ETH zürich

Anticipating the European Supercomputing Infrastructure of the Early 2020s

Thomas C. Schulthess

European Commission President Jean-Claude Juncker

"Our ambition is for Europe to become one of the top 3 world leaders in high-performance computing by 2020"

27 October 2015

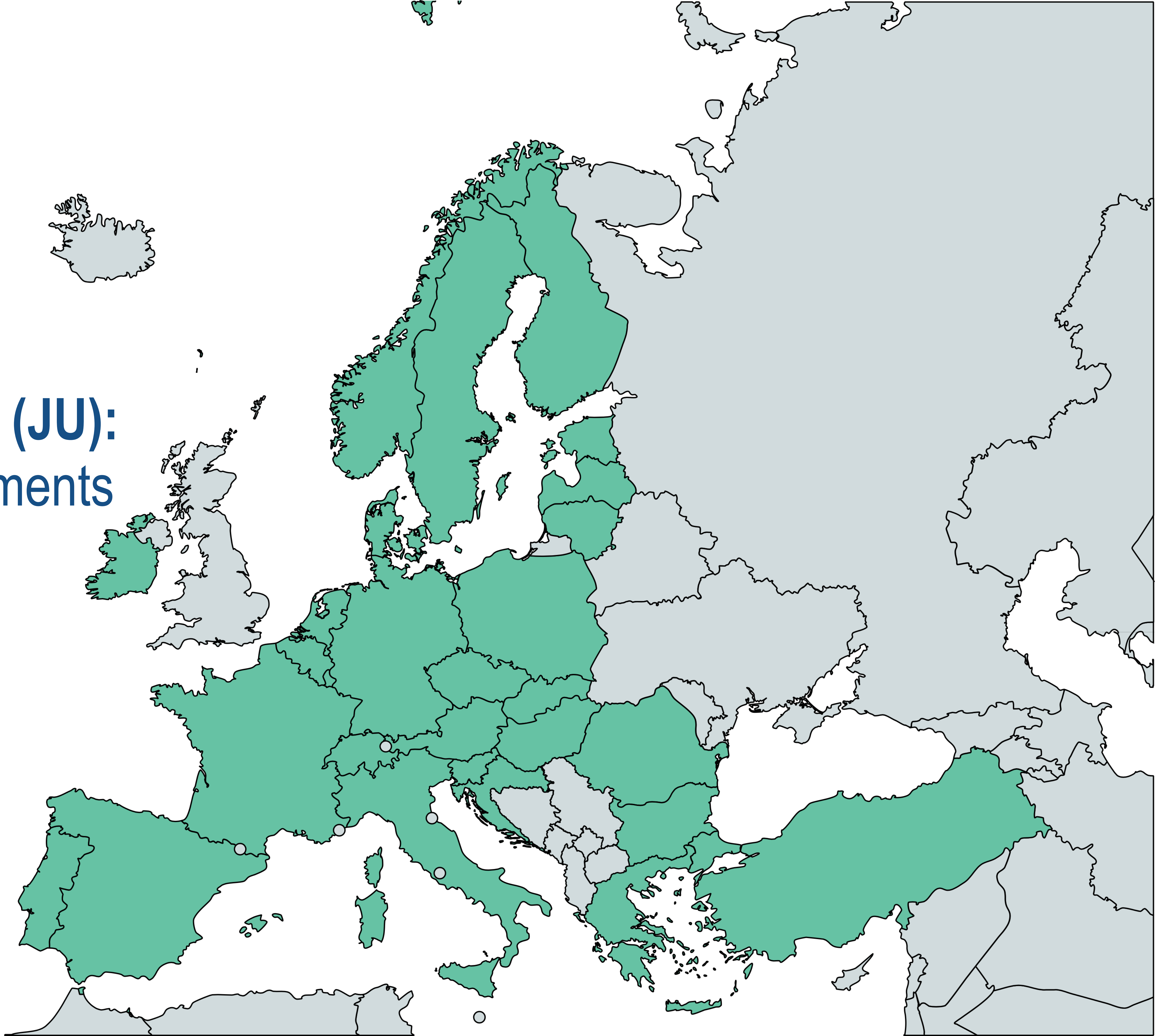
European Cloud Initiative (ECI) by the EC [COM(2016) 178, 04/2016]

- Help create a digital single market in Europe
- Create incentives to share data openly & improve interoperability
- Overcome fragmentation (scientific & economic domains, countries, …)
- Invest in European HPC ecosystem
- Create a dependable environment for data-producers & users to re-use data

ETH zürich

EuroHPC Joint Undertaking (JU):

A legal entity for joint procurements between states and the European Commission



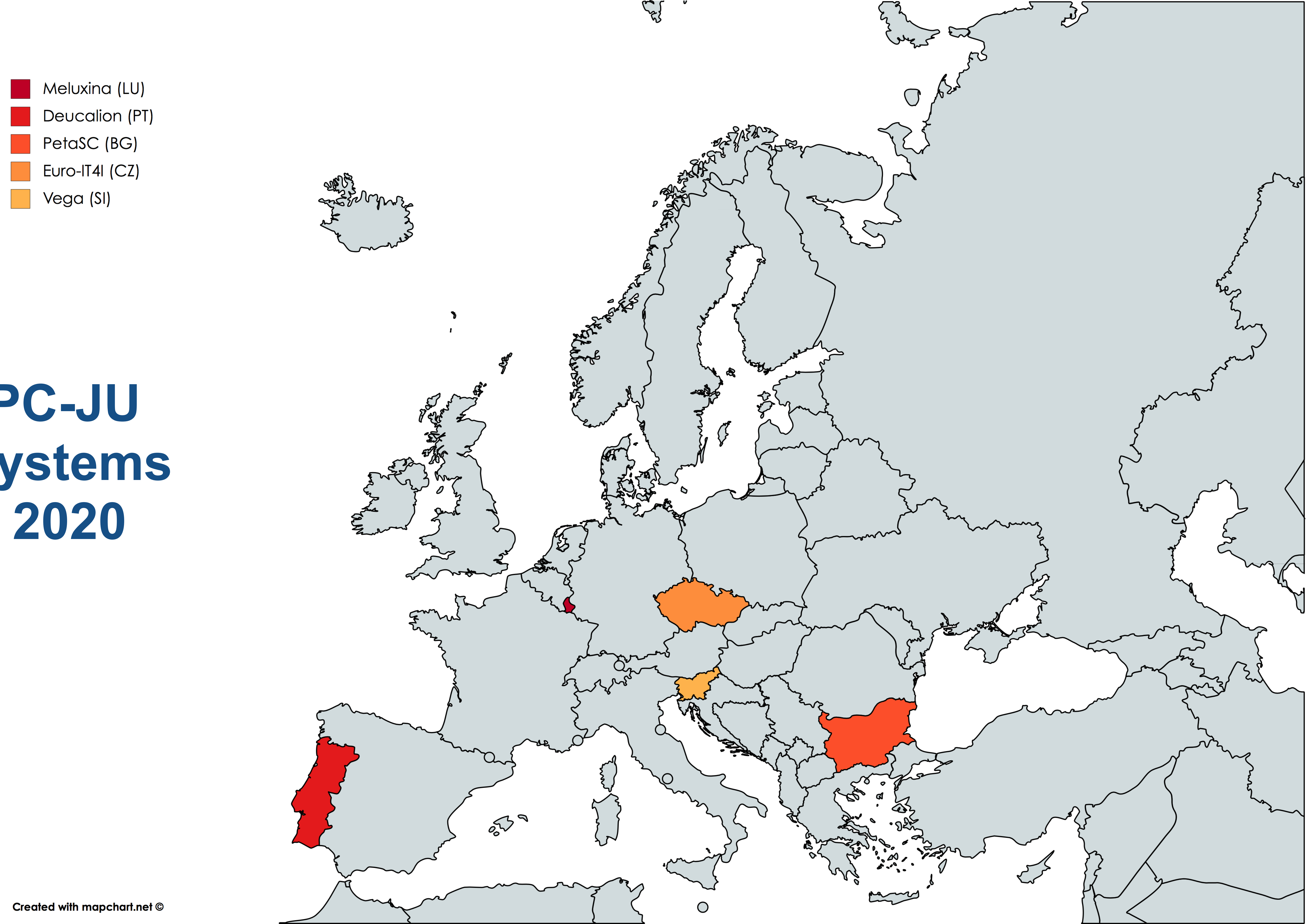
Members in June 2019

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

Created with mapchart.net ©

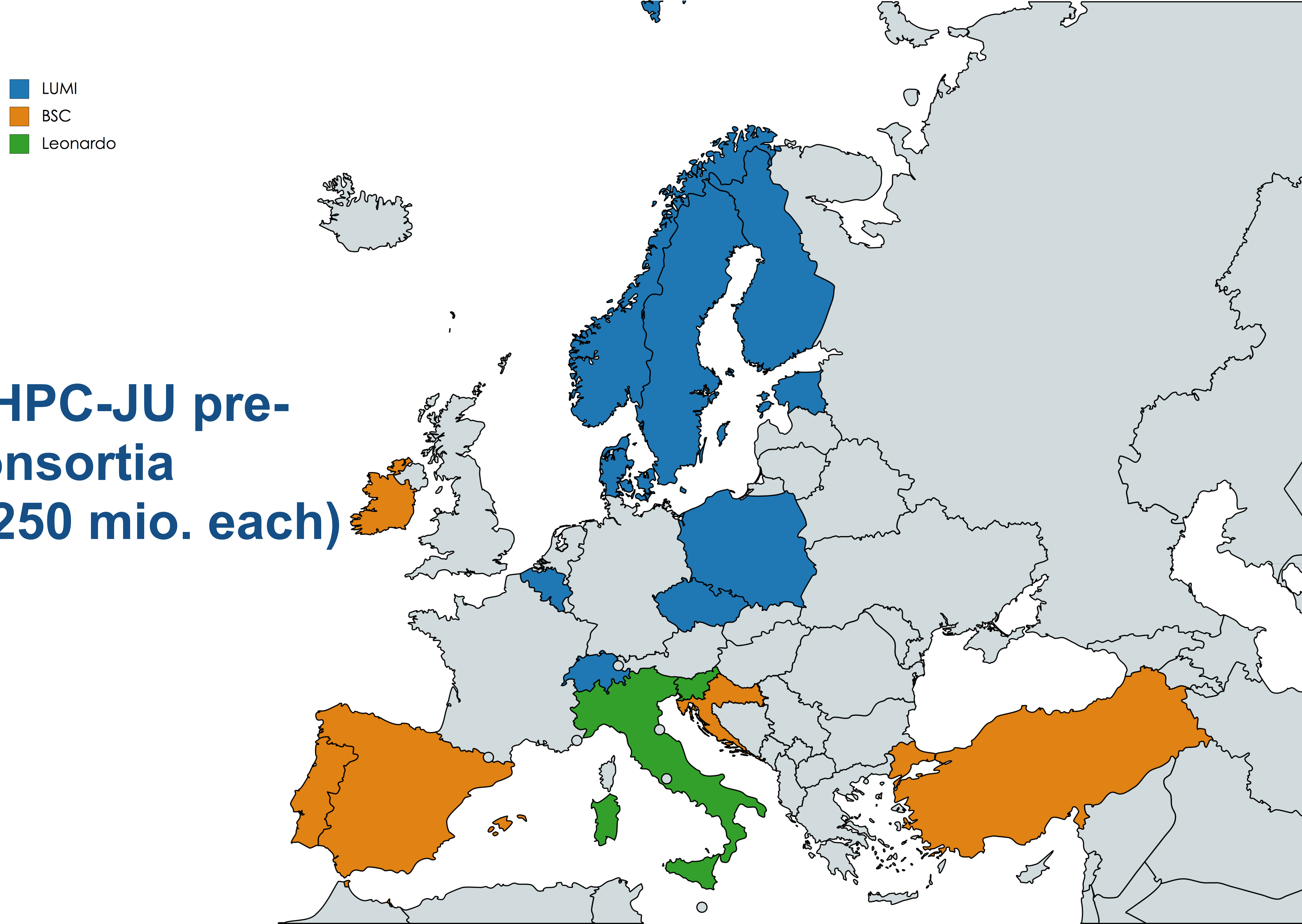
ETH zürich

Five EuroHPC-JU Petascale systems Installed by 2020



CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



Three EuroHPC-JU pre-exascale consortia (TCO ~200-250 mio. each)

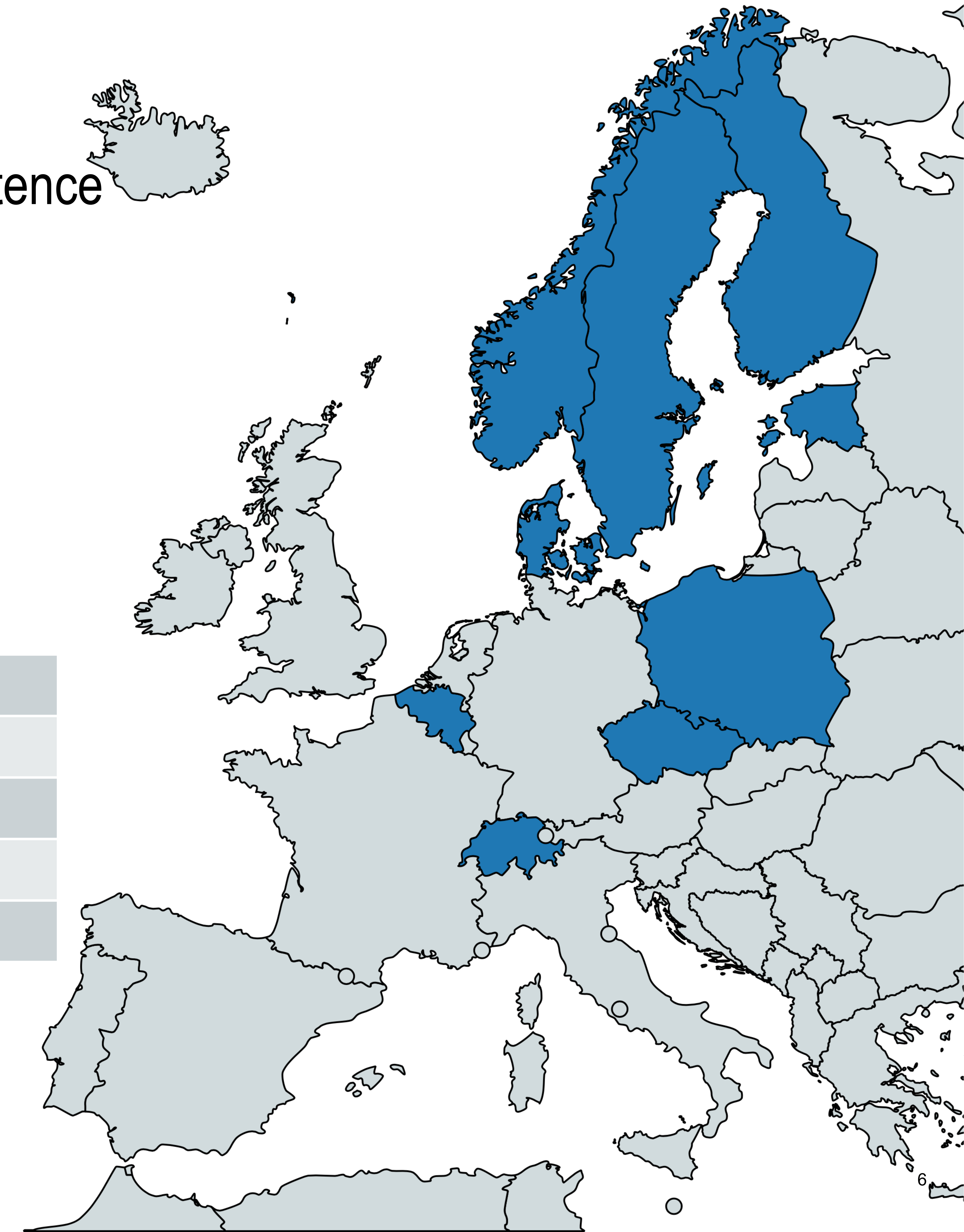
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

LUMI Consortium

- Large consortium with strong national HPC centres and competence provides a unique opportunity for
  - knowledge transfer;
  - synergies in operations; and
  - regionally adaptable user support for extreme-scale systems
- National & EU investments (2020-2026)

Finland	50 M€	Norway	4 M€
Belgium	15.5 M€	Poland	5 M€
Czech Republic	5 M€	Sweden	7 M€
Denmark	6 M€	Switzerland	10 M€
Estonia	2 M€	EU	104 M€

Plus additional investments in applications development



ETH zürich

EuroHPC-JU members

- EuroHPC-JU not committed
- 5 Petascale sites (2020)
- 3 PreExascale consortia (2020-2021)
- 2 Exascale sites (2022-2023)

Strong commitment towards a European HPC ecosystem!

Created with mapchart.net ©

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

Kajaani Data Center (LUMI)

2200 m² floor space, expandable up to 4600 m²

100% free cooling @ PUE 1.03

100% hydroelectric energy up to 200 MW

One power grid outage in 36 years

Extreme connectivity:
Kajaani DC is a direct part of the Nordic backbone; 4x100 Gbit/s in place; can be easily scaled up to multi-terabit level

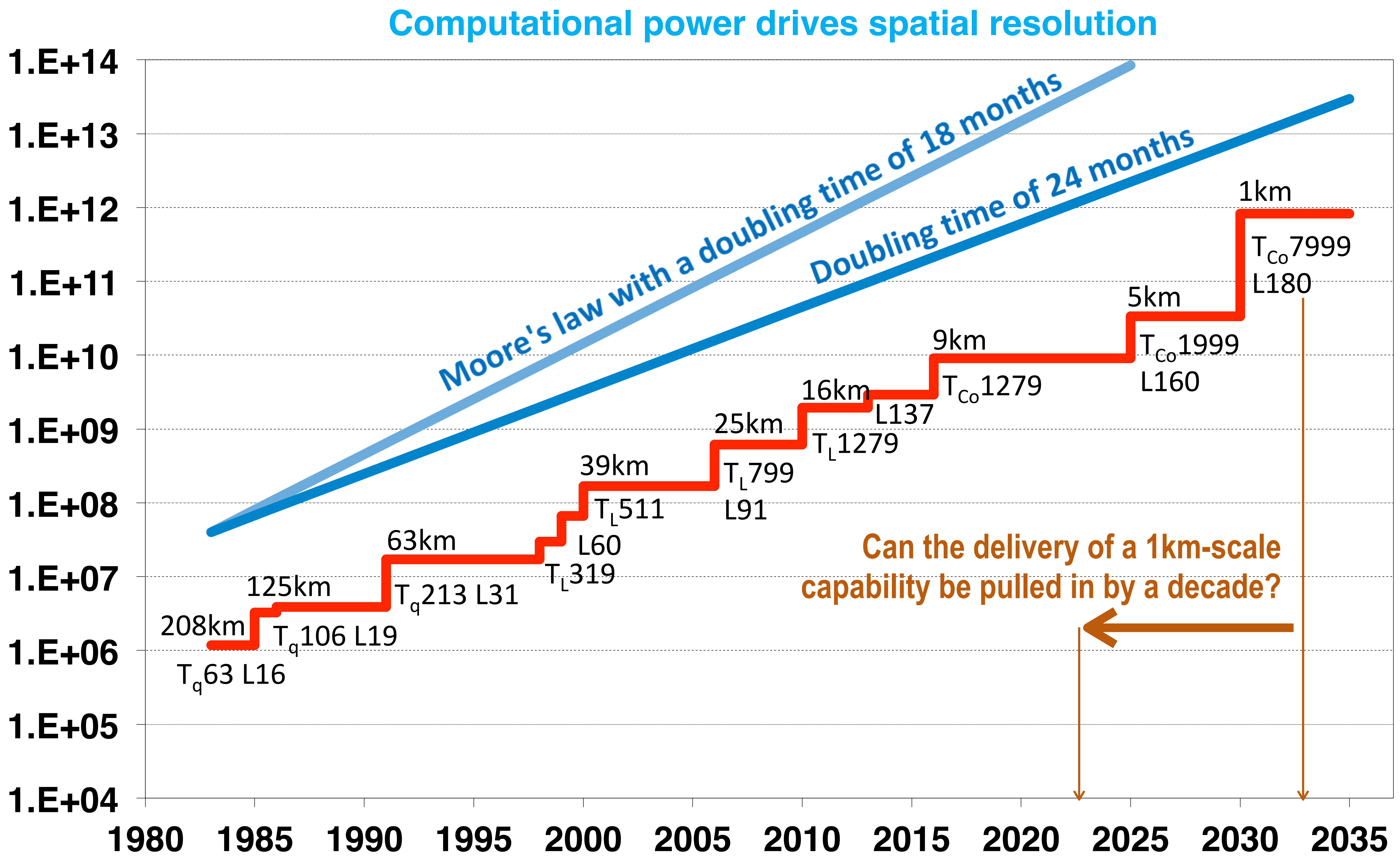
Waste heat reuse: effective energy price 35 €/MWh, negative CO2 footprint: 13500 tons reduced every year

Zero network downtime since the establishment of the DC in 2012

CSCS vision for next generation systems

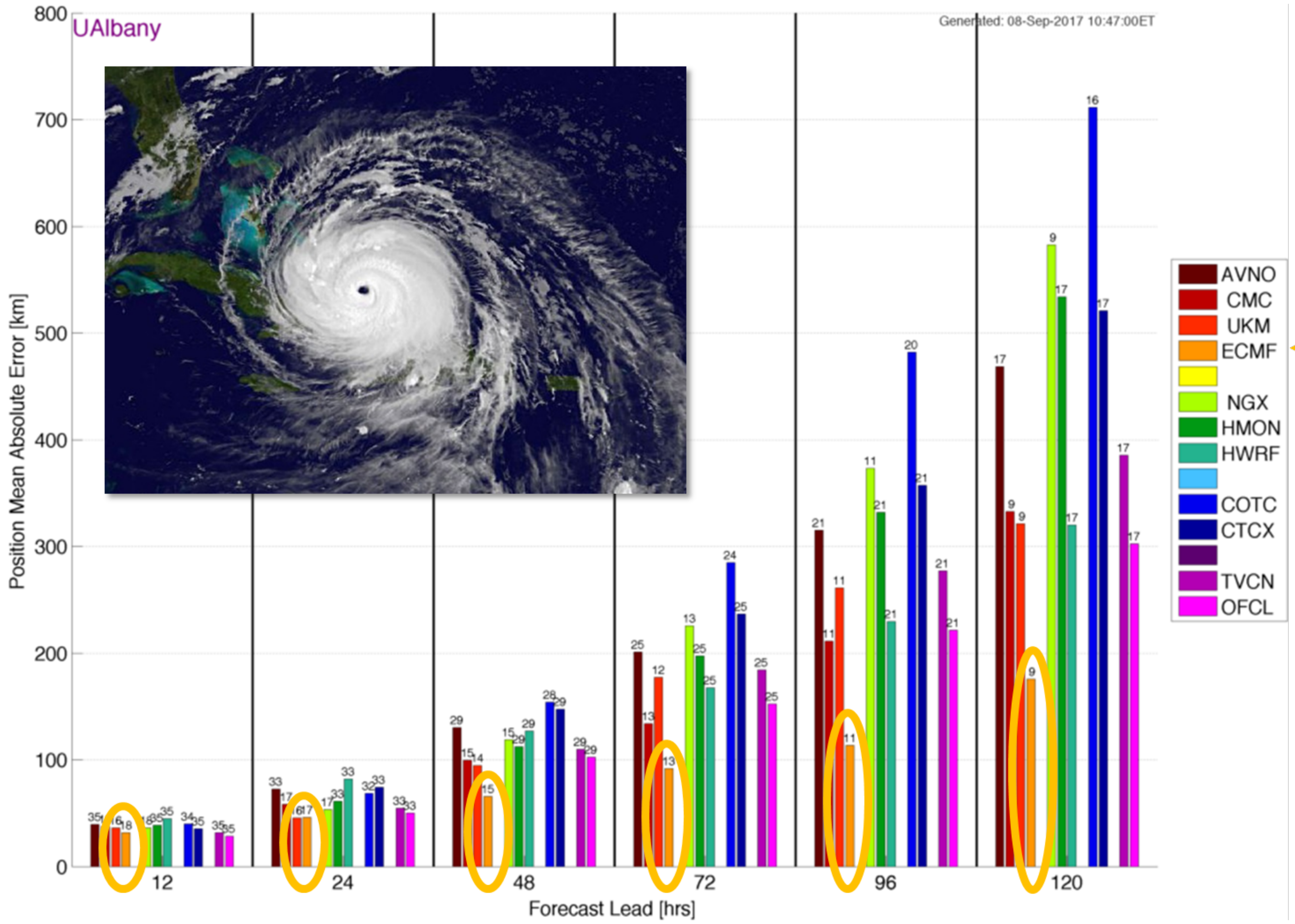
Pursue clear and ambitious goals for successor of Piz Daint

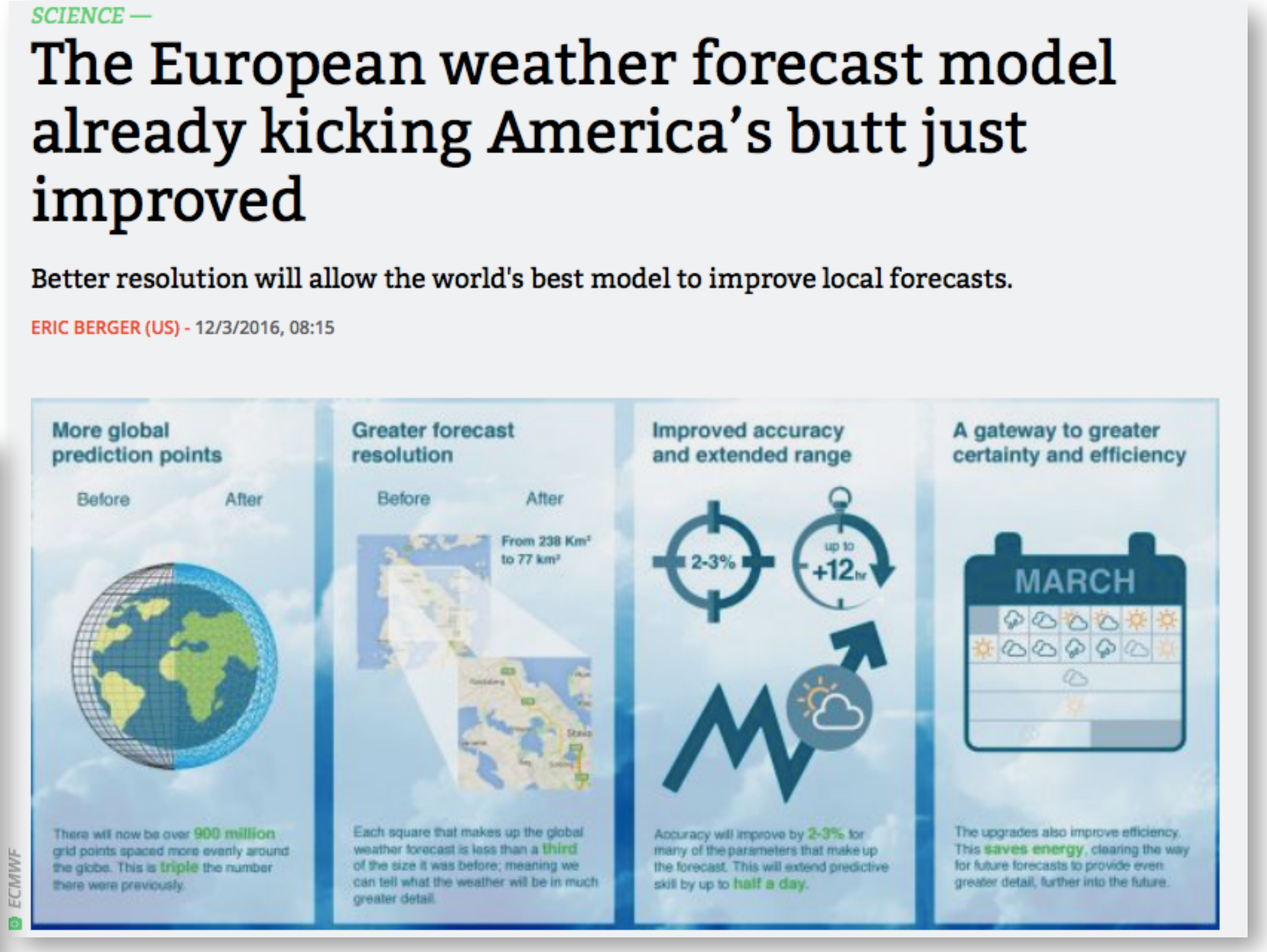
- Performance goal: develop a **general purpose system (for all domains)** with enough performance to run “exascale weather and climate simulations” by 2022, specifically,
  - Run global model with 1 km horizontal resolution at one simulated year per day throughput on a system with similar footprint at Piz Daint;
- Functional goal: **converged Cloud and HPC services** in one infrastructure
  - Support most native Cloud services on supercomputer replacing Piz Daint in 2022
  - In particular, focus on software defined infrastructure (networking, storage and compute) and service orientation



Source: Christoph Schär, ETH Zurich, & Nils Wedi, ECMWF

Leadership in weather and climate




SCIENCE —

The European weather forecast model already kicking America's butt just improved

Better resolution will allow the world's best model to improve local forecasts.

ERIC BERGER (US) - 12/3/2016, 08:15


NATIONAL GEOGRAPHIC | PHOTO OF THE DAY | TV | LATEST STORIES |

Why Are Europeans Better at Predicting Weather?

Wednesday's snow no-show in Washington was another misfire by U.S. forecasters.

By **Peter Miller**, for National Geographic News

PUBLISHED MARCH 8, 2013


SUBSCRIBE

SCIENTIFIC AMERICAN.

THE SCIENCES MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODCASTS BLOGS STORE

SUSTAINABILITY

Are Europeans Better Than Americans at Forecasting Storms?

European and U.S. models frequently make different predictions about weather and storm tracks, including that of Hurricane Joaquin. Here's why


ars TECHNICA BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE FORUMS

SCIENCE —

At times during Harvey, the European model outperformed humans

NOAA's new hurricane model, the HMON, performed terribly.

ERIC BERGER - 9/5/2017, 2:30 PM

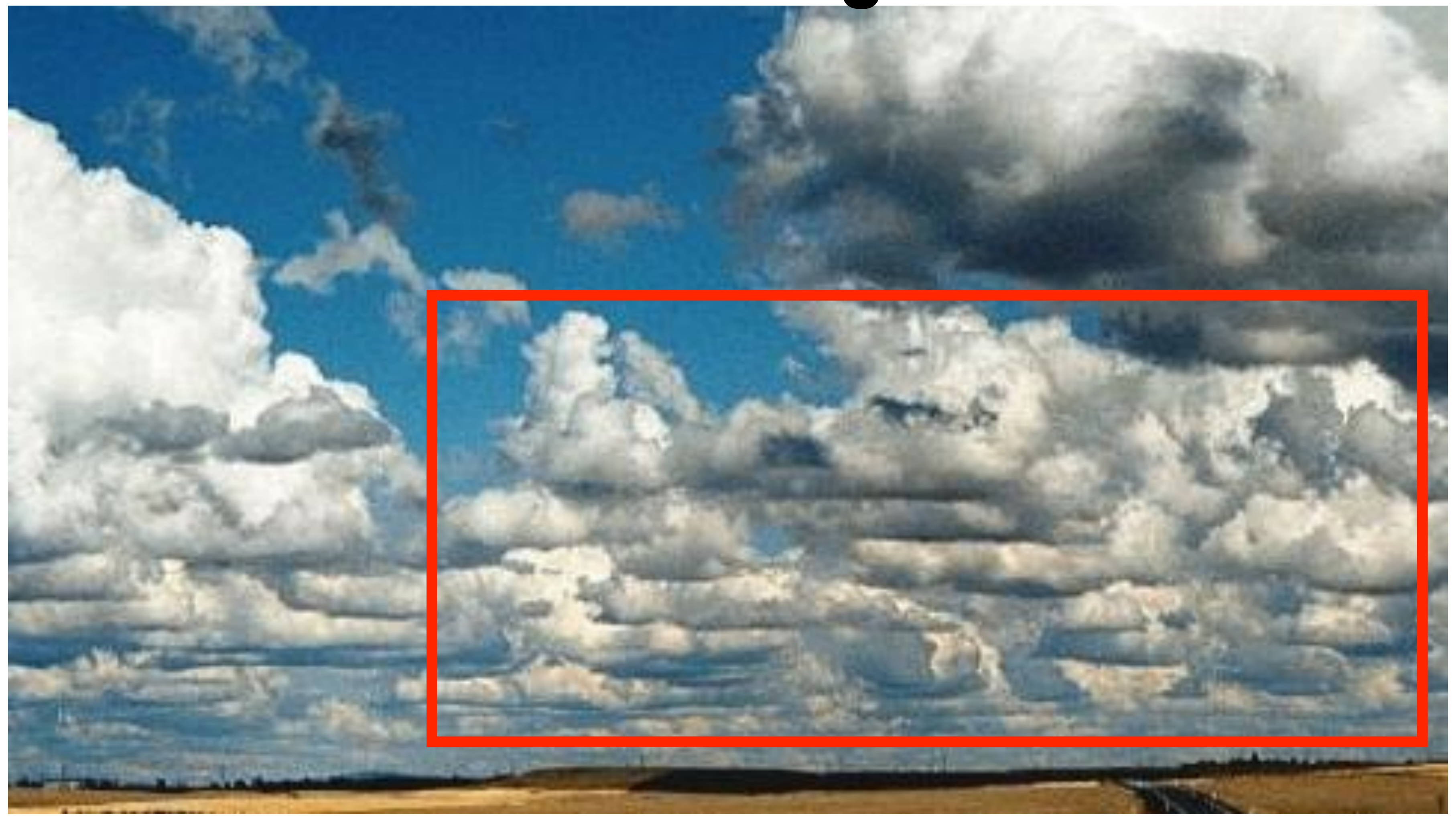
Enlarge / Which model did the best job of forecasting Harvey has a hurricane? The European model, of course.

European model may be the best - but far away from sufficient accuracy and reliability!

Peter Bauer, ECMWF

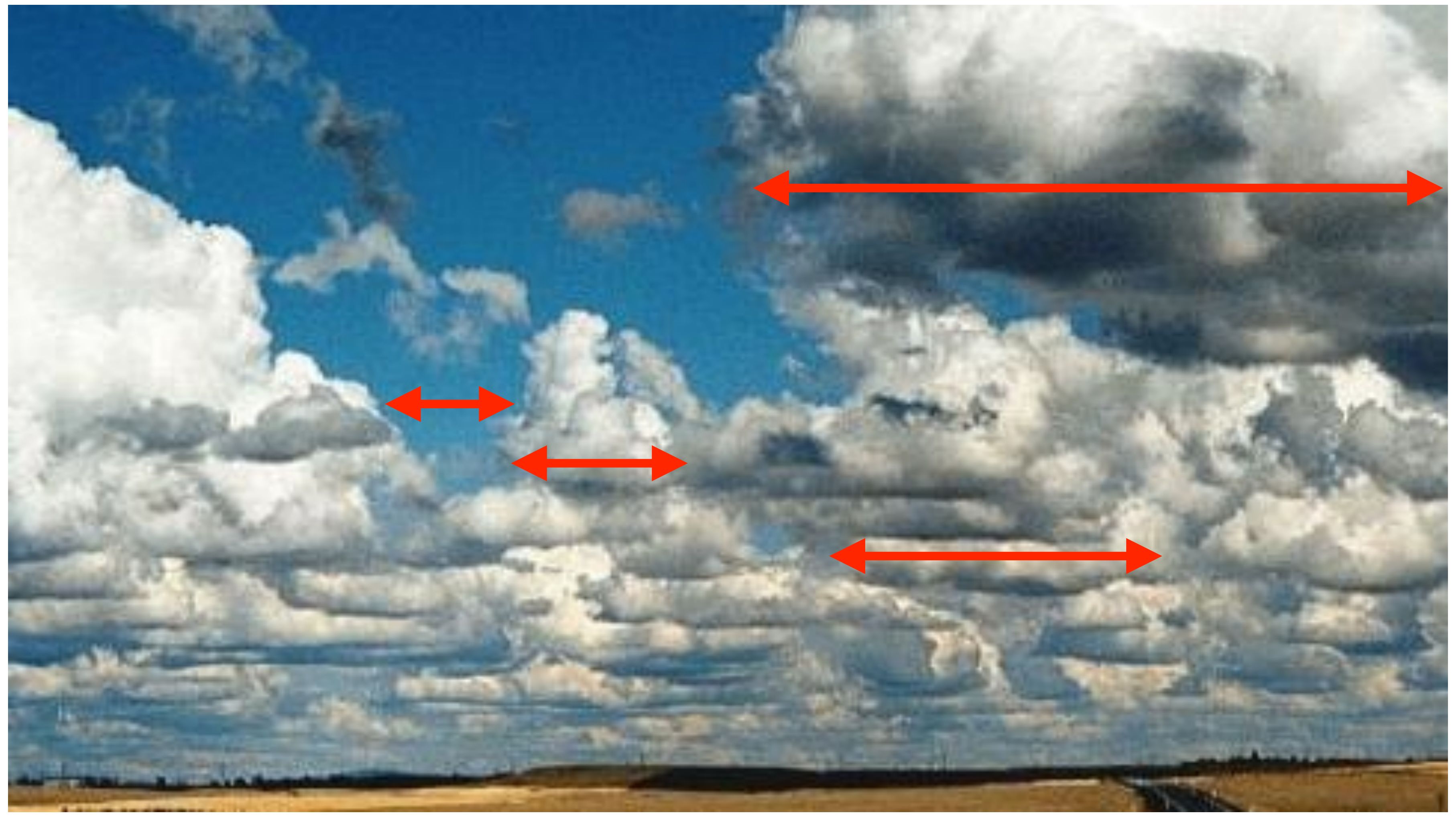
Resolving convective clouds (convergence?)

Bulk convergence



Area-averaged bulk effects upon ambient flow:
E.g., heating and moistening of cloud layer

Structural convergence

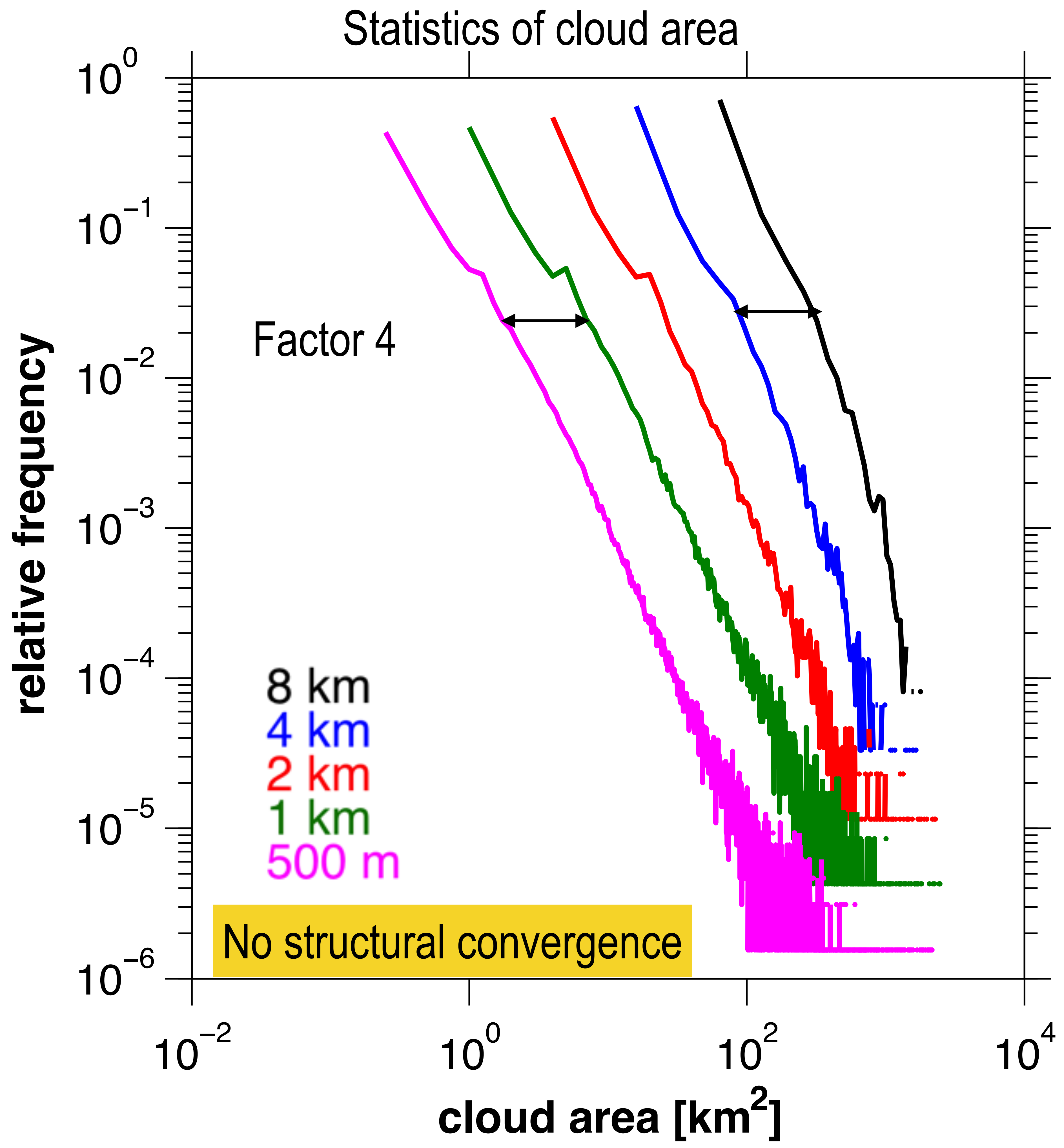


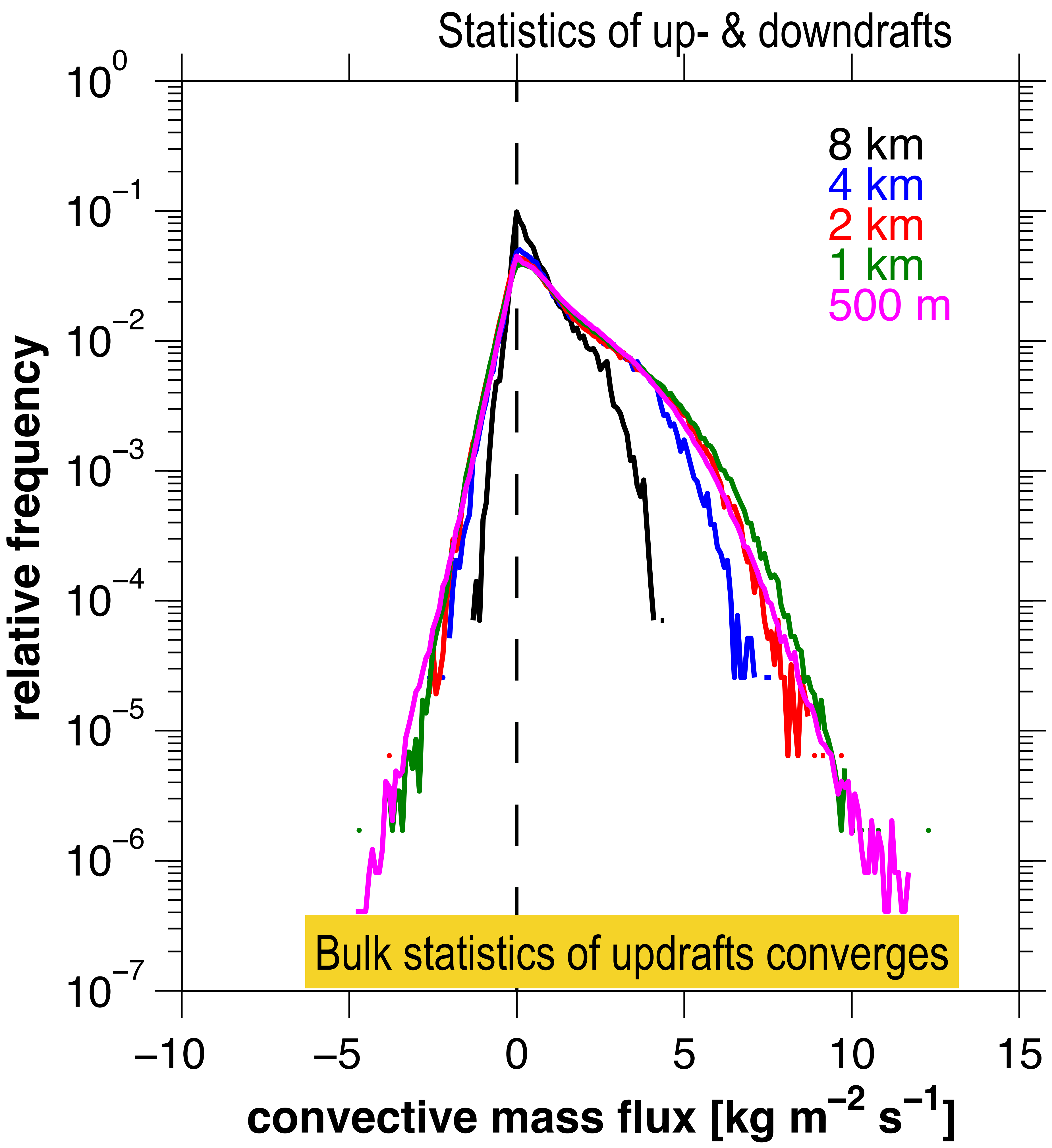
Statistics of cloud ensemble:
E.g., spacing and size of convective clouds

Source: Christoph Schär, ETH Zurich

Structural and bulk convergence

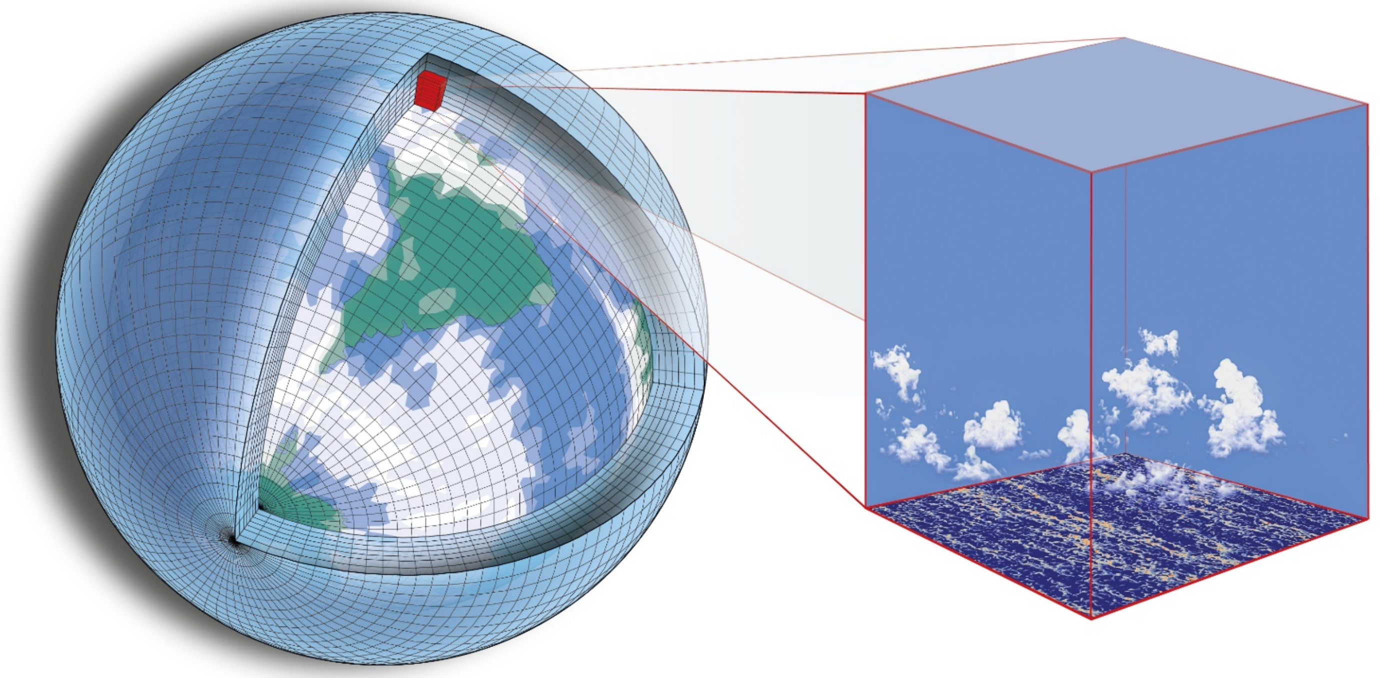
(Panosetti et al. 2018)



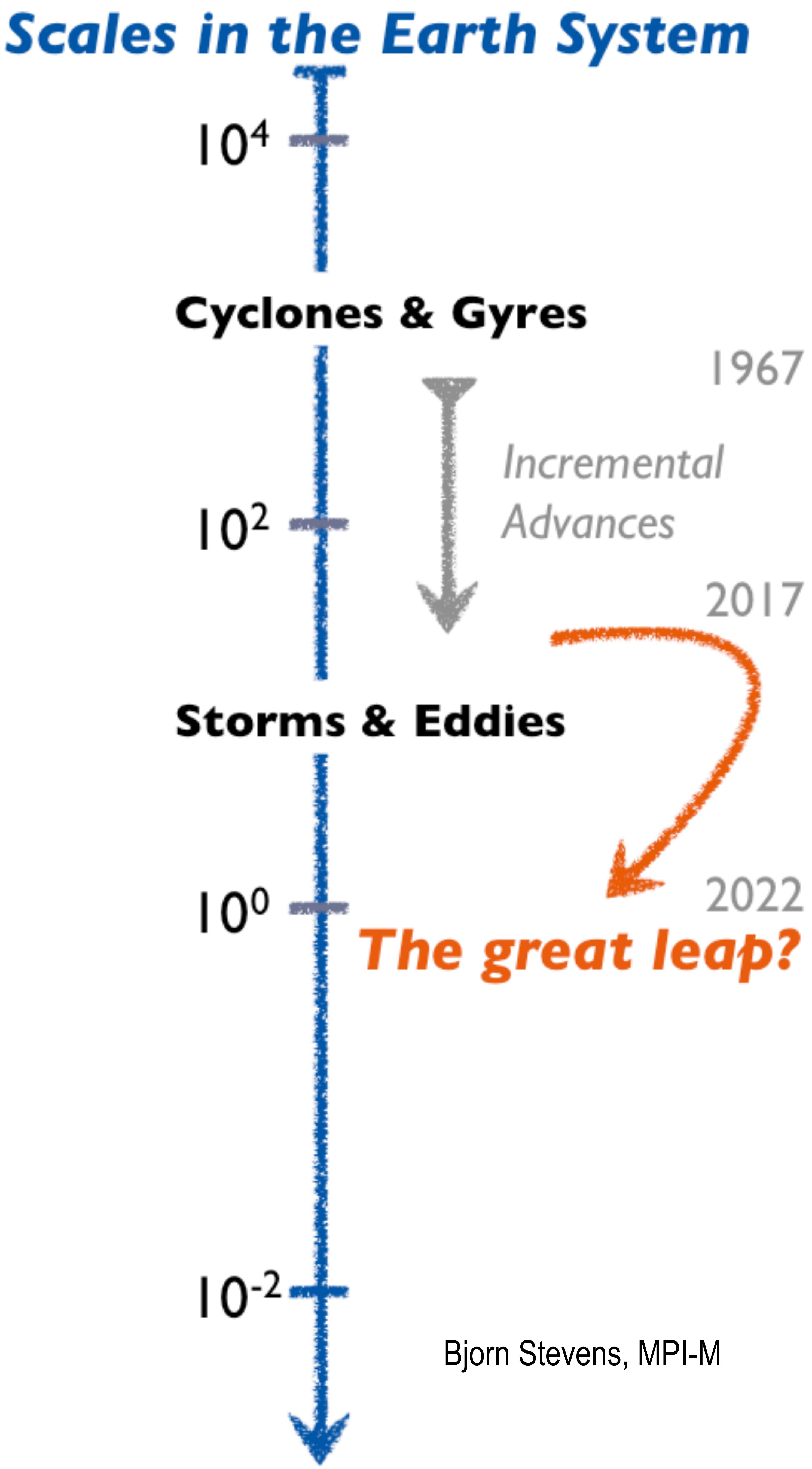


Source: Christoph Schär, ETH Zurich

What resolution is needed?



- **There are threshold scales in the atmosphere and ocean**: going from 100 km to 10 km is incremental, 10 km to 1 km is a leap. At 1km
  - it is no longer necessary to parametrise precipitating convection, ocean eddies, or orographic wave drag and its effect on extratropical storms;
  - ocean bathymetry, overflows and mixing, as well as regional orographic circulation in the atmosphere become resolved;
  - the connection between the remaining parametrisation are now on a physical footing.
- **We spend the last five decades in a paradigm of incremental advances.** Here we incrementally improved the resolution of models from 200 to 20km
- **Exascale allows us to make the leap to 1 km.** This fundamentally changes the structure of our models. We move from crude parametric presentations to an explicit, physics based, description of essential processes.
- **The last such step change was fifty years ago.** This was when, in the late 1960s, climate scientists first introduced global climate models, which were distinguished by their ability to explicitly represent extra-tropical storms, ocean gyres and boundary current.



Bjorn Stevens, MPI-M

Our “exascale” goal for 2022

Horizontal resolution	1 km (globally quasi-uniform)
Vertical resolution	180 levels (surface to ~100 km)
Time resolution	Less than 1 minute
Coupled	Land-surface/ocean/ocean-waves/sea-ice
Atmosphere	Non-hydrostatic
Precision	Single (32bit) or mixed precision
Compute rate	1 SYPD (simulated year wall-clock day)

Running COSMO 5.0 & IFS ("the European Model") at global scale on Piz Daint

Scaling to full system size: ~5300 GPU accelerate nodes available

Running a near-global (±80º covering 97% of Earths surface) COSMO 5.0 simulation & IFS

> Either on the hosts processors: Intel Xeon E5 2690v3 (Haswell 12c).

> Or on the GPU accelerator: PCIe version of NVIDIA GP100 (Pascal) GPU

The baseline for COSMO-global and IFS

	Near-global COSMO[15]		Global IFS[16]	
	Value	Shortfall	Value	Shortfall
Horizontal resolution	0.93 km (non-uniform)	0.81×	1.25 km	1.56×
Vertical resolution	60 levels (surface to 25 km)	3×	62 levels (surface to 40 km)	3×
Time resolution	6 s (split-explicit with sub-stepping)*	–	120 s (semi-implicit)	4×
Coupled	No	1.2×	No	1.2×
Atmosphere	Non-hydrostatic	–	Non-hydrostatic	–
Precision	Single	–	Single	–
Compute rate	0.043 SYPD	23×	0.088 SYPD	11×
Other (e.g., physics, ...)	microphysics	1.5×	Full physics	–
Total shortfall		101×		247×

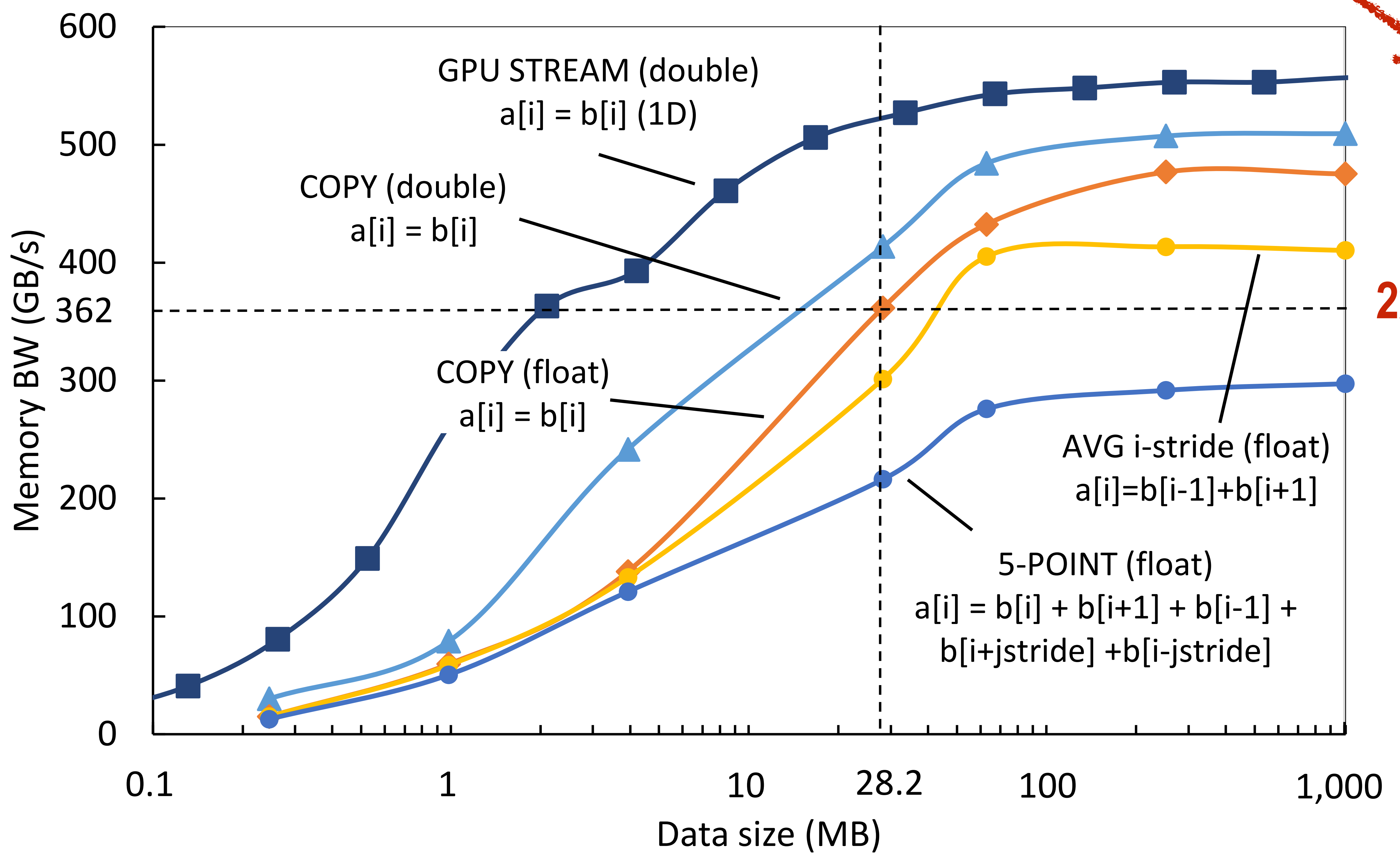
Memory use efficiency

Fuhrer et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-230, published 2018

Necessary data transfers — 0.88 — Achieved BW

$$MUE = \text{I/O efficiency} \cdot \text{BW efficiency} = \frac{Q}{D}\frac{B}{\hat{B}} = 0.67$$

Actual data transfers — 0.76 — Max achievable BW (STREAM)



0.55 w. regard to peak BW

2x lower than peak BW

Can the 100x shortfall of a grid-based implementation like COSMO-global be overcome?

1. Icosahedral/octahedral grid (ICON/IFS) vs. Lat-long/Cartesian grid (COSMO)
   - 2x fewer grid-columns
   - Time step of 10 ms instead of 5 ms

   4x

2. Improving BW efficiency
   - Improve BW efficiency and peak BW
   - (results on Volta show this is realistic)

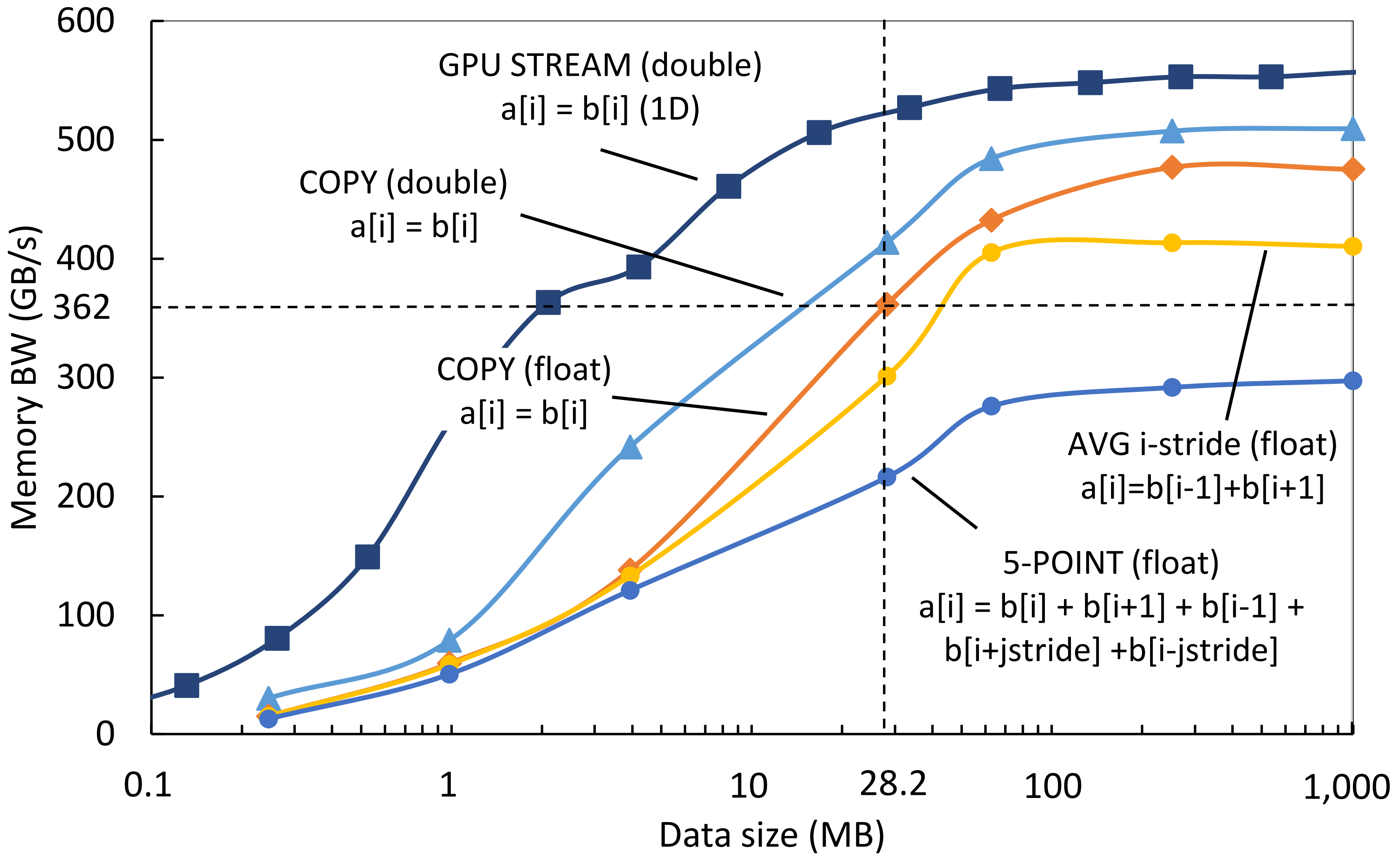
   2x

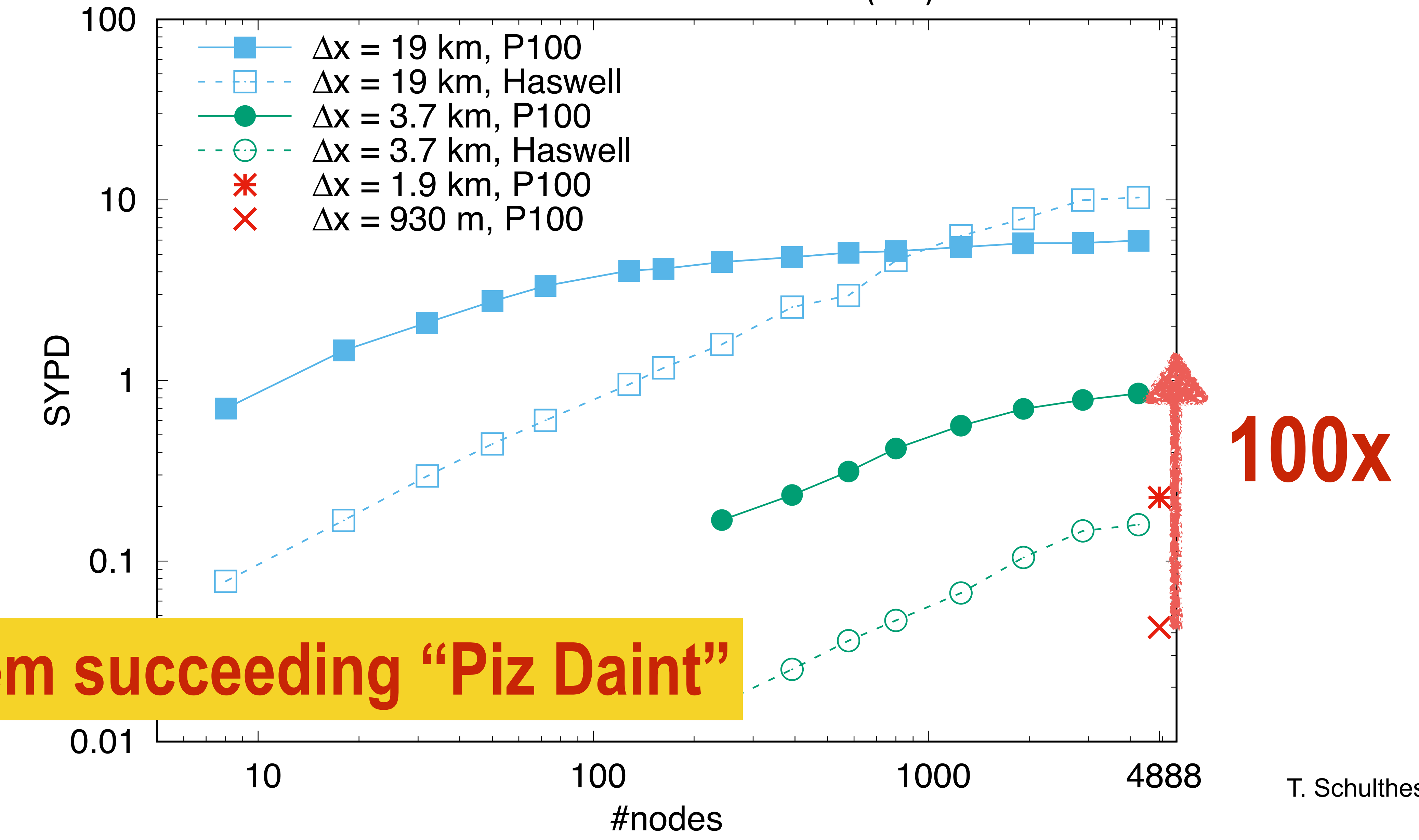
3. Strong scaling
   - 4x possible in COSMO, but we reduced available parallelism by factor 1.33

   3x

4. Remaining reduction in shortfall
   - Numerical algorithms (larger time steps)
   - Further improved processors / memory

   4x





But we don't want to increase the footprint of the 2022 system succeeding "Piz Daint"

What about ensembles and throughput for climate? (Remaining goals beyond 2022)

1. Improve the throughput to 5 SYPD

Change the architecture from control flow to data flow centric (reduce necessary data transfers)

$$MUE = \text{I/O efficiency} \cdot \text{BW efficiency} = \frac{Q}{D}\frac{B}{\hat{B}}$$

Necessary data transfers → Q; Actual data transfers → D; Achieved BW → B; Max achievable BW → $\hat{B}$

2. Reduce the footprint of a single simulation by up to factor 10-50

We may have to change the footprint of machines to hyper scale!

Much of the data present here was from this article

Race to Exascale Computing

Theme Article

Reflecting on the Goal and Baseline for Exascale Computing: A Roadmap Based on Weather and Climate Simulations

Thomas C. Schulthess
ETH Zurich, Swiss National Supercomputing Centre

Peter Bauer
European Centre for Medium-Range Weather Forecasts

Nils Wedi
European Centre for Medium-Range Weather Forecasts

Oliver Fuhrer
MeteoSwiss

Torsten Hoefler
ETH Zurich

Christoph Schär
ETH Zurich

***Abstract*—We present a roadmap towards exascale computing based on true application performance goals. It is based on two state-of-the art European numerical weather prediction models (IFS from ECMWF and COSMO from MeteoSwiss) and their current performance when run at very high spatial resolution on present-day supercomputers. We conclude that these models execute about 100–250 times too slow for operational throughput rates at a horizontal resolution of 1 km, even when executed on a full petascale system with nearly 5000 state-of-the-art hybrid GPU-CPU nodes. Our analysis of the performance in terms of a metric that assesses the efficiency of memory use shows a path to improve the performance of hardware and software in order to meet operational requirements early next decade.**

Digital Object Identifier 10.1109/MCSE.2018.2888788
Date of publication 24 December 2018; date of current version 6 March 2019.

SCIENTIFIC COMPUTATION WITH precise numbers has always been hard work, ever since Johannes Kepler analyzed Tycho Brahe's data to

Collaborators on Exascale (climate)


Tim Palmer (U. of Oxford)


Bjorn Stevens (MPI-M)


Peter Bauer (ECMWF)


Oliver Fuhrer (MeteoSwiss)


Nils Wedi (ECMWF)


Torsten Hoefler (ETH Zurich)


Christoph Schar (ETH Zurich)

Thank you!