Klemens Barfus, Ronald Queck, Ahmed Homoudi, Dánnell Quesada-Chacon, Lena Marie Müller, Luise Wanner and Matthias Mauder

Institute of Hydrology and Meteorology, Chair of Meteorology

Modelling across scales - Application of HPC to meteorological and climatological problems

ZIH Colloquium, Dresden, 02.06.2022
Forecast factory of L. F. Richardson with 64,000 human computers solving equations (Weather forecasting by Numerical Process, 1922).
Chair of Meteorology (Prof. Matthias Mauder)

Working groups:

• Land use and boundary layer processes
• Water and carbon budgets
• Regional climate and climate change
• Urban climate
• Capacity development
• 1 professor, 1 senior professor, 3 technical and administrative employees, 14 researchers / postdocs, 9 PhD students
History of HPC usage at the Chair of Meteorology

2002: DWD Lokalmodell with 7km resolution

2005: Cloud microphysics modelling and 3D radiative transfer simulations

2010: COSMO-CLM regional climate simulations for the Ukraine

2010: Analysis of GCM ensembles for the Arabian Peninsula, Ukraine and Brazil

2012: Flow disturbance around a flux measurement tower (OpenFOAM)

2014: First implementation of Weather Research and Forecast Model (WRF) on HPC

...
Characteristic scales of atmospheric phenomena
Double-ITCZ in General Circulation Models

- more than 30% of the global precipitation is produced within the Intertropical Convergence Zone (ITCZ).

- in climate models, there is a spurious double ITCZ bias.

- detection of ITCZ in model data is based on convergence and linkage to the precipitation field.

- influence of Double-ITCZ on climate projections over land?

- our work found that the double ITCZ is a typical phenomena but it is more intense and frequent in climate models.
Double-ITCZ in General Circulation Models

- Reanalysis & Climate model data (10 GCMs) analysed: ~ 6TB
- Processing one model output with the analysis algorithm takes around 24hrs.
- Preprocessing with CDO (preferred format is NetCDF)
- The algorithm is written in R, Python, and Fortran 90. The latter was introduced to optimize the computation time.

Reanalysis: model run for the past with assimilation of measurement data
Moisture sources for precipitation events on the Arabian Peninsula

Precipitation = Water Vapour + Aerosols + Cooling Mechanism

Recycling ratio of moisture [%] (Dirmeyer et al, 2009)
Moisture sources for precipitation events on the Arabian Peninsula

Backward trajectories for the flood event in Tabuk City, KSA, starting from 2021-02-05 4 UTC to 2021-01-24 4 UTC.
Moisture sources for precipitation events on the Arabian Peninsula

Resources needed to produce this example:

- Download ERA5 data: 3 Days (data are tape stored).
- Preparation of ERA5 on HPC: 16 hrs. using CDO.
- Running Lagranto model: 15 mins

Data and code for the overall project:

- Satellite data: ~ 3 TB, 22 years, dt= 30 mins, 11 km
- Reanalysis data: ~ 175 TB, 1 hr, 37 km, global extent
- Lagranto model: from ETH Zürich, written in shell, Perl, and Fortran
Downscaling of climate model output

Problem: Scale mismatch between the resolution of General Circulation Models (~100 – 200 km) and the resolution needed for climate change impact assessment (less than a few km), such as agriculture, energy, floods and ecosystems.

Statistical Downscaling:
T, Q, U, V, P, ...
(GCM, different levels, gridcell)

Transfer function (e.g. regression, NN)

Precipitation
(from station)

Dynamical Downscaling:
Nesting of regional model
Downscaling of climate model output with AI

Implementation overview:

Instead of convolutions:
Downscaling of climate model output with AI

• only example with usage of GPU
• excellent results (convolutional neural network of Bano-Medina (2020) as benchmark)
• Tensorflow, Python, R, Climate4R
• focus on repeatability
• currently under review: https://gmd.copernicus.org/preprints/gmd-2022-14/
Deep convection
DCUA - Analysis of the modification of deep convection over urban areas using radar data and mesoscale models

- radar cell tracking (20 years, dt = 5 min, dx = 1km, domain: Germany)
- input data: 3.9 TB, computation time: 500 days, output: 25 GB
- Python and Fortran
Deep convection - parametrized

Parametrization: influence of unresolved processes on model (prognostic) variables

dx = 50 km
Deep convection – resolved (Convection-permitting modelling – CPM)

dx = 1 km
DCUA - Analysis of the modification of deep convection over urban areas using radar data and mesoscale models

Simulation with Local Climate Zones for description of the urban surface

- ensemble approach needed

Simulation with urban surface replaced by natural vegetation
DCUA - Analysis of the modification of deep convection over urban areas using radar data and mesoscale models

Some numbers:

- WRF, 48 nodes, 36 hours simulation time, 5 minutes output: between 48 and 192 hours (dependent on microphysics scheme)
- output: 320 GB

WRF on GPU?

- “ESiWACE, the Centre of Excellence in Simulation of Weather and Climate in Europe – initiative to bring CPU models onto GPUs”
- interesting problem
- not clear what the performance gain will be (up to factor of 10)
- recoding needs to be done with developers of WRF code
- (existing GPU version of WRF with reduced number of parametrization schemes)
Convection-permitting modelling on climate time scales

- CPM modelling on climate time scales (> 30 years) not feasible
- Pseudo Global Warming approach
- Statistical-dynamical downscaling: Predictor (Global Model) → Logistic Regression → Event

![Convective days vs Year graph]

- fast moving cells
- slow moving cells
- 11 years average

Year: 1980, 2000, 2020, 2040, 2060, 2080, 2100

Convective days: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Convective cells in Saxony in future scenario (ECHAM6, RCP8.5)
Future of climate modelling

- ICOsahedral Non-hydrostatic atmosphere model (ICON)
- seamless modelling is possible
- simulation of 4 days in 2013 with $dx = 100$ m over Germany (Heinze et al, 2016)
Future of climate modelling

- goal: global simulations in convection-permitting scale (1 km) in 1 SYPD (simulated years per wall-clock day)

- currently (2020): 0.043 SYPD on 4888 nodes of Piz Daint (dx = 0.93 km) and 0.23 SYPD (dx = 2 km) (all experiments with COSMO model – rectangular grid)

- strategies: usage of GPUs, Domain Specific Languages, ...

- problem: huge amount of data: storage vs. recalculation with restart files (reproducability required, FAIR principles?)

Micro Scale: Simulation of surface – atmosphere interactions
Micro Scale: The exchange between surface and atmosphere is turbulent

Problem: heterogeneous surfaces disturbs homogeneous turbulence

Spectrum of turbulence - measured and simulated

![Figure from Schlegel et al. 2010](image)

LES for the ICOS measurement site “Tharandter Wald”
OpenFOAM® Simulation on HPC, Bull/ATOS Taurus
**Micro Scale: Simulation of surface – atmosphere interactions**

**Challenges**
- Turbulent exchange between surface and atmosphere
  - energy balance and mass balances, evapotranspiration, CO2, NOx, particles
- Near-surface atmospheric conditions
  - wind, heat, air pollution

**Selected Applications**
- Global Carbon Sequestration
- Urban climate

**Tools relevant for HPC**
- Large Eddy Simulation (LES)
  - OpenFOAM® >> PALM-4U

**LES for the ICOS measurement site “Tharandter Wald”**
OpenFOAM® Simulation on HPC, Bull/ATOS Taurus

Figure from Schlegel et al. 2010
Micro Scale: What method should we use?

DFG Program: Metström
Project: TurbEFA
Turbulent exchange between forest and atmosphere

Comparison of
• Measurements
• Wind tunnel
• Boundary Layer Modelling (RANS)
• Large-Eddy Simulation (LES)
at ICOS measurement site „Tharandter Wald“

TurbEFA: an interdisciplinary effort to investigate the turbulent flow across a forest clearing.

picture: Queck 2009
Micro Scale: Why LES?

- **LES** (Large Eddy Simulation)
- **TKE** (turbulent kinetic energy) in m²/s²

**Measurements, Wind Tunnel, BLM, LES**

**Vertical momentum exchange / flux**

**PAD in m²·m⁻³**
**Micro Scale: Why Large-Eddy Simulation (LES)**

**DFG Program: Metström**

**Project: TurbEFA - Turbulent exchange between surface and forest and atmosphere**

Wind tunnel not possible to capture the full range of the turbulence spectrum

**Boundary Layer Modelling** of the mean flow (RANS) underestimates the turbulent exchange in heterogeneous terrain

**Large-Eddy Simulation**

All relevant turbulent structures are resolved → best correspondence between simulation and measurements


---

**Measurements: real world data**

**Simulation of mean wind, fluxes**

**Simulation of restricted turbulence**

**Simulation of relevant turbulence**
Micro Scale: **TOOLS – Large Eddy Simulation with PALM-4U**

PALM was developed by the PALM working group (Prof. Raasch → Prof. Maronga, Uni Hannover) and several contributors.

It is a turbulence-resolving LES model specifically designed to run on massively parallel computer architectures.

**PALM-4U is currently developed and evaluated within the BMBF program “Urban Climate under Change” (http://www.uc2-program.org)**

PALM-4U is an advanced and state-of-the-art meteorological modeling system for atmospheric and oceanic boundary layer flows.

Figure: [https://palm.muk.uni-hannover.de/trac/wiki/palm4u](https://palm.muk.uni-hannover.de/trac/wiki/palm4u) (modified)
Micro Scale: BMBF Program “Urban Climate Under Change”

Simulations for major German cities

**Berlin:** Urban Heat Island
   → Aggregation of single local influences

**Stuttgart:** Air Pollution

**Hamburg:** Wind

... 

**Dresden:** Effect of vegetation

**Computation time:**

Example case Berlin 48 h real time computed on Cray-XC40 of the North-German Computing Alliance (HLRN)

8000 CPUs × 350 h ~ 3 million CPU hours
Micro Scale: BMBF Program “Urban Climate Under Change”

Simulations for major German cities

- **Berlin**: Urban Heat Island
  - Aggregation of single local influence

- **Stuttgart**: Air Pollution

- **Hamburg**: Wind

- **Dresden**: Effect of vegetation

**Computation time:**

Example case Berlin 48 h real time computed on Cray-XC40 of the North-German Computing Alliance (HLRN)

8000 CPUs \(\times\) 350 h \(\sim\) 3 million CPU hours

Includes two child domains (3 km \(\times\) 3 km, resolution 2 m)

Figure: Maronga et al. (2018) ICUC10, 5D.1
Micro Scale: BMBF Program “Urban Climate Under Change”

Simulations for major German cities

Berlin: Urban Heat Island
  → Aggregation of single local influences

Stuttgart: Air Pollution

Hamburg: Wind

...  

Dresden: Effect of vegetation

Computation time:

Example case Berlin 48 h real time computed on Cray-XC40 of the North-German Computing Alliance (HLRN)

8000 CPUs × 350 h ~ 3 million CPU hours

Particle concentration in the metropolitan area of Berlin simulated with PALM-4U

Figure: Maronga et al. (2018) ICUC10, 5D.1
Micro Scale Project Examples:
“Urban Climate Under Change”
SP TUD: “Urban vegetation and combined stress factors”

**Parent Domain:** centre of Dresden

**Child Domain:** Strehlen

**Real time:** 24 h pre-run + 24 h LES

**Boundary Conditions:** hot summer day, autochtonic conditions, light wind from West

**Used Module:** Urban Surfaces, Land Surfaces, Radiation, Vegetation, Biometeorology

**Computed on Bull/ATOS Taurus**

**Computation time:** ~ 30 000 CPUh (288 – 1200 CPUs)

**Memory usage:** 2.5 TB
Micro Scale Project Examples: Evaluation of PALM-4U, Site Dresden

Air temperature in °C, summer day 16:30 MEZ in z ~ 5 m (bias corrected)

Parent Domain, 6000 m × 6000 m × 4000 m
dx = dy = dz = 10 m

Child Domain 2, 1200 m × 1200 m
z ~ 1 m, dx = dy = dz = 2 m

Mobile Measurements Trajectorie
18 × 3,5 km in 2 days

min(ta) = 28 °C
max(ta) = 36 °C

Leaflet © OpenStreetMap contributors © CAIPO
Leaflet © OpenStreetMap contributors © CAIPO
Micro Scale Project Examples: **Evaluation of PALM-4U, Site Dresden**

![Daily course of air temperature in °C on the measurement trajectory](image)

- **Air temperature in °C**
- **UTC**

- **Measurements**
- **Simulation**
- **Maximum**
- **Minimum**
Micro Scale Project Examples: “Urban Climate Under Change”
SP TUD: “Urban vegetation and combined stress factors”

Task:
How to model vegetation structure effectively?
What resolution is the necessary?

Approach:
LES for a well investigated measurement site.
Test remote sensing techniques (Laser scans).
LES for different grid sizes using a high quality object model?

Total computation time: 100 000 CPU hours
Total memory usage: 10 TB
Effective modelling of vegetation structur

- Generation of vegetation models by terrestrial laser scanning.
- What is the necessary grid resolution?

(grey trees are private and not registered in the city catalog)

0.2 m
Effective modelling of vegetation structure

- Generation of vegetation models by terrestrial laser scanning.
- What is the necessary grid resolution?

(gray trees are private and not registered in the city catalog)

0.5 m

Figure: Anne Bienert, Chair of Photogrammetry
Effective modelling of vegetation structure
- Generation of vegetation models by terrestrial laser scanning.
- What is the necessary grid resolution?
(grey trees are private and not registered in the city catalog)
Micro Scale Project Examples:

**Exchanges processes triggered by the surface heterogeneity**

**Question:** How does the surface affect secondary circulations and what is their influence on the energy balance gap?

**Aim:** development of simplified models,

**Approach:**

- **Artificial surfaces** with randomly distributed patches applying controlled variation of parameters,
- **Idealized LES** of different surface fluxes of sensible and latent heat and CO2,
- **Relation** between surface conditions and energy balance gap
  - 4 heterogeneity scales + 1 homogeneous surface
  - 7 atmospheric conditions
  - 2 ensemble runs per combination → 70 simulations

**Computed on Cray-XC40 (HLRN)**

- Total computation time: 29 120 cpu hours
- Total memory usage: 5 TB
Micro Scale Project Examples:
Exchanges processes triggered by the surface heterogeneity

The attempt to simulate reality

CHEESEHEAD19
(Chequamegon Heterogeneous Ecosystem Energy-balance Study Enabled by a High-density Extensive Array of Detectors)

Goal? investigate influence of ecosystem heterogeneity on atmospheric transport processes in the boundary layer (energy-balance gap, secondary circulations)

What? Large number of in situ and remote sensing instruments over a 10x10 km² area in northern Wisconsin, USA

Why LES? Gaps in latent and sensible heat fluxes can be investigated separately, CO₂ flux can be investigated
Micro Scale Project Examples: Exchanges processes triggered by the surface heterogeneity

Set up:

<table>
<thead>
<tr>
<th>domain</th>
<th>lx * ly * lz (km)</th>
<th>dx * dy * dz (m)</th>
<th>nx * ny * nz</th>
<th>Number of ncpu along X,Y</th>
<th>Total ncpu</th>
<th>Time step (s)</th>
<th>grid points</th>
</tr>
</thead>
<tbody>
<tr>
<td>coarse</td>
<td>30 * 30 * 6</td>
<td>120 * 120 * 80</td>
<td>250 * 250 * 75</td>
<td>18, 16</td>
<td>288</td>
<td>0.6</td>
<td>4687500</td>
</tr>
<tr>
<td>medium</td>
<td>24 * 24 * 1</td>
<td>30 * 30 * 20</td>
<td>1000 * 1000 * 50</td>
<td>30, 21</td>
<td>630</td>
<td>0.6</td>
<td>50000000</td>
</tr>
<tr>
<td>fine</td>
<td>12 * 12 * 0.5</td>
<td>6 * 6 * 4</td>
<td>200 * 200 * 125</td>
<td>25, 50</td>
<td>1250</td>
<td>0.6</td>
<td>5000000</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2168</td>
<td></td>
<td>59687500</td>
</tr>
</tbody>
</table>

Resources for 1 simulation:
Computation time: 2168 cpu x 87 hours = 188616 cpu hours
Memory usage: 1 Tb (during simulation ~ 8 Tb)

Total resources: 8 ensemble runs of 2 time periods

Total computation time: 3 million CPU hours
Total memory usage: 16 TB
Computation on Cray-XC40 of the HLRN
Micro Scale: PALM-4U, Performance

PALM shows excellent scaling which was tested for up to 50 000 processor cores. Tests have been performed on the Cray-XC40 of the North-German Computing Alliance (HLRN) (https://palm.muk.uni-hannover.de)

Problems:

• PALM-4U is written in Fortran, pre and post processing is realized in Python and R, data input and output uses NetCDF. The MPI causes sometimes problems on HPCs.
• The intel compiler is recommended, but in our current installation on the HPC in Dresden, it only runs using the GNU compiler
• Data output on SSD is often limited
• Parallel data output on SCRATCH causes writing errors => Solution: combining the current state of all CPUs in a single restart file for saving

Figure: https://palm.muk.uni-hannover.de/trac/wiki/doc/tec/parallel
Synopsis

What is achieved

• Different codes/models have been applied to solve a range of meteorological problems
• Large data amounts had to be processed
• The number of runs or computational demand of the processes requires parallel processing on HPCs
• HPC usage on all levels (bachelor to postdoc level)

Further objectives

• PALM-4U: evaluation, module development (interaction plant-atmosphere)
• Model nesting: WRF <> PALM-4U
• establish Pseudo Global Warming framework
• making use of cloud radar and additional satellite data (CloudSat, EarthCARE)
• ICON ?
• ...

Potential collaborations with ZIH

• Optimization of the code performance
• Data management
• Visualisation of simulation results