

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

Introduction

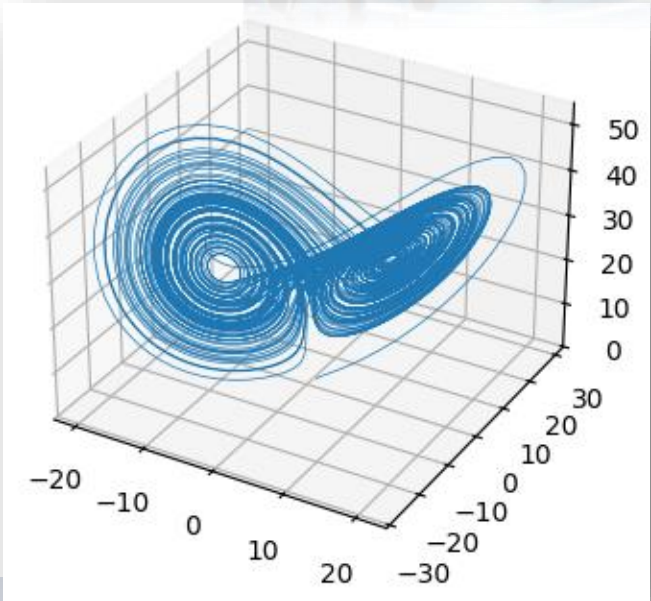
I	Programming and working with Data in the Time of AI	i
II	General Coding Rules and Strategy	iv
III	6 Days Python and AI	xi



Day 1: Python as Workhorse

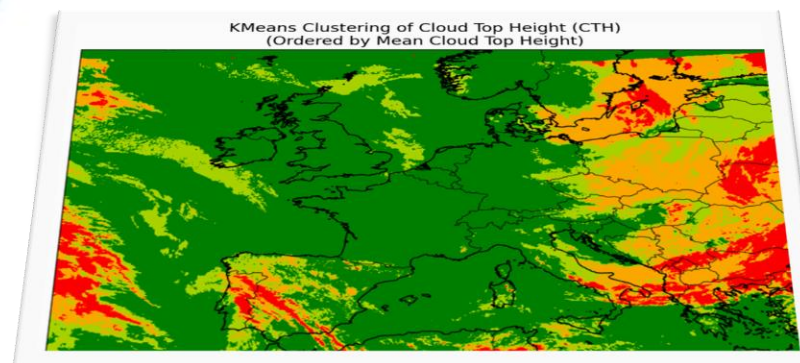
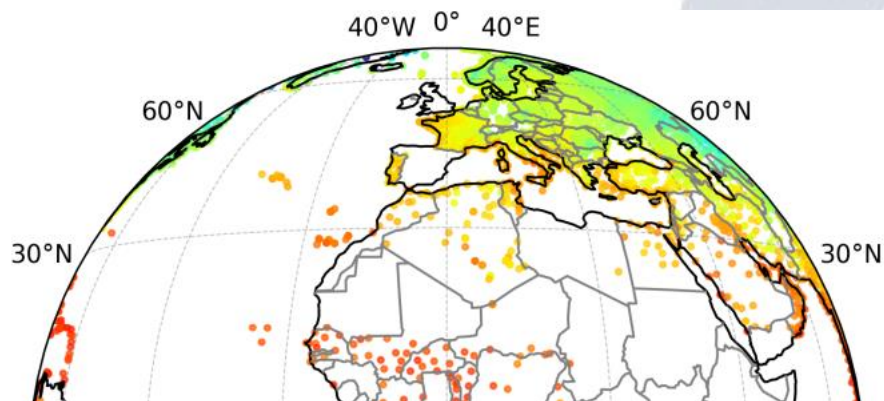
1 Python Basics

1.1	Install, Virtual Environment, Pip und Import	1
1.2	Managing Dependencies with requirements.txt	1
1.3	Introduction to NumPy	3
1.4	Generating Plots based on Matplotlib	7
1.5	Functions	9
1.6	Python Essentials	12



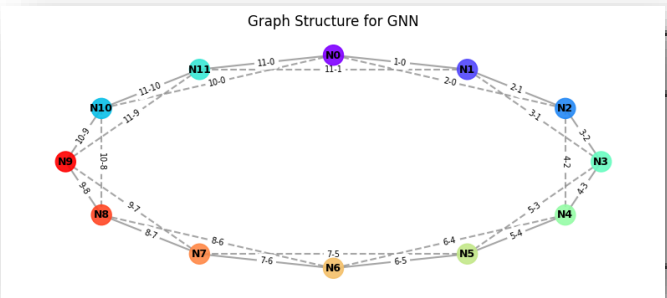
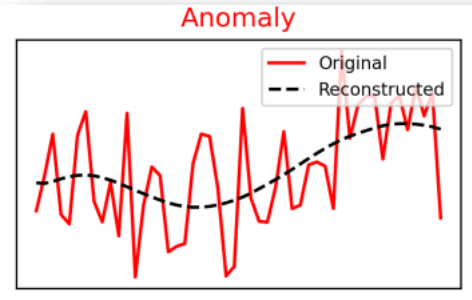
Contents Day 1

2	Jupyter Notebooks, APIs and Servers	22
2.1	Introduction to Jupyter Notebooks	22
2.2	Introduction to APIs: A Key Principle in Code Development	29
2.3	Making API Requests with requests	31
2.4	Fortran Integration using ctypes as API	38
3	Eccodes for Grib, Opendata, NetCDF, Visualization	40
3.1	Downloading ICON Model GRIB Files from DWD Open Data Server	41
3.2	The Grib Library eccodes	43
3.3	Accessing SYNOP Observation Files from NetCDF	53
3.4	Analysing AIREP Feedback Files in NetCDF Format	57



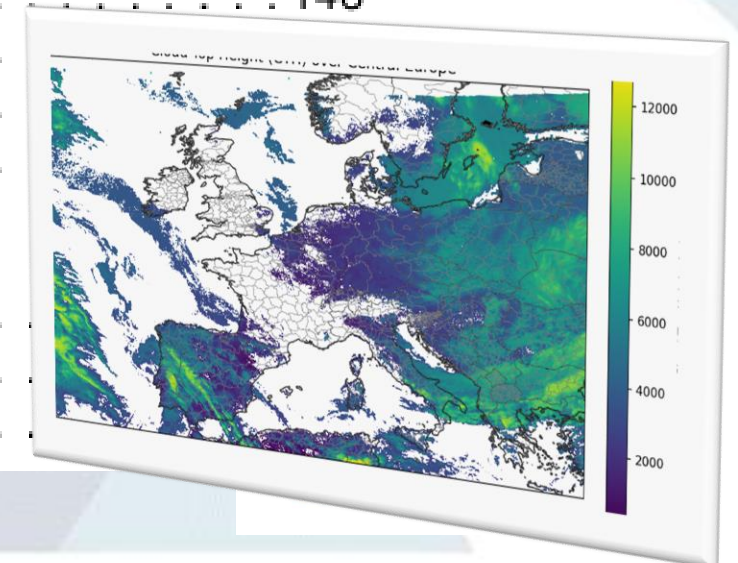
Contents Day 2

Day 2: AI/ML Basic Introduction	64
4 Basics of Artificial Intelligence and Machine Learning (AI/ML)	65
4.1 AI and ML - Basic Ideas	65
4.2 Torch Tensors - Basics and Their Role in Minimization	68
4.3 PyTorch Fundamentals - Model, Loss, and Optimizer	69
4.4 Simple Neural Network Training Example	74
4.5 Gradient Field and Decision Boundary	78
5 Neural Network Architectures	82
5.1 Feed Forward Networks	82
5.2 Graph Neural Networks	87
5.3 Applying Convolutional Neural Networks for Function Classification	92
5.4 LSTM-Based Anomaly Detection in Sensor Data	97
6 Large Language Models	106
6.1 LLM Network as Sequence-to-Sequence Machines via Transformer Models	106
6.2 Implementing and Training a Simple Transformer-Based LLM	110
6.3 Install Your Own LLM, Chat with it and Develop Applications	115



Contents Day 3

Day 3: LLM RAG, Python Packages, Multi-Modality	126
7 LLM with Retrieval-Augmented Generation (RAG)	127
7.1 Preparing Documents	128
7.2 Generating Embeddings for Documents	130
7.3 Using an LLM Locally or with OpenAI for Response Generation	133
7.4 Saving and Reloading the Vector Database, Collecting Search Originals, Chunking long Documents	138
8 Python Packages	143
8.1 Review of the Python Standard Library	143
8.2 Xarray - Multi-dimensional labeled Data	146
8.3 Pandas - Data Frames and Analysis Package	
8.4 SciPy Scientific Computing, Optimization and Statistics	
8.5 Scikit-Learn - Machine Learning, Classification, Regression	
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	



Contents Day 4-6

Day 4: GPUs, AI Agents, Services and Impact

10 Further ML Architectures and Topics

- 10.1 Diffusion Networks
- 10.2 Using GPUs for Training
- 10.3 Dynamic Graphs in Neural Networks for Observation Processing

11 Agents and Coding with LLM

- 11.1 Introduction to Automated Coding with LLM
- 11.2 Survey of Agent Frameworks
- 11.3 Example 1: Using LangChain for Code Execution
- 11.4 Example 2: Using Auto-GPT to Automate Tasks
- 11.5 Generated Code: Fetching and Plotting a 2m Temperature Field
- 11.6 Building a Vector Database as Package

12 LLMs for Geosciences, Weather, and Climate

- 12.1 The LLM AI Interface and Framework DAWID
- 12.2 AI-Assisted Feature Detection in Weather and Climate Data
- 12.3 Automated Weather Report Generation and Interpretation

Day 4: ML Architectures, Agents and LLM Framework DAWID

Day 5: MLFlow, MLOps, Fine Tuning LLMs

Day 6: Anemol, AICON and AI Data Assimilation

Day 5: LLM Maturity and Operations

13 MLFlow - Managing and Monitoring Training

- 13.1 Setting up MLFlow
- 13.2 Monitoring Training
- 13.3 Comparing Experiments
- 13.4 Managing Parameters

14 MLOps - Operations

- 14.1 Introduction to MLOps: Principles and Workflow
- 14.2 Model Deployment and Monitoring
- 14.3 CI/CD for Machine Learning: Automation and Reproducibility
- 14.4 Scalability and Infrastructure: Kubernetes, Cloud, and On-Premise Solutions

15 Fine-Tuning LLMs

- 15.1 Introduction to Finetuning Large Language Models
- 15.2 Dataset Preparation and Preprocessing
- 15.3 Techniques and Strategies for Finetuning
- 15.4 Evaluation and Deployment of Finetuned Models

Day 6: AI Model and AI Data Assimilation

16 Anemol – AI-Based Weather Modeling

- 16.1 Introduction to Anemol
- 16.2 Core Architecture and AI Components
- 16.3 Training Anemol with Historical Weather Data

17 Model Emulation and AICON

- 17.1 The AICON Training Dataset
- 17.2 The AICON Setup, Grid and Graph Network
- 17.3 How AICON Hierarchical Training works
- 17.4 AICON Runs Verification

18 AI Data Assimilation

- 18.1 Introduction to AI-VAR
- 18.2 Observation Processing
- 18.3 Training and Applications

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



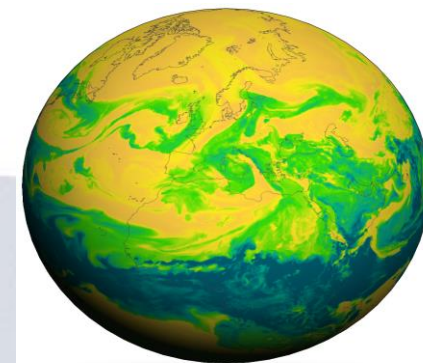
Building AI Expertise with Python and AI Workflows

To effectively integrate AI, we must ensure our **teams are skilled** in both AI methods and operational



- Establishing structured learning paths for key AI techniques relevant to weather and climate modeling.
- Using Python libraries such as numpy, eccodes, netcdf, and xarray for handling large meteorological datasets.
- Training teams in machine learning frameworks such as TensorFlow, **PyTorch**, and Hugging Face Transformers.
- Setting up end-to-end AI workflows in **Jupyter-based environments**, covering data ingestion, training, validation, and inference.
- Encouraging collaboration between meteorologists, model developers, and AI experts to foster cross-disciplinary innovation workflows.

Replacing and Hybridizing Forecasting Systems with AI

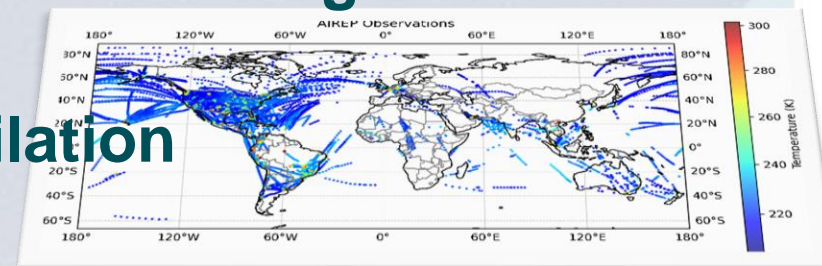


Some forecasting components will be fully replaced by AI, while others will integrate AI as a hybrid solution. Key shifts include:

- **AI-Based Nowcasting**: AI-driven short-term weather predictions using real-time observational data (e.g., satellite, radar, sensors), enhancing or replacing conventional nowcasting techniques.
- **Neural Weather Models**: Deep learning models trained on historical data can generate competitive forecasts at lower computational cost.
- **Hybrid AI-NWP Models**: AI enhances physics-based forecasting through bias correction, uncertainty quantification, and ensemble optimization.
- **Machine Learning for Subgrid Processes**: AI could improve or replace empirical parameterizations in turbulence, cloud physics, and convection models.
- **Automated Impact Forecasting**: AI-driven models provide direct risk assessments for extreme weather events, minimizing reliance on manual interpretation.

AI in Data Assimilation and Learning Directly from Observations

- Machine Learning for **Observation Processing**: AI-driven quality control of observational data, filling data gaps and detecting sensor anomalies.
- AI-Based **Data Assimilation**: AI improving assimilation processes by optimizing observation ingestion.
- Deep Learning for Data Assimilation: AI learning complex relationships between observations and model states, accelerating assimilation workflows.
- End-to-End AI Data Ingestion: Future AI models trained **directly on observational datasets**, potentially reducing reliance on classical assimilation techniques.
- Self-Learning Systems: AI dynamically adjusting to new data, improving continuously without manual recalibration.



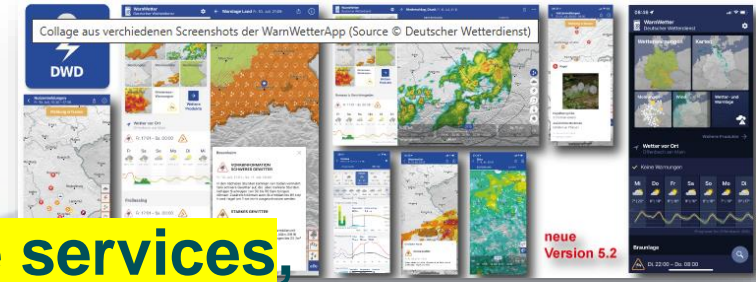
Using AI for Code Refactoring and Model Development

AI also modernizes modeling **workflows**, improving efficiency in research and development:

- **Refactoring Legacy Code:** AI-assisted tools improving Fortran, C++, and Python models for better maintainability and performance.
- **Automated Model Optimization:** AI **tuning hyperparameters** and optimizing computational performance.
- **AI-Assisted Scientific Discovery:** AI identifying new climate and weather patterns in large datasets.
- **AI-Generated Documentation and Testing:** Automating documentation and generating validation tests for numerical models.

```
mo_nwp_diagnosis.f90 122.94 KiB
1  ! ICON
2  !
3  ! -----
4  ! Copyright (C) 2004-2024, DWD, MPI-M, DKRZ,
5  ! Contact information: icon-model.org
6  !
7  ! See AUTHORS.TXT for a list of authors
8  ! See LICENSES/ for license information
9  ! SPDX-License-Identifier: BSD-3-Clause
10 ! -----
11
12 ! Diagnosis of physics after physics call
13
14 !-----
15 #include "omp_definitions.inc"
16 #include "consistent_fma.inc"
17 !-----
18
19 MODULE mo_nwp_diagnosis
20
21
22     USE mo_kind,                ONLY: wp
23
24     USE mo_impl_constants,      ONLY: itccov, it
25                                     min_rlccl
26     USE mo_impl_constants_grf,  ONLY: grf_bdywid
27     USE mo_loopindices,         ONLY: get_indice
28     USE mo_exception,          ONLY: message, m
29     USE mo_model_domain,       ONLY: t_patch
30     USE mo_model_config,       ONLY: iqv, iqc,
31                                     iqni, iqq,
32     USE mo_grid_config,         ONLY: n_dom, n_d
33     USE mo_timer,              ONLY: ltimer, ti
34     USE mo_nonhidden_tunes     ONLY: t_nb_ncon
```

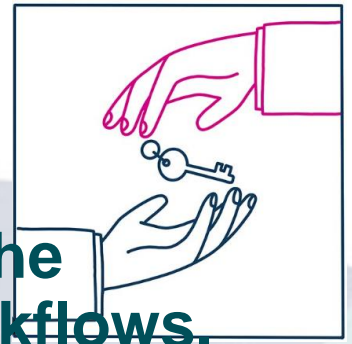
Transforming Services and User Interaction with AI



AI enables new ways to deliver **weather and climate services**, improving automation, personalization, and accessibility:

- **AI-Generated Weather Reports:** Natural language generation models translating raw data into meaningful insights for different user groups.
- **Conversational Forecasting Assistants:** AI chatbots and voice assistants allowing users to interactively query weather and climate predictions.
- **Real-Time Impact Forecasting:** AI models directly linking weather forecasts to risks in agriculture, energy, transportation, and disaster management.
- **AI-Powered Data Visualization:** Interactive AI tools allowing users to explore, manipulate, and interpret complex weather datasets.

A Clear **Migration Strategy** for AI Transformation



1. **AI Readiness Assessment**: Identify areas where AI provides the highest impact while ensuring compatibility with existing workflows.
2. **Pilot AI Replacements**: **Test** AI-based forecasting models **in parallel** with traditional methods before full adoption.
3. **Hybrid Deployment Strategy**: Introduce AI-driven improvements in stages, ensuring **fallback options** are in place.
4. **AI Validation and Trust Building**: Develop **transparent evaluation** metrics for AI models to ensure trust and reliability.
5. **Workforce Training and Knowledge Transfer**: Enable teams to transition smoothly from traditional methods to AI-driven solutions.
6. **Continuous AI Governance**: Establish guidelines for AI model retraining, performance monitoring, and ethical considerations.

Limitations and Responsible Use of AI

While AI offers transformative opportunities in forecasting, modeling, and service delivery, it is crucial to acknowledge its current limitations and apply it with scientific caution:

- **Data Requirements**: Most AI models rely on large, high-quality datasets and perform poorly in data-sparse or non-stationary environments.
- **Lack of Physical Consistency**: AI predictions may violate conservation laws or produce unrealistic results in rapidly evolving scenarios.
- **Limited Interpretability**: Unlike traditional models, many AI systems operate as black boxes, making it difficult to understand or trace their internal reasoning.
- **Bias and Overfitting**: Biased or unbalanced training data can lead to flawed predictions, while overfitting to historical data may reduce adaptability to changing climate conditions.
- **Need for Rigorous Validation**: AI models must be continuously monitored, validated, and benchmarked to ensure stability, fairness, and scientific reliability. Validation needs metrics and scores beyond traditional forecasting metrics.
- **Complementary Role**: AI should be seen as an enhancement to—not a replacement for—physics-based modeling, supporting a hybrid approach for trustworthy innovation.

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