

Python and Machine Learning for Weather, Climate and Environment

Roland Potthast

with contributions by Stefanie Hollborn, Jan Keller, Thomas Deppisch,
Mareike Burba, Matthias Mages, Sarah Heibutzki, Marek Jacob, Florian
Prill, Tobias Göcke, Nastaran Najari



Contents Day 1 Morning

Introduction

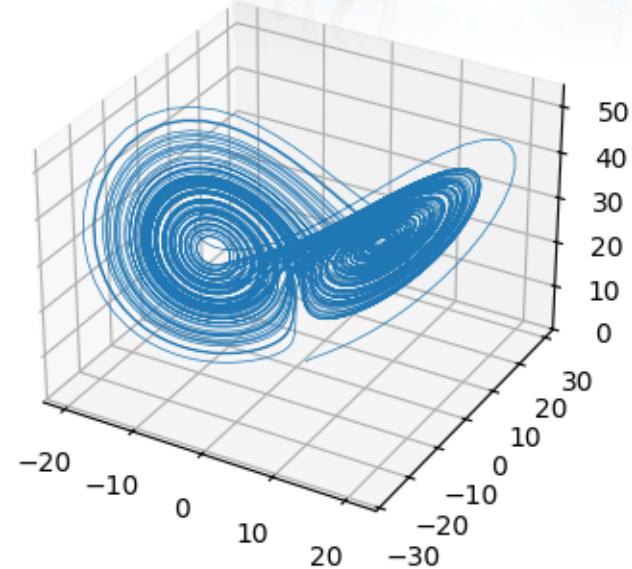
- I Programming and working with Data in the Time of AI
- II General Coding Rules and Strategy
- III 5 Days Python and AI

1 Python Basics

- 1.1 Install, Virtual Environment, Pip und Import
- 1.2 Managing Dependencies with requirements.txt
- 1.3 Introduction to NumPy
- 1.4 Generating Plots based on Matplotlib
- 1.5 Functions
- 1.6 Python Essentials

2 Jupyter Notebooks, APIs and Servers

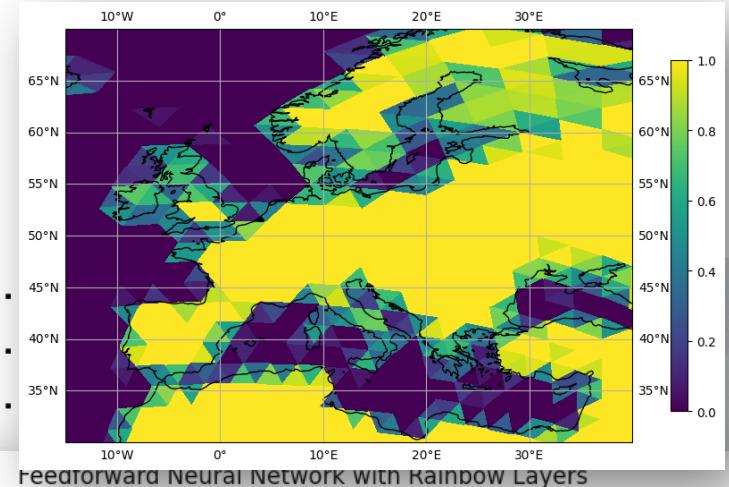
- 2.1 Introduction to Jupyter Notebooks
- 2.2 Introduction to APIs: A Key Principle in Code Development
- 2.3 Making API Requests with requests
- 2.4 Fortran Integration using ctypes as API



Contents Day 1 Afternoon

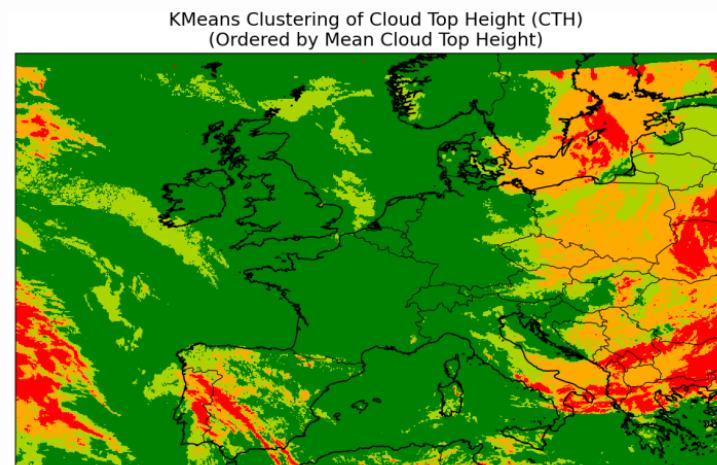
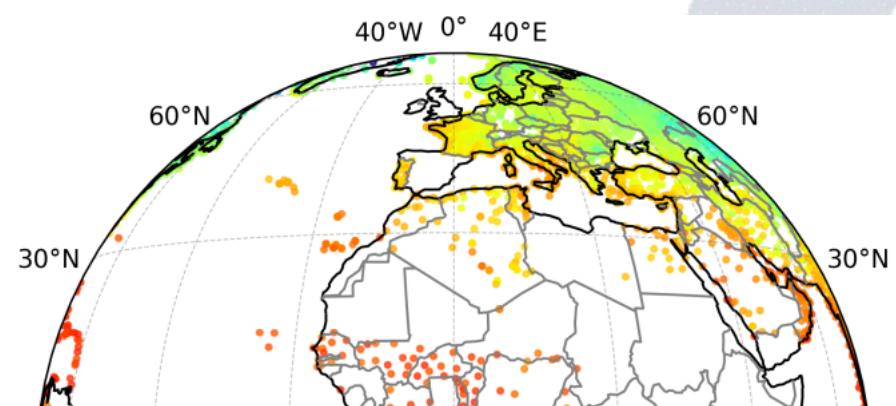
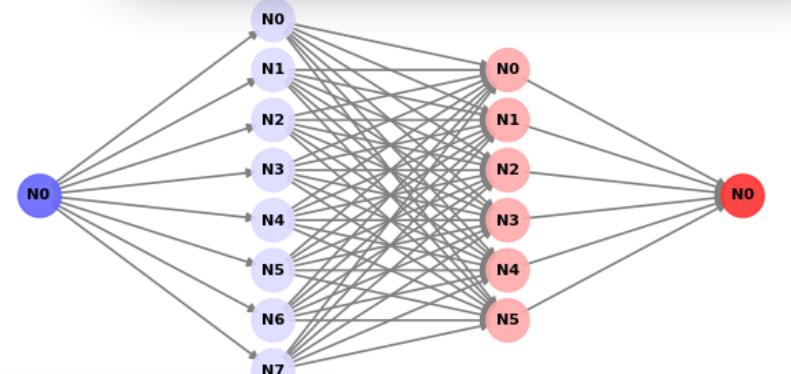
3 Eccodes for Grib, Opendata, NetCDF, Visualization

- 3.1 Downloading ICON Model GRIB Files from DWD Open Data Server
- 3.2 The Grib Library eccodes
- 3.3 Accessing SYNOP Observation Files from NetCDF
- 3.4 Analysing AIREP Feedback Files in NetCDF Format



4 Basics of Artificial Intelligence and Machine Learning (AI/ML)

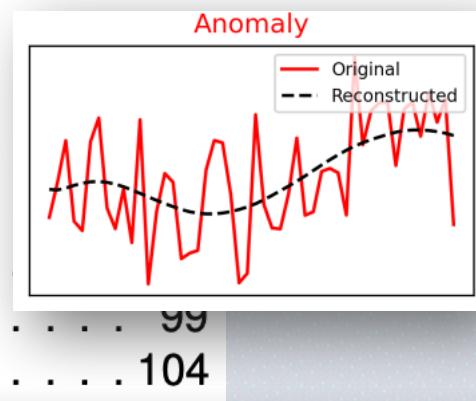
- 4.1 AI and ML - Basic Ideas
- 4.2 Torch Tensors - Basics and Their Role in Minimization
- 4.3 PyTorch Fundamentals - Model, Loss, and Optimizer
- 4.4 Simple Neural Network Training Example
- 4.5 Gradient Field and Decision Boundary



Contents Day 2 Morning

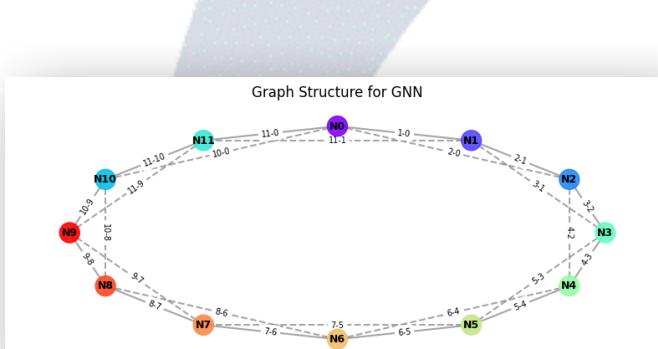
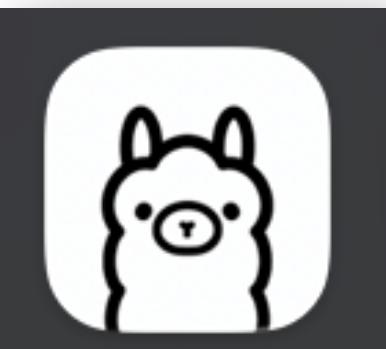
5 Neural Network Architectures

- 5.1 Feed Forward Networks
- 5.2 Graph Neural Networks
- 5.3 Applying Convolutional Neural Networks for Function Classification
- 5.4 LSTM-Based Anomaly Detection in Sensor Data



6 Large Language Models

- 6.1 LLM Network as Sequence-to-Sequence Machines via Transformer Model
- 6.2 Implementing and Training a Simple Transformer-Based LLM
- 6.3 Install Your Own LLM, Chat with it and Develop Applications



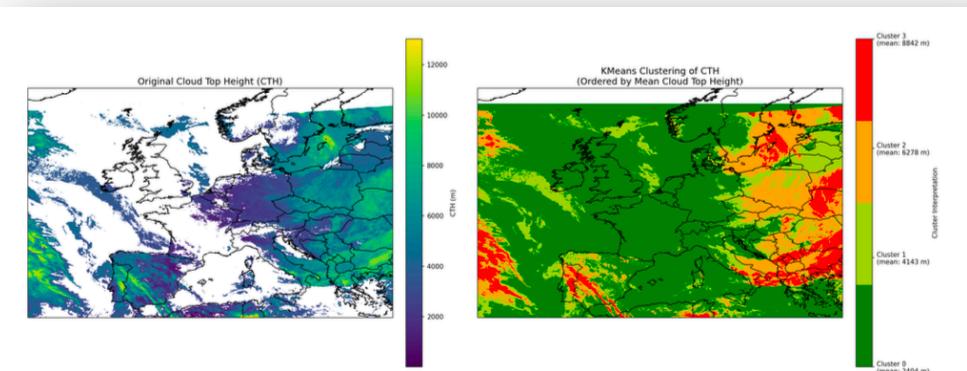
Contents Day 2 Afternoon

7 LLM with Retrieval-Augmented Generation (RAG)

- | | | |
|-----|--|-----|
| 7.1 | Preparing Documents | 140 |
| 7.2 | Generating Embeddings for Documents | 140 |
| 7.3 | Using an LLM Locally or with OpenAI for Response Generation | 140 |
| 7.4 | Saving and Reloading the Vector Database, Collecting Search Originals, Chunking long Documents | 145 |
| 7.5 | End-to-End AI-Powered Answer Pipeline | 145 |

8 Python Packages

- | | | | | | | | |
|-----|---|---|---|---|---|---|---|
| 8.1 | Review of the Python Standard Library | . | . | . | . | . | . |
| 8.2 | Xarray - Multi-dimensional labeled Data | . | . | . | . | . | . |
| 8.3 | Pandas - Data Frames and Analysis Package | . | . | . | . | . | . |
| 8.4 | SciPy Scientific Computing, Optimization and Statistics | . | . | . | . | . | . |
| 8.5 | Scikit-Learn - Machine Learning, Classification, Regression | . | . | . | . | . | . |



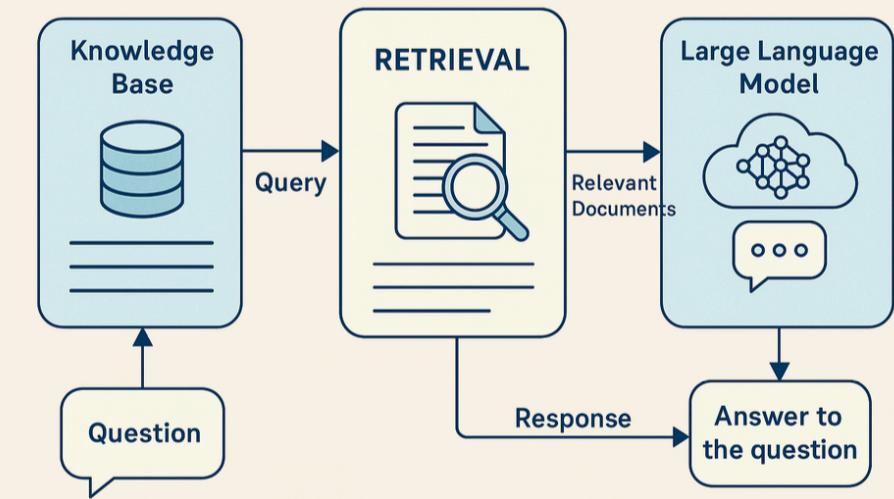
Generating Responses with OpenAI

```
1 # Step 6: Use OpenAI to discuss results
2 def chat_with_openai(query):
3     retrieved_docs = query_vector_db(query,k=20)
4     context = "\n\n".join([f"File: {file[0]}\nContent: {doc}" for doc, file in
5     retrieved_docs])
6
7     response = client.chat.completions.create(
8         model="gpt-4o-mini", # using the gpt-4o-mini model
9         messages=[
10             {"role": "system", "content": "You are a helpful assistant."},
11             {"role": "user", "content": f"Based on the following documents, answer
12             the question:\n{context}\n\nQuestion: {query}"}
13         ]
14     )
15 
```

. 140

. 145

Retrieval-Augmented Generation (RAG)



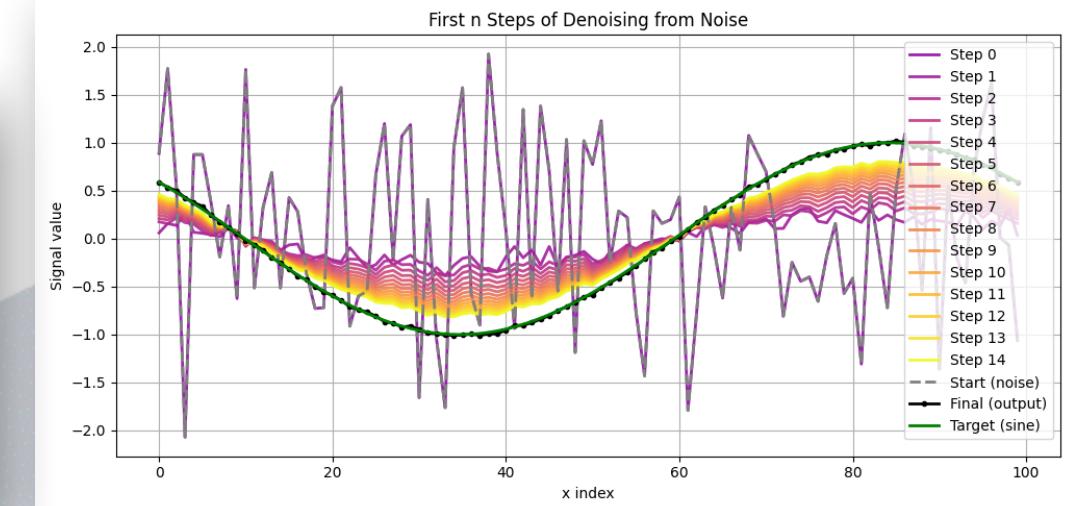
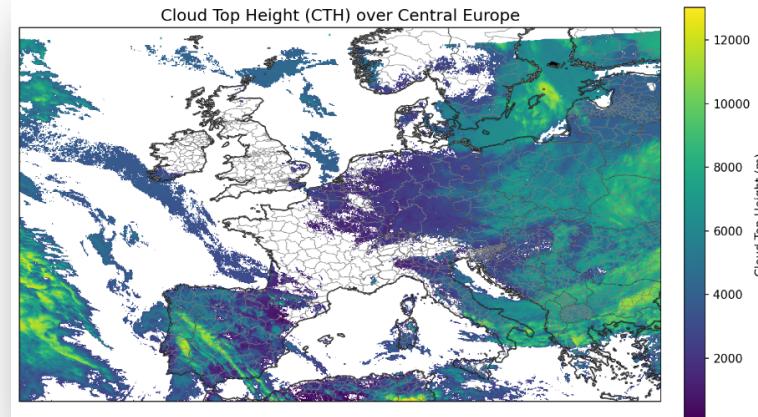
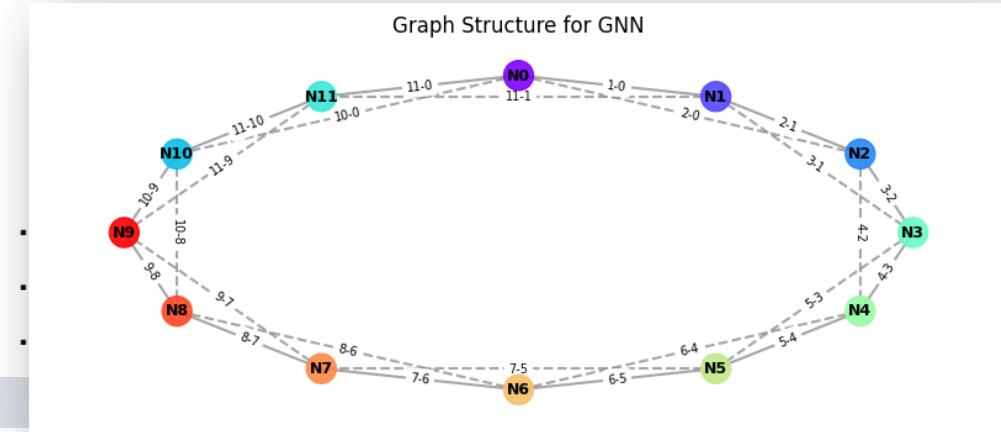
Contents Day 3 Morning

9 Multimodal LLMs

| | |
|--|--|
| 9.1 Fundamentals of Multimodal Large Language Models | |
| 9.2 Radar Data Access and AI Interpretation | |
| 9.3 Cloud Top Height as a Multimodal AI Application | |

10 Diffusion and flexible Graph Networks

| | |
|--|-----|
| 10.1 Diffusion Networks | 176 |
| 10.2 Flexible Graph Networks for Learning from Sparse Observations | 189 |
| 10.3 Exploring Graph Structures in Detail | 196 |
| 10.4 PyTorch Lightning and PyTorch Geometric | 201 |



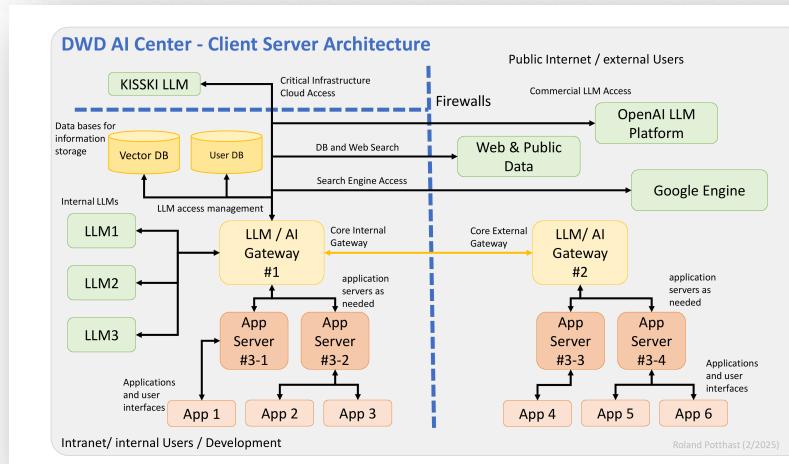
Contents Day 3 Afternoon

11 Agents and Coding with LLM

- 11.1 Introduction to Automated Coding with LLMs
- 11.2 Survey of LLM-Based Agent Frameworks
- 11.3 Using LangChain for Code Design and Execution
- 11.4 LangGraph-Based Forecast Assistant

12 DAWID, LLMs and Feature Detection

- 12.1 The DAWID Frontend: Upload and Interaction Interface
- 12.2 DAWID Backend Architecture
- 12.3 AI based Feature Detection for Fronts



Screenshot of the DAWID beta interface. The top navigation bar shows tabs for SynCast_D, SynCast UI, 6020 WebC, Jupyter Notebooks, Marc Raute, and DWD AI Int. The address bar is eumetnet-ai.eu/dawid/index.php. The session information shows User: roland.pothast@dwd.de, Session ID: a5b73566964fa24f61c88ea0ce3f7a54, and a New Session button.

The main area displays a conversation between the user and DAWID:

You asked:
Tell me about ECMWF

DAWID:
► Model
ECMWF — short overview

- What it is: The European Centre for Medium-Range Weather Forecasts (ECMWF) is an independent intergovernmental organisation that develops and runs a leading global numerical weather prediction system and provides a wide range of forecast products and reanalyses to national meteorological services, researchers and the public.
- Mission and role: produce high-quality medium-range (and longer) weather forecasts, develop the Integrated Forecasting System (IFS), advance data assimilation and modelling research, and provide global datasets for operational use and science.

Screenshot of the DAWID beta interface showing a weather forecast result. The text says:

You asked:
what is the weather tomorrow in Reading? 12pm

DAWID:
► Model
► Function get_weather_forecast Results
► Model

Here's the forecast result for Reading tomorrow (2025-11-28) at 12:00:

ICON-EU 2m Temperature (°C)
2025-11-27 00:00 UTC +036h; forecast for: 2025-11-28 12:00 UTC

A map of Europe showing temperature in degrees Celsius. A color scale from -20 to 40 is provided. The map shows a gradient from blue (lower temperatures) to red (higher temperatures), with most of Europe colored in shades of green and yellow, indicating temperatures between 0°C and 20°C.

- Notes:
- The plot above shows the temperature evolution up to 12:00 on 2025-11-28.
 - If you'd like exact numeric values (temperature in °C, precipitation probability, wind speed/direction, cloud cover) or a short plain-language summary (e.g., "sunny, light breeze, ~X°C"), tell me which parameters and I'll fetch them.

Contents Day 4 Morning

13 MLFlow – An easy way of Managing and Monitoring Training

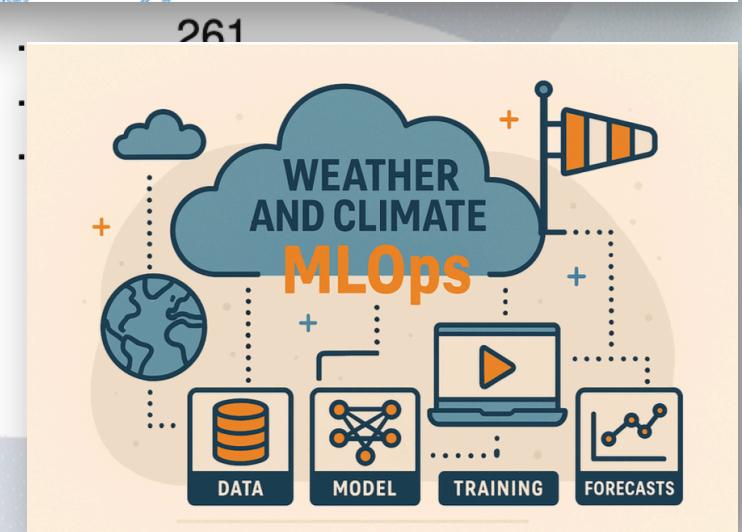
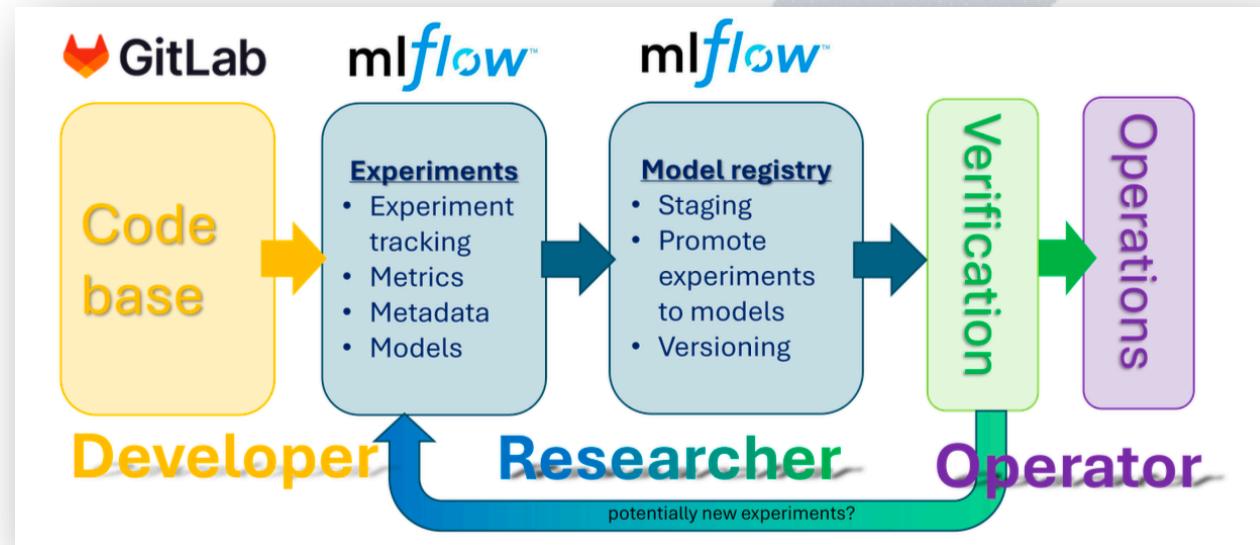
- 13.1 Introduction to MLFlow
- 13.2 Logging ML Experiments with MLFlow
- 13.3 Running an MLFlow Server
- 13.4 Advanced Features and Model Management

The screenshot shows the MLflow web interface version 2.16.2. The top navigation bar includes 'Experiments' and 'Models'. The main area is titled 'aicon_prototype_v3' and indicates it 'Uses anemoi version from January 2025 <https://gitlab.dkrz.de/aicon/anemoi/anemoi-core>'. The 'Experiments' tab is active, displaying a list of experiments such as 'hyperparameter_study', 'airfronts_felix', 'seven', etc. The 'Runs' tab is also visible. Below the experiments list, there is a table of runs with columns for 'Run Name' and 'Created' date.

| Run Name | Created |
|---|-------------|
| rfl5_h4_3h_2011_2021_val2022_clpprecmax | 13 days ago |
| rfl5_h4_3h_2011_2021_val2022_clpprec | 13 days ago |
| rfl6_h5_3h_2011_2021_val2022_noclp | 13 days ago |
| rfl6_h5_3h_2011_2021_val2022_noclp | 13 days ago |
| rfl6_h5_3h_2011_2021_val2022_clpprec | 14 days ago |
| rfl6_h5_3h_2011_2021_val2022_clampstd | 14 days ago |
| rfl5_h4_3h_2011_2021_val2022_clampstd | 16 days ago |
| rfl5_h4_3h_2011_2021_val2022_noclamp | 16 days ago |

14 MLOps - Development and Operations Integrated

- 14.1 DevOps and MLOps – Foundations and Motivation
- 14.2 Containerization and Reproducibility
- 14.3 DevOps at DWD – Numerical Weather Prediction as a Forerunner



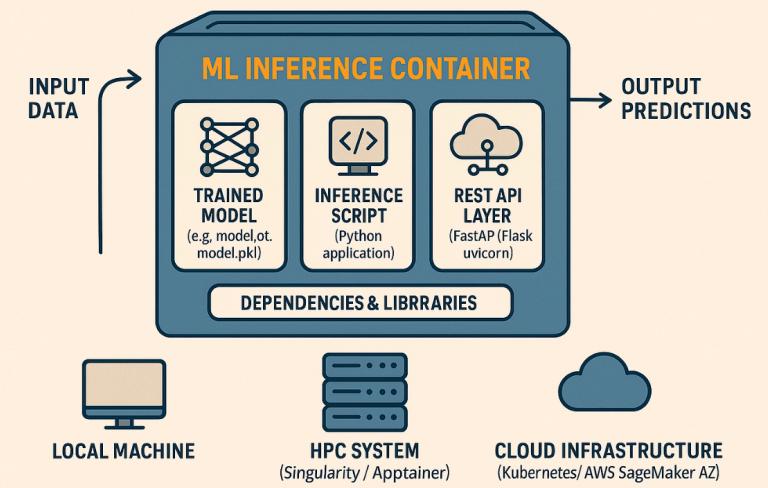
Contents Day 4 Afternoon

15 CI/CD – Continuous Integration and Deployment of ML Code

- 15.1 Overview and Motivation
- 15.2 Tools and Frameworks for CI/CD
- 15.3 Testing with Pytest
- 15.4 CI/CD Runners and Cloud Integration
- 15.5 CI/CD for ICON, AICON, and Anemoi

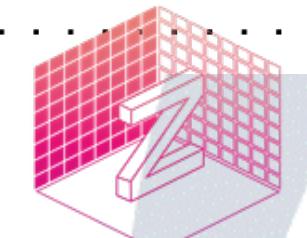
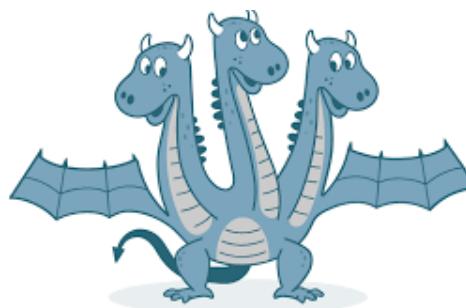
Zur Seite 304

CONTAINER-BASED ML INFERENCE SYSTEM

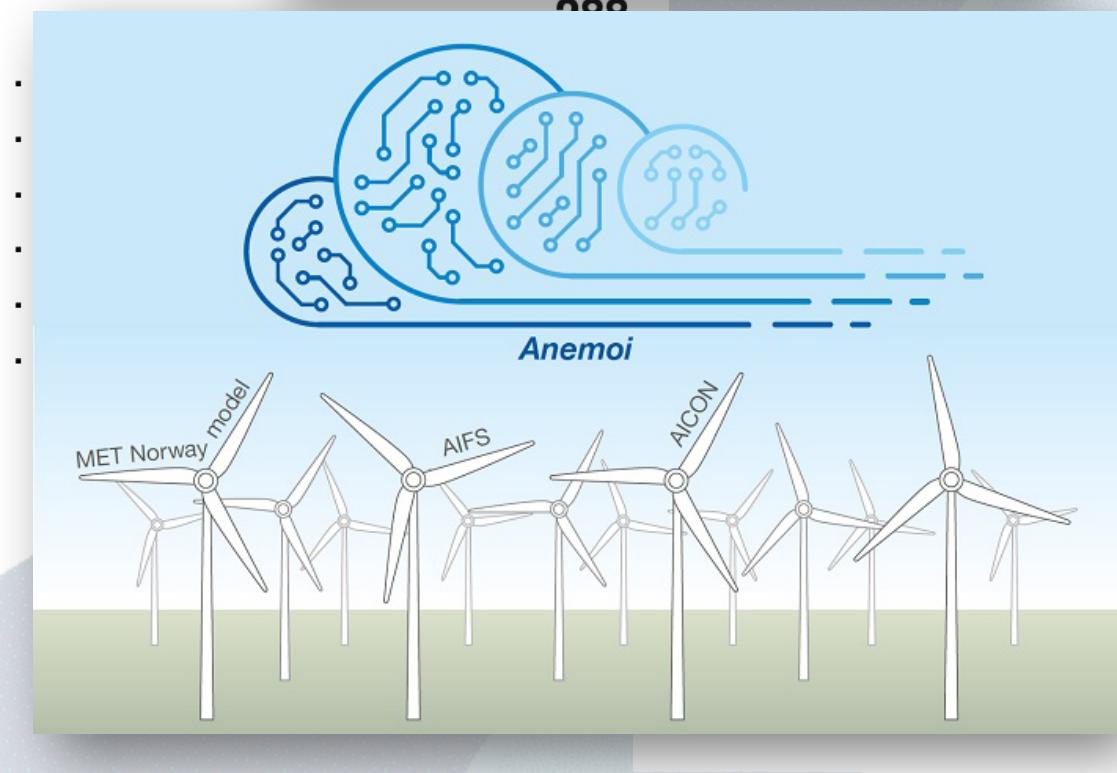


16 Anemol – AI-Based Weather Modeling

- 16.1 Yaml, Hydra and OmegaConf
- 16.2 Introduction to Anemol
- 16.3 ZARR, ERA and Datasets for Anemol
- 16.4 Building an Icosahedral Graph with Anemol
- 16.5 Hands-On Datasets, Validation, Training With Anemol
- 16.6 Training Pipeline in Anemol



Zarr



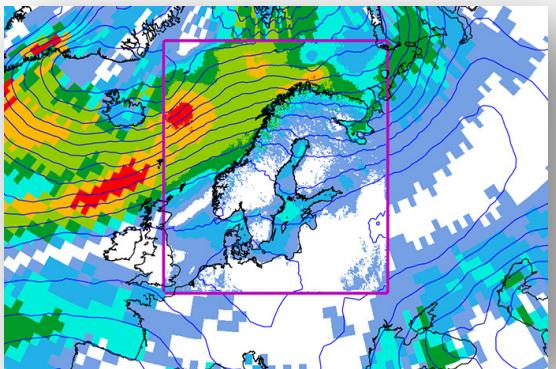
Contents Day 5 Morning

17 Model Emulation and AICON

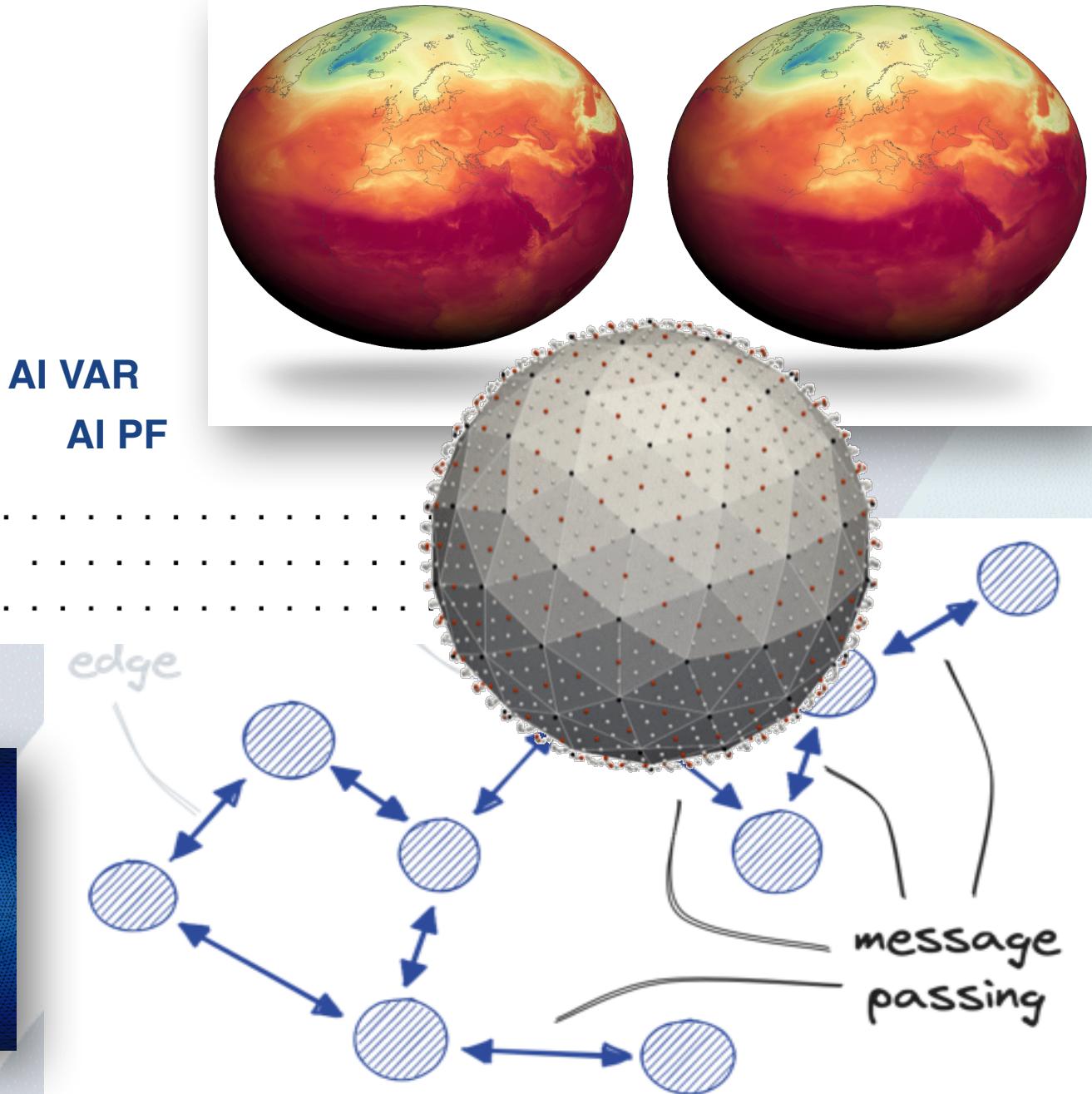
- 17.1 AIFS - the ECMWF flagship model
- 17.2 AICON - trained on ICON-DREAM
- 17.3 BRIS - stretched grid approach

18 AI Data Assimilation

- 18.1 Introduction to AI-based Data Assimilation
- 18.2 Training for AI-based Data Assimilation (AIDA)
- 18.3 Info Zur Seite 333



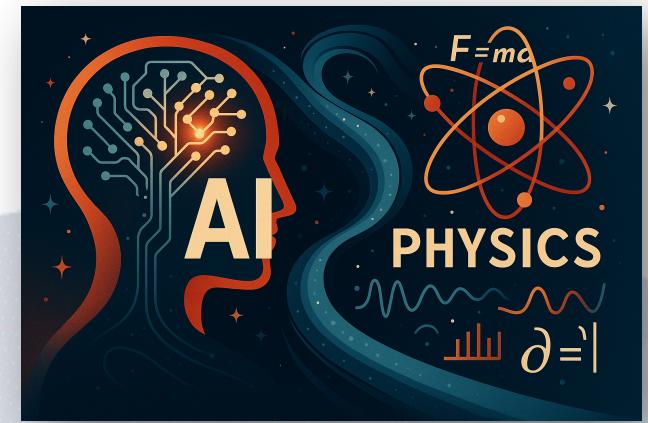
10



Contents Day 5 Afternoon

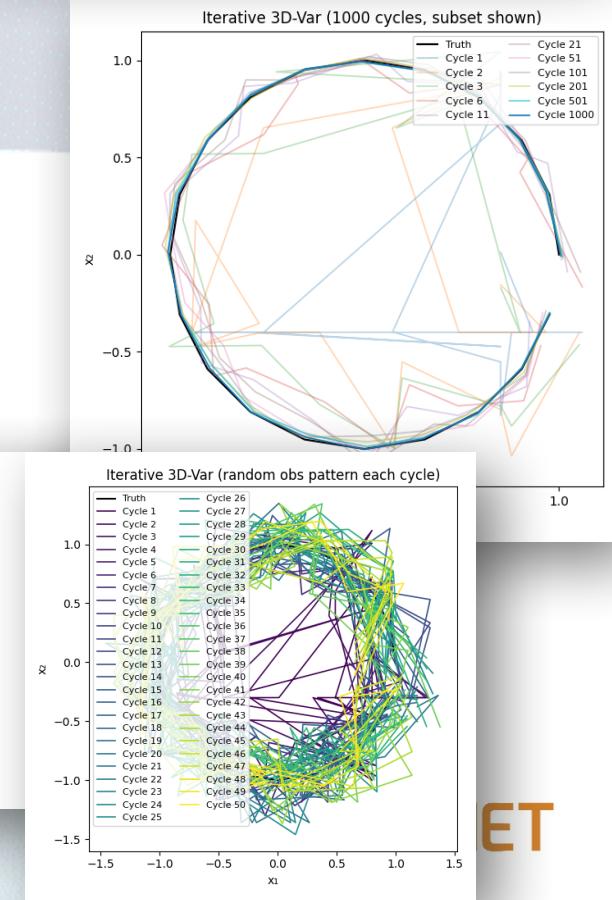
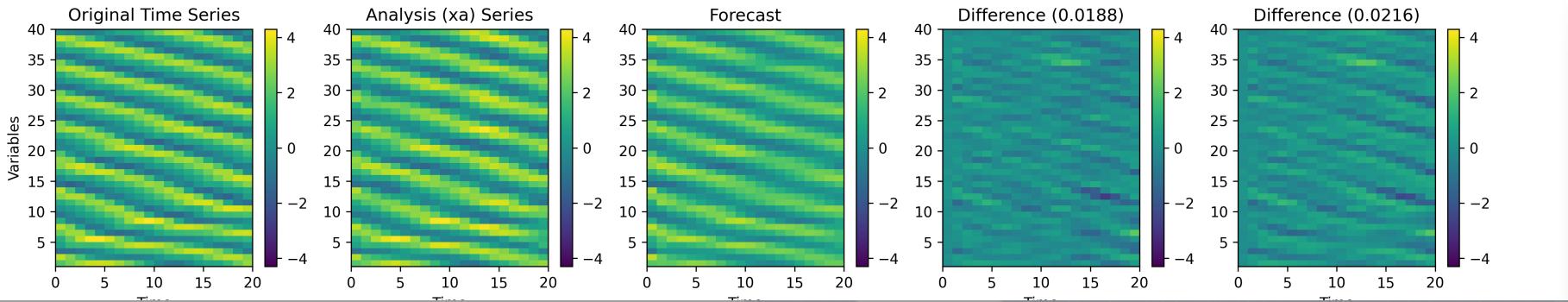
19 AI and Physics

- 19.1 Methods to extract governing equations from learned models,
- 19.2 Physics-informed neural networks and physics-regularised loss functions,
- 19.3 Embedding conservation laws (mass, energy, momentum) in the architecture



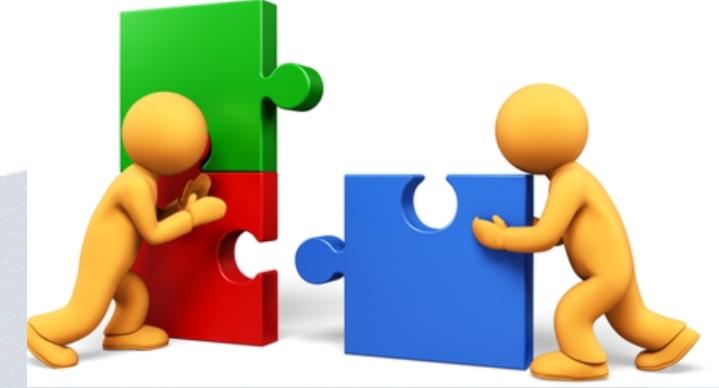
20 Learning from Observations Only

- 20.1 Direct Observation Prediction
- 20.2 ORIGEN Observation based Reconstruction and Inference for Generative Emulation of Nonlinear Systems
- 20.3 The Interplay of Human Insight, Invention, Modelling and Machine Learning



Tutorial Philosophy

- 1) Learn to master things **yourself**
- 2) Work in small groups & help **each other**
- 3) Work from **elementary** understanding towards larger packages
- 4) Learn **simple** examples first
- 5) We need to understand the **whole range** of ML/AI techniques today in context





Be careful, look closely!

Lets get into it now ...

CONTACT DETAILS

Prof. Dr. Roland Potthast
E-AI Programme Manager

EIG EUMETNET
European Meteorological Services' Network
www.eumetnet.eu

EIG EUMETNET Secretariat
c/o L'Institut Royal Météorologique de Belgique
Avenue Circulaire 3
B-1180 Brussels, Belgium

Registered Number 0818.801.249 - RPM Bruxelles

Tel: 0049 69 8062 2713

Email: Roland.Potthast@dwd.de

Dr. Marek Jacob
E-AI Programme Coordination