Philipp Neumann, Joachim Biercamp, PanagiotisAdamidis
Deutsches Klimarechenzentrum (DKRZ)

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.
Motivation: Cloud-Resolving Schemes

- High level of detail $\rightarrow$ 100-1000m horizontal resolution
- Less parameterisation $\rightarrow$ In the limit, we know the equations!
- Challenges: very compute/memory/data intensive!

Courtesy by Leonhard Scheck (LMU Munich), Bastian Kern (DLR) P. Adamidis (DKRZ); BMBF project HD(CP)2
Overview

1. Motivation: Cloud Resolving Schemes
2. ESiWACE
   1. Overview and Goals
   2. High-resolution Demonstrators
   3. Scalability, Usability, Exploitability
3. Summary
ESiWACE: Overview

- ESiWACE = Centre of Excellence in Simulation of Weather and Climate in Europe
- Funded by H2020, e-Infrastructures „Centres of Excellence for computing applications“
- Duration: Sep 2015 – Aug 2019
- ESiWACE leverages two European networks:
  - European Network for Earth System Modelling (ENES)
  - European Centre for Medium-Range Weather Forecasts (ECMWF)
ESiWACE: Goals and Structure

WP1 Governance and engagement
WP2 Scalability
   Global high resolution model demonstrators
WP3 Usability
WP4 Exploitability
WP5 Management and dissemination

ESiWACE substantially improves efficiency and productivity of numerical weather and climate simulation on high-performance computing platforms by supporting the end-to-end workflow of global Earth system modelling.
ESiWACE: Goals and Structure

WP1 Governance and engagement
WP2 Scalability
Global high resolution model demonstrators
WP3 Usability
WP4 Exploitability
WP5 Management and dissemination

ESiWACE substantially improves efficiency and productivity of numerical weather and climate simulation on high-performance computing platforms by supporting the end-to-end workflow of global Earth system modelling.
ESiWACE: Goals and Structure

WP1 Governance and engagement
WP2 Scalability
Global high resolution model demonstrators
WP3 Usability
WP4 Exploitability
WP5 Management and dissemination

ESiWACE substantially improves efficiency and productivity of numerical weather and climate simulation on high-performance computing platforms by supporting the end-to-end workflow of global Earth system modelling.
ESiWACE: Goals and Structure

- Establishment of list of all European institutions supporting ESiWACE, including contact points and key topics of interest
- Collection of purpose, features, governance and further aspects of each software package that is of common interest to the climate and weather modelling community
- Interactions with ETP4HPC, PRACE and other CoEs/projects/activities: EXDCI, EPIGRAM, Montblanc, PoP, ESCAPE, IS-ENES2, HD(CP)2, NEMO consortium, EC-EARTH consortium,...
- Surveys on community software and new ESiWACE developments
- ...

WP1 Governance and engagement
WP2 Scalability
WP3 Usability
WP4 Exploitability
WP5 Management and dissemination

ESiWACE substantially improves efficiency and productivity of numerical weather and climate simulation on high-performance computing platforms by supporting the end-to-end workflow of global Earth system modelling.
Long-Term Strategy and Sustainability – What if ...

Time line

ESiWACE

Related projects (ESCAPE, euroEXA, HD(CP)² ....)

Extreme Scale Demonstrators CoEs 2nd round

ETP 4 HPC

EPECC prototype?
**ESiWACE: The Demonstrator Approach**

- **Global high-resolution simulations to demonstrate the computability** of weather/climate predictions
  - 1km resolution atmosphere-only, 10km ocean-atmosphere
  - no need to parametrize gravity waves, precipitating convection, ocean eddies
  - higher fidelity of high-impact regional events
- Implementation and operation of required infrastructures
- Long-term vision: extreme-scale robust high-resolution simulations in 50 member ensemble at 100-1000 forecast days per day
- Codes: IFS, NEMO, ICON, EC-EARTH, MPI-ESM2
High-resolution Predictions with ICON

- ICON: Icosahedral Non-hydrostatic
- Global, unstructured grid: created via successive refinement of icosahedron

Courtesy by Michael Böttinger
High-resolution Predictions with ICON

- ICON: Icosahedral Non-hydrostatic
- Global, unstructured grid: created via successive refinement of icosahedron

Courtesy by Michael Böttinger

5km
High-resolution Predictions with ICON

- ICON: Icosahedral Non-hydrostatic
- Global, unstructured grid: created via successive refinement of icosahedron

2.5km

Courtesy by Michael Böttinger
High-resolution Predictions with ICON

- ICON: Icosahedral Non-hydrostatic
- Global, unstructured grid: created via successive refinement of icosahedron

Courtesy by Michael Böttinger

1.2km
High-resolution Predictions with ICON

- ICON: Icosahedral Non-hydrostatic
- Global, unstructured grid: created via successive refinement of icosahedron

Courtesy by Michael Böttinger

1.2km
High-resolution Predictions with ICON

- ICON: Icosahedral Non-hydrostatic
- Global, unstructured grid: created via successive refinement of icosahedron

1.2km

Courtesy by Michael Böttinger
Challenge: Scalability

1D-Toy problem „Burger’s equation“
- Domain size: \( N = 2,097,152 \)
- Platform: Mistral@DKRZ, 1 node=2x18 Broadwell cores (E5-2695v4)
- Impl: OpenMP/MPI with communication hiding
- Performance: ca. 80% (single-node) mem-bw

3D complex weather/climate code
- Domain size: \( N = O(10^{11}) \) and increasing at decreasing time step size
- Platform: Mistral@DKRZ, Exascale system vs. long-term software
- Impl: anything that performs and integrates well
- Performance: (less than) 5-10% peak

P. Neumann et al: ESiWACE
ESiWACE: Co-Design

- Trends in hardware evolution ≠ Application evolution → Performance gap
- Co-design try to bridge this gap to improve application performance
ESiWACE: Co-Design

• Bull/Atos:
  – Insight into current/future hardware and how to extract optimal performance
  – Understanding performance impacts in ESM applications and workflows

• Allinea:
  – Exploitation and usage of tools (debuggers, profilers) in production environment
  – Understanding the requirements of ESM such as performance reproducibility
  – Obtaining insight into ESM workflows

• Seagate:
  – Strategic value in helping to understand the specific storage and data needs of ESM community
  – Bolstering the overall landscape of object storage based solutions for HPC
  – Expanding on feature sets and pave the path towards exascale

• ... and beyond: Hardware vendor workshops (jointly organized with IS-ENES2)
ESiWACE: Co-Design

- **Bull/Atos:**
  - Insight into current/future hardware and how to extract optimal performance
  - Understanding performance impacts in ESM applications and workflows

- **Allinea:**
  - Exploitation and usage of tools (debuggers, profilers) in production environment
  - Understanding the requirements of ESM such as performance reproducibility
  - Obtaining insight into ESM workflows

- **Seagate:**
  - Strategic value in helping to understand the specific storage and data needs of ESM community
  - Bolstering the overall landscape of object storage based solutions for HPC
  - Expanding on feature sets and pave the path towards exascale

- **... and beyond:** Hardware vendor workshops (jointly organized with IS-ENES2)

---

**Co-Design:** Bridging the gap in a CoE, will be a chasm for extreme-scale demonstrators!
Excerpt refactoring list, HD(CP)^2

- compute decomposition (fixed by using distributed algorithm)
- compute local halo information (fixed by rewriting algorithm)
- generate local grid partition (fixed by using distributed data structures; based on shared mem.)
- store decomposition information (fixed by rewriting data structures)
- store gather communication pattern (fixed by using two-phase gather algorithm)

Scalability: Local High-Resolution

Experiment hdcp2_lam_240m on JUQUEEN

- old (r12357)
- refactored (hdcp2-20140929, r20180)
- refactored (hdcp2-refactor-20150105, r20510)
- optimal (hdcp2-refactor-20150105, r20510)
Scalability: Local High-Resolution

- Germany at 120m resolution
- NWP-LEM + 42 270 720 horizontal cells, 160 levels
- Parallel efficiency of 71-80% at 131 072 cores (JUQUEEN), corresponding to
- $322.5 \times 160 = 51600$ process-local (volumetric) cells
Scalability: Local High-Resolution

- IFS: Hydrostatic; ICON: Non-hydrostatic
- Desire for exascale: If I had 10M cores,...
  - ...I could solve 1.25km global simulations at 440 days/day
  - ...if we can retain scalability with 2080 local cells (33 horizontal cells)
  - ...I’d have trouble with big data: 20 TB/forecast day or 8800TB/compute day
Scalability: Local High-Resolution

- IFS: Hydrostatic; ICON: Non-hydrostatic
- Desire for exascale: If I had 10M cores,...
  - ...I could solve 1.25km global simulations at 440 days/day
  - ...if we can retain scalability with 2080 local cells (33 horizontal cells)
  - ...I’d have trouble with big data: 20 TB/forecast day or 8800TB/compute day
- Good news:
  - IFS: 40% efficiency gains by double -> single precision at acceptable accuracy
  - IFS: 13% eff gains concurrent execution of radiation with other components
  - NEMO: 8% eff gains from vectorization, communication and memory access
  - NEMO: 5% eff gains from hybrid parallelization
- Still: What we rather need is a 10-50x speedup...
Usability: Supporting the Admins

Handbooks on installation of models and environments for end-to-end-workflows: Using Package Manager Spack

• Extension of package collection: cdo, grib-api, libemos, magics, ncl, cmor, uuid

• Improvement of existing packages: harfbuzz, pango, qt, libtiff, pixman, libjpeg-turbo, gmp, python, py-netcdf, environment-modules, hdf5

• Core functionality improvements

• Deployment successes:
  – Duplication of model environment within Spanish HPC network (BSC -> Altamira, Univ de Cantabria)
  – Reduction in software setup time from O(1 week) to O(2days)
Handbooks on installation of models and environments for end-to-end-workflows:
- Using Package Manager Spack
  - Extension of package collection: cdo, grib-api, libemos, magics, ncl, cmor, uuid
  - Improvement of existing packages: harfbuzz, pango, qt, libtiff, pixman, libjpeg-turbo, gmp, python, py-netcdf, environment-modules, hdf5
- Core functionality improvements
- Deployment successes:
  - Duplication of model environment within Spanish HPC network (BSC > Altamira, Univ de Cantabria)
  - Reduction in software setup time from $O(1 \text{ week})$ to $O(2 \text{ days})$

Usability: Supporting the Admins

cf. www.esiwace.eu, Section Deliverables

A Living Document on the Specification of a Standard Recommendation for an ESM System Software Stack

Kim Serradell
Sergey Kosukhin
Grenville Lister
Reinhard Budich
Usability: Workflow Support

Optimal scheduling and exploitation of parallelism within complex workflows by use of Cylc suites

- Client-server interaction enhancements
- New built-in support for parameterized tasks for cleaner workflow
- Performance and feature enhancement to the task/job management subsystem
- Performance and efficiency improvements to the suite validation, runtime and memory usage
- Significant growth in the number of automated tests

...
Exploitability: The Data Challenge

- Business model for storing and exploiting high volume data
  - Requirements capture
  - Modelling of the performance, cost, and resilience for storage architectures in data centres → considering alternative scenarios for architectures of data centres and their impact
  - Fine-grained simulation of hierarchical storage systems with a focus on tape → prototype implemented

- ESD middleware design for new storage layout
  - Goal: Provide novel data-specific layout interfaces and support various backends, including object stores
  - Collaboration with HDF group and project SAGE
Summary

• ESiWACE – Joining forces for weather and climate simulations at extreme scale
  – ISC, June 2017, Frankfurt: BoF session on cloud resolving models
  – PASC, June 2017, Lugano: Minisymposia in weather & climate tracks
• Scalability: Current models suggest $O(1-10)$ SYPD at extreme scale (not counting in I/O...)
  → ICON: Scalability for local and global high-resolution simulations
  → IFS: see talk on ESCAPE in next session
• Usability:
  – Simplify system and application setup using Spack
  – Cylc support and development for enhanced workflow solutions
• Exploitability: ESD middleware design and business models for storing/handling data
• Contacts: esiwace@lists.dkrz.de, www.esiwace.eu

Acknowledgement:
ESiWACE has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 675191. Parts of the work were funded by the Federal Ministry of Education and Research in Germany through the research programme High Definition Clouds and Precipitation for Advancing Climate Prediction (HD(CP)2). The authors gratefully acknowledge the computing time granted by the German Climate Computing Centre and Jülich Supercomputing Centre.