Operational NWP at DWD

Life before Exascale

Ulrich Schättler
Department for Numerical Modeling (FE1)
Deutscher Wetterdienst, Germany

With contributions from many colleagues!

Contents

New Hardware at DWD: NEC SX-Aurora Tsubasa

Procurement in 2019

General Settings:

- In 2019 DWD asked for a replacement of the hardware with an upgrade of computing power by a factor of about 3.0 in a first phase (2020-2022).
- In a second phase, an additional upgrade was desired, but it was left to the vendor, how much could be delivered within the financial limits (58 M€ for 2019-2023).
- „Hardware“ consists of
  - servers for the models and for pre- and postprocessing (operational / research);
  - data servers and storage systems (operational / research);
  - (smaller) compute server and storage system for the Military Service.
- Research system has to be installed in a different location (disaster backup).

Additional condition:

- Thanks to the work done by colleagues from MeteoSwiss and CSCS, COSMO can run on GPUs.
- Work for porting ICON to GPUs is ongoing. So a GPU based computer was not yet an option for DWD.

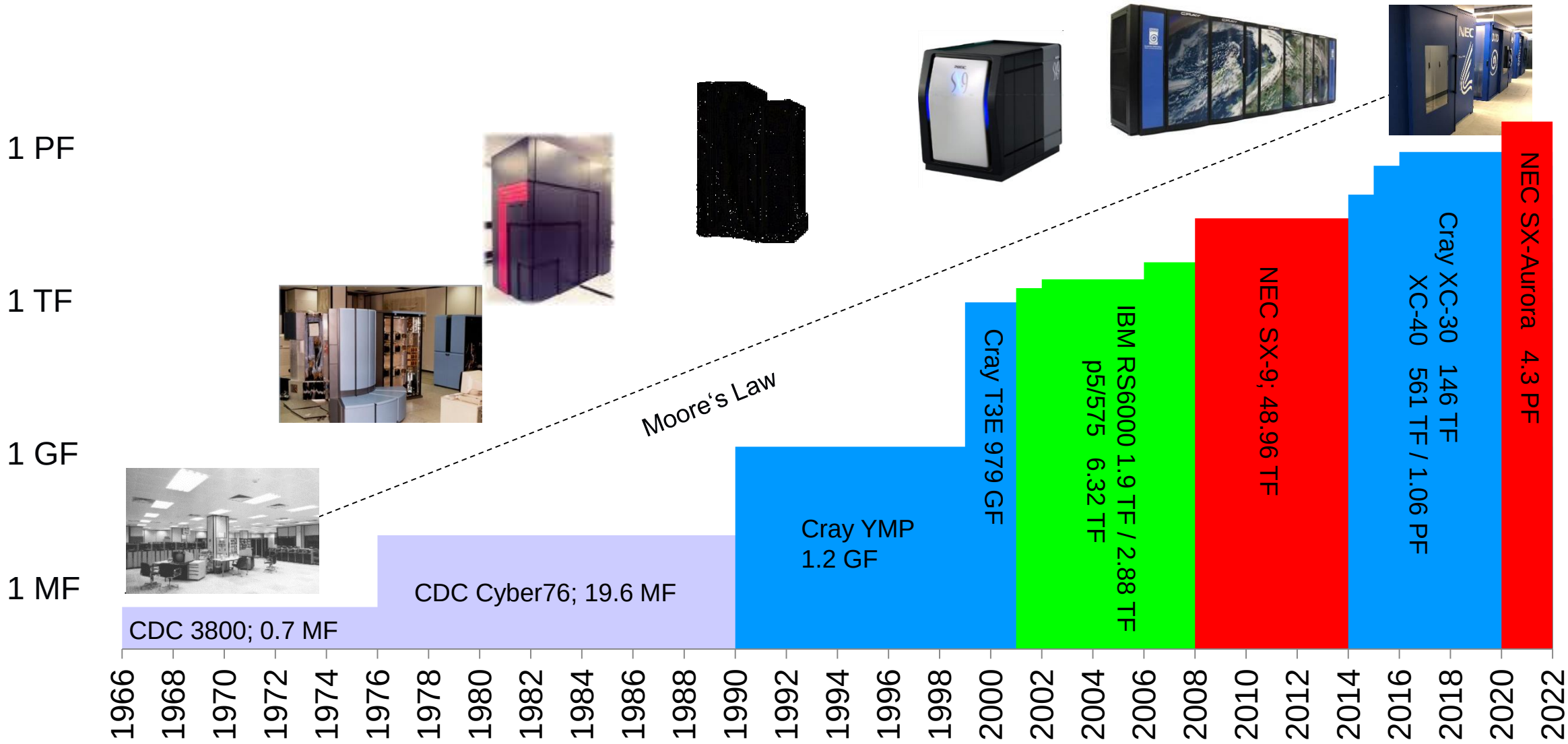
And the Winner is…

NEC with SX-Aurora Tsubasa

- NEC offered a duplication of performance in Phase 2 (total upgrade factor of 6.0). Competitors were way below 5.
- At the same time, SX-Aurora Tsubasa has the lowest power consumption. Typical power expected for Phase 2 is 777 kW. Competitors were between 899 and 1060 kW (with lower upgrade factor).
- From the latest Green500 list (June 2021; Phase 1): 2 entries for SX-Aurora Tsubasa, NEC, Deutscher Wetterdienst (Research and Operational Cluster)

Rank in Green 500	Rank in TOP 500	#cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)	Efficiency (GFlops/W)
49	105	18,688	3,870.1	5,605.6	648	5.972
50	124	14,336	3,250.4	4,282.1	565	5.752

DWD Computers 1966-2021



SX-Aurora Tsubasa

The building block of the NEC vector system is a node, which consists of

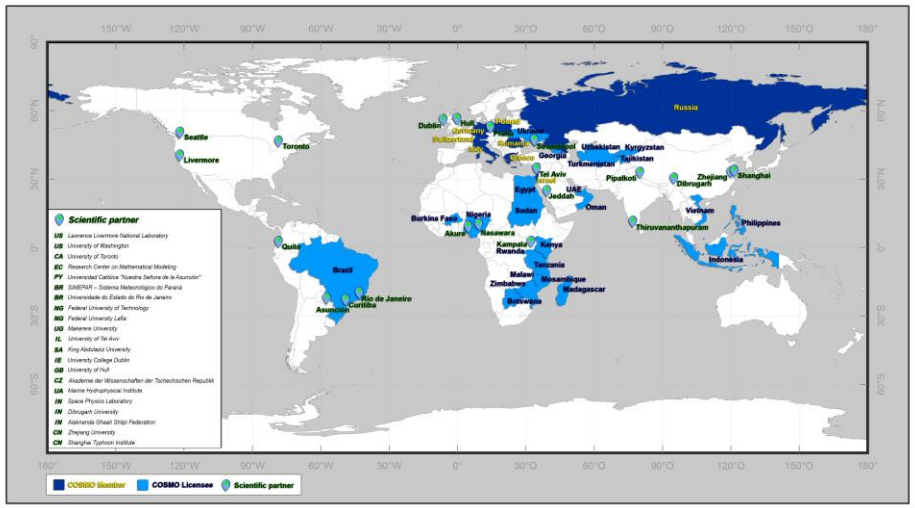
- a vector host (VH): a scalar CPU (24-core AMD Rome; 2.8 GHz; 256 GB memory)
- 8 vector engines (VE): SX-Aurora 1 TSUBASA Typ 10AE (PCIe card) with
  - 8 vector cores: 1.5 GHz; 304.1 GF/s (DP); 608.3 GF/s (SP) per core
  - 48 GB HBM2 3D-stacked memory (6 GB / core; 1.35 TB/s bandwidth)

Phase	Operational			Research			Available
	VH	VE	Cores	VH	VE	Cores	
0	178	1424	11392	232	1856	14848	12/2019
1	224	1792	14336	292	2336	18688	12/2020
2	325	2600	20800	424	3392	27136	12/2022

Modifications to the Operational NWP Suite

Time To Say Good-Bye

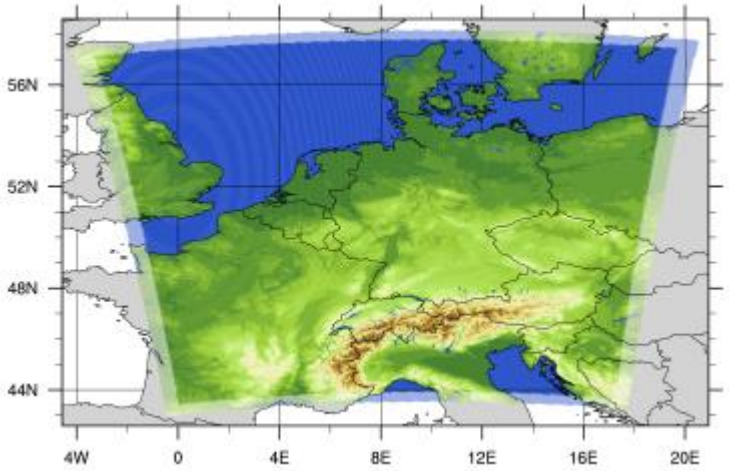
After more than 21 years of operational regional forecasting, the COSMO-Model will retire.



- It has been switched off at DWD and has been replaced with ICON-LAM (limited area mode) on 10th of February 2021.
- Work has started at all consortium partners to migrate their operational workload also to ICON-LAM.
- Licensees will be ported later. There are still some license issues to be clarified.

Phase 1: Moderate Upgrades

ICON-D2 (which replaced COSMO-D2) now does a 48 hour forecast instead of 27 hours (with same resources).



In the last months the following measures were tested:

- ➔ Increasing the vertical resolution in the tropopause and stratosphere for the global deterministic model from 90 to 120 levels. Horizontal resolution still 13/6.5 km.

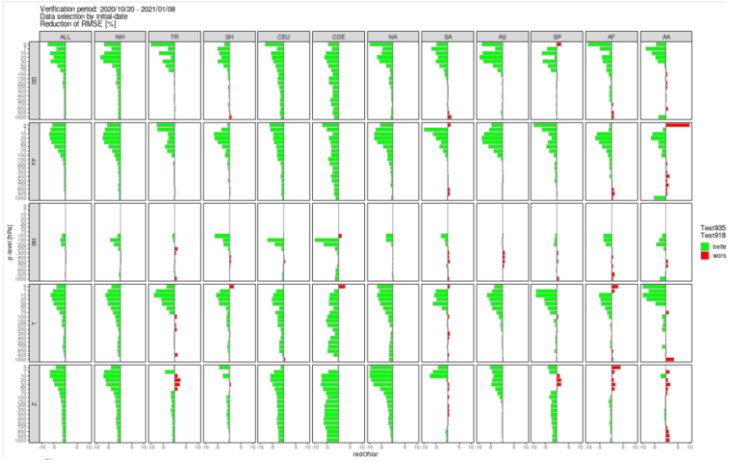
  Needs 48 instead of 36 VE.

- ➔ Increasing the horizontal resolution of the EPS (incl. nest) from 40/20 km to 26/13 km. And the vertical resolution from 90 / 60 levels to 120 / 74 levels.

  Needs 8 instead of 2 VEs per member.

Will become operational soon.

Scorecard for DET Tests



Phase 2: SINFONY

Goal of the second upgrade is a seamless prediction of the atmospheric state, in particular for the prediction of small-scale, severe weather events.

- A new prediction system is developed in the project SINFONY (2019-2022):

  „Seamless INtegrated FOrecastiNg sYstem“.

- Products are created as a blend of nowcasting and very short-range forecasts.

  (see: https://www.dwd.de/EN/research/researchprogramme/sinfony_iafe/sinfony_en_node.html)

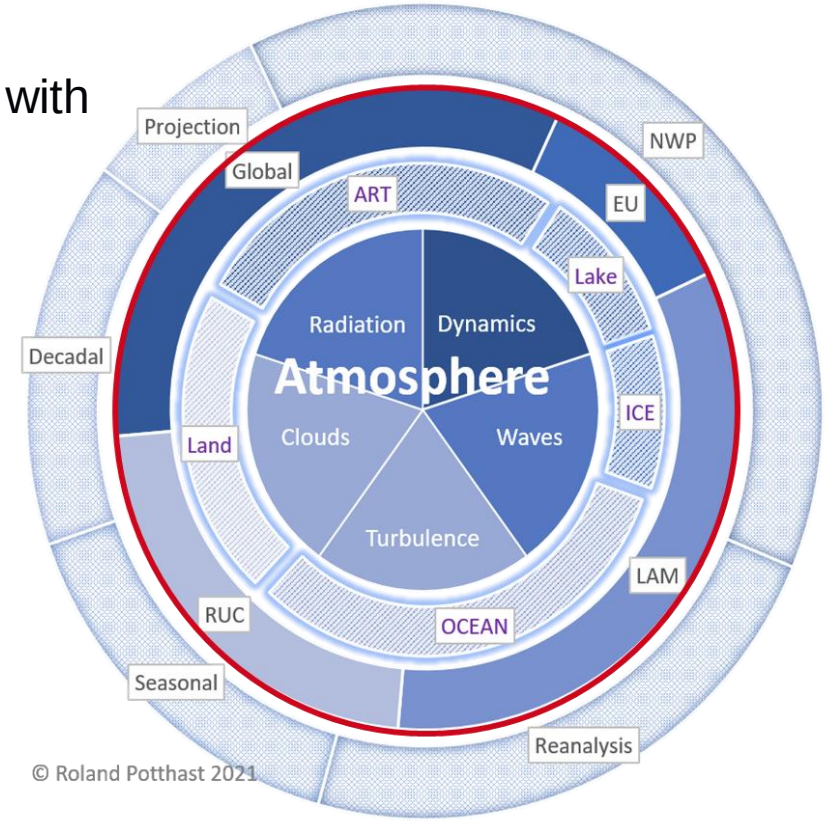
- An hourly Rapid Update Cycle based on an ICON limited area mode ensemble with 2 / 1 km grid size will be implemented.

Building up and testing the system is already ongoing.

ICON

ICON Modelling Framework

- Nonhydrostatic dynamical core on icosahedral-triangular Arakawa-C grid with mass-related quantities at cell circumcenters.
- Local mass conservation and mass-consistent transport
- Two-way nesting with capability for multiple overlapping nests.
- Limited-area mode available.
- Available components: NWP, climate, ocean, land.
- At the moment 2 different physics packages for NWP and climate mode.
- Runs in global NWP mode at DWD since January 2015.



- ICON-ART by KIT: Simulation of aerosols, trace gases and pollen.

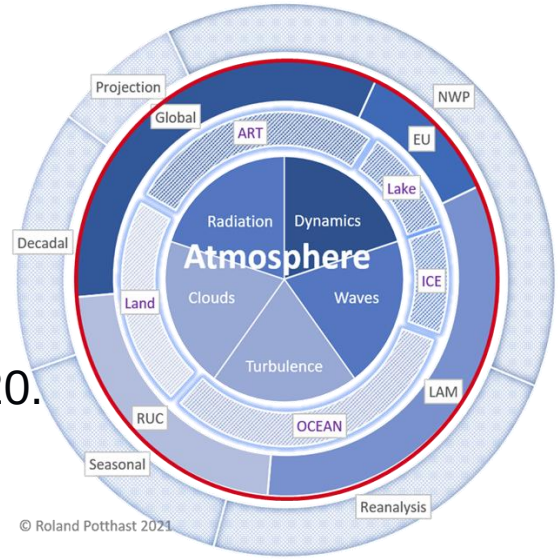
- Thanks to the work by CSCS, MPI-M, DKRZ and MCH, parts of ICON (dynamics, parts of climate physics and parts of NWP physics) are already running on GPUs.

ICON Governance

- Joint development project originally between DWD and the Max-Planck-Institute for Meteorology.
- Further development and governance with partner institutes:
  - German Climate Computing Center (DKRZ)
  - Institute of Meteorology and Climate Research (IMK) at KIT: ICON-ART
  - Center for Climate Systems Modelling (ETH Zürich: joined earlier this year)
- ICON Directors Board (ICON-D5) and ICON Coordination Group (ICON-C5) consist of members from these 5 partners.
- Associated groups:
  - Consortium for small-scale modelling (COSMO)
  - Climate limited-area modelling community (CLM)
  - Swiss National Computing Center (CSCS)
- Licenses are available for research and (soon) for NHMS for operational duties.

ICON Projects

Some projects with (major) DWD contributions (besides standard developments):



ICON-Seamless:

- NWP plus climate forecasting with ICON-A-O-L and Data Assimilation.
- Initiated by the directors of DWD, MPI-M, KIT and DKRZ in October 2020.
- Expert Groups (Atmosphere, Ocean, Land, Data Assimilation) started work in 12/2020.

ICON-Consolidated: Software Redesign (Interfaces and Structures)

- Interface Redesign Workshop in Q4/2021
- Software Concept Design Workshop in Q4/2021

Technical Projects: („Computer Science“ Projects)

- PASC-ENIAC: (CH) Enabling the ICON model on heteorogeneous architectures.
- COSMO PP IMPACT: ICON on Massively Parallel Architectures.
- IAFE-GPU: DWD project to support ICON and Data Assimilation implementation on GPUs

ICON Projects

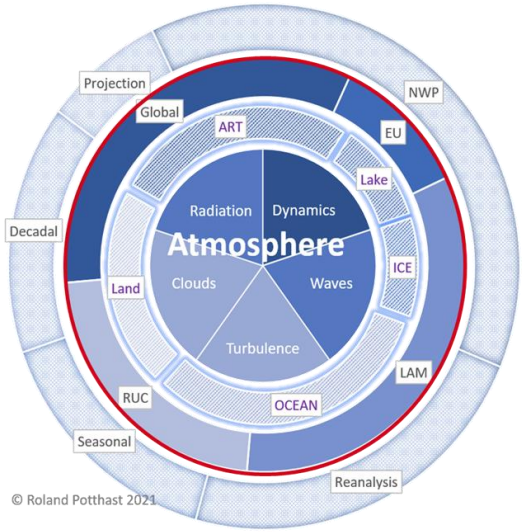
Some projects with (minor) DWD contributions, initiated by climate colleagues:

EXCLAIM: Extreme scale computing and data platform for cloud-resolving weather and climate modeling)

- ➔ Swiss interdisciplinary open ETHZ project that aims to develop an ICON-model based infrastructure capable of running kilometer-scale climate simulations.
- ➔ Brings together computer scientists with data and domain experts.
- ➔ See: https://exclaim.ethz.ch

(Pre-)Warmworld

- ➔ BMBF Programme (German Federal Ministry for Education and Research).
- ➔ Contribution to ICON refactoring and consolidation / storm resolving GCM.



These last projects sure need (at a certain point in time):

Exascale Computing

Exascale Computing

According to experts, Exascale Computing needs

- ➔ next generation of supercomputers: new hardware architectures, accelerators such as GPUs,…
- ➔ new programming paradigms / languagues: Cuda, HIP, Kokkos, Julia, DSLs,…

So how far can you get with

- ➔ Standard Programming Language + MPI + OpenMP / OpenACC?
- ➔ At least you can try.

- ➔ Fortran + MPI + OpenACC is used to port ICON to GPUs. Lately DWD joined these efforts.
- ➔ GPU machine available at DWD: Up to now only our Linux workstations, with NVIDIA (PGI) compiler 21.2. This is a good platform for development, of course not for testing.
  - ➔ Device: NVIDIA Quadro P400
  - ➔ Memory: 2 GB

Very Early Results

Some first time measurements for ICON-LAM on the workstation, but also on PIZ Daint, CSCS.

- ➔ Workstation: small domain with 1788 cells; 1 hour of forecast
- ➔ Piz Daint: ICON-D2 with 2.949.120 cells; 1 hour of forecast

	Workstation		Piz Daint	
	1 CPU nproma=16	1 GPU nproma=1788	32 CPUs nproma=16	32 GPUs nproma=8192
Total	21.390	22.489	1156.176	678.565
nh solve	6.237	3.191	608.619	17.830
NWP Microphysics	0.488	0.367	52.789	0.873
NWP Turbulence	0.566	1.450	74.444	16.825
NWP Convection	0.699	3.372	43.744	25.660

- ➔ Porting and optimization is still ongoing.
- ➔ Convection has several vertical nested loops (ascent / descent of particles), which result in a lot of kernels.

Coming Back to Exascale

Tests done by Günther Zängl (DWD): ICON (global) on SX-Aurora:

Time for a 7.5 days forecast for three different resolutions:

	R3B7N8 (13 / 6.5 km) (oper. setup)	R3B8 (6.5 km) (no nest!)	R3B9 (3.25 km) (no nest!)
# cells (2D)	2.949.120	11.796.480	47.185.920
# levels	90 / 60	120	120
time step (s)	120 / 60	60	30
# nodes / cores	4.5 / 288	15 / 960	75 / 4800
Time (s)	2897	4905	10975

Scaling tests for R3B8 (only 3 days forecast)

# nodes	# cores	Time (s)
15	960	2062
30	1920	1074
60	3840	566
120	7680	322

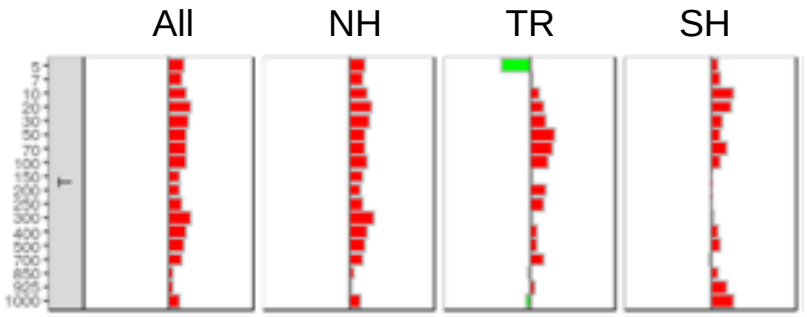
It‘s a long way to Exascale!

And timings are only part of the story…

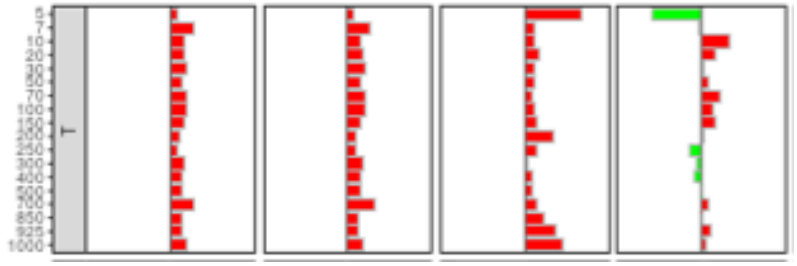
Verification against Radiosondes

here: Temperature

13 vs. 6.5 km
(green: 6.5 km better)



6.5 vs. 3.25 km
(green: 3.25 km better)



Comments by Günther:

- With explicitly simulated convection, some aspects that are well known to be notoriously misrepresented by parameterizations improve: intensity spectrum and diurnal cycle of convection, organization and propagation of mesoscale convective systems.
- However, a huge cold bias arises that has its maximum in the upper tropical troposphere. It entails a planetary-scale redistribution of mass that compromises the pressure forecast over the whole globe.
- Convection-permitting is not convection-resolving!
- Leaving the convective gray zone means entering the gray zone of turbulence: the list of open issues will remain long.

Conclusions

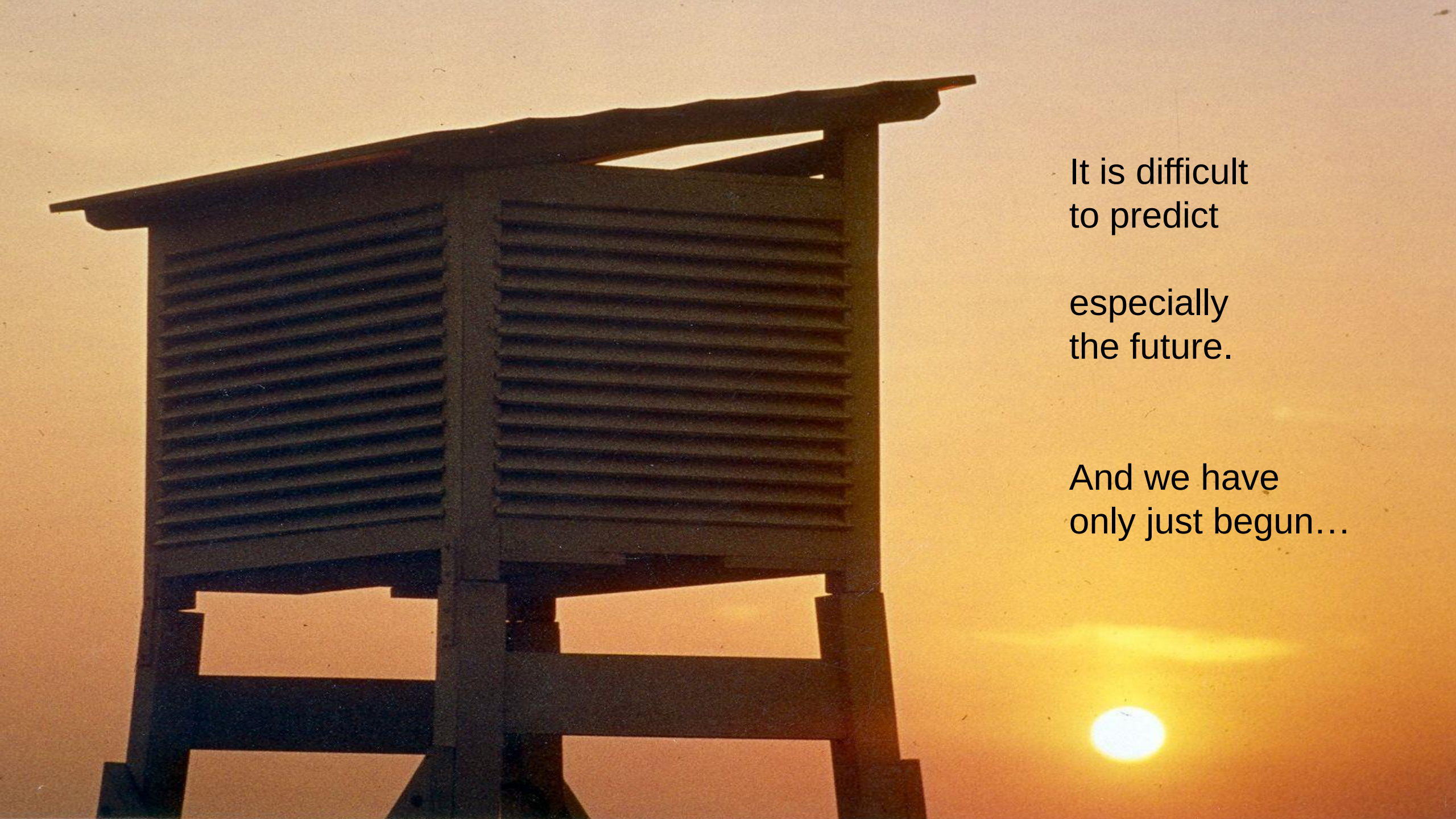
Conclusions

- DWD‘s operational NWP suite is now running on an energy-efficient powerful vector architecture.

- For NWP at DWD, exascale computing is not necessary tomorrow:
  - To really benefit from global convection-permitting configurations, substantial development work on the physics package will have to be done in the upcoming decade.
  - Nonetheless we understood the message: „The Times They Are a-Changin“ (Bob Dylan)

- ICON partners running climate applications might have a different view on the timeline.

- Exascale Projects have to tackle both:
  - The technical aspect of implementation.
  - The improvement of the model for very high resolutions.



It is difficult
to predict

especially
the future.

And we have
only just begun…