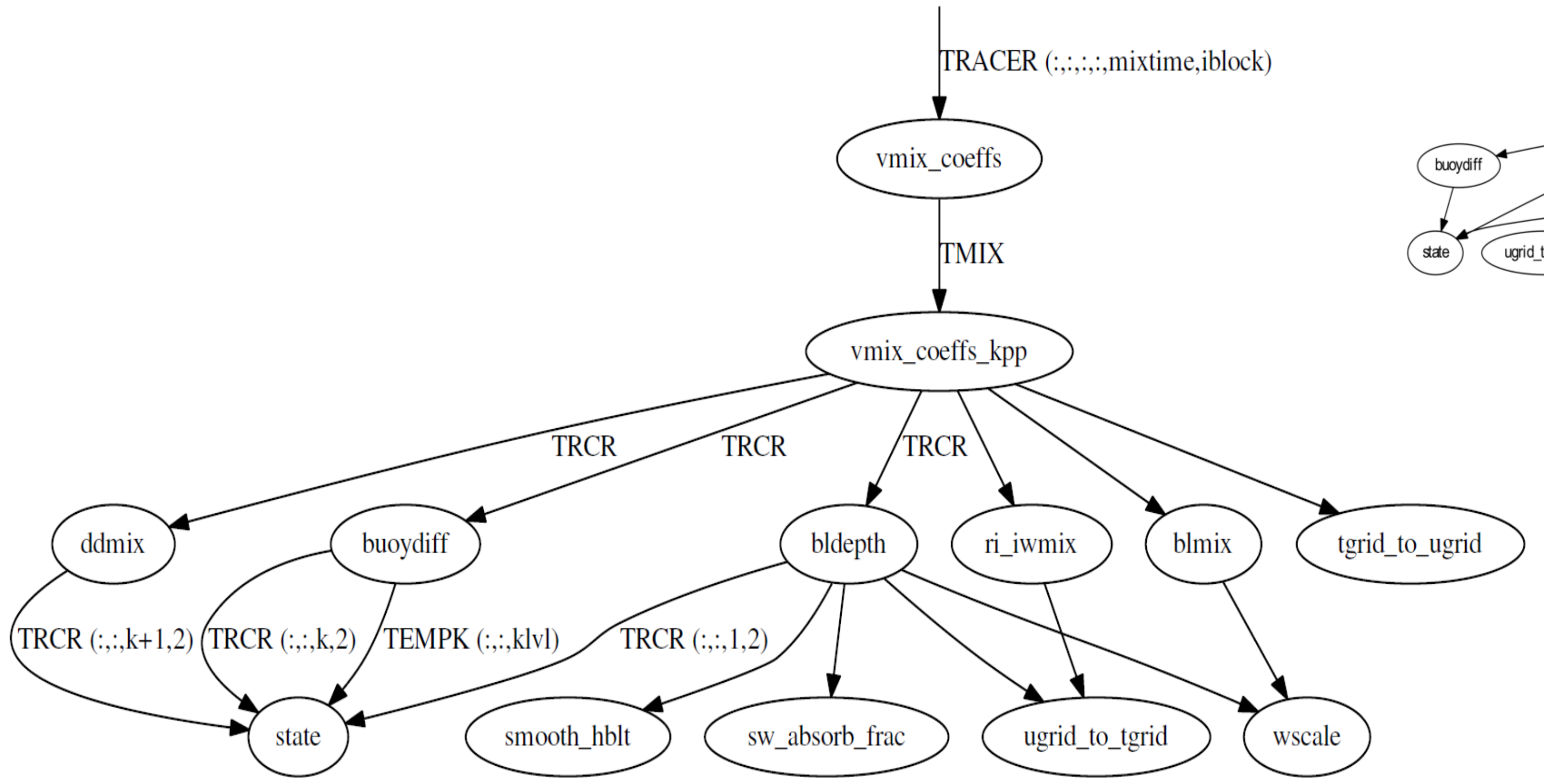


Fig. 5. Pipeline performance for the AMD HD7970, SKA1 scenario.

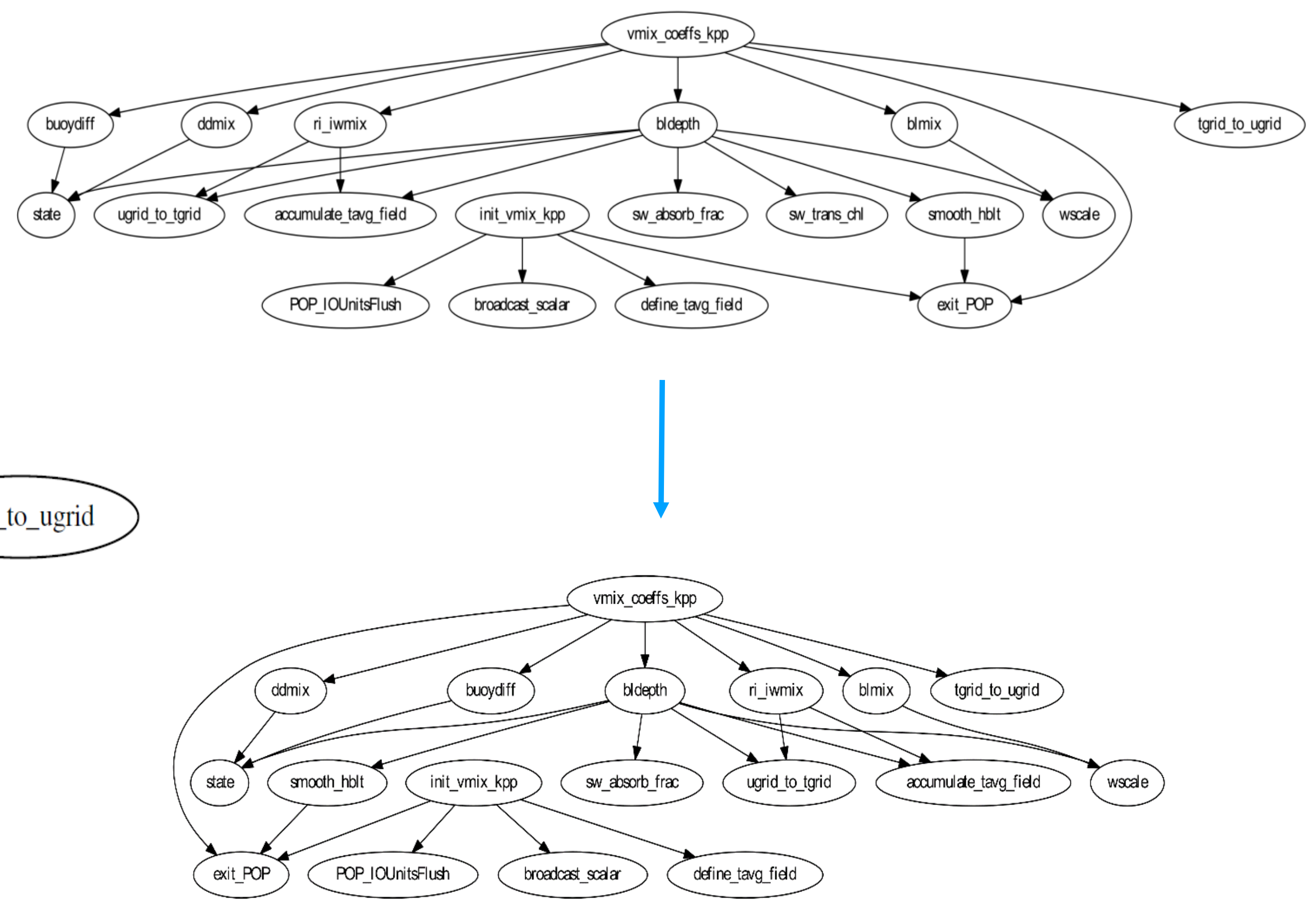
Sclocco et al 2015

Marver

source code analysis



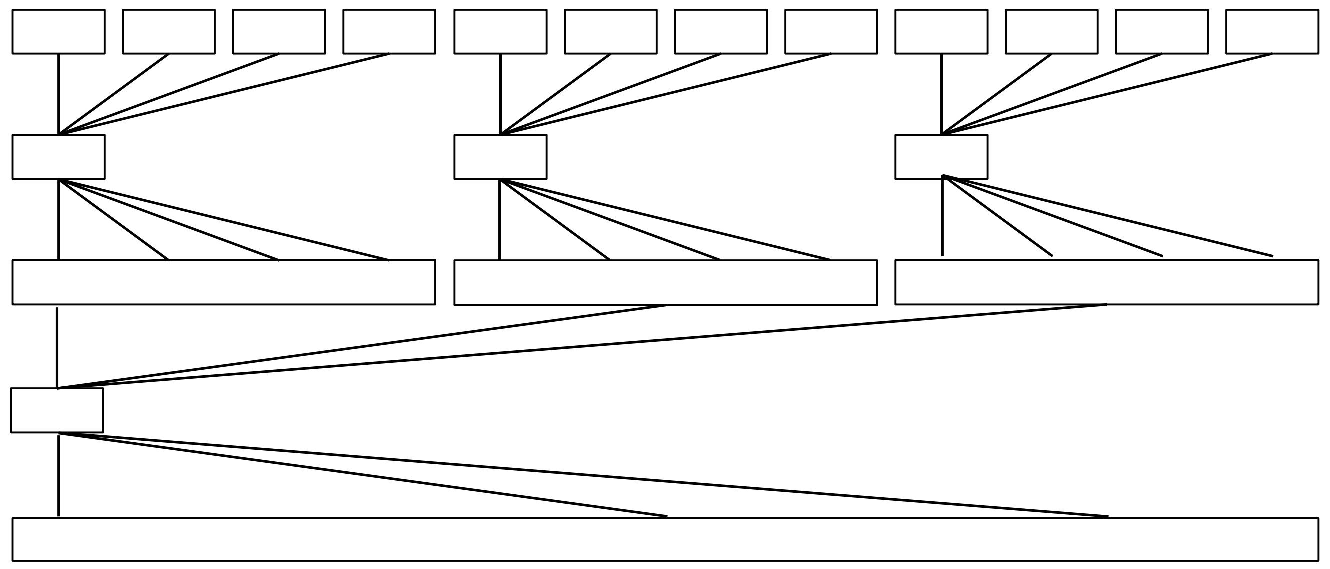
transformation

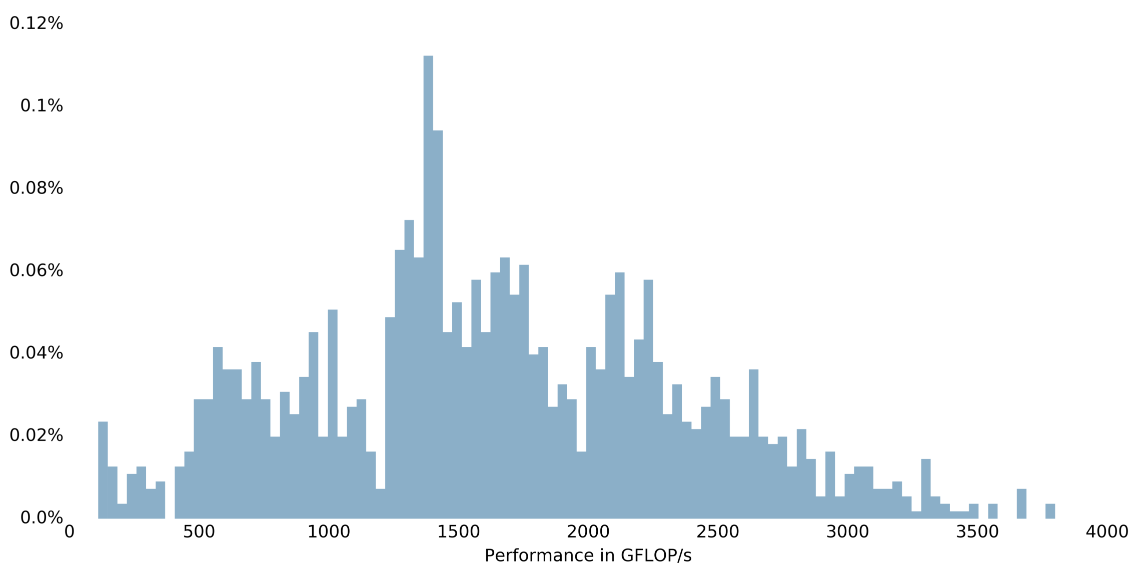


source-to-source translation

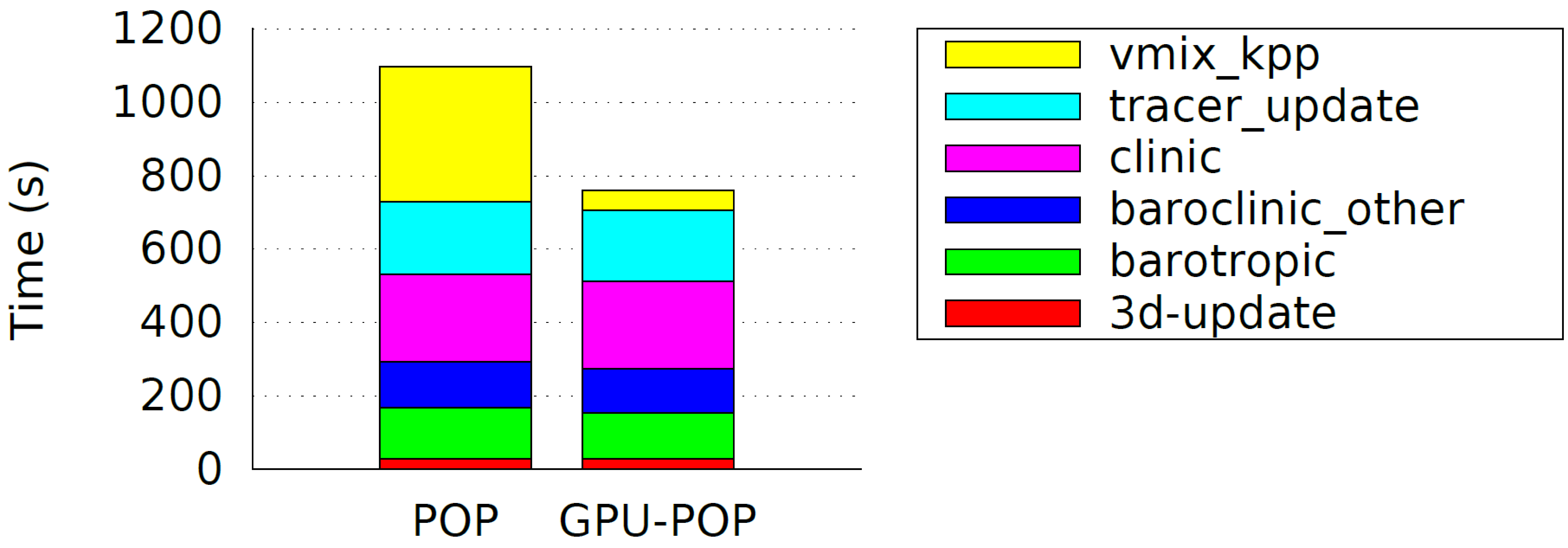
```
VVC(:,:,k) = merge(WORK2, c0,
                   (k < KMU(:,:,bid)))

VVC(i,j,k) = ((k < KMU(i,j,bid) ) ?
                   WORK2 : c0);
```

Kernel tuner






```
kernel_string = """
__global__ void vector_add(float *c, float *a, float *b,
int n) {
    int i = blockIdx.x * block_size_x + threadIdx.x;
    if (i<n) {
        c[i] = a[i] + b[i];
    }
}"""

n = numpy.int32(1e7)
a = numpy.random.randn(n).astype(numpy.float32)
b = numpy.random.randn(n).astype(numpy.float32)
c = numpy.zeros_like(b)
args = [c, a, b, n]
params = {"block_size_x" : 512 }

answer = kernel_tuner.run_kernel("vector_add",
kernel_string, n, args, params)
assert numpy.allclose(answer[0], a+b, atol=1e-8)
```



Dijktsra, Bal, Werkhoven et al

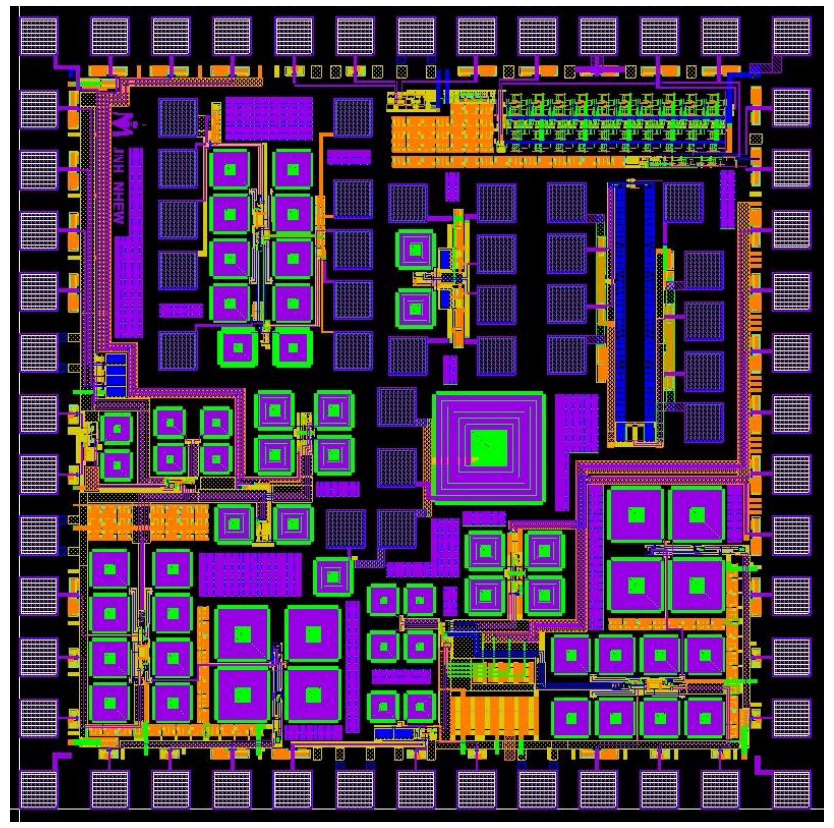
Reconfigurable circuit (no instruction set!)

Very low latency

Built in floating point operations

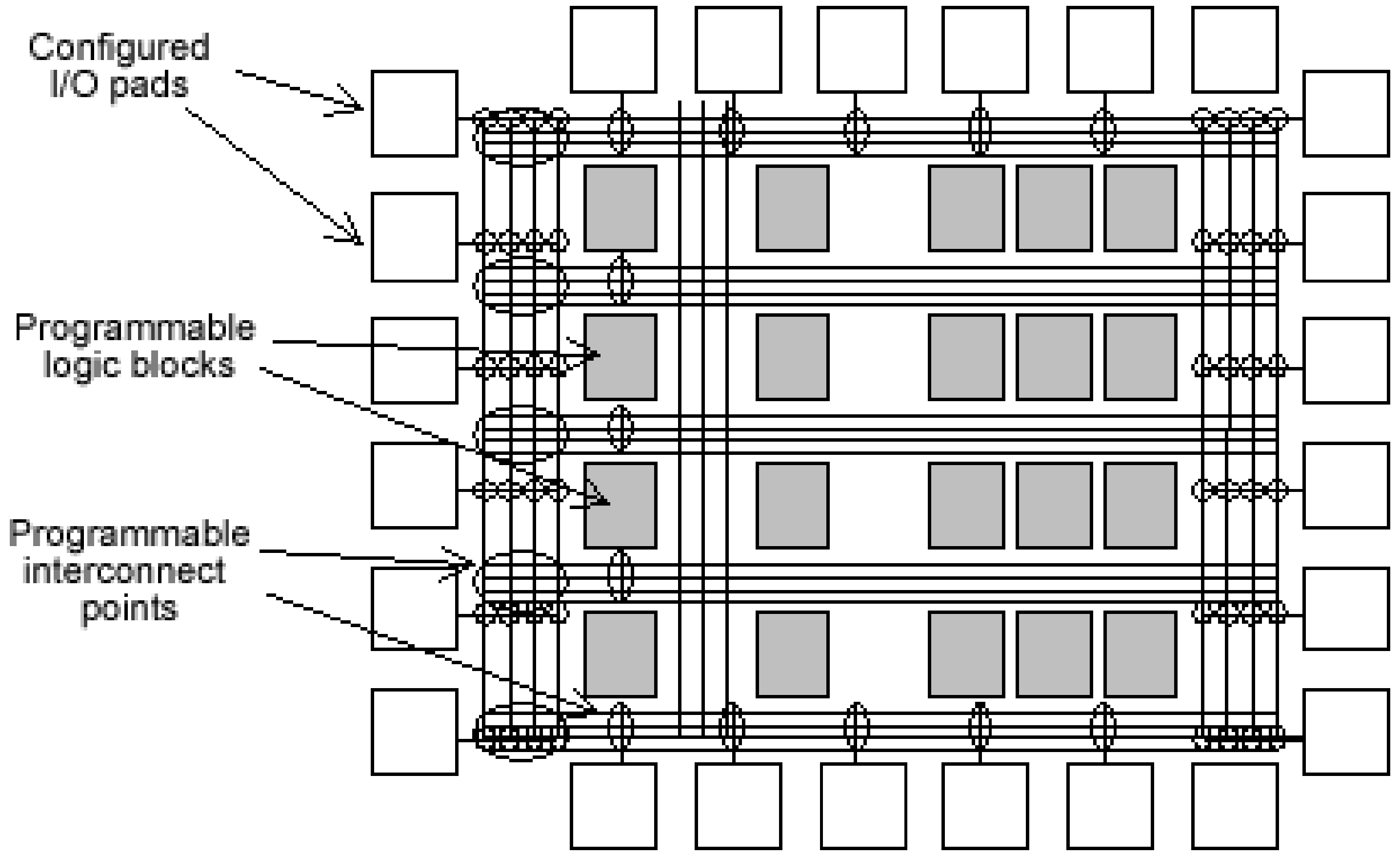
CPU on FPGA board (high bandwidth)

Gigabit Ethernet on board

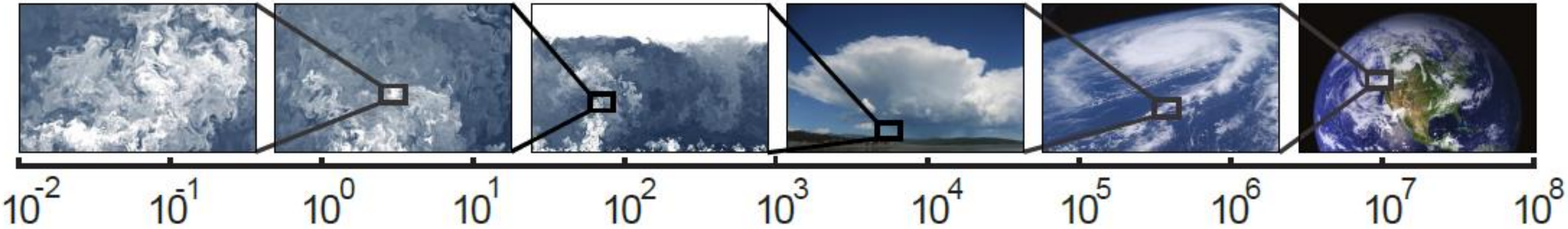


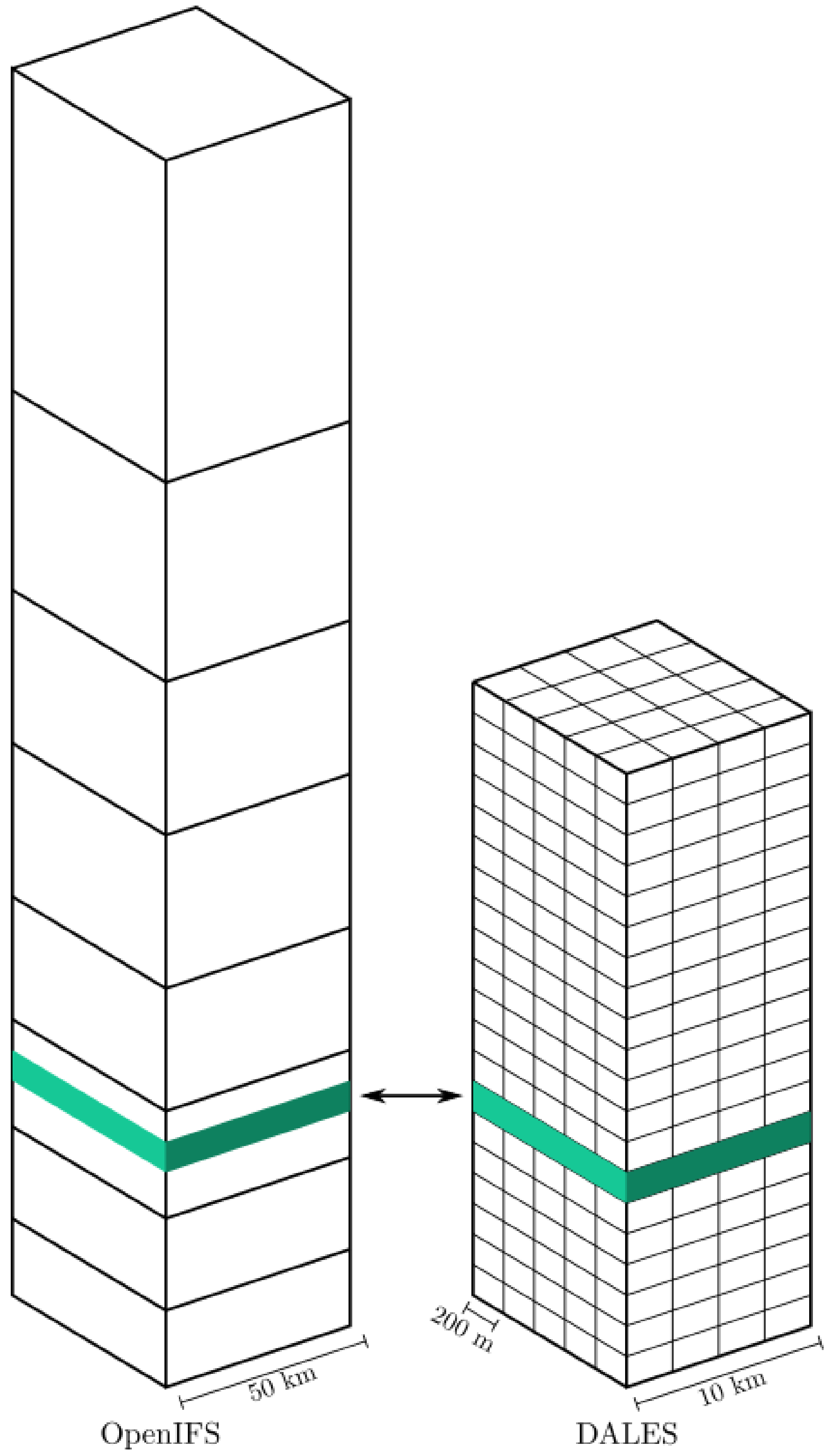
Compute kernel

8-20 hrs



Superparameterization, downscaling and machine learning

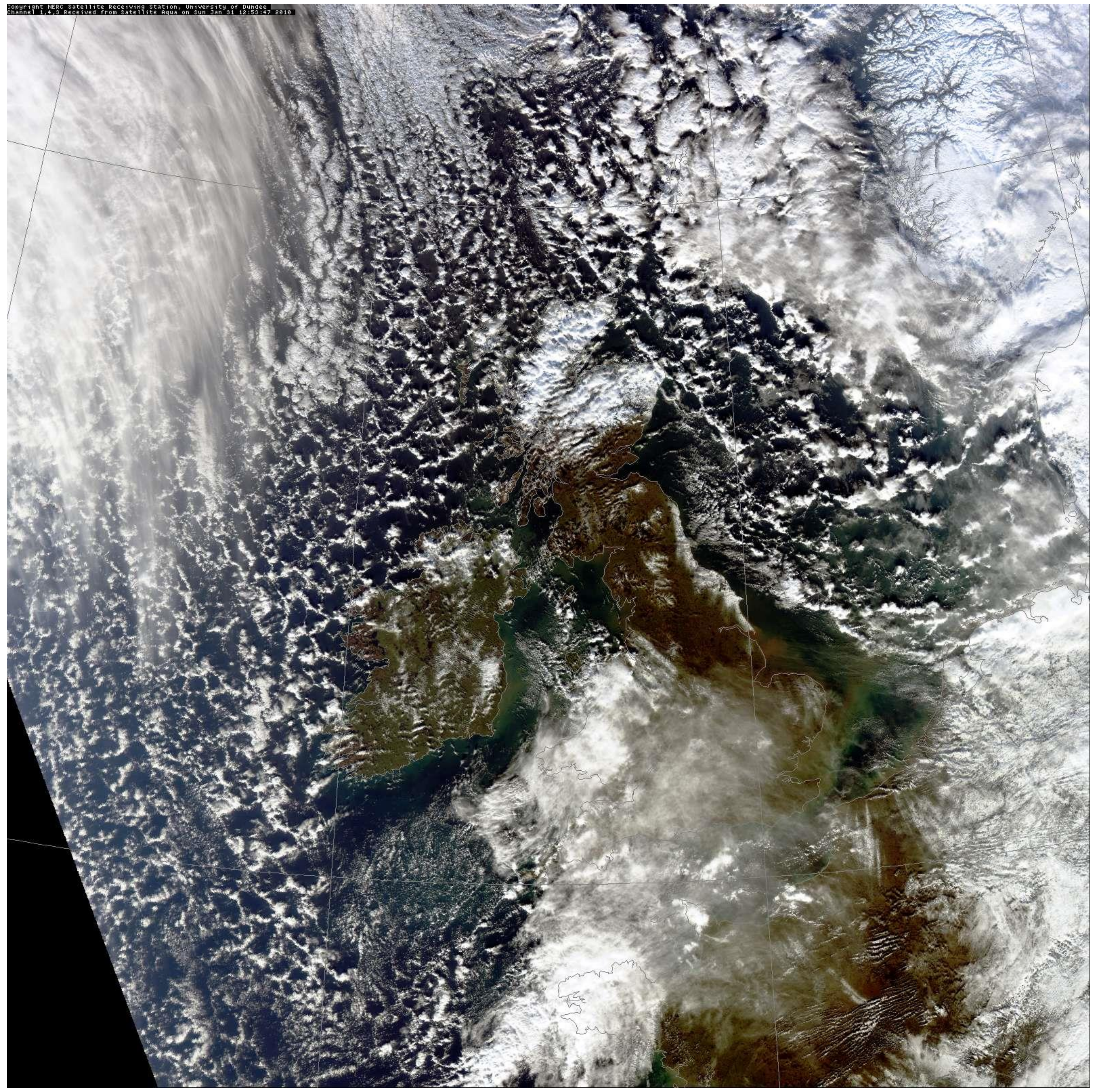


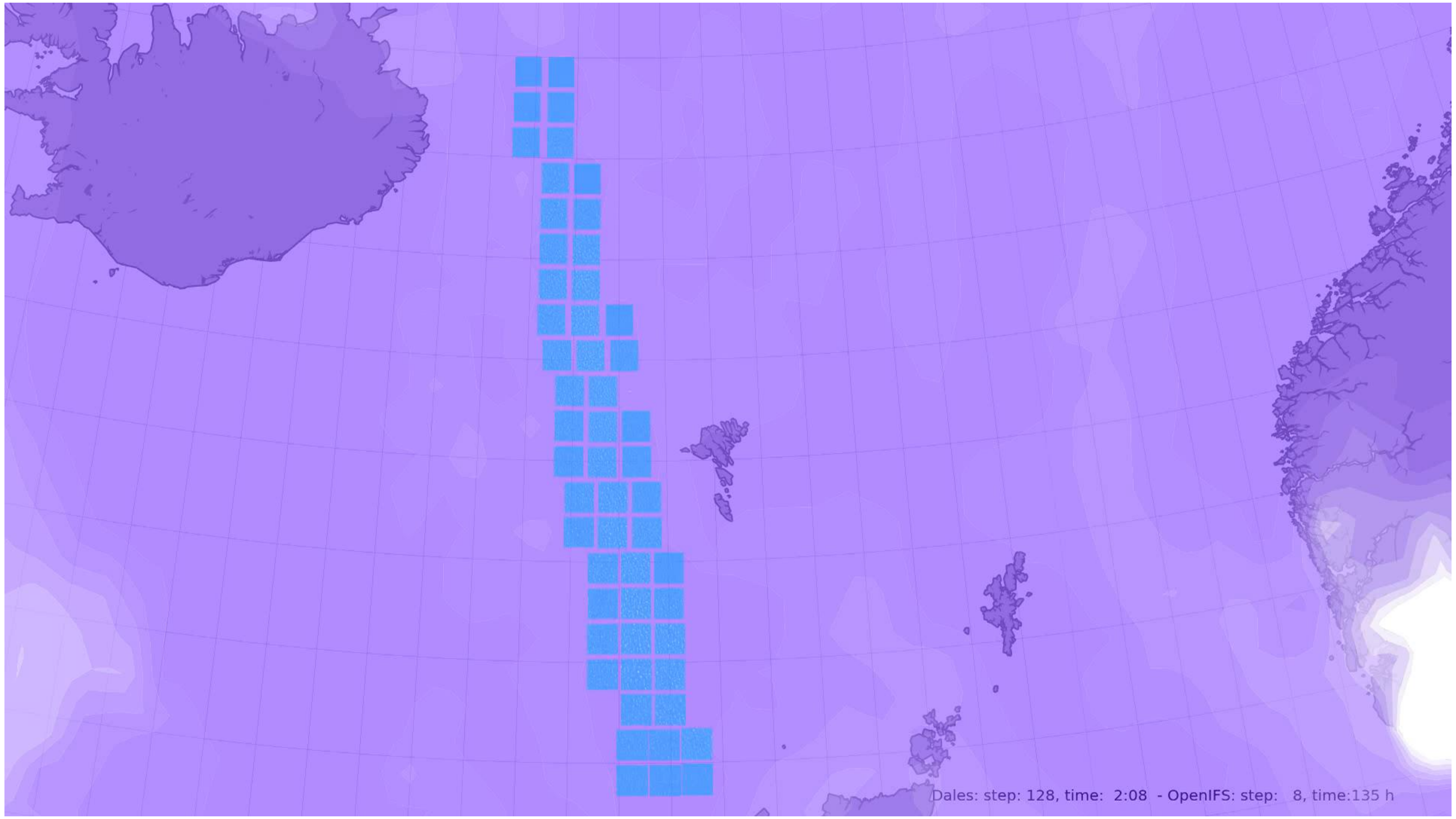


Gijs van Oord (NLeSC), Frederik Jansson (CWI), Pier Siebesma (TUDelft), Daan Crommelin (CWI)

Project output: Cloud-resolving climate models

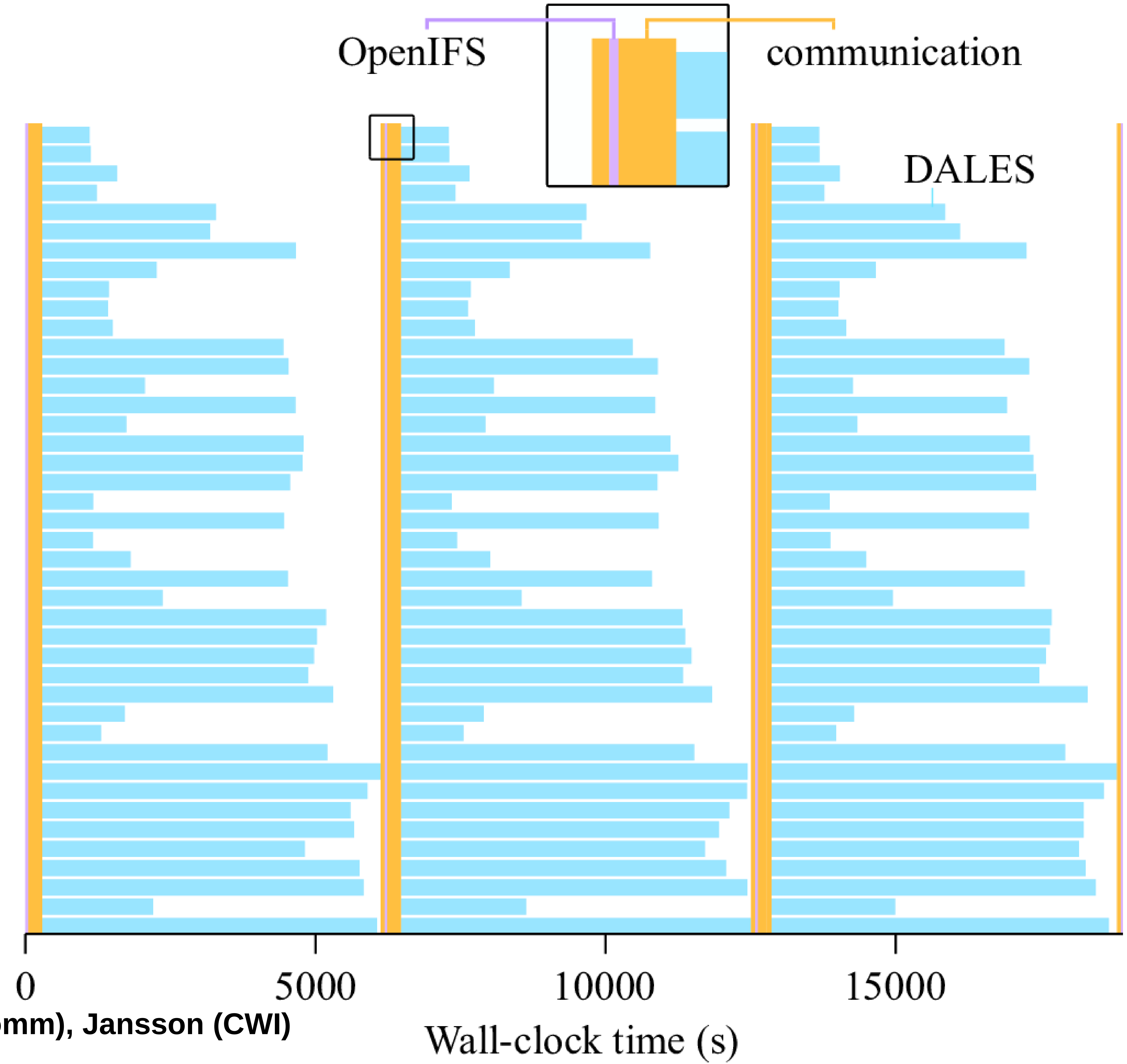
van Oord (NLeSC, pers comm), Jansson (CWI)





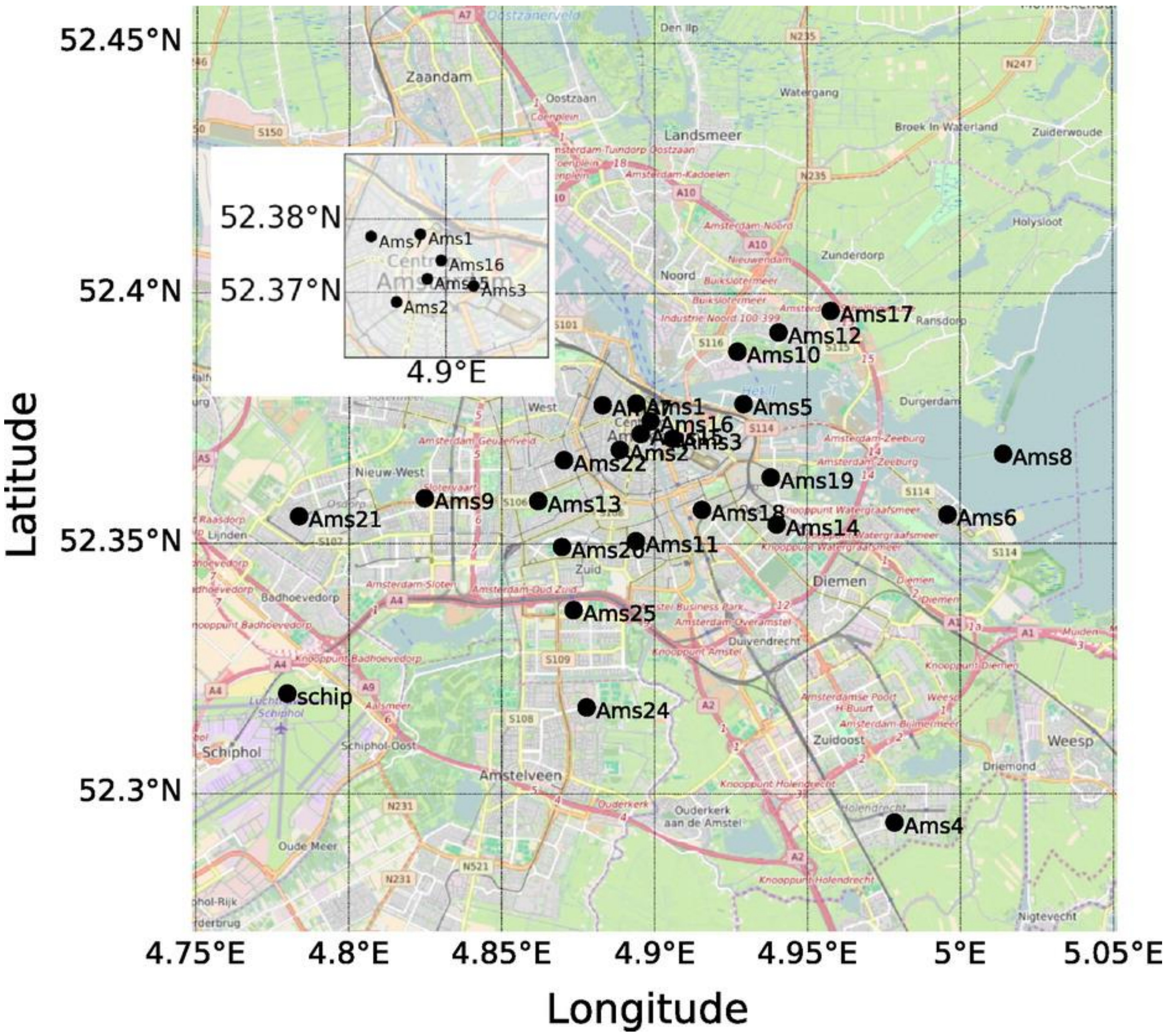
van Oord (NLeSC, pers comm), Jansson (CWI)

Load balancing OpenIFS and DALES



van Oord (NLeSC, pers comm), Jansson (CWI)

Computing and data challenge: nowcasting and short term forecasting at local scale



Latitude
Longitude
52.45°N
52.4°N
52.35°N
52.3°N
4.75°E
4.8°E
4.85°E
4.9°E
4.95°E
5°E
5.05°E
52.38°N
52.37°N
4.9°E
Ams7
Ams1
Ams16
Ams15
Ams3
Ams2
Ams17
Ams12
Ams10
Ams5
Ams22
Ams8
Ams19
Ams21
Ams9
Ams13
Ams18
Ams6
Ams14
Ams20
Ams11
Ams25
schip
Ams24
Ams4
Zaandam
Oostzaan
Landsmeer
Den Ilp
Watergang
Broek In Waterland
Zuiderwoude
Holysloot
Zunderdorp
Noord
Ransdorp
Durgerdam
West
Nieuw-West
Zuid
Diemen
Duivendrecht
Badhoevedorp
Schiphol
Schiphol-Oost
Amstelveen
Zuidoost
Weesp
Driemond
Ouderkerk aan de Amstel
Oude Meer
Nigtevecht

Downscaling

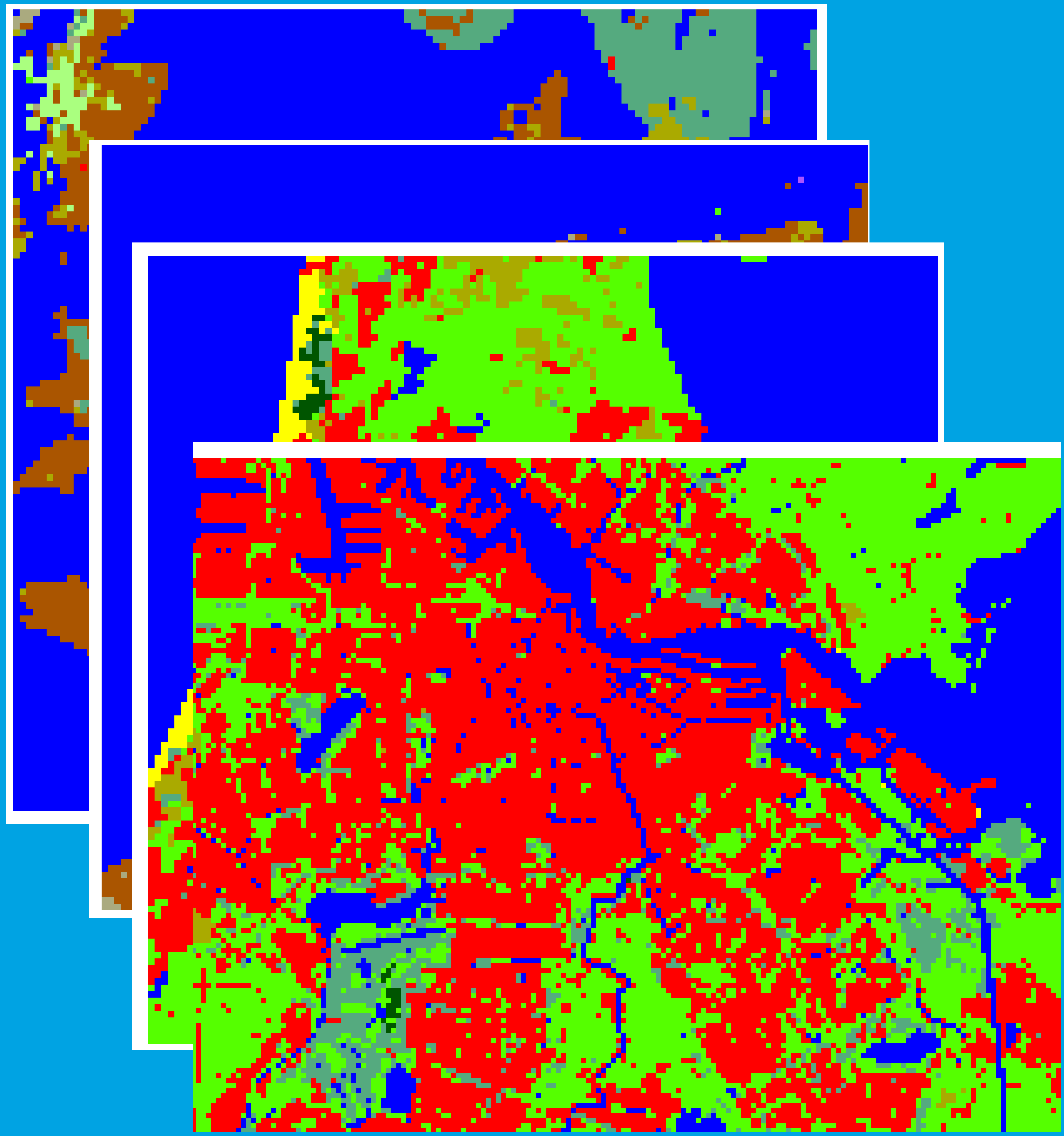
Daily forecasts
WRF3.5 + urban module (SLUCM)
48 hour runs, 24 hour spin-up

Domain 1: 12.5km
default setup

Domain 2: 2.5km
default setup

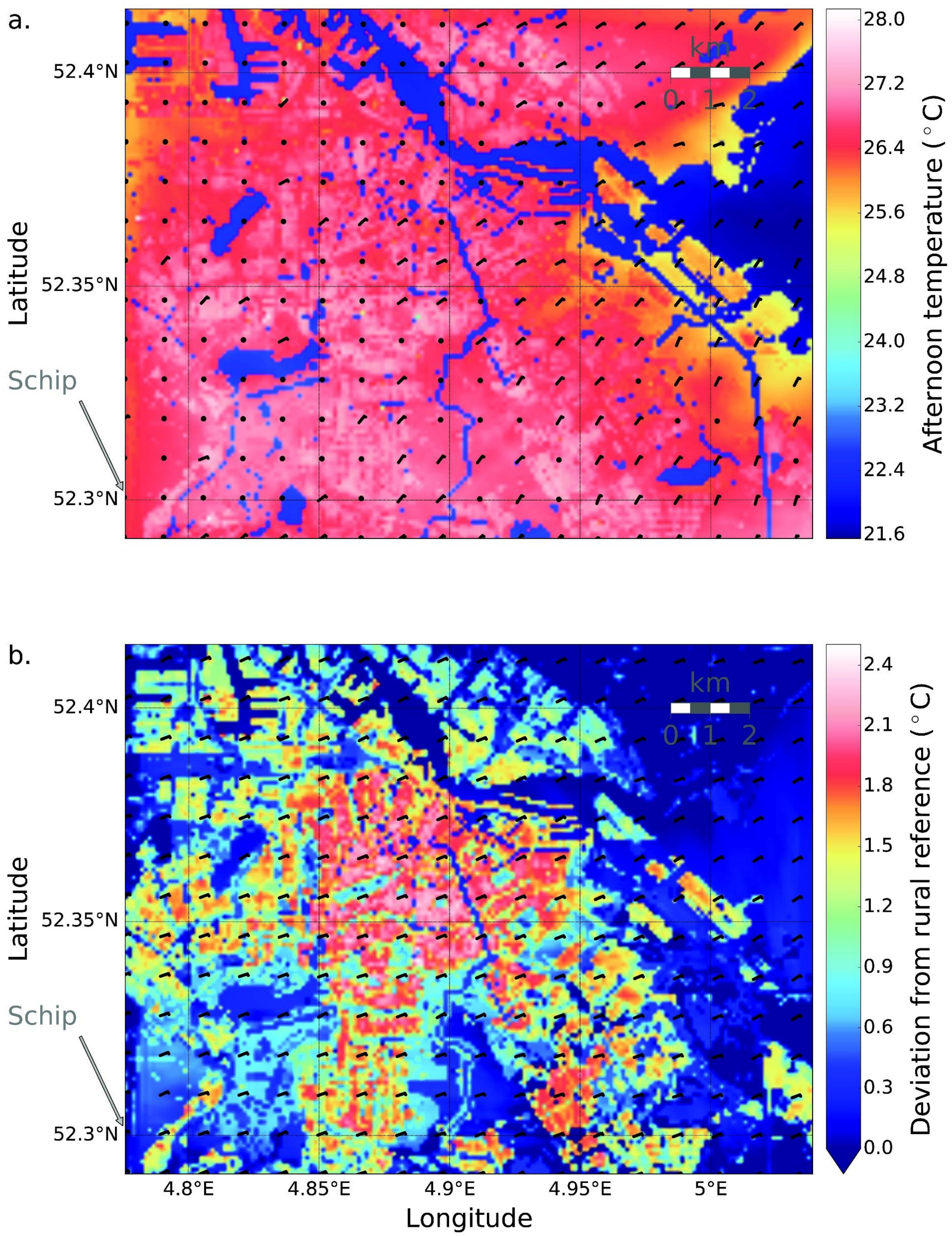
Domain 3: 500m
hi-res landuse,
Rijkswaterstaat river temperatures

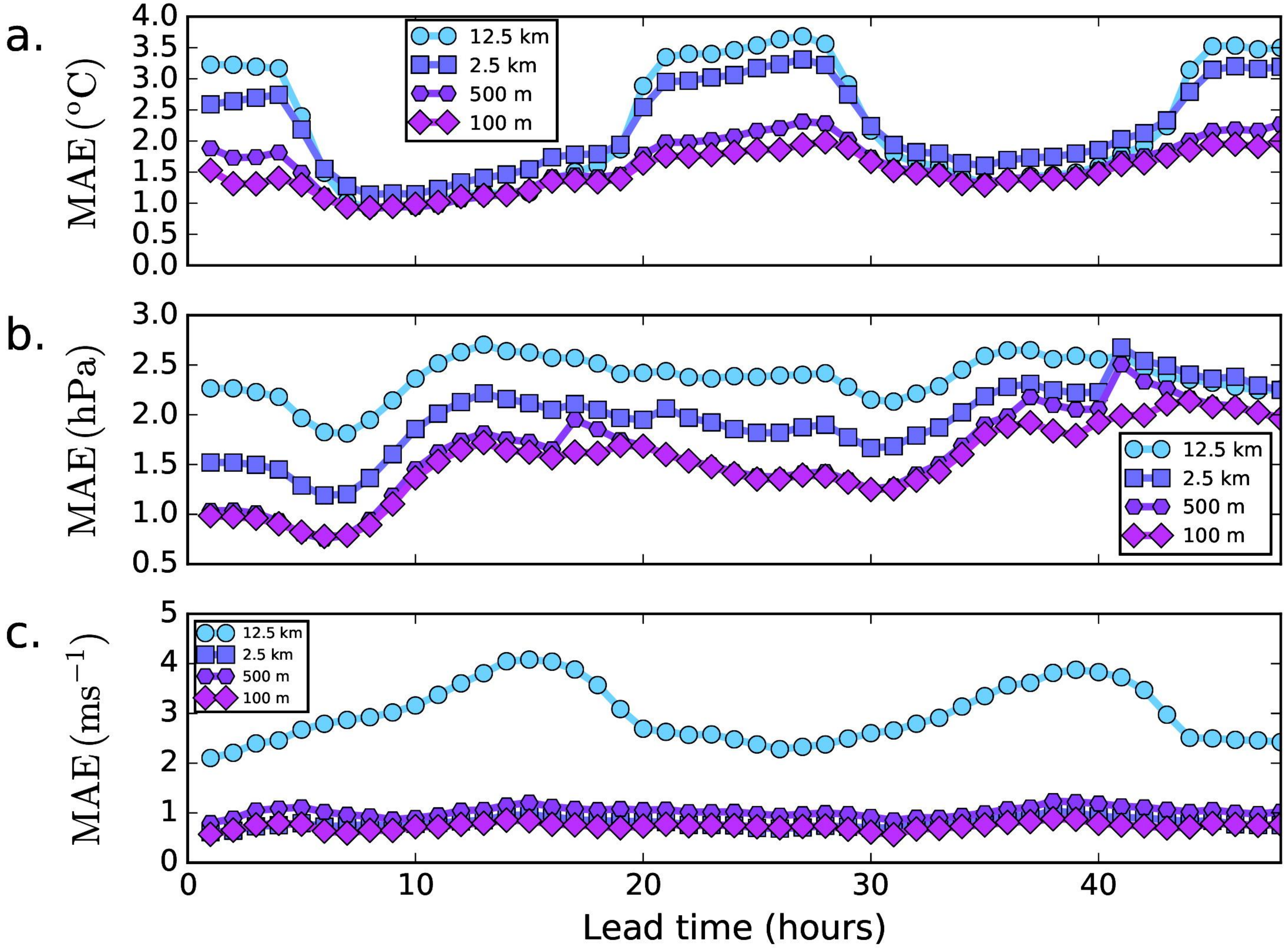
Domain 4: 100m
Rijkswaterstaat river temperatures,
TOP10NL, satellite imagery, AHN2
(height map), CBS data



Attema et al, IEEE eScience, 2015

Short range weather forecasting at street level





Ronda et al BAMS 2017

Vector of model parameters, computable $\boldsymbol{\theta}_c$ (e.g. high res models) and non-computable $\boldsymbol{\theta}_n$

$$\theta = (\theta_c, \theta_n)$$

$\boldsymbol{\theta}$ in parameterization schemes of climate model (ς), that forms a map parameterized by time t, that takes the parameters $\boldsymbol{\theta}$ to the state variables *x. And state variables are related to observables y*

$$x(t) = \varsigma(\theta, t)$$

$$y(t) = \varkappa\big(x(t)\big)$$

Actual observation ($\tilde{y}$)and observable mismatch (note, y depends on θ, but $\tilde{y}$ does not, so mismatch can be used to learn θ) :

$$y - \tilde{y}(t)$$

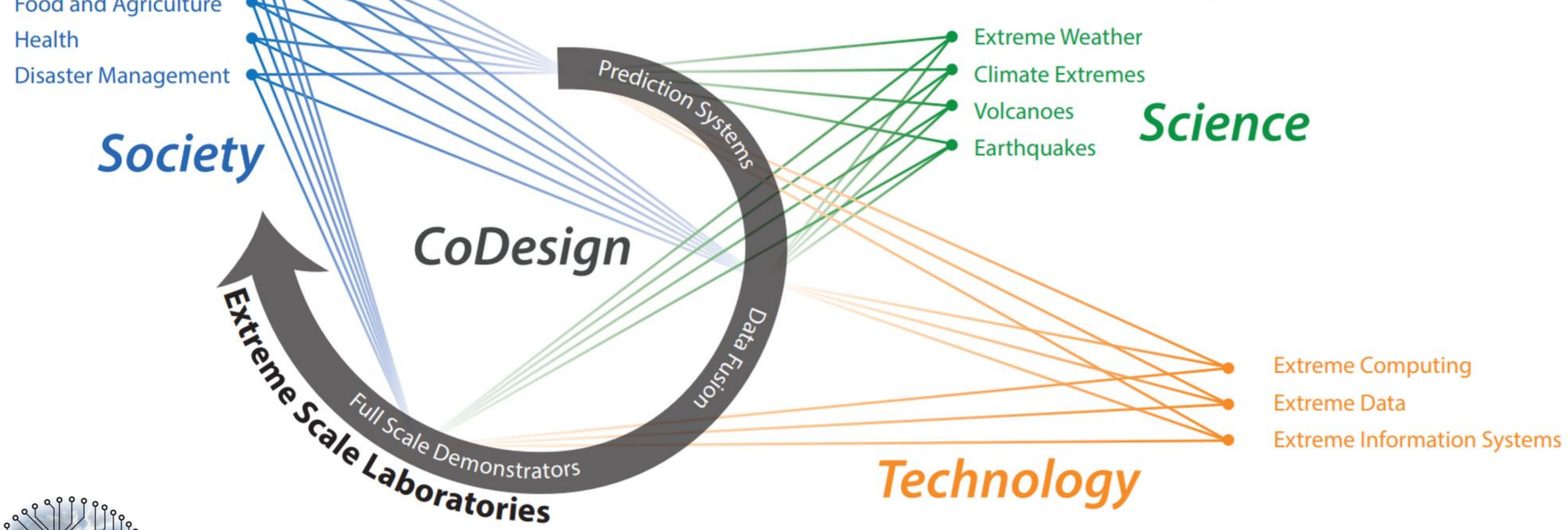
$$J_0 = \frac{1}{2}\|\langle f(y)\rangle_T - \langle f(\tilde{y})\rangle_T\|^2_{\Sigma_y}$$

High-resolution simulations nested in a climate model may be viewed as a time-dependent map *C* from the state variables *x* of the climate model to simulated state variable $\tilde{z}$. The variable *z* in the climate model depends on all parameters θ and again the mismatch can be used to learn the non computable parameters (a similar cost function can be defined as for *y*),

$$\tilde{z}(t) = C(\theta_n, t; x)$$

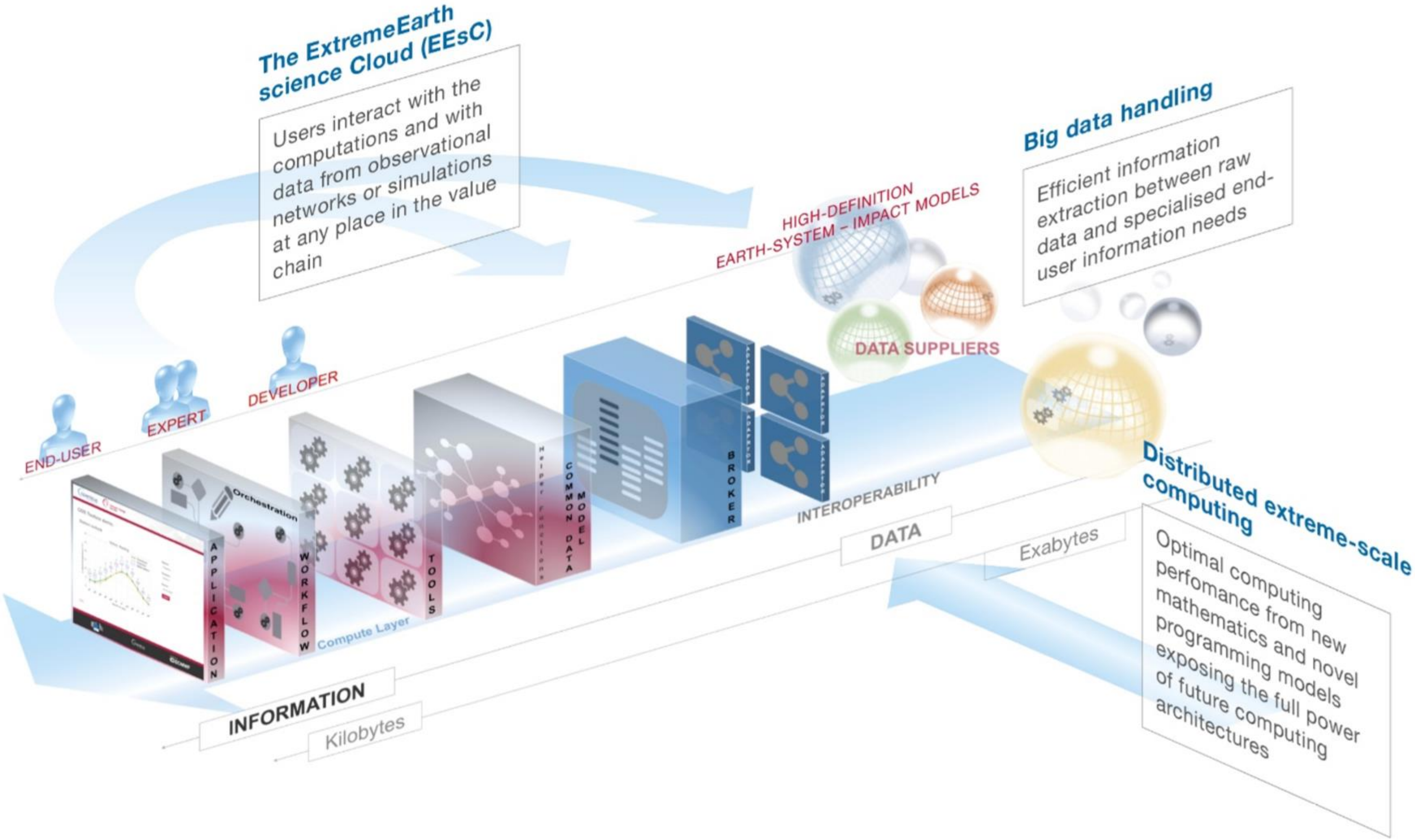
$$z(t) = s(\theta, t; x)$$

ExtremeEarth will co-design the technological transformation necessary for step changes in the ability of European society to anticipate weather and climate extremes, earthquakes and volcanoes; to collect and integrate Earth system data; and to access and adapt this information to the needs of an expanding user community.



ExtremeEarth

What e-infrastructure does it take?



A step-change in domain-specific, distributed high-performance computing for the simulation and prediction of Earth-system extremes.

A step-change in domain-specific, distributed big data handling for the simulation and prediction of Earth-system extremes, and for exploring the full range of information from simulations and observation

User interaction enabled by a domain-specific, integrated information system towards the ExtremeEarth science Cloud (EEsC)

14th International IEEE eScience Conference

29 October – 1 November 2018
Amsterdam, the Netherlands

www.eScience2018.com

END


5th National eScience Symposium. Photography: Elodie Burrillon

Program overview

Monday	workshops & tutorials
Tuesday	plenary session & conference dinner
Wednesday	focused sessions & reception & poster session
Thursday	plenary session

Workshops Monday

- Generic components of the High Energy Physics eScience ecosystem
- International Workshop on Sustainable Software for Science (WSSPE)
- Research Objects
- Handling Uncertainties in Big Data

5th National eScience Symposium. Photography: Elodie Burrillon

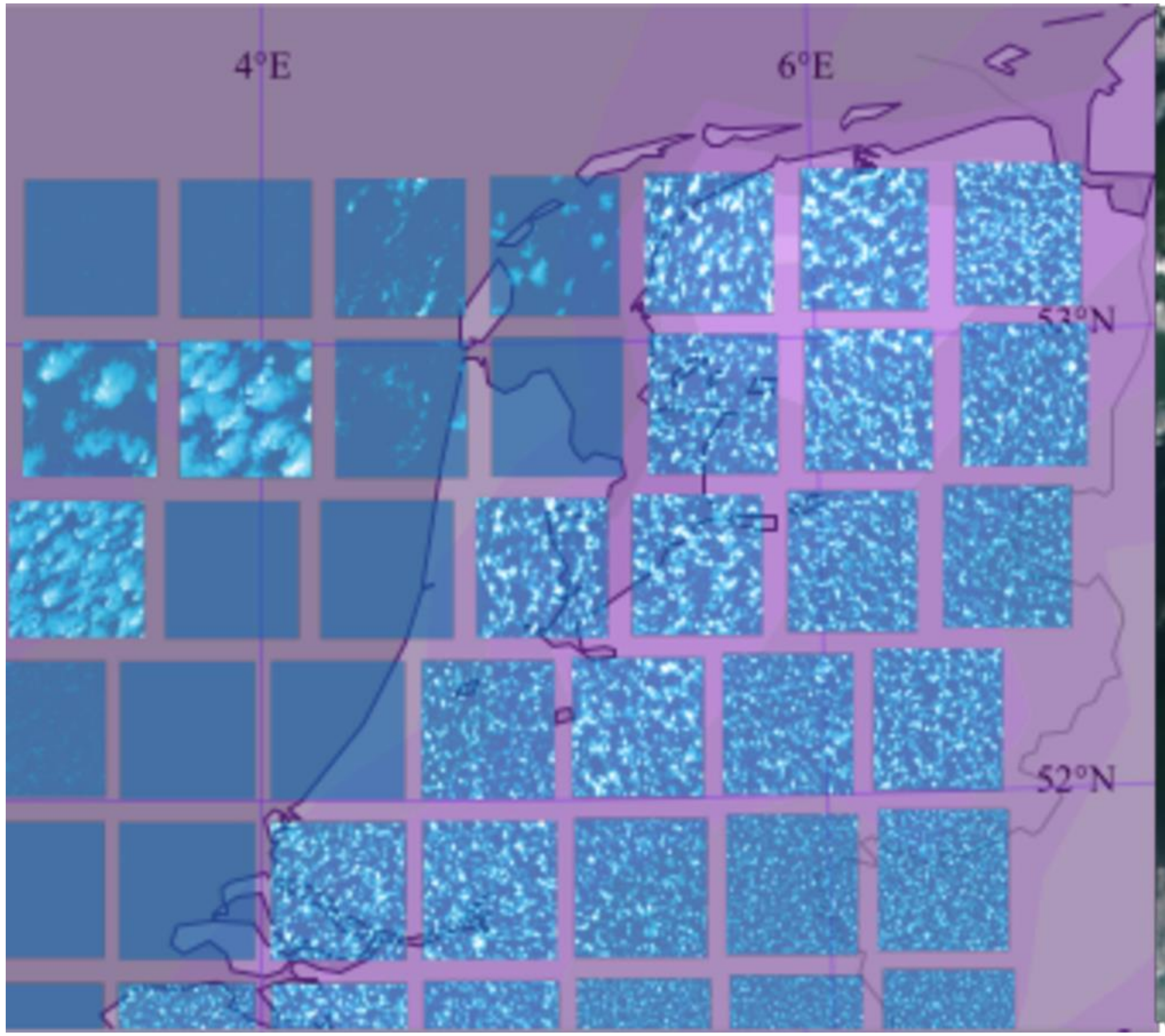
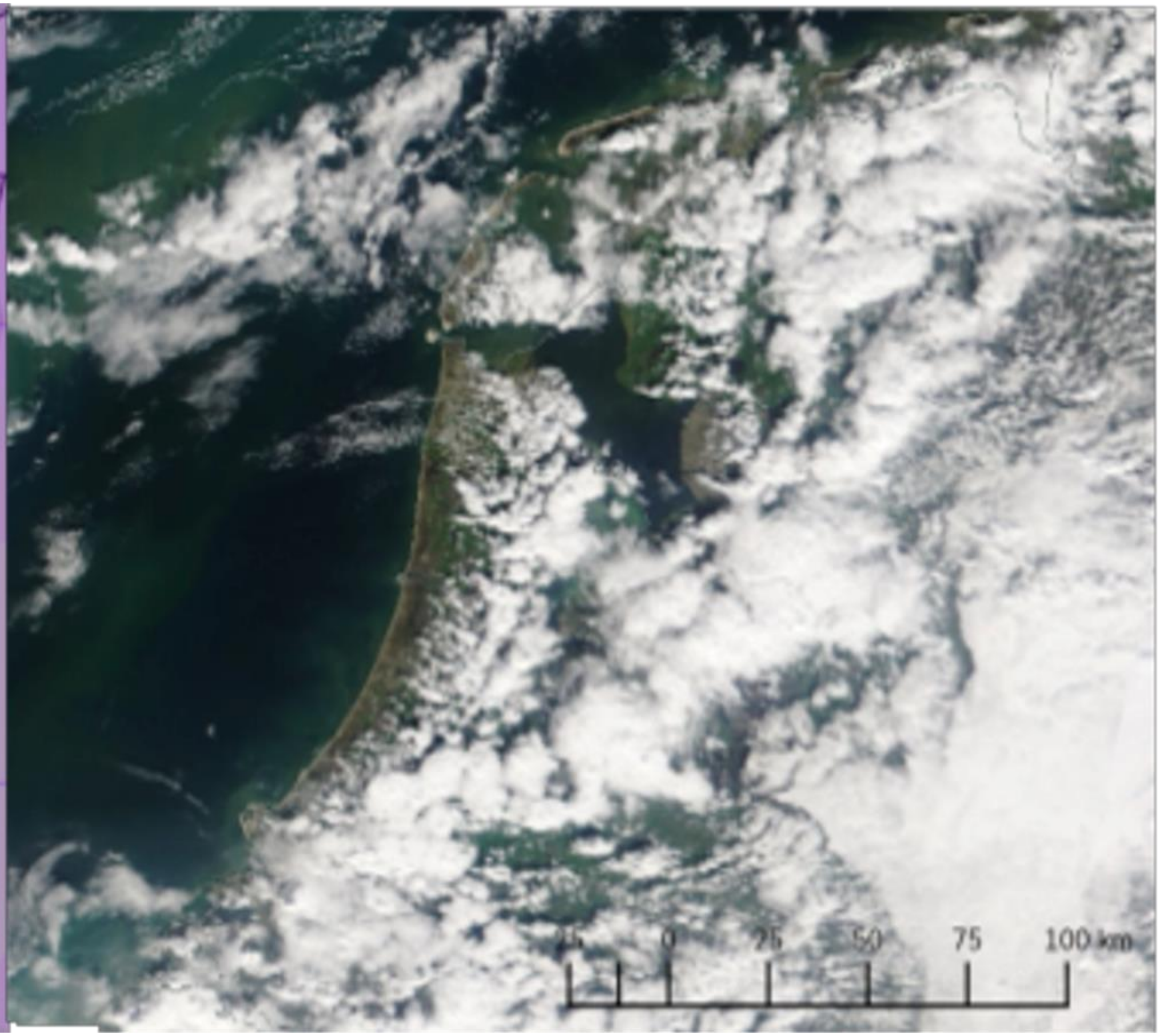
Workshops Monday

- eScience-FAIR science by PLAN-E
- Platform driven e-infrastructure innovations
- Tutorial: Contemporary Peer Code Review in Scientific Software Development

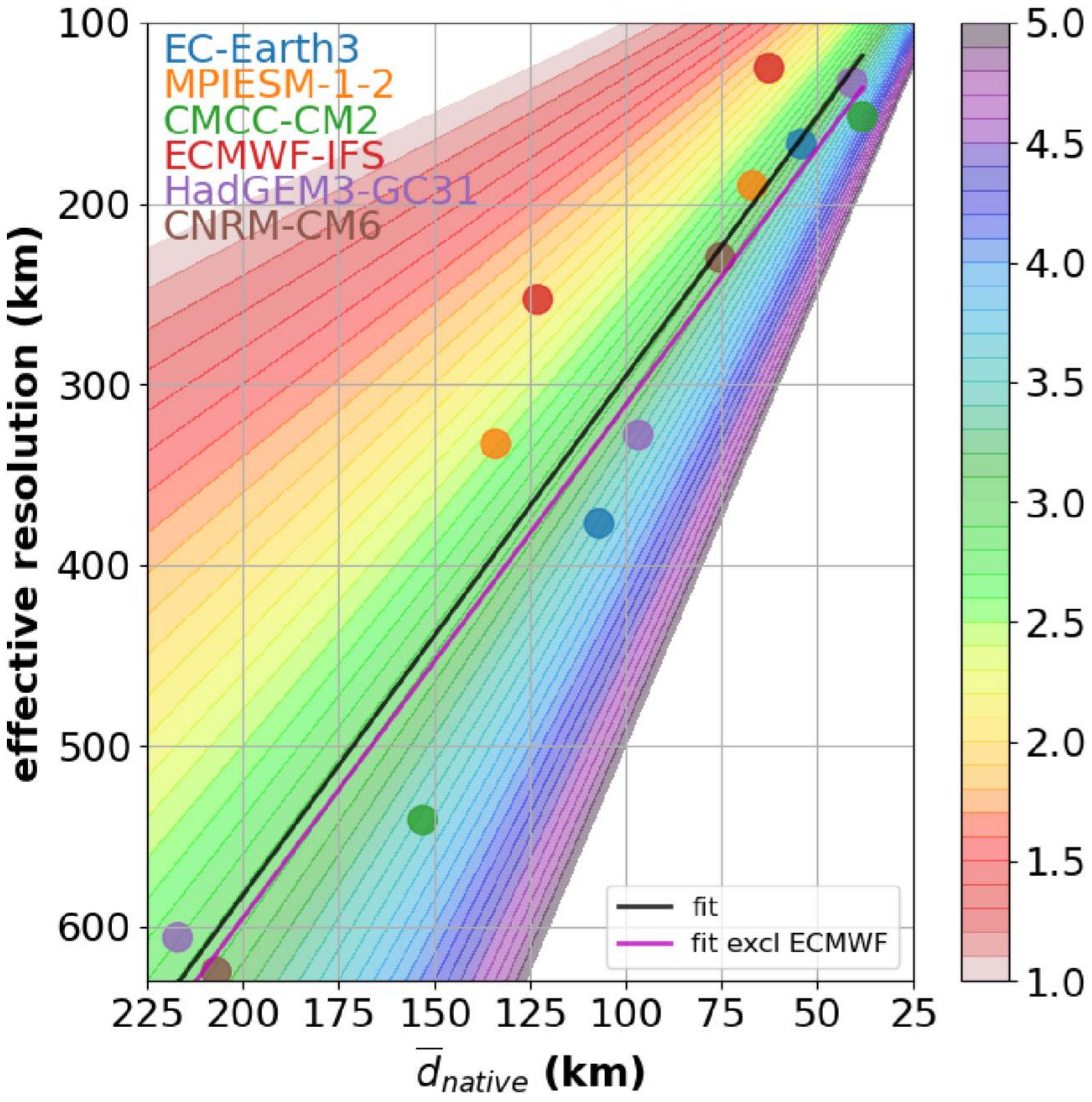
Multitrack day with focused sessions

- Data Handling and Analytics for Health
- Exascale Computing for High Energy Physics
- Weather and Climate science in the digital era
- Advances in eScience for the Humanities and Social Sciences
- Computer Science: Tools & Infrastructure

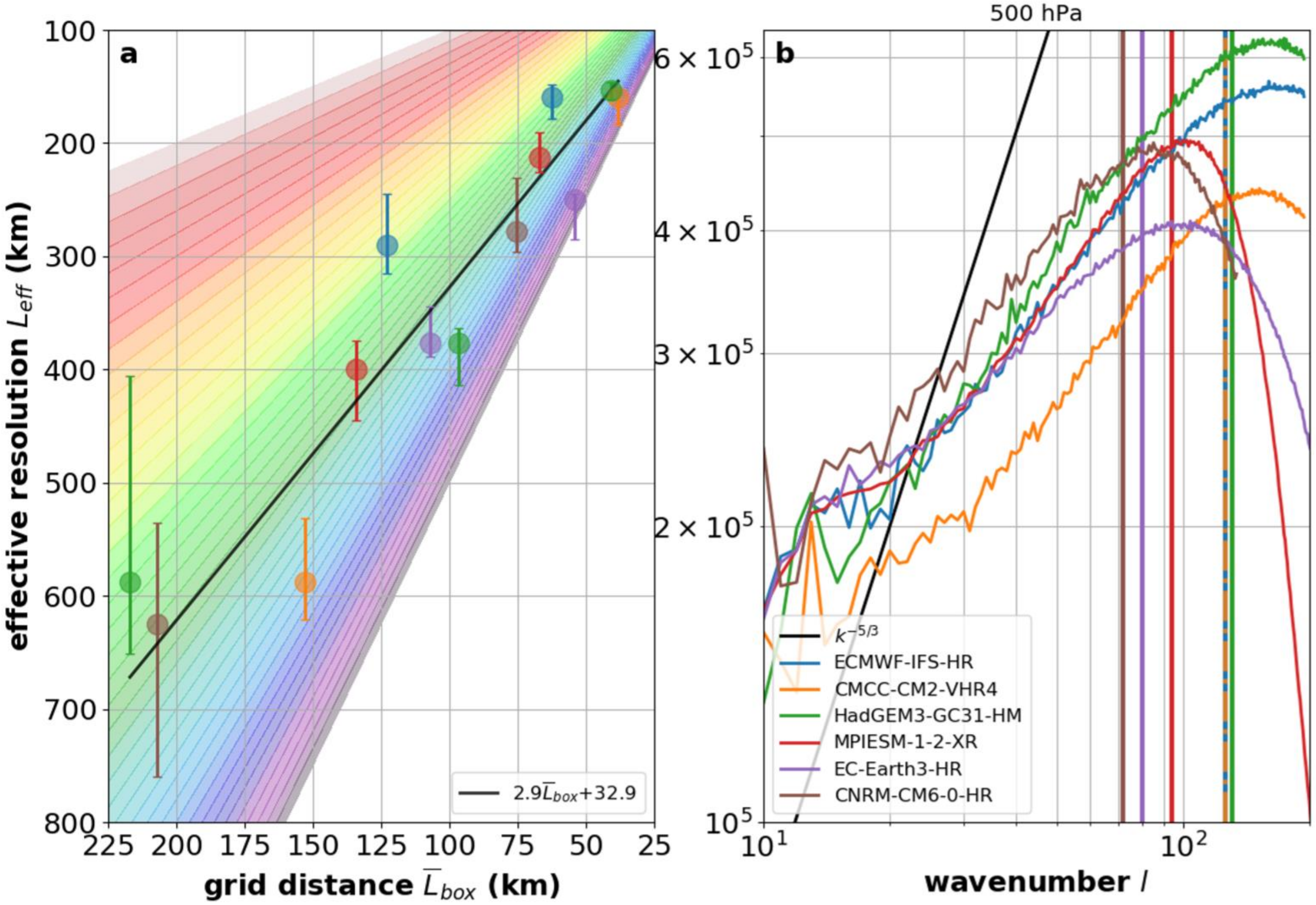
Venue Mövenpick Hotel Amsterdam City Center, the Netherlands


4°E
6°E
53°N
52°N

0
25
50
75
100 km

Effective resolutions



Klaver et al in prep



a
effective resolution L_{eff} (km)
grid distance $\overline{L}_{box}$ (km)
$2.9\overline{L}_{box}$+32.9
100
200
300
400
500
600
700
800
225
200
175
150
125
100
75
50
25
b
500 hPa
6×10^5
4×10^5
3×10^5
2×10^5
10^5
10^1
10^2
wavenumber l
$k^{-5/3}$
ECMWF-IFS-HR
CMCC-CM2-VHR4
HadGEM3-GC31-HM
MPIESM-1-2-XR
EC-Earth3-HR
CNRM-CM6-0-HR

Easy to use:

Can be used directly on existing kernels and code generators

Inserts no dependencies in the kernels

Kernels can still be compiled with regular compilers

Supporting:

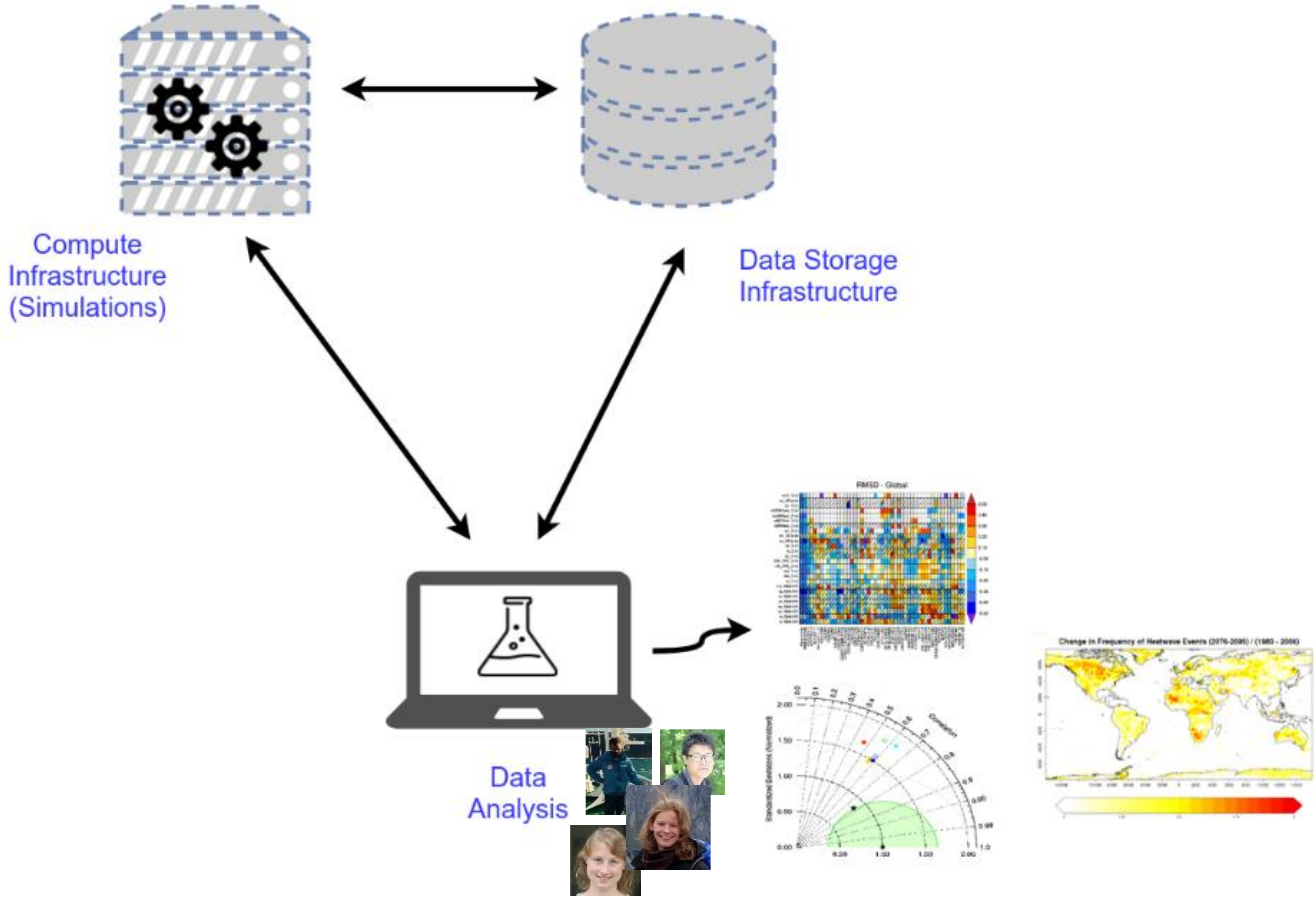
Large number of effective search optimizing algorithms

Output verification for auto-tuned kernels and pipelines

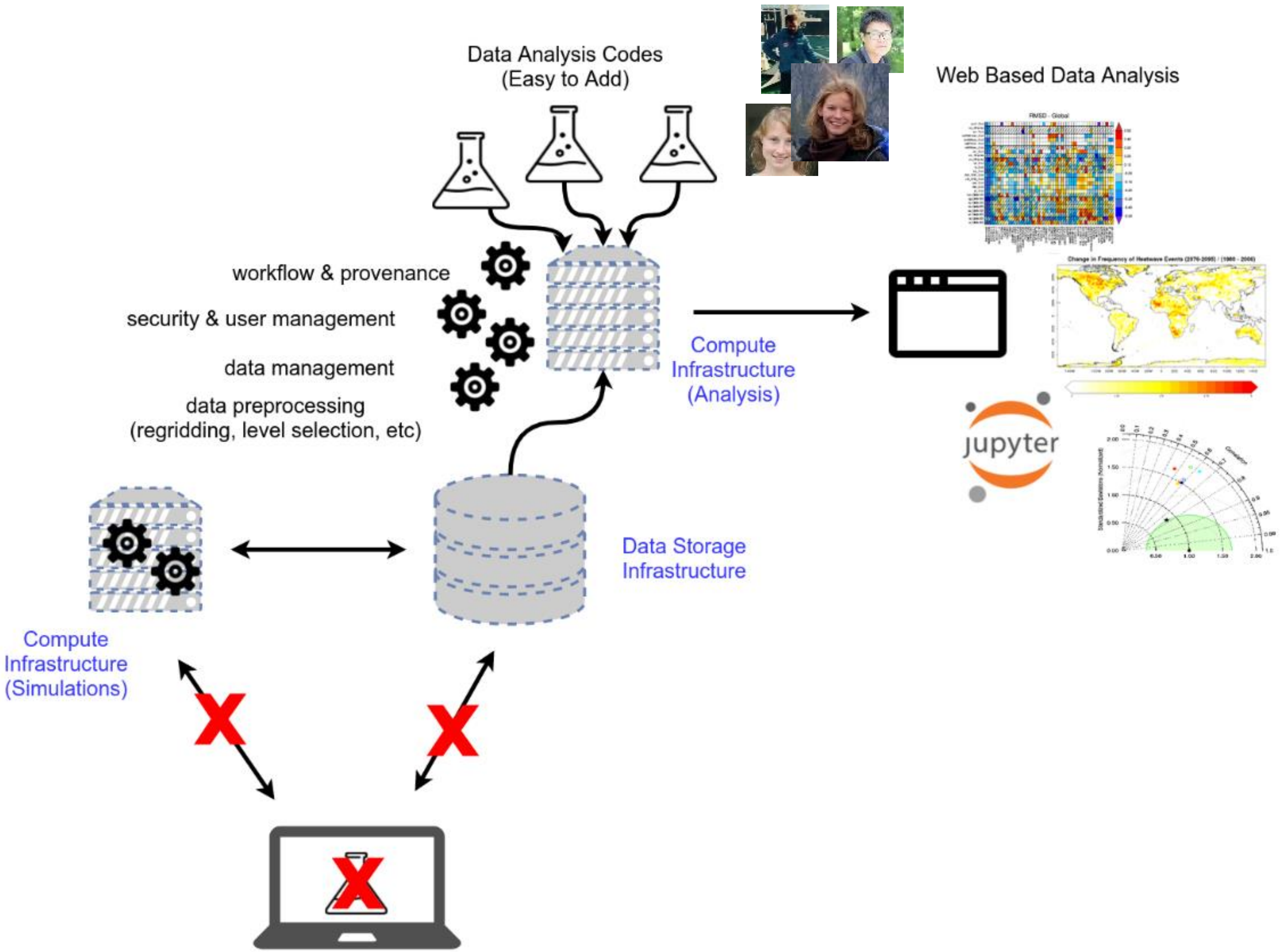
Tuning parameters in both host and device

Python-based unit testing of GPU code

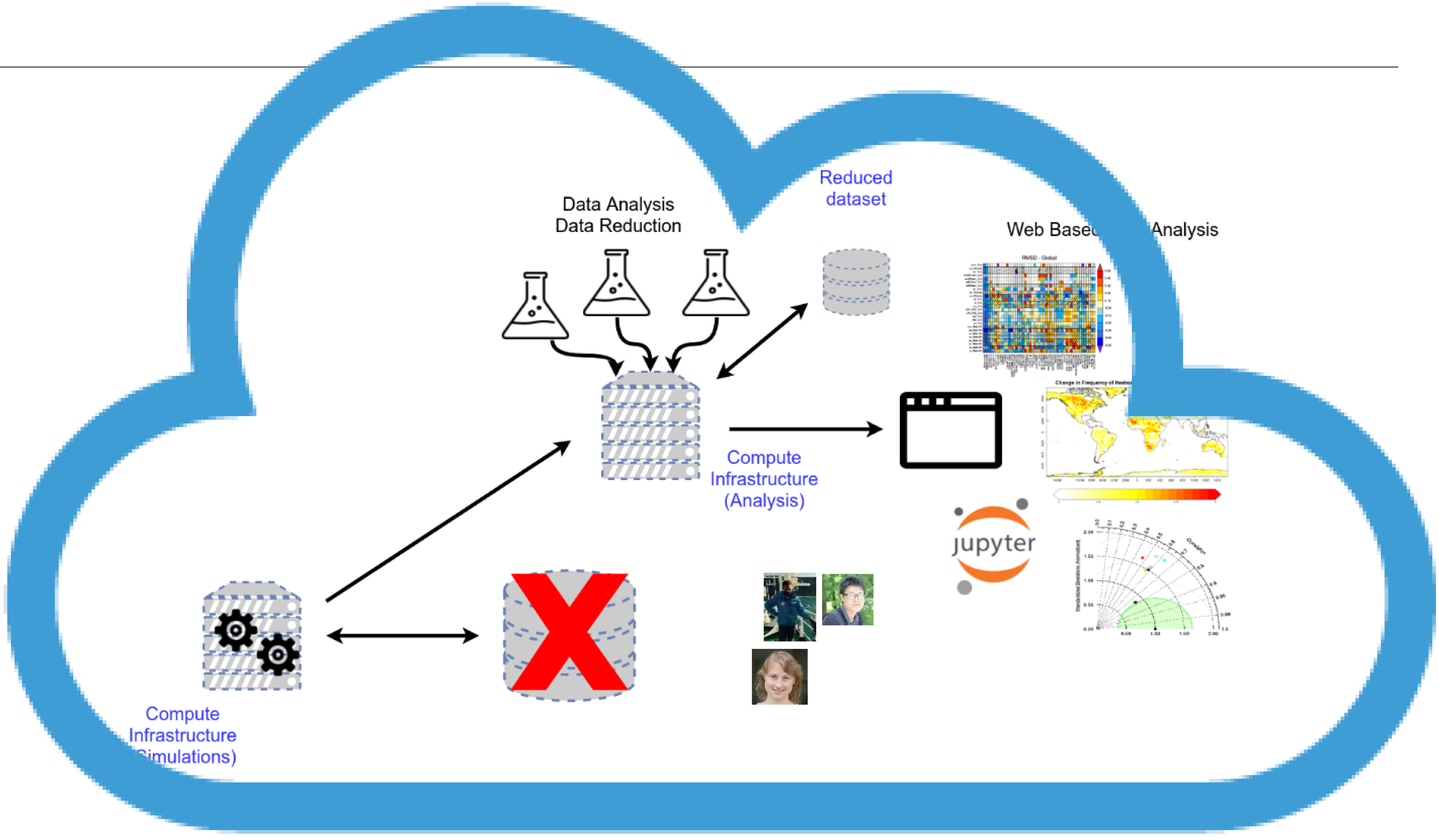
http://benvanwerkhoven.github.io/kernel_tuner/



Compute
Infrastructure
(Simulations)
Data Storage
Infrastructure
Data
Analysis
RMSD - Global



Data Analysis Codes
(Easy to Add)
Web Based Data Analysis
workflow & provenance
security & user management
data management
data preprocessing
(regridding, level selection, etc)
Compute
Infrastructure
(Analysis)
jupyter
Data Storage
Infrastructure
Compute
Infrastructure
(Simulations)


Data Analysis
Data Reduction
Reduced dataset
Web Based Analysis
Compute Infrastructure (Analysis)
jupyter
Compute Infrastructure (Simulations)

Let’s stay in touch

+31 (0)20 460 4770

info@esciencecenter.nl

www.esciencecenter.nl

blog.esciencecenter.nl

eScienceCenter

eScienceCenter

linkd.in/1j2uS8S

w.hazeleger@esciencecenter.nl