https://deeprain-project.de

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
Förder-Kennzeichen 01 IS 18O47 A-E

DeepRain
– Improving local scale rainfall prediction through deep learning

Martin G. Schultz[1], Jessica Ahring[1], Peter Baumann[2], Rüdiger Busche[3], Rita Glowienka-Hense[5], Bang Pham Huu[2], Andreas Hense[5], Jan Keller[4], Felix Kleinert[1], Lukas Leufen[1], Johannes Leugering[3], Vlad Merticariu[2], Dimitar Misev[2], Pascal Nieters[3], Gordon Pipa[3], Jonas Rebstadt[3], Bernhard Reichert[4], Mathilde Romberg[1], Erik Schaffernicht[4], Sabine Schröder[1], Olaf Stein[1], Susanne Theis[4], Sebastian Villaroya-Fernandez[2]

1 JÜLICH Forschungszentrum, Mitglied der Helmholtz-Gemeinschaft
2 JACOBS UNIVERSITY
3 UNIVERSITÄT OSNABRÜCK
4 DWD Deutscher Wetterdienst, Wetter und Klima aus einer Hand
5 UNIVERSITÄT BONN

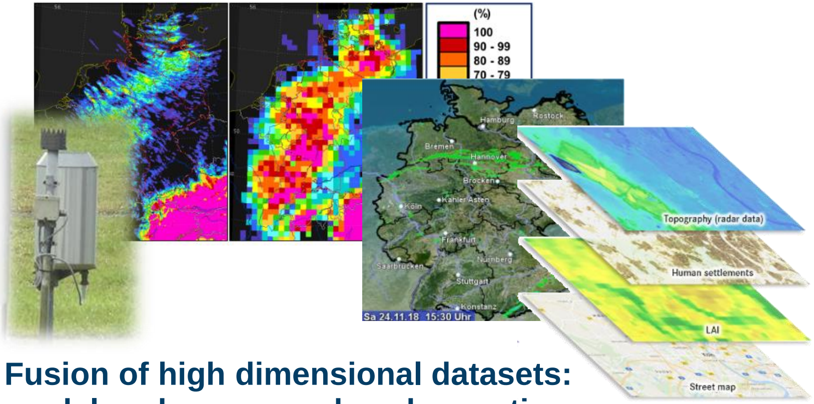
DeepRain motivation and objectives

(1) Current regional (and future global) operational regional weather forecast models (will) operate at grid sizes around 2 km permitting the **simulation of deep, heavy convection** which can have severe economic and environmental consequences

(2) Due to numerical reasons the spatial localization is accurate to about 3-5 times the grid size leaving room to exploit further downscaling and to increase the pointwise forecast skill

We want to explore to what extent deep learning methods can help to improve precipitation forecasts by including fine scale orography and radar observations which are not yet used to their full potential.

Novel data use, ML application, validation, and technology

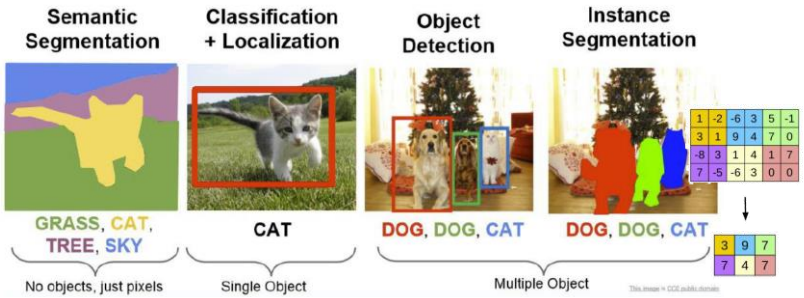


Fusion of high dimensional datasets: model, radar, orography, observation

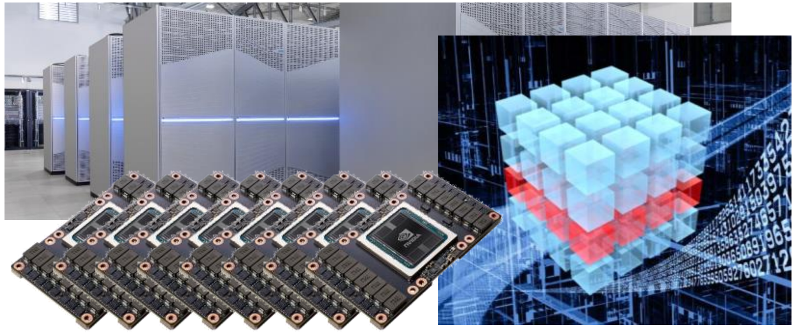
Partial correlation:

$$\rho_{12\cdot 3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}}$$

Develop validation methods, added value of ML compared to alternatives



Apply modern deep learning methods

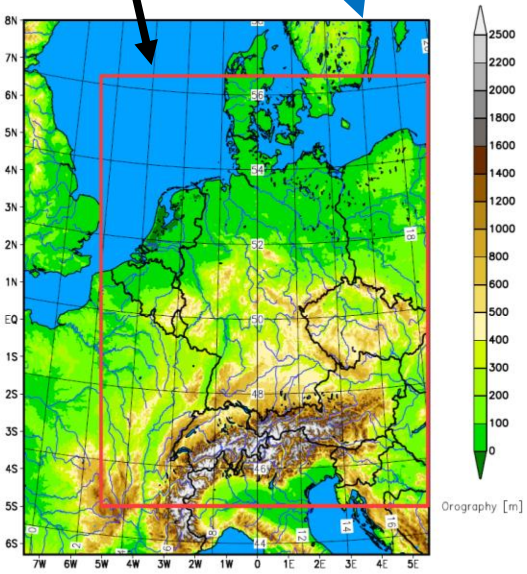


Operability: Develop high-performance workflows

„Big“ Datasets: DWD COSMO-DE-EPS data archive and radar data

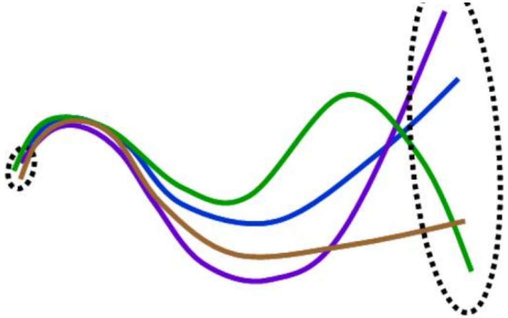
Data management established: so far ~30 TB copied from DWD to Jülich

Model Domain
COSMO-DE-EPS (COSMO-D2-EPS)



DWD NWP Data Archive
COSMO-DE-EPS (COSMO-D2-EPS)

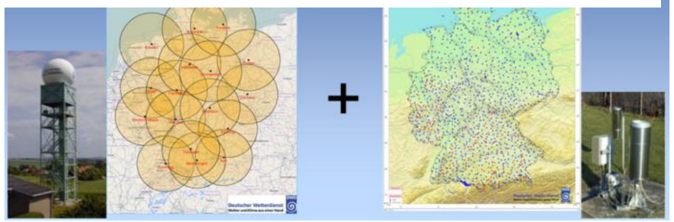
- Archive range 2011-2018 (2017-2019)
- Archive size 440 TB (200 TB) total: ~2 mio. files
- **Grid size 2.8 km (2km)**
- 8 model runs (starts) per day (00, 03, 06,… 21 UTC)
- Hourly Lead times: +00, +01, +02,… +21 (27, 45) hours
- 20 Ensemble (Monte Carlo) members



DWD RADOLAN Radar Data

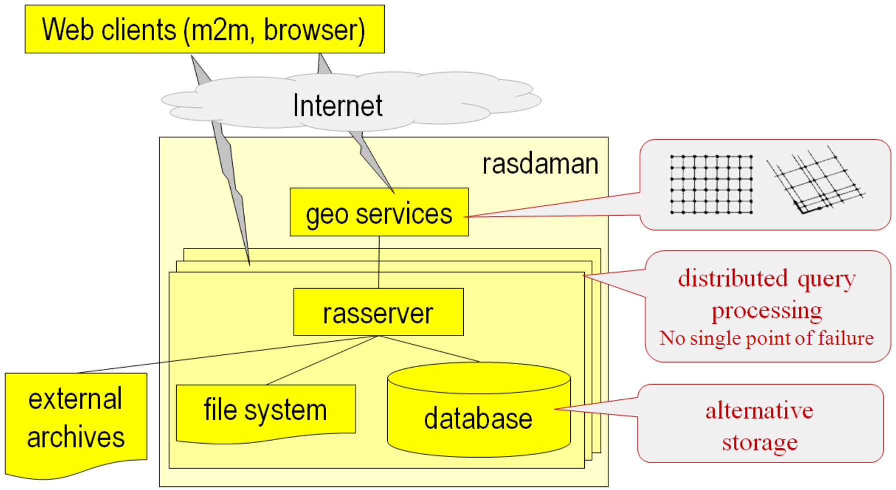
- Radar based quantitative precipitation estimation product
- Rain gauge adjusted quantitative radar composite with hourly precipitation amounts (0.1 mm resolution) and a **spatial resolution of 1 km²**
- Covers Germany in real-time (with 30 min. delay) every 10 min.

RADOLAN
DWD Radar Network + Rain Gauge Data



Data and work flow, the RASDAMAN array DB

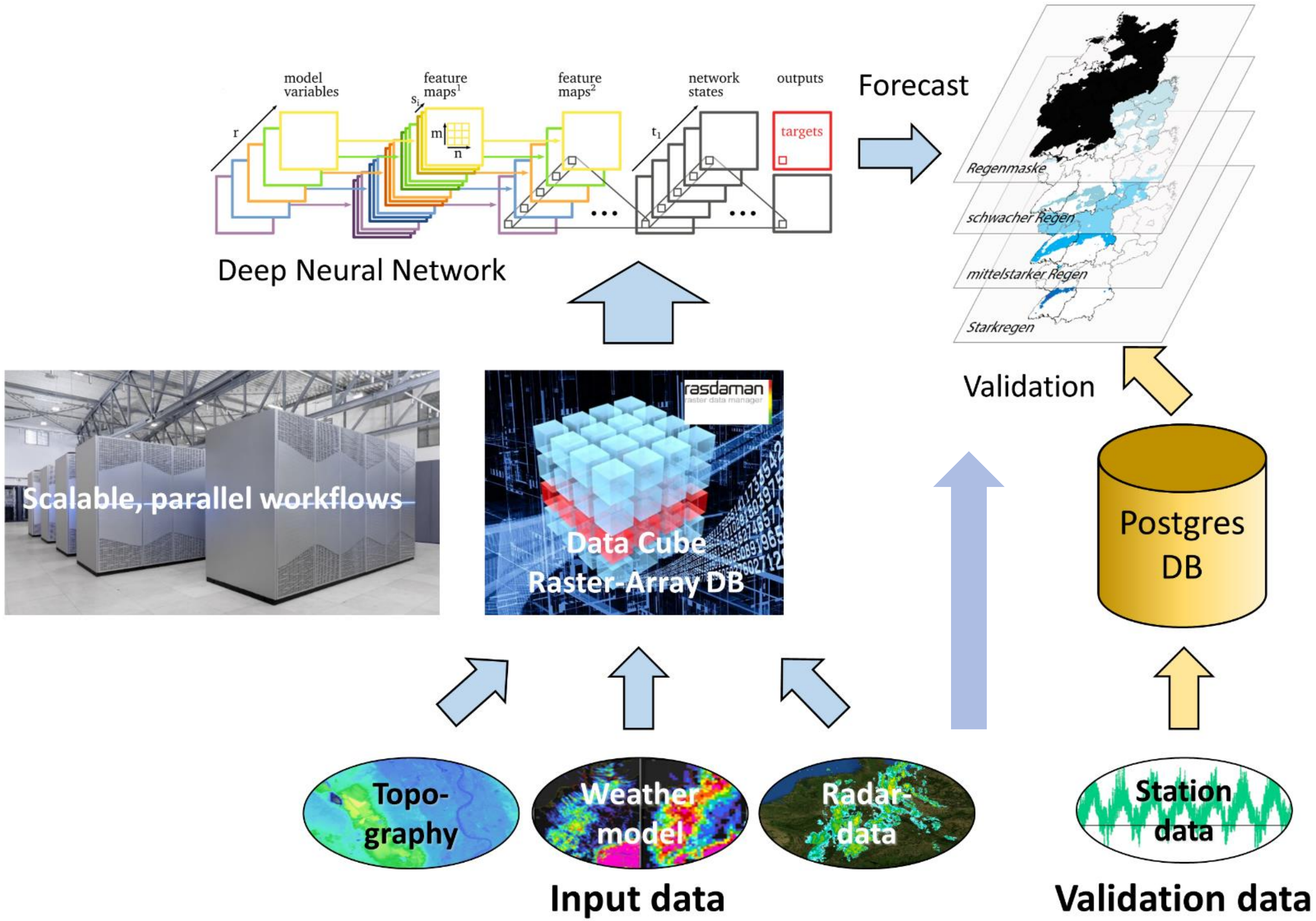
- Problem: for smooth data analysis, analysis-ready data are needed
  - **„Big Data“ part in the project**:
    (regular & irregular) x / y / z / t_{cal} / t_{lead} / MC grids
    = spatio-temporal datacubes
- Approach: Provide datacube view and functionality
  - instead of large number of files in specialized formats
- **Tool used: rasdaman array DBMS**
  - SQL-style querying on massive n-D arrays
  - Standardized as Part 15 of ISO SQL
  - Variety of clients supported through open OGC standards



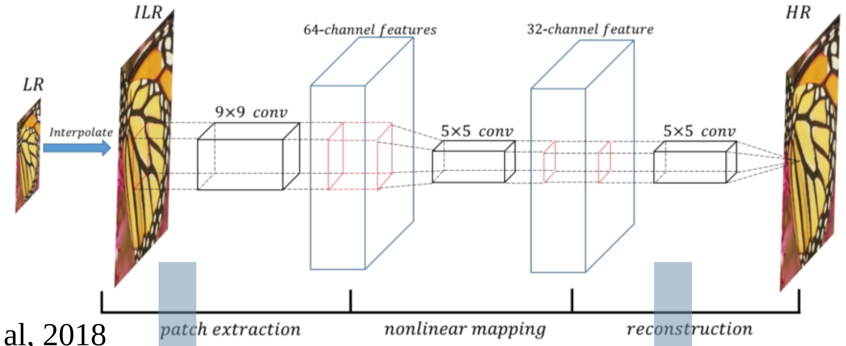
Research question: How can Array Databases make ML on Big Datacubes easier, faster, more scalable?

Data import functioning; first extraction queries ready

Data system architecture

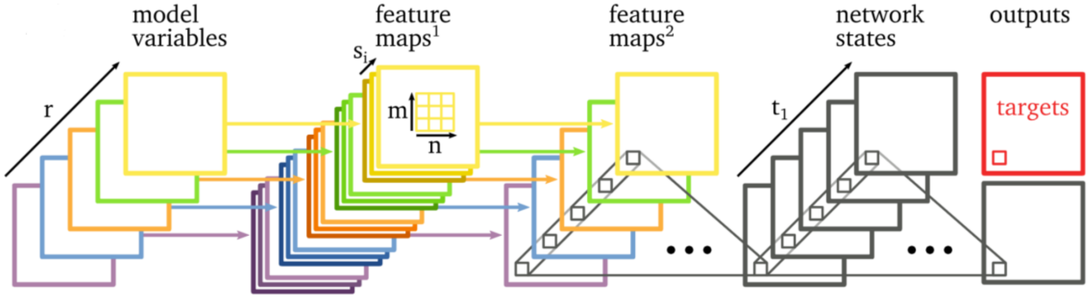


Adapt deep learning (convolutional) networks to meteorological data



Yang et al, 2018

- Super-Resolution in Deep Learning: Bad interpolation, then learn the correction.
- High dimensional RGB inputs and spatially invariant mapping

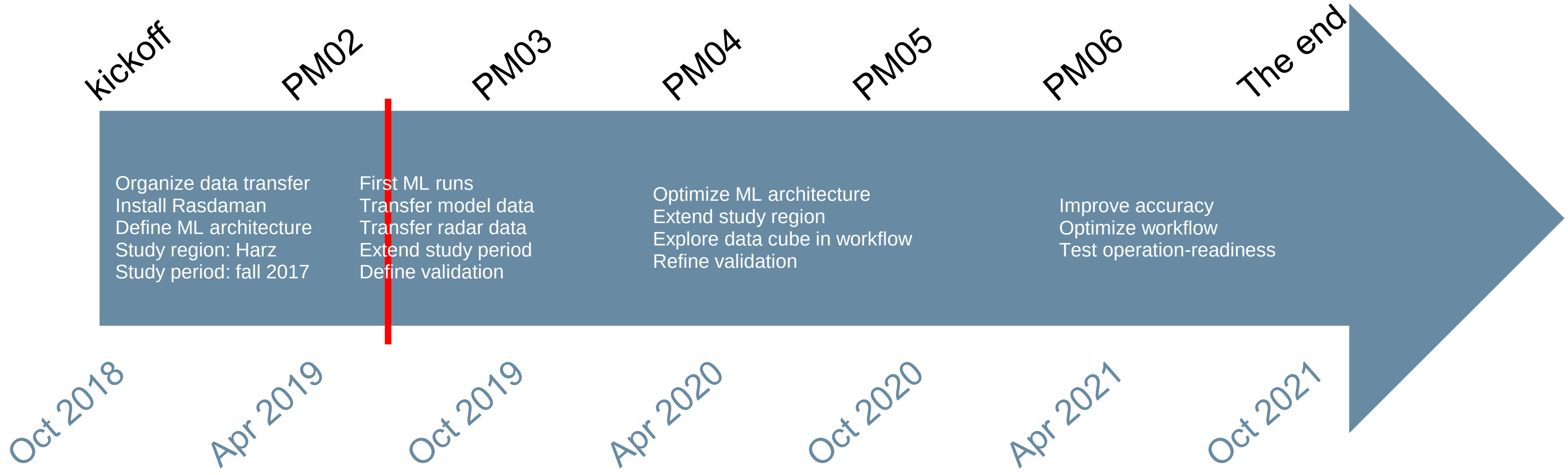


Model concept ready

For meteorological data:

- Much higher feature depth **r**
- **Input: Model | Target: Radar**
- Spatial invariance only if geographical features in the input
- High-dim. data but fewer data-points for models with big #parameters
- Constraints dictated by Physics (?)
- Evaluation vs alternative approaches
  - Specify the (ML) Added Value

Project roadmap and status



- Project is generally on schedule; slight delays due to data management issues and hiring
- Data organisation and access to storage and compute resources in place
- Highly motivated team