Performance Modeling and Scalability for Global High-Resolution Weather and Climate Predictions

Philipp Neumann, Joachim Biercamp, Panagiotis Adamidis, Niklas Röber
Deutsches Klimarechenzentrum (DKRZ)

Luis Kornblueh, Matthias Brück
Max-Planck-Institut für Meteorologie

Daniel Klocke
Deutscher Wetterdienst

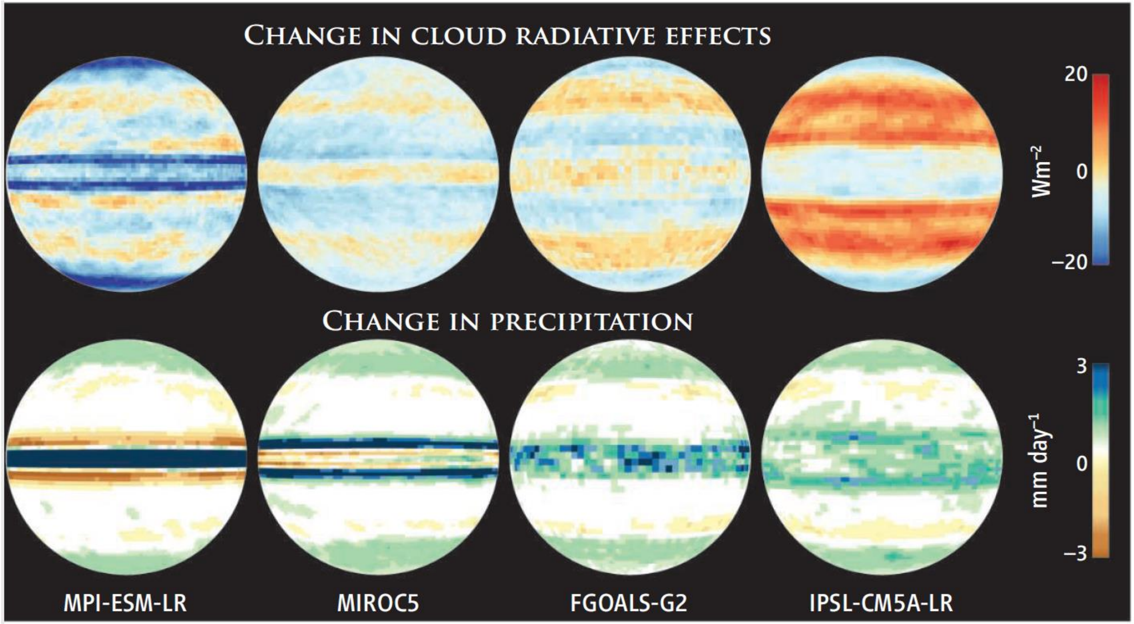
…and the members of ESiWACE and DYAMOND

ESiWACE: Centre of **E**xcellence in **Si**mulation of **W**eather **a**nd **C**limate in **E**urope

- Goals
  - Substantially improve efficiency and productivity of weather and climate models
  - Prepare models for exascale systems
    → scalability and performance analysis, tuning, …
- ESiWACE: **Kilometre-scale demonstrators (prototypical)**
  → ICON, IFS, NEMO, EC-Earth
- ESiWACE2: Towards production-ready models at pre-exascale
- Read more: Website: www.esiwace.eu
  ESiWACE newsletters: www.esiwace.eu/newsletter

ICON Aqua-Planet Experiment @1.25km

- **1.25km resolution,**
  335 544 320 horiz. cells,
  45 vert. levels
- **1408 nodes,**
  2MPI x 18 OpenMP
- **Throughput: 1.8 SDPD**,
  no IO
- Benchmark (160km-5km)
  available at:

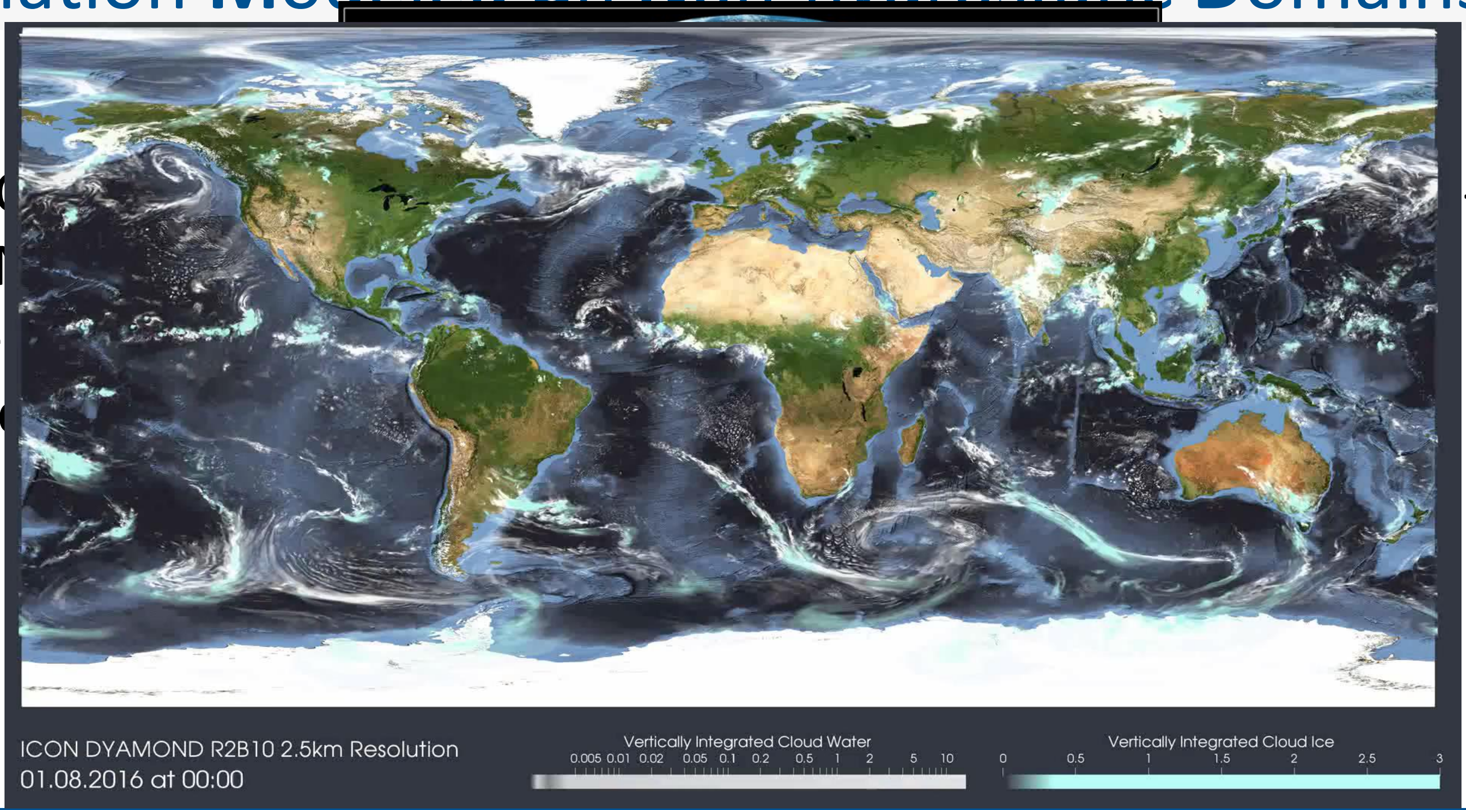


Stevens, Bony (2013). Science 340 (6136), 1053-1054

https://redmine.dkrz.de/projects/icon-benchmark/wiki/Instructions_on_download_execution_and_analysis_ICON_Benchmark_v160

DYAMOND: **Dy**namics of the **A**tmospheric General Circulation **M**odeled **o**n **N**on-hydrostatic **D**omains

- Inte
  → IC -H
- ICO
- Scie
  Rea



Outline

1. **Accessible resolutions**
2. **Semi-analytical Performance modeling for hi-res runs**
3. **Performance modeling and prediction with sparse grids**

Read more:

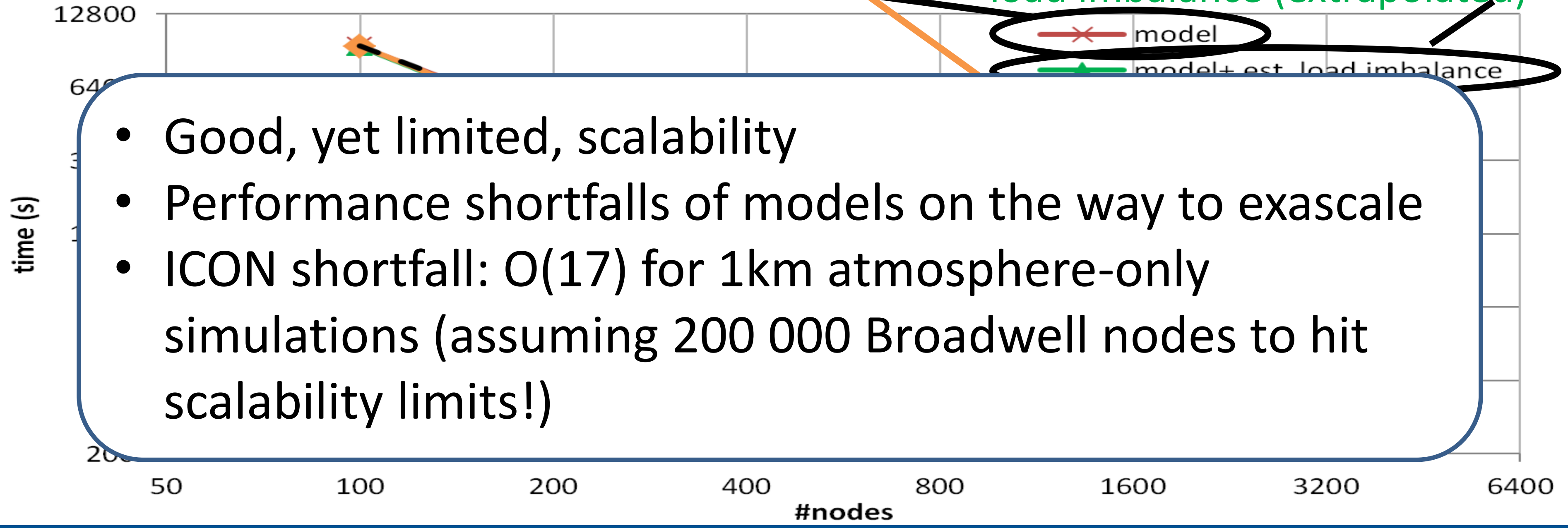
1. P. Neumann et al.
   Phil. Trans. R. Soc. A. 377:20180148, 2019
2. T. Schulthess et al.
   IEEE Computing in Science & Engineering 21(1):30-41, 2018
3. B. Stevens et al. DYAMOND (Submitted)

Performance Models for Scalability Predictions at Exascale (ICON-DYAMOND 5km)

Model 1: compute +communication

ICON-DYAMOND 5km Performance

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



- Good, yet limited, scalability
- Performance shortfalls of models on the way to exascale
- ICON shortfall: O(17) for 1km atmosphere-only simulations (assuming 200 000 Broadwell nodes to hit scalability limits!)

Performance Modeling with Sparse Grids: Objectives

- **Multi-parameter influence on computational performance**
  - → **computational:** OpenMP/MPI decomposition, loop-blocking, vector lengths, …
  - → **algorithmic:** time step, number of iterations, error control/tolerance,…
  - → **all aforementioned categories for every model subcomponent**
    - **→ high-dimensional parameter space**

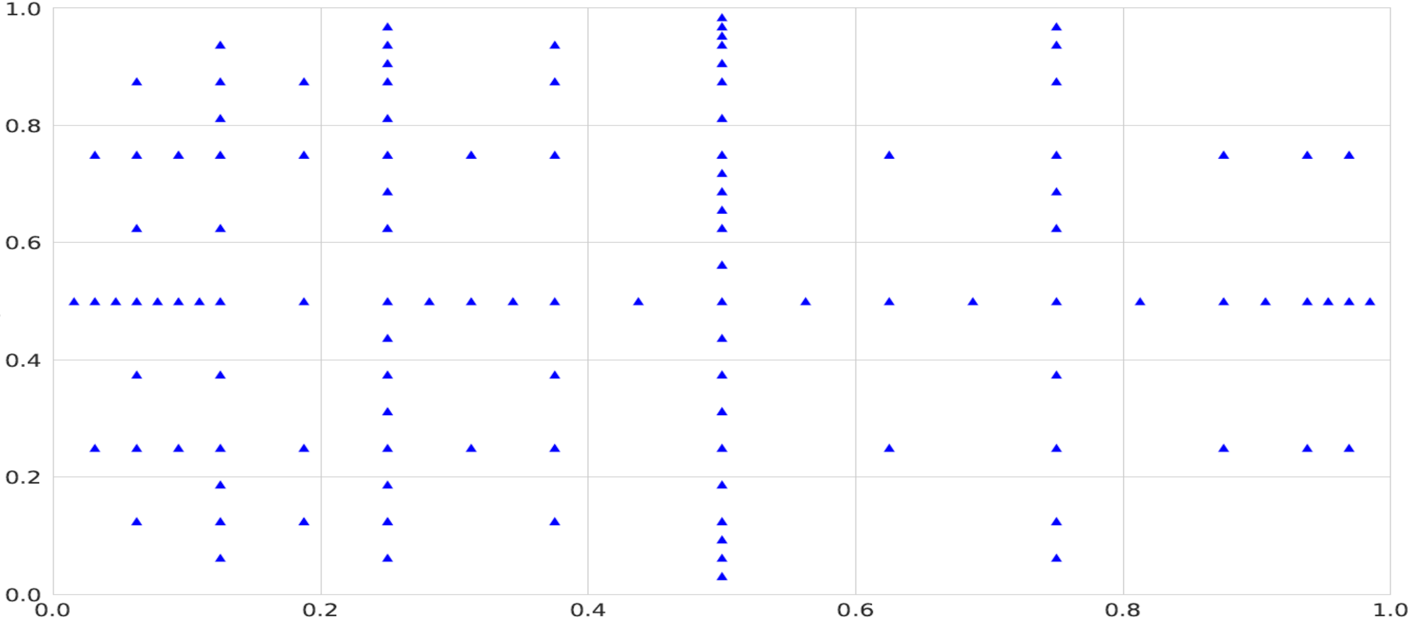
- **Objectives: performance estimate for complex ESMs…**

  …to gain insight into (wanted or unwanted) hotspots

  …to improve scheduling (relevant to workflows?)

- **Approach:** **Regression on high-dimensional parameter space via adaptive sparse grids**

Sparse Grids



Theorem 1 *For the interpolation error of a function* $f \in H^2_{0,mix}$ *in the sparse grid space* $V^s_{0,n}$ *holds*

$$||f - f^s_n||_2 = \mathcal{O}(h^2_n \log(h^{-1}_n)^{d-1}).$$

J. Garcke.
Sparse grids in a nutshell

SG: O(N(log N)$^{d-1}$) points

Full grid: O(N^{d}) points

Regression on Sparse Grids in a Nutshell

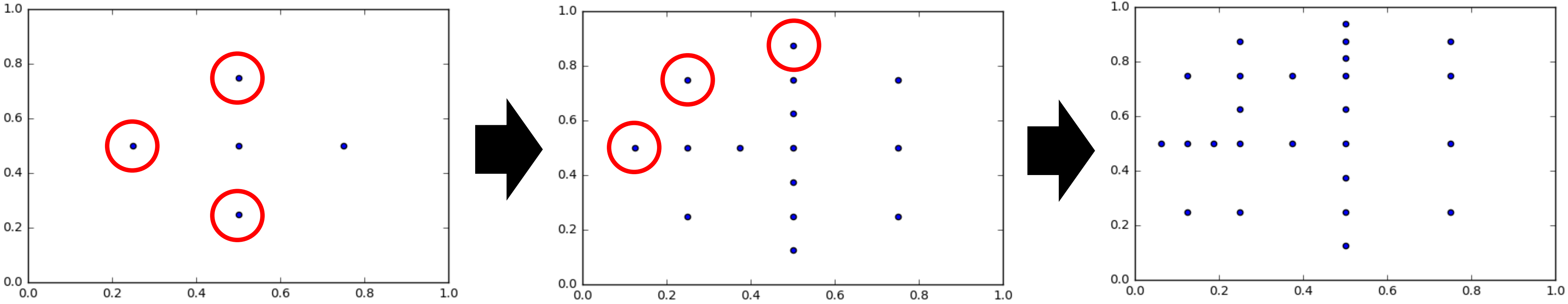
- Define linear hat function per sparse grid point
  → defines function space V_n
- Solve regression problem on run time data y_j, measured for parameter combination x_j:

$$u = \arg\min_{v \in V_n} \left(\frac{1}{M} \sum_{j=1}^{M} (y_j - v(\vec{x}_j))^2 + \lambda C(v) \right)$$

  with $v(\vec{x}) := \sum_i \alpha_i \varphi_i(\vec{x})$
- Results in linear system: $\left(\frac{1}{M} B B^\top + \lambda \mathbb{I} \right) \vec{\alpha} = \frac{1}{M} B \vec{y}$

Evaluation Procedure

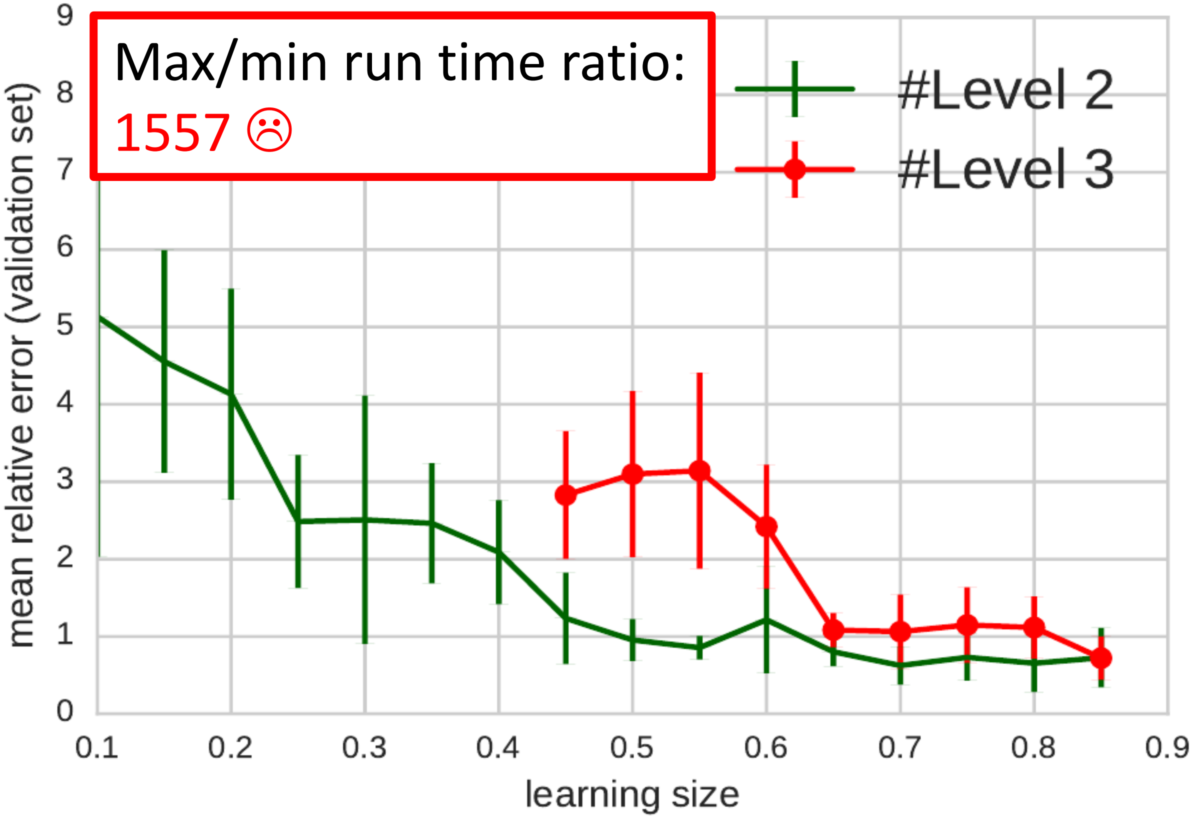
- Data splitting:
  Use s % of data for learning and 1-s % for validation
- Mean relative error:
  - Start from one data split
  - Compute and average relative errors for this data split
  - Repeat this procedure for 10 data splits and average errors
- Consider different initial sparse grid level refinements (level-2 and level-3 grids)
- Apply local refinement ($r=3$ refinement iterations, $m=3$ refinements/it.)



Example: Particle Simulation (5D Parameter Space)

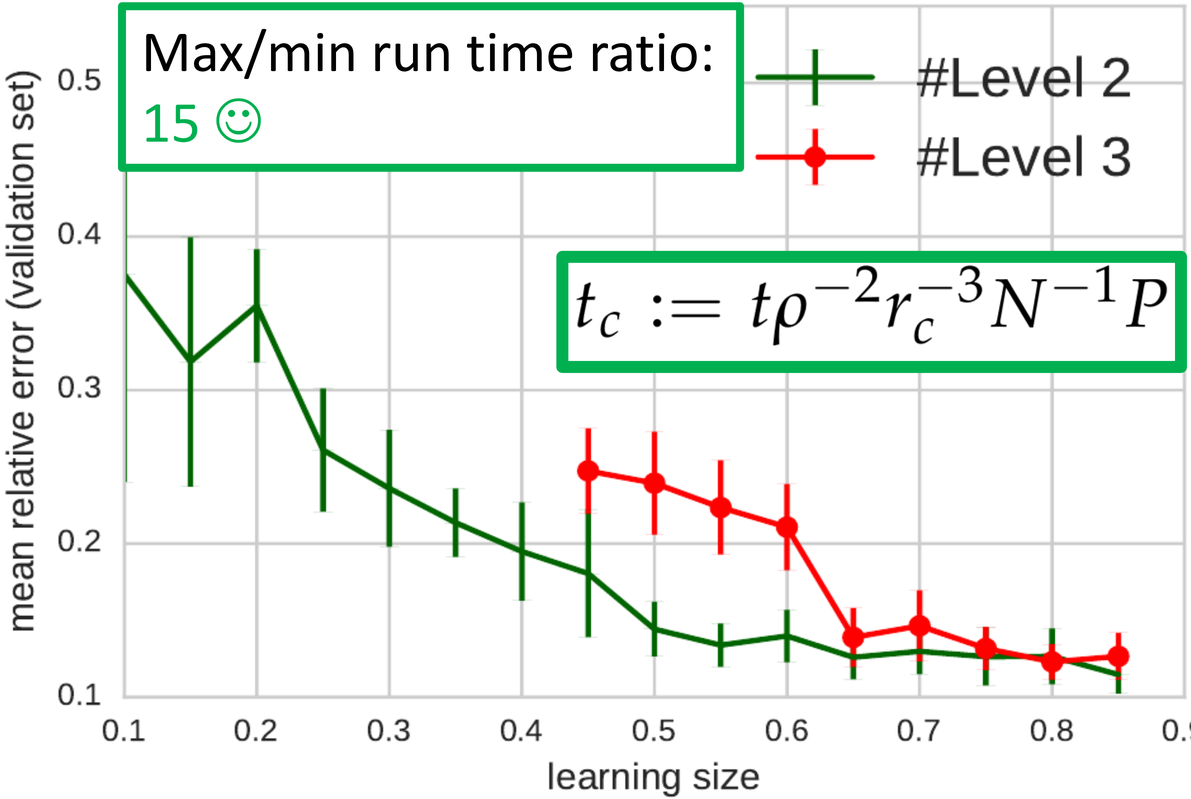
- Parameters:
  particle density ρ,
  number of particles N,
  cut-off radius r_c,
  blocksize,
  no MPI processes P
- Random sampling of run time space
  → 357 samples

Max/min run time ratio:
1557 ☹



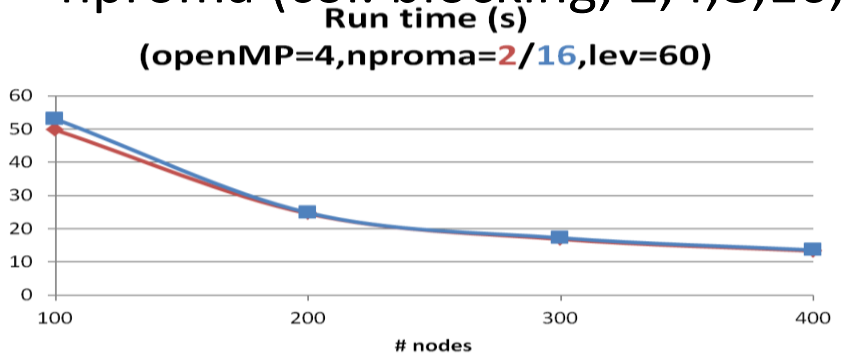
Example: Particle Simulation (5D Parameter Space)

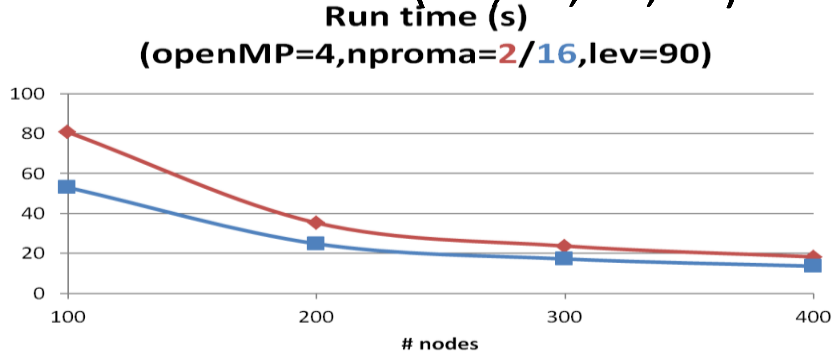
- Parameters:
  particle density ρ,
  number of particles N,
  cut-off radius r_c,
  blocksize,
  no MPI processes P
- Random sampling of run time space
  → 357 samples

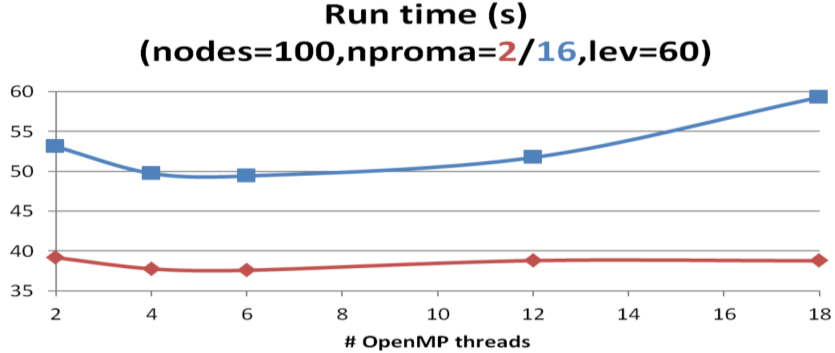


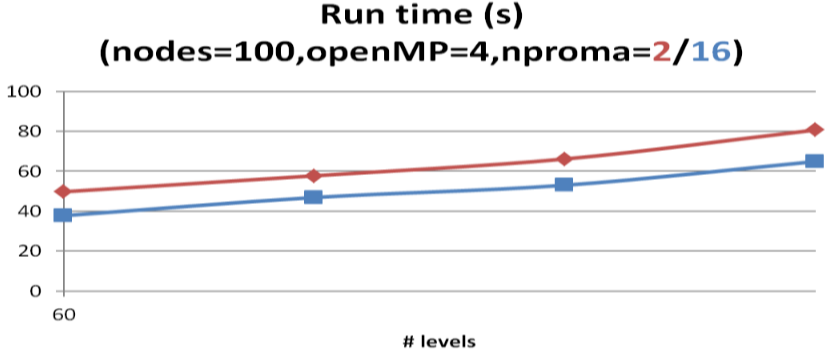
Example: ICON-DYAMOND 5km Runs (4D Space)

- Parameters:
  # OpenMP threads (2,4,6,12,18), # nodes (100,200,300,400),
  nproma (col. blocking; 2,4,8,16,32), # vertical levels (60,70,80,90)



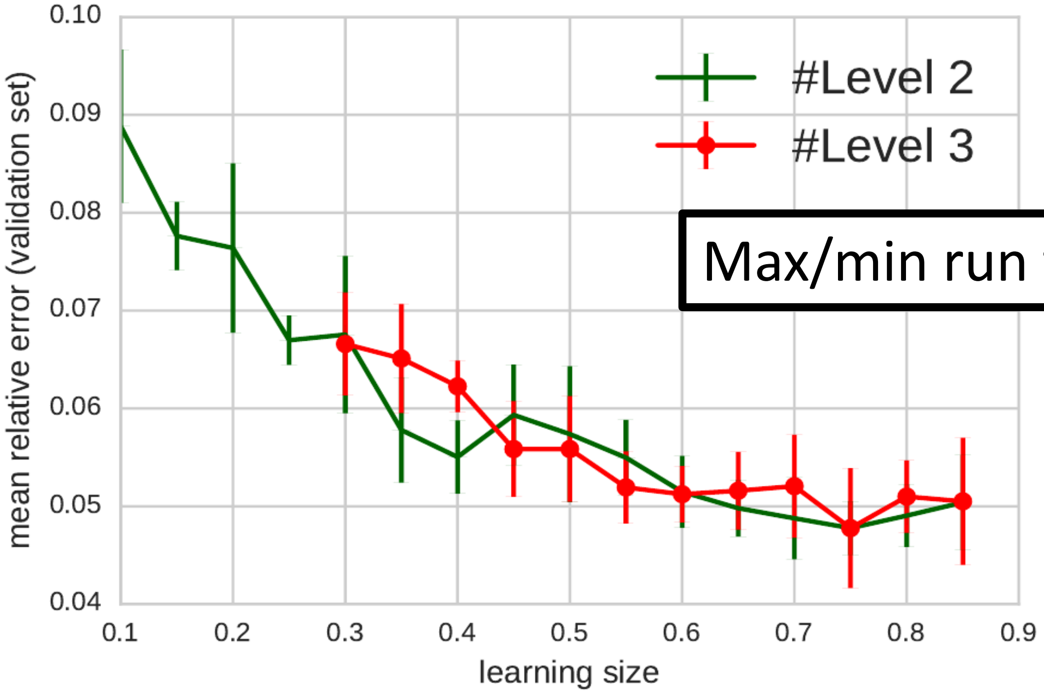






Example: ICON-DYAMOND 5km Runs (4D Space)

- Parameters:
  # OpenMP threads (2,4,6,12,18), # nodes (100,200,300,400), nproma (col. blocking; 2,4,8,16,32), # vertical levels (60,70,80,90)



Max/min run time ratio: 2

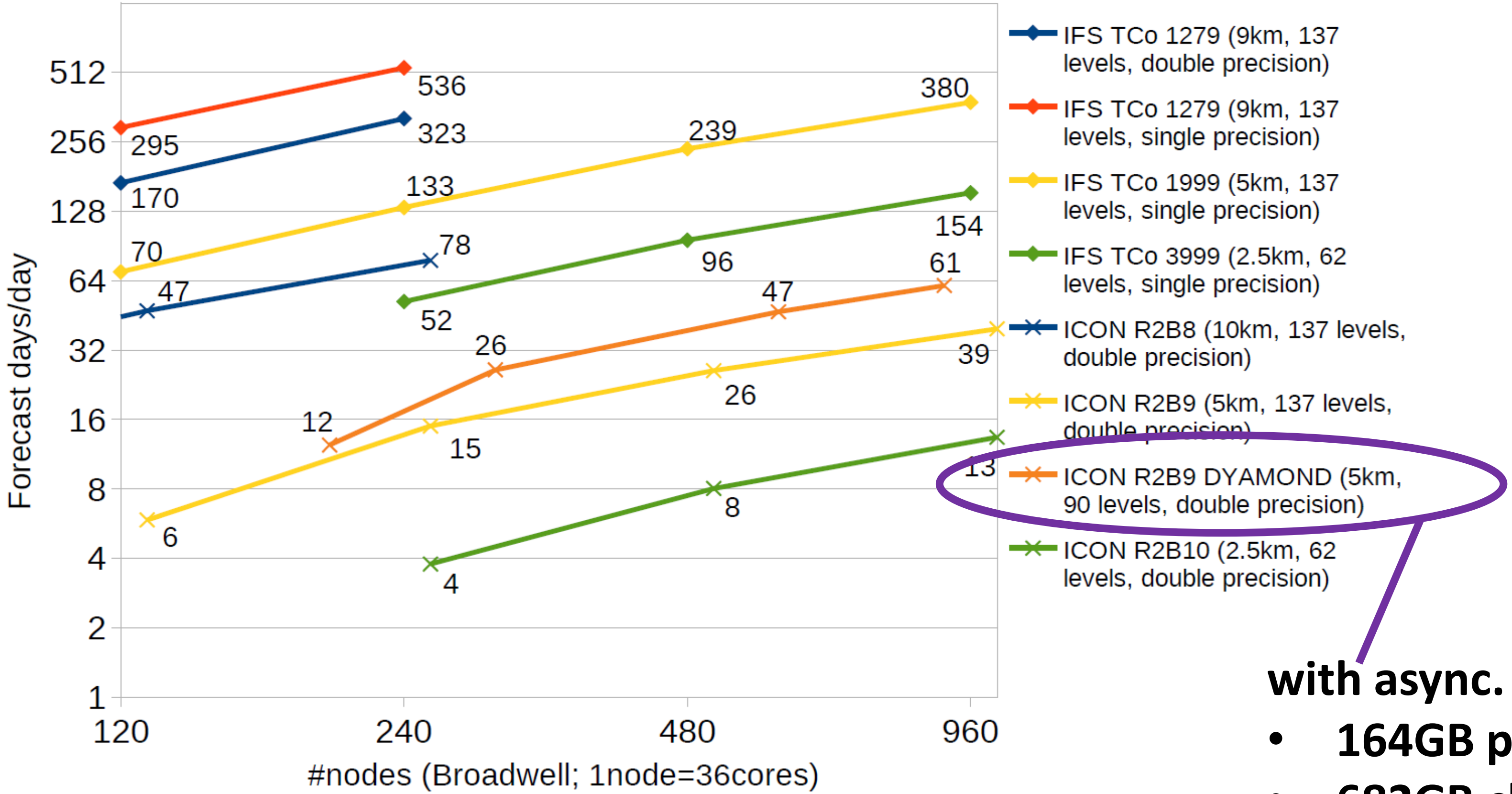
Summary

- DYAMOND/ESiWACE: Towards production-ready scalable global hi-res modeling
  - → scalability and performance
  - → scientific insight and model intercomparison
- Performance shortfall of global high-resolution m… (sub)-kilometre-scale simulations
  - → factor O(17) for ICON, similar for other mod…
  - → ***this factor is (quasi-)independent from the supercomputer's size!***
- Scalability investigation and prediction via performance modeling
  - → ***semi-analytical model for ICON-5km describes model's scaling behaviour well***
- Performance prediction for arbitrary parameters
  - → ***sparse grid regression for high-dimensional parameter spaces works well***
  - → future: comparison with neural networks, Gaussian processes

2nd DYAMOND-ESiWACE Hackathon
19-21 June 2019, Mainz/Germany
Registration before 1 May at:
esiwace.eu->Events->Hackathon…

P. Neumann acknowledges ESiWACE. ESiWACE 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 European Commission is not responsible for any use that may be made of the information it contains.

BACKUP Scalability



with async. IO:

- **164GB per sim. day**
- **682GB checkpoint**