The ESiWACE project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 675191

This material reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.
1. Introduction

2. ESiWACE selected results

3. ESiWACE2 plans
The weather & climate community has a “nearly infinite“ need for computing capability and computing capacity:

- We could and would do better science if we had faster (better) HPC

This community has a growing problem with HPC

- Systems get broader not faster (in fact the may get slower)
- It is increasingly difficult to make progress in leveraging new systems
- The market is driven by cell phones and deep learning
The European HPC ecosystem

Access to best HPC for industry and academia
EXCELLENT SCIENCE E-INFRASTRUCTURES

Specifications of exascale prototypes. Technological options for future systems.

Collaboration of HPC Supercomputer Centres and application CoEs. Provision of HPC capabilities and expertise.

Identify applications for co-design of exascale systems. Innovative methods and algorithms for extreme parallelism of traditional/emerging applications.

Excellence in HPC

+ national and institutional funding

Centres of Excellence

Joachim Biercamp, DKRZ - ESIWACE
ESIWACE (phase 1) objectives

- ESIWACE will substantially improve the efficiency and productivity of numerical weather and climate simulation on high-performance computing platforms.

- ESIWACE will support the end-to-end workflow of global Earth system modelling for weather and climate simulation in high performance computing environments.
Selected results
Global high-resolution model demonstrators:

- **Demonstrate computability of:**
  - 1km global Atmosphere, IFS, 2017
  - 1km global Atmosphere, ICON, 2017
  - 1km global Ocean, NEMO, 2019
  - at least 10km global, coupled ESM, EC-EARTH, 2019
  - target 1km global, coupled ESM, ICON-ESM, 2019

- **Demonstrate that size and communality of the problem justifies a coordinated approach and strategic investment**

Developing the next generation of climate and weather codes and related environments for exascale is a long-standing issue, beyond what ESiWACE can do in 4 years, but to which ESiWACE will help paving the way.
ICON-APE @1.25km and IFS @1.3km

Aqua planet simulations:
From Stevens, Bony (2013). Science 340 (6136), 1053-1054

- 1.25km resolution,
  335 544 320 horizontal cells, 45 vertical levels
- 1408 nodes, hybrid 2MPI x 18 OpenMP configuration
  -> ca. 50.000 cores
- throughput: 1.8 simulated days per day (no IO)

benchmark (160km - 5km) available at:
https://redmine.dkrz.de/projects/icon-benchmark/wiki/
Performance Models for Scalability Predictions at Exascale

Model 1: compute + communication

Model 2: compute + communication + load imbalance (extrapolated)

ICON-DYAMOND 5km Performance

- Good, yet limited, scalability
- Estimated performance shortfalls of models on the way to exascale
- **ICON shortfall**: $O(17)$ for 1km atmosphere-only simulations
- No matter how many compute nodes we have, we’ll hit the scalability limit, (here ~ 250 000 nodes)
The cost profile of a 1.25km (non-hydrostatic) IFS atmosphere simulation Piz Daint

**Example: TCo7999 L62 (~1.25km)**

- 63% comms
- 37% compute
- 17% GP_DYNAMICS
- 18% SI_SOLVER
- 60% SP_TRANSFORMS
- 4880 MPI tasks x 12 threads
- 32 FC/day ~ 0.088 SYPD
- single precision / FLT
- ~191.74 MWh / SY

Based on the Piz Daint, Swiss Cray XC50 Haswell, Aries interconnect, ~5000 nodes total
DYAMOND
DYnamics of the
Atmospheric general circulation
Modeled On
Non-hydrostatic Domains

Identifying similarities and differences that emerge at storm resolving scales (1 km to 5 km) as compared to traditional (hydrostatic-scale) representations of the atmospheric circulation;

Open to all interested international groups
Technical support through ESIWACE

https://www.esiwace.eu/services/dyamond
<table>
<thead>
<tr>
<th>Model</th>
<th>Horizontal Resolution</th>
<th>Vertical level</th>
<th>Model top</th>
<th>#columns</th>
<th>data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAM</td>
<td>4km</td>
<td>74</td>
<td>37km</td>
<td>42,467,328</td>
<td>38TB</td>
</tr>
<tr>
<td>FV3</td>
<td>3.25km</td>
<td>75</td>
<td>3hPa</td>
<td>56,623,104</td>
<td>62TB</td>
</tr>
<tr>
<td>ICON</td>
<td>2.5km</td>
<td>90</td>
<td>75km</td>
<td>83,886,080</td>
<td>132TB</td>
</tr>
<tr>
<td>ICON (varying/fixed SST)</td>
<td>5km</td>
<td>90</td>
<td>75km</td>
<td>20,971,520</td>
<td>8/28Tb</td>
</tr>
<tr>
<td>NICAM</td>
<td>3.5km</td>
<td>78</td>
<td>50km</td>
<td>41,943,042</td>
<td>28TB</td>
</tr>
<tr>
<td>NICAM</td>
<td>7km</td>
<td>78</td>
<td>50km</td>
<td>10,485,762</td>
<td>7TB</td>
</tr>
</tbody>
</table>

✓ = highest resolution of that model and all 40 days transferred to DKRZ
Zoom on tropical storm 'Howard' after 36h forecast time

ICON 5 km  ICON 2.5 km  NICAM 7 km  SAM 4 km  FV3 3.25 km

NASA worldview

Matthias Brueck (MPI)
Finding, analysing and solving MPI communication bottlenecks in Earth System models

Journal of Computational Science

Oriol Tintó Prims, Miguel Castrillo, Mario C. Acosta, Oriol Mula-Valls, Alicia Sanchez Lorente, Kim Serradell, Ana Cortés, Francisco J.Doblas-Reyes

Highlights

• There was not a well established method to find MPI issues in Earth Science models.
• A performance analysis of ESMs can lead to productive and efficient optimizations.
• State-of-the-art ESMs still use algorithms not suitable for massive-parallel systems.
JDMA: a prototype tape library for advanced tape subsystems

- JASMIN Data Migration App(lication)
- A multi-tiered storage library
  - Provides a single API to users to move data to and from different storage systems
  - HTTP API running on webserver, database records requests and file metadata
  - Command line client which interfaces to HTTP API
- Multiple storage “backends” supported:
  - Amazon S3 (Simple Storage Solution) for Object Stores and AWS
  - FTP, also for tape systems with a FTP interface
  - Elastic Tape – a proprietary tape system based on CASTOR
- Backends have a “plug-in” architecture:
  - Extra backends can be added by writing the plug-in
- A number of daemons (scheduled processes) carry out the data transfer
  - Asynchronously
  - On behalf of the user
JDMA: a prototype tape library for advanced tape subsystems

Figure 2: System structure of the JDMA showing the client, HTTP API residing on the webserver, the central database containing information about the migrations and requests, the worker processes which carry out the migrations and the storage backends.
Exchange and collaborate: ENES HPC workshops

The Series of ENES HPC-workshops is co-organized by ESIWACE and IS-ENES with the ENES HPC task force.
ESiWACE2 plans
weather & climate computing as driver for HPC

| 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |

**FETHPC:**
1. Weather and climate benchmarks, and IO (HPCW)
2. Demonstration of novel programming models (DSL)
3. Data aware numerical methods

→ Feasibility of new concepts, computability

**EINFRA:**
1. Full-sized weather and climate models for EuroHPC
2. Community testing of novel programming models (DSL)
3. Data handling workflows, data analytics research

→ Adaptation of leading models to (pre-)exascale
→ Strategy for achieving full-sized requirements

**FETFLAG:**
1. Full-sized applications with required speed/volume and power footprint
2. Ingestion of downstream applications, all ensembles
3. Domain-specific, distributed computing capability, interactive workflows

→ Redesign entire prediction philosophy

**EuroHPC:**

- EuroHPC-1&2 pre-exascale
- EuroHPC-3&4 exascale

Joachim Biercamp, DKRZ - ESIWACE
Reading, 25 Sept 2018
1. **Enable leading European weather and climate models to leverage the available performance of pre-exascale systems with regard to both compute and data capacity in 2021.**

2. **Prepare the weather and climate community to be able to make use of exascale systems when they become available.**
ESiWACE2 work packages

WP1    Cutting Edge Resolution in Earth system modelling
WP2    Establish and watch new technologies for the community
WP3    HPC services to prepare the weather and climate community for the pre-exascale
WP4    Data Handling at Scale
WP5    Data post-processing, analytics and visualisation
WP6    Community engagement and Training
WP7    Coordination, Management and Dissemination
WP1: Cutting Edge Resolution in Earth system modelling

WP1 will develop coupled weather and climate models in unprecedented technical quality and performance as well as the organisational framework to assess their scientific performance.

Lead: Peter Bauer ECMWF; Joachim Biercamp, DKRZ

• Extend the ESiWACE demonstrator approach to production type configurations: For fixed SYPD (=1) push resolution as high as technically feasible. Tentative goal:
  – EC-Earth: 16 km (TL1279) atmosphere coupled to a 1/12 degree (~8 km) ocean
  – ECMWF: 5 km (TCo1999) atmosphere coupled to a ¼ degree (25 km) ocean
  – ICON-ESM: 5 km atmosphere coupled to a 5 km ocean, aiming at higher resolutions for the ocean
  – The IPSL model: 10 km atmosphere coupled to a 1/12 degree (~8 km) ocean

• Extend the DYAMOND idea and provide the necessary infrastructure
WP2 will establish, evaluate and watch new technologies to prepare climate and weather simulation for the exascale era.

**Lead:** Rupert Ford, STFC; Carlos Osuna, MeteoSwiss

- Establish DSLs in the community
- Evaluate Concurrent Components to improve performance
- Evaluate Containers to port Earth system models to new hardware
- Watch emerging technologies
WP3 will develop and provide services to improve performance and portability of climate codes with respect to existing and upcoming tier1 and tier0 computers.

Lead: Ben van Werhoven, NLeSC; Erwan Raffin, Bull/ATOS

- Open call for service requests to organise support for existing Earth system models that target the European pre-exascale systems planned for 2021.
  - Model portability and refactoring
  - Coupling, IO and workflows

- Weather and climate benchmarking
  - “HPCW” (V1.0 developed by ESCAPE-2)
The HPCW – Benchmark

ESCAPE
Identify Weather and Climate Dwarfs (motifs)

ESCAPE-2
Extend the dwarf concept, include DSL

ESiWACE
Demonstrate very high scalability ($O \ 10^6 - 10^7$)

ESiWACE2
support (pre)-exascale production workflows

Roadmap to cloud resolving models

The HPCW benchmark

The EuroHPC Systems Co-design Roadmap to cloud resolving models

Joachim Biercamp, DKRZ - ESiWACE
Reading, 25 Sept 2018
WP4 and WP5: Data Handling at Scale

**WP4** will provide the necessary toolchain to handle data at pre-exa-scale and exa-scale, for single simulations, and ensembles

**Lead:** Bryan Lawrence, UREAD; Julian Kunkel, UREAD

**WP5** will enhance the tools to analyse and visualise these data

**Sandro Fiore, CMCC; Niklas Röber, DKRZ**
WP6: Community engagement and Training

WP6 will link ESiWACE2 to the weather and climate community it serves on the one hand and to the European HPC ecosystem on the other hand.

Lead: Sylvie Joussaume, CNRS-IPSL; Sophie Valcke, CERFACS

- Community engagement
  - HPC Workshops
  - HPC task force
  - Interface to PRACE

- Training and Schools
  - IO and HPC awareness
  - DSL
  - C++ for HPC
  - OASIS3-MCT
  - High performance Data Analytics
  - Docker
  - Summer school in HPC for weather and climate
Thank you for your attention
ENDE