# AI STRATEGY FOR EARTH SYSTEM DATA

CLICK ME

K I : S T E

CLICK ME

AI-PLATFORM


DATA-DRIVEN EARTH SYSTEM RESEARCH

E-LEARNING PLATFORM



JÜLICH Forschungszentrum
RWTHAACHEN UNIVERSITY
UNIVERSITÄT BONN
Universität zu Köln
52n
amb
Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit

This is a non-linear presentation.

Please go to slide 1 and click on the icons you would like to explore.

internal hyper-link

LINK to external page

# INTRODUCTION TO KI:STE

Supported by:

Federal Ministry
for the Environment, Nature Conservation
and Nuclear Safety

based on a decision of the German Bundestag

Artificial intelligence (AI) methods are currently experiencing rapid development and are also being used more frequently in the context of environmental data. However, this is often done in the context of isolated solutions. The systematic use of modern AI methods is not yet established in Earth system sciences.

The project KI:STE (AI strategy for Earth system data) closes this discrepancy with a sophisticated strategy that combines the development of diverse AI applications on different aspects of Earth system research with a strong training and network concept. It creates the technical prerequisites to make high-performance AI applications on environmental data portable for future users and to establish environmental AI as a key technology. Deliverables are an Earth-AI-Platform and a corresponding e-learning platform.

KI:STE is funded by the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety and is supported by Zukunft-Umwelt-Gesellschaft GmbH.

Grant agreement: 67KI2043

Project Partners

Top Page

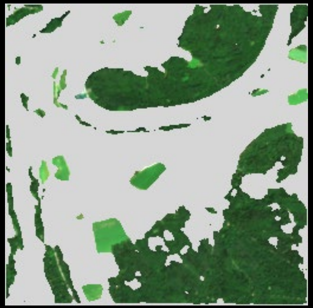
# WILDERNESS (by Timo)

AI finding concepts for wilderness

**original**

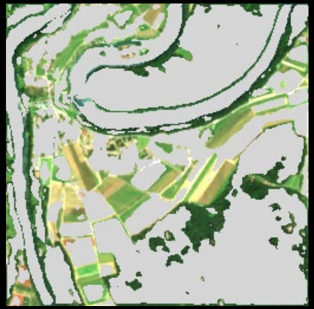
**water & manmade**

water bodies
man-made structures
rocky and sandy surfaces



**forests**

different types of forest
larger vegetation



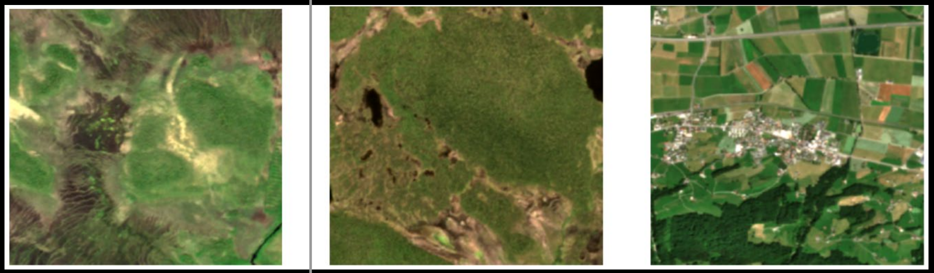
**wasteland**

very diverse wastelands
rough surfaces

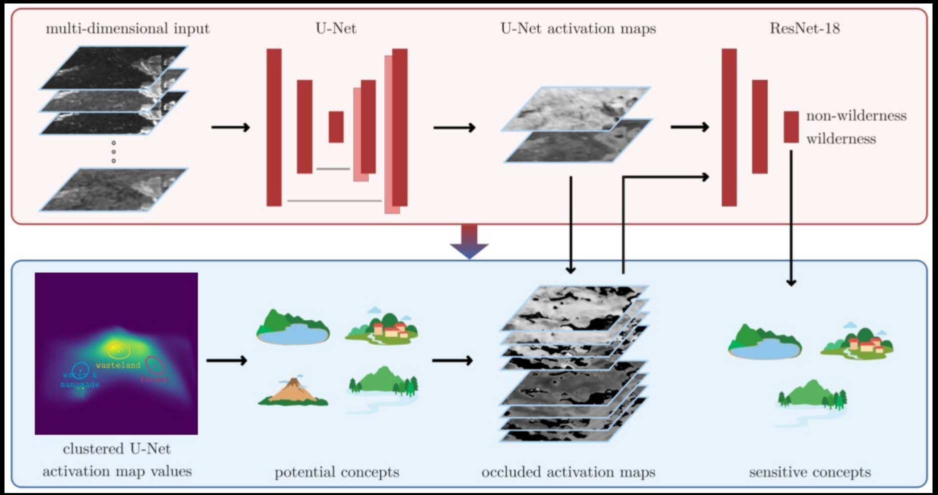
K I : S T E

# WILDERNESS

- Research goal:
  - We want to find sensitive concepts for wilderness by classifying Sentinel-2 images into (non-)protected areas.
- Methodology:
  - We combine a U-Net and a ResNet18. After training the network, we cluster the U-Net's activation maps to find potential concepts.
- First results:
  - Due to occluding the concept's areas in the activation map, we find two sensitive concepts for non-wild areas (water & manmade, forests) and one slightly sensitive concept for wild areas (wasteland).
  - Email Timo for further details or if you are interested in the paper



protected non-protected

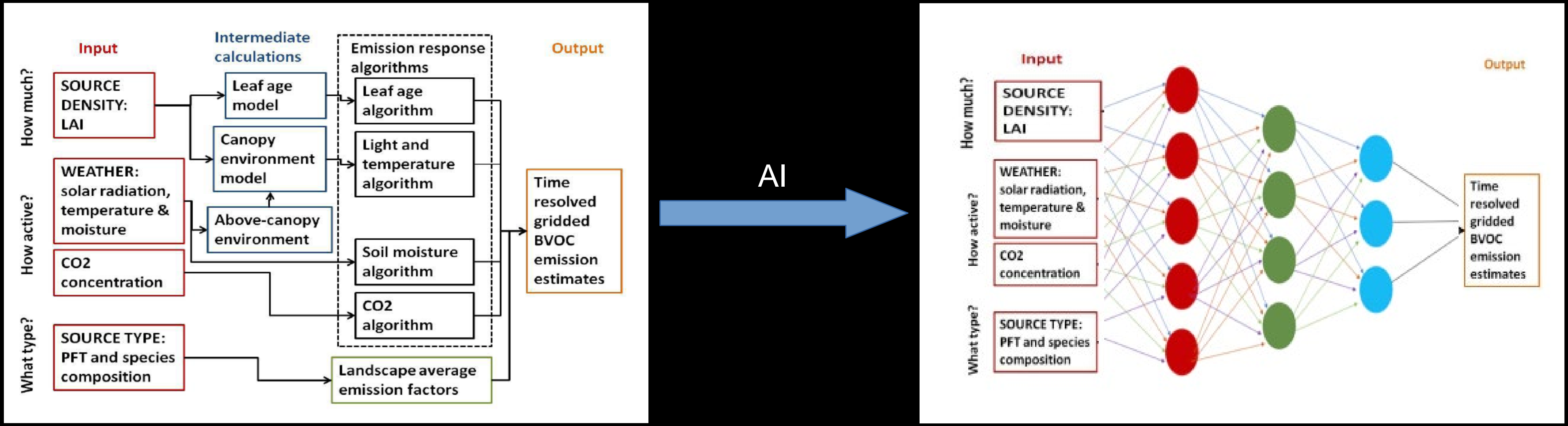



Top Page

# BIOGENIC EMISSIONS (by Ankit)

Numerical modeling appproach for biogenic emissions with MEGAN (Guenther et. al. 2012)

Approach using unsupervised learning on heterogeneous input data (this work)



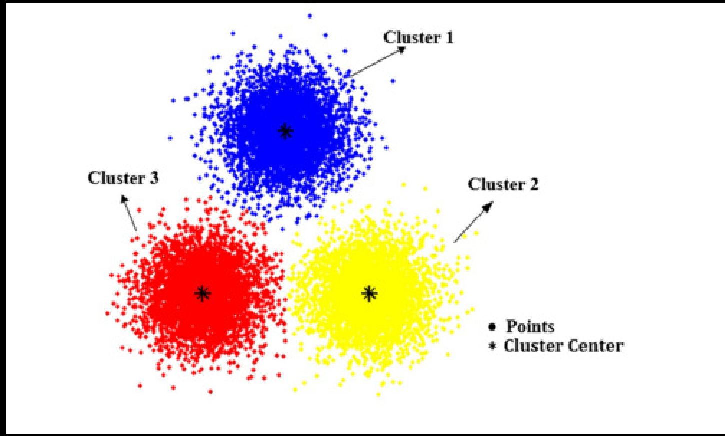

Abbreviations:

| | |
|---|---|
| BVOC | Biogenic Volatile Organic Compounds |
| LAI | Leaf Area Index |
| PFT | Plant Functional Type |

More Details

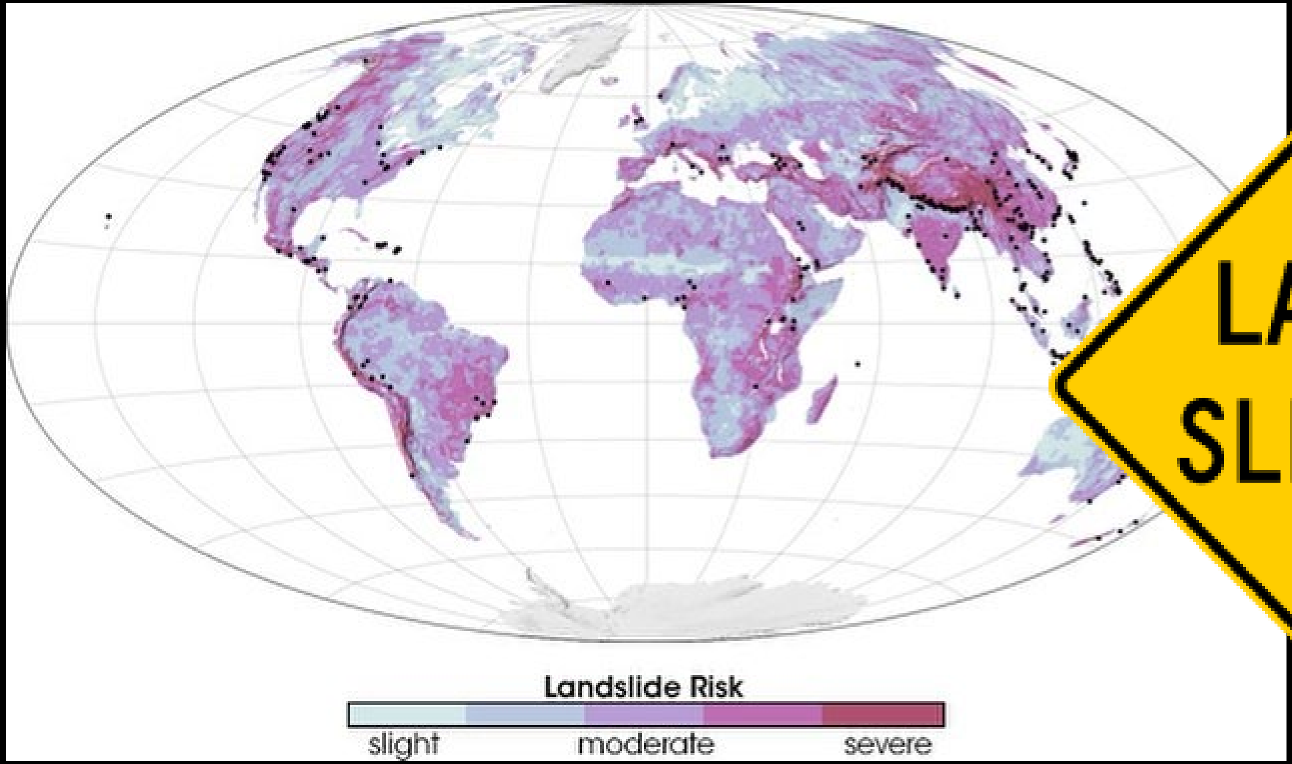
# BIOGENIC EMISSIONS (by Ankit)

- Research Questions:
  - How can we use state-of-the-art unsupervised methods in estimating biogenic emissions?
  - How does the sensitivity of a data-driven method compare to existing numerical methods?
  - Can we obtain parameters by reverse-engineering to improve numerical emission models?
- Methodology:
  - In this study, we will use two quite different types of datasets: satellite images and gridded reanalysis data. Convolutional Neural Networks (CNNs) and Graph Neural Networks (GNNs) can learn information from local patches rather single pixel/grid point information.
  - Moreover, in addition to the spatial information captured by the CNNs and GNNs, we can add recurrent cells to our architecture to encode temporal relations.
  - Our data is unlabeled, so we need to explore unsupervised techniques.




Source: Sentinel, Zhang et al (2017)

# HAZARD PREDICTION (by Ann-Kathrin)





# HAZARD PREDICTION (by Ann-Kathrin)

Motivation:

- Natural hazards pose risks to society and infrastructure worldwide.
- Hazard mapping provides a way of identifying high risk areas by compiling, combining and analyzing various data sources.
- The resulting hazard maps are essential tools for decision and policy makers.

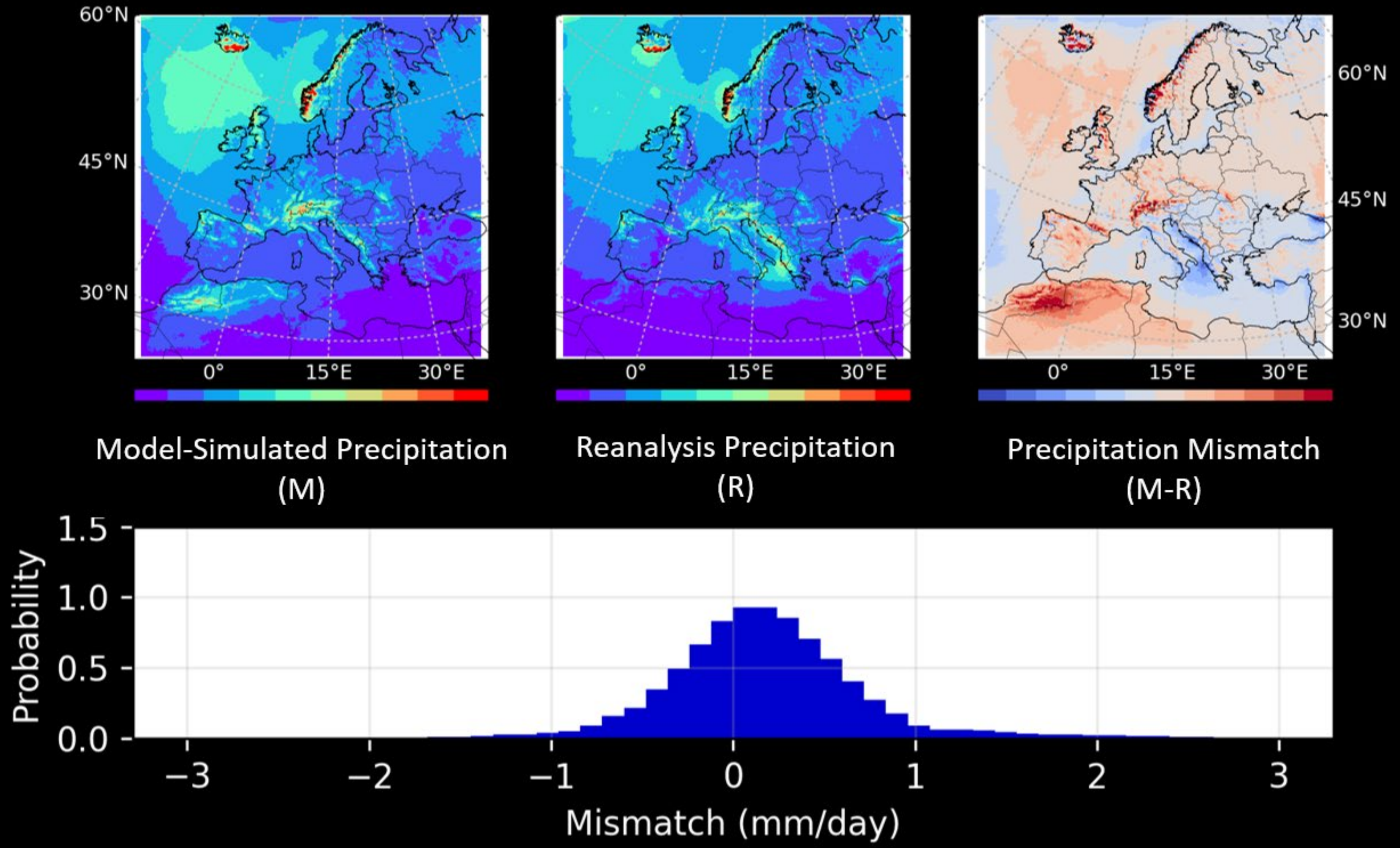
Research goals:

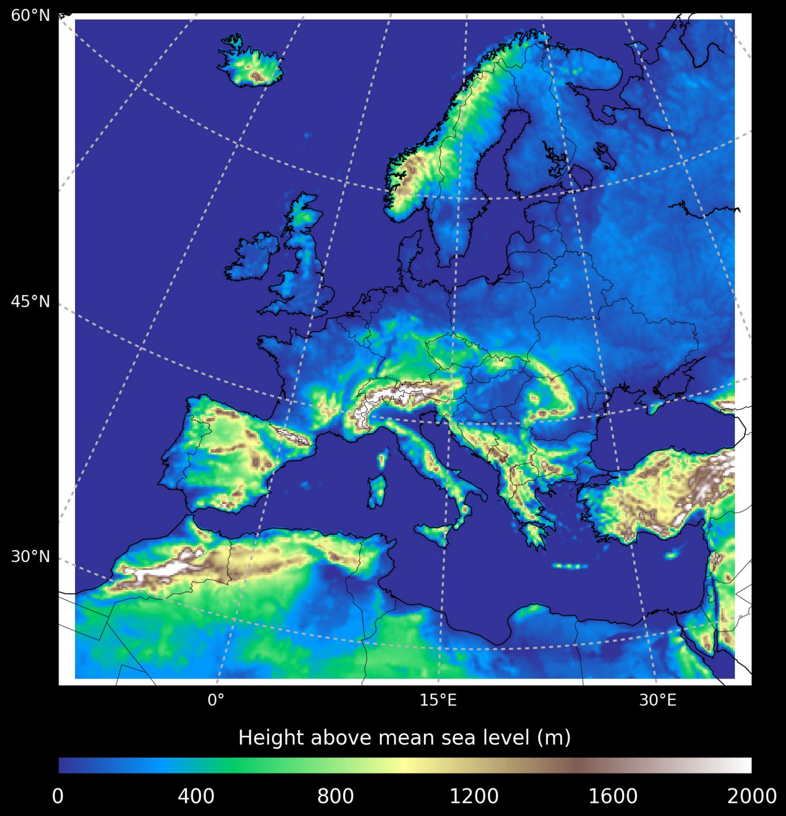
Development of a systematic and flexible approach to hazard mapping based on data, for which it is necessary to:

- handle heterogeneous data sources (e.g. remote sensing and field data)
- develop a hazard mapping workflow that incorporates both physics-based models and machine learning algorithms
- investigate the propagation of uncertainties through the workflow

Initially, the exemplary hazard of landslides is considered and other hazards will be included in later stages of the project. The work shall be guided by the FAIR principles.

# HYDROMETEORLOGICAL EXTREMES (by Kaveh)




**Study Domain**

**Learning from Model-Reanalysis Mismatches**

KI:STE

More Details

# HYDROMETEORLOGICAL EXTREMES (by Kaveh)

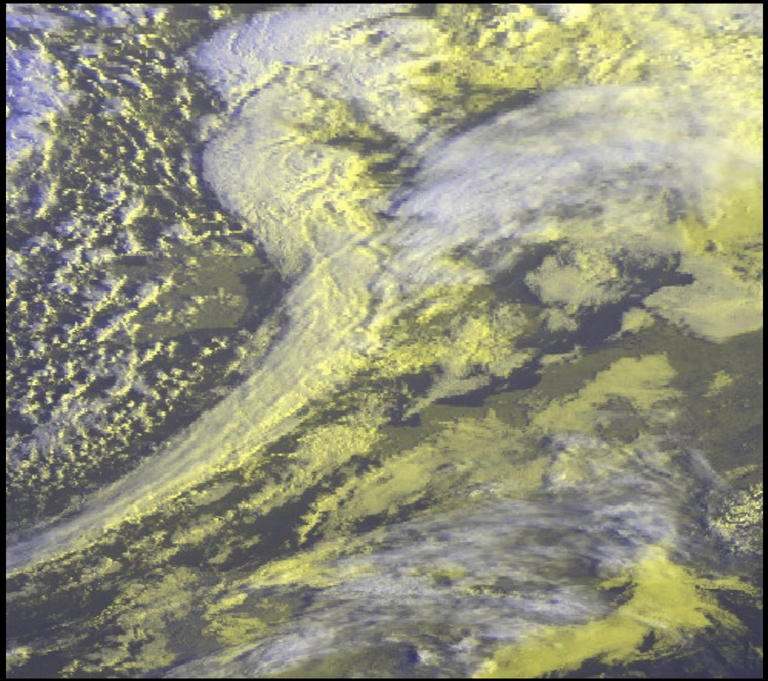
- Research Questions:
  - How can we make use of the state-of-art AI methods for merging modelled and reanalysis/observational data?
  - How is the performance of AI-merged data in quantification of hydrometeorological extremes (e.g., floods and droughts)?

- Methodology:
  - By using Convolutional Neural Networks (CNN), we want to predict the mismatches between the simulated hydrometeorological variables by our terrestrial model, and the reference data, which can be a reanalysis and/or observation data.
  - CNNs are known to efficiently capture the spatial information from all the neighboring pixels in the domain. Hydrometeorological data, e.g. precipitation data, includes spatial correlations which can be captured by using CNNs.

# CLOUD VARIABILITY (by Dwaipayan)

Enable machines to learn directly from the vast amount of information available from Meteo-sat third generation geostationary satellites, rather than just from training data created specifically by manual labelling

In order to understand the physical information from environmental data such as ‘clouds’ we want the AI to have a more deeper sense of understanding beyond what is specified in manually labelled dataset



**Label this image**

1. Clouds
2. Patchy Clouds
3. Shallow Cumulus Clouds
4. ?
5. ?

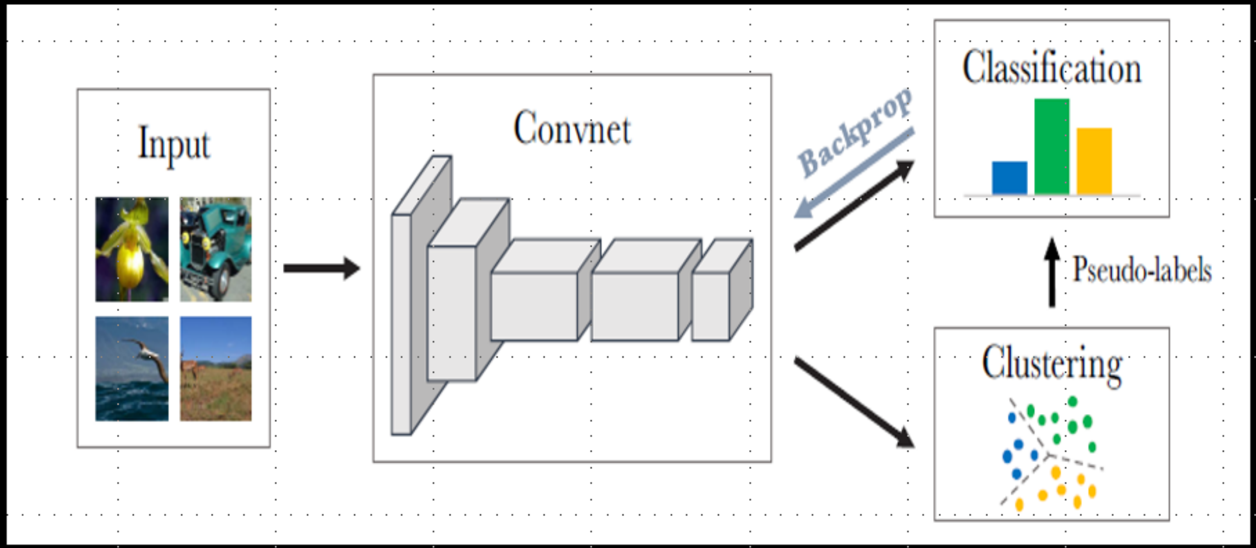
# CLOUD VARIABILITY (by Dwaipayan)

Research motivation

- If we label the clouds then we are bringing the biasness inherently into the learning stage of AI
- In earth science domain we want to understand how objects are interacting or how physics plays a role into the task

AI methodology

- Self - Supervised Learning




Data Set

- Current geostationary satellite data provide moderate spatial and temporal resolution with much finer resolution in space of time – thus large data streams – to become available with the upcoming Meteosat Third Generation.

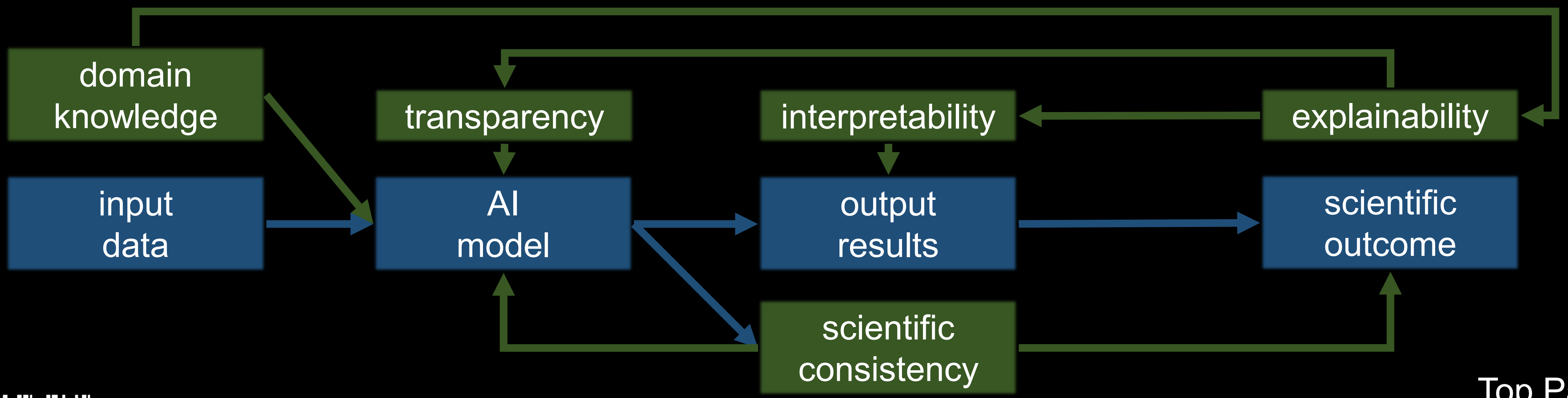
KI:STE

Top Page

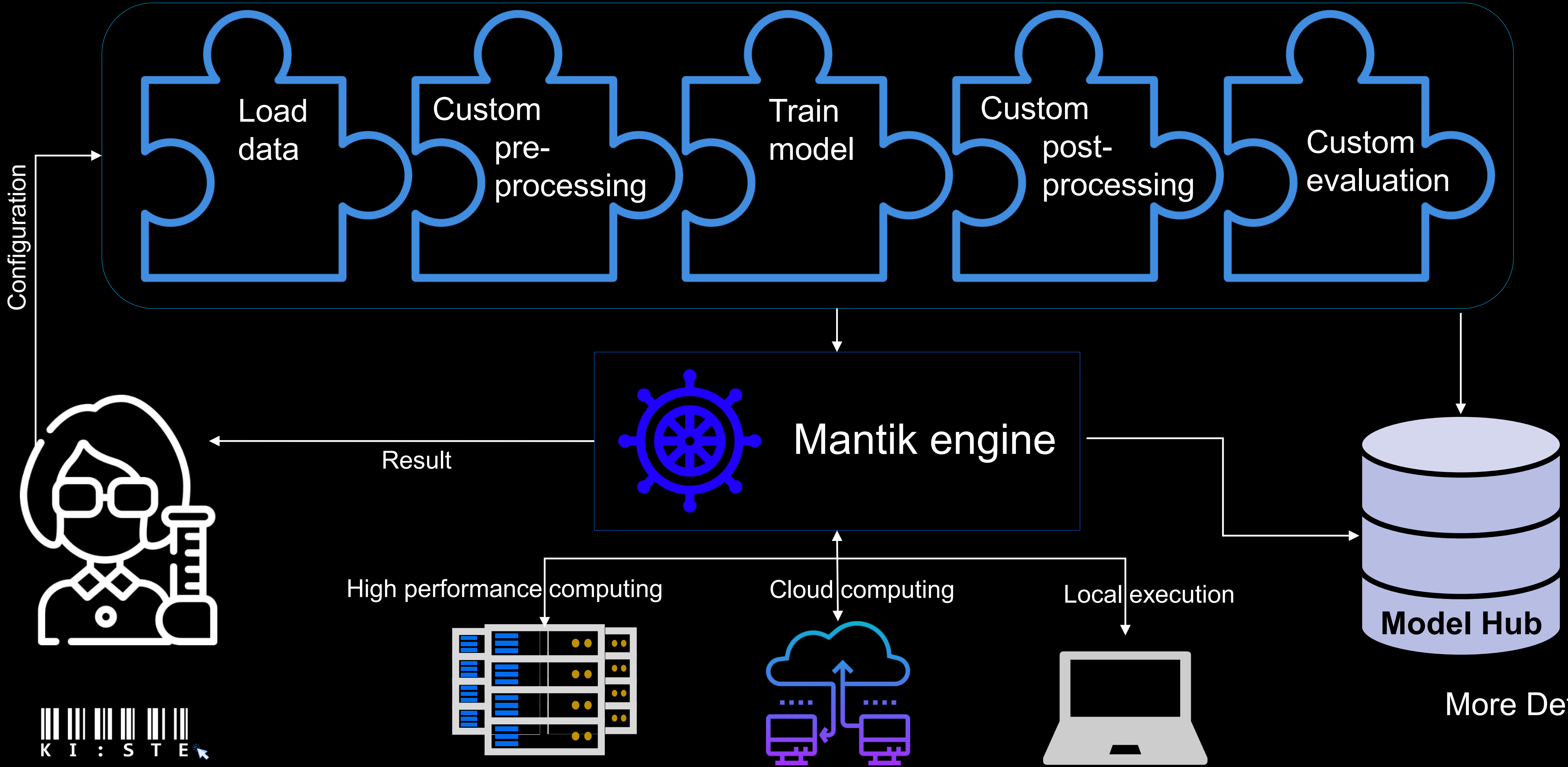
# XAI: EXPLAINABLE AI

- The usage of machine learning in the natural sciences is driven by the goal to obtain novel scientific insights and discoveries from observational or simulated data.
- Thus, domain knowledge is needed to gain explainability and enhance scientific consistency.
- For details, please refer to [Roscher et al. 2020]

AI-driven workflow with XAI components (adapted from Roscher et al. 2020, Fig. 1)



# AI-PLATFORM


Load data
Custom pre-processing
Train model
Custom post-processing
Custom evaluation
Configuration
Mantik engine
Result
Model Hub
High performance computing
Cloud computing
Local execution
K I : S T E

# AI-PLATFORM

Take a closer look at our AI-Platform vEGU21 display [here](#).

- The KI:STE AI-Platform is currently under development.
- The essence of our platform is to create a user-friendly software so researchers can use data-driven methods and machine learning with less overhead.
- For implementation and software related details, please check [Thomas' presentation](#).


Top Page

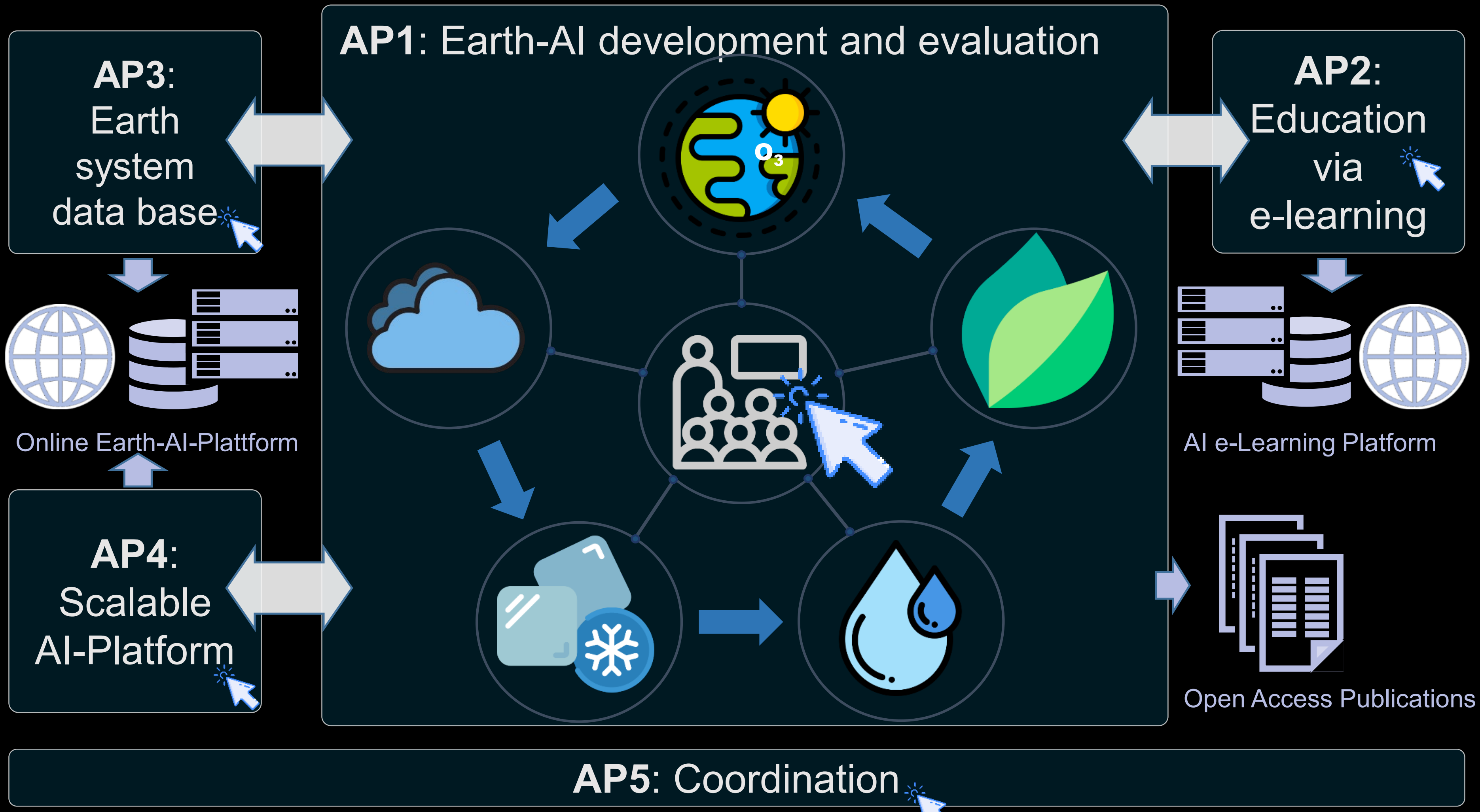
# E-LEARNING PLATFORM

- The KI:STE e-learning platform aims to facilitate getting into machine learning for students and scientists with a Earth system science background.
- Common introductions to machine learning focus on using algorithms and datasets used in industrial applications. The properties of these datasets are different from Earths system data, which usually is not ideally identically distributed and includes spatial correlation and temporal auto-correlation.
- Our e-learning platform will introduce machine learning from Earth science point of view along with algorithms which are currently used in Earth system science.
- Moreover, it will also teach how to use the KI:STE AI platform where many technical obstacles to machine learning, such as workflow design, were already set up.

KI:STE

Top Page

# KI:STE PROJECT WORK PACKAGES




Top Page

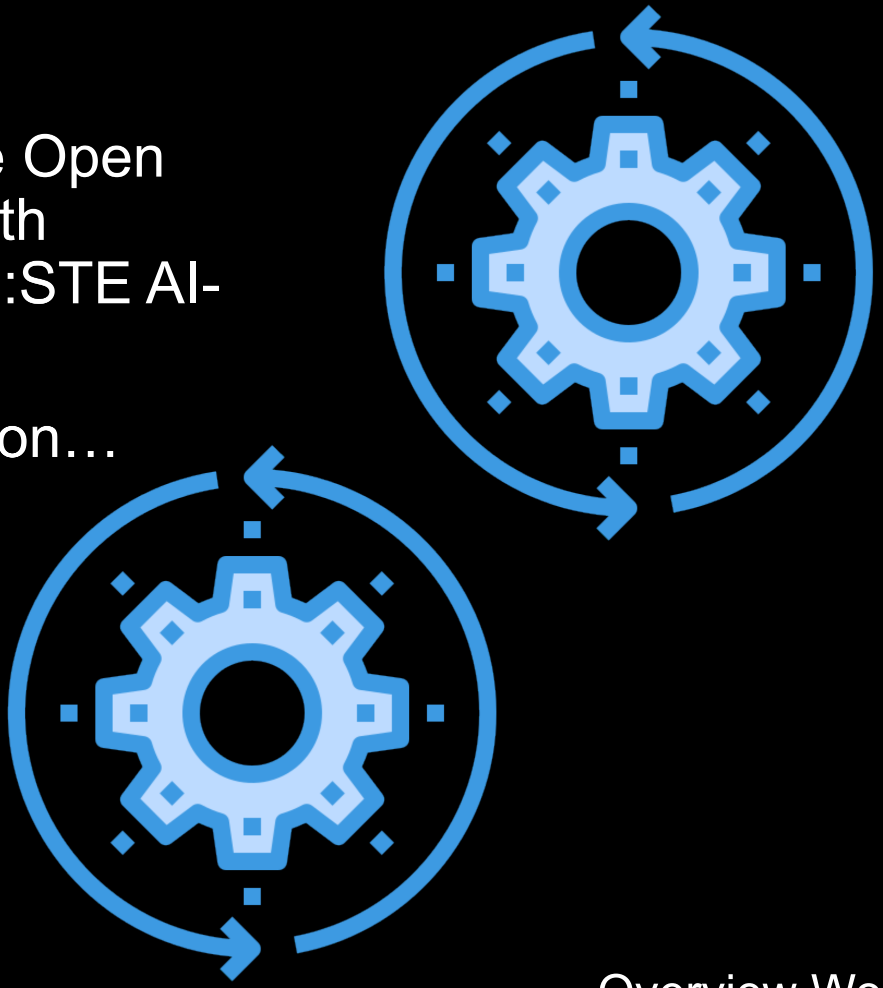
# KI:STE PROJECT WORK PACKAGE 1

- Work package 1 consists in five PhD theses written in different Universities and Institutes.
- Supervisors work in RWTH Aachen, University of Bonn, University of Cologne and Forschungszentrum Jülich (JSC, IBG-3)
- All PhD theses focus on different Earth sub-systems (clouds, snow/ice, water, plants and atmospheric chemistry).
- The topics are interconnected leaving much room for collaborations.
- All PhD theses use AI, namely, machine learning methods and some also use the same datasets for different purposes.
- We agreed on focusing on Europe or European countries for the analysis.
- Moreover, explainable AI methods to understand why machine learning works or does not work. We also hope to make new discoveries from the datasets.

Top Page

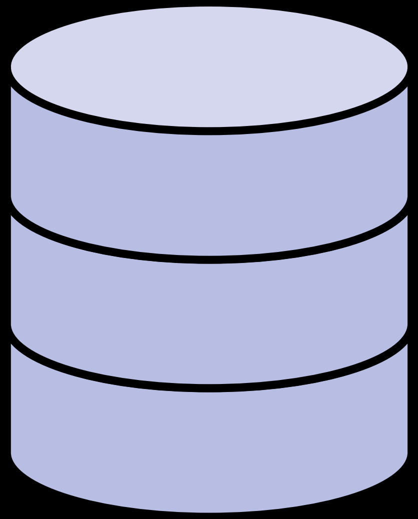
# KI:STE PROJECT WORK PACKAGE 2

- The goal of work package 2 is to create a Massive Open Online Course (MOOC) explaining AI used for Earth System Science and introduce the usage of the KI:STE AI-Platform.
- There is no prototype available yet, but coming soon…



Overview Work Packages

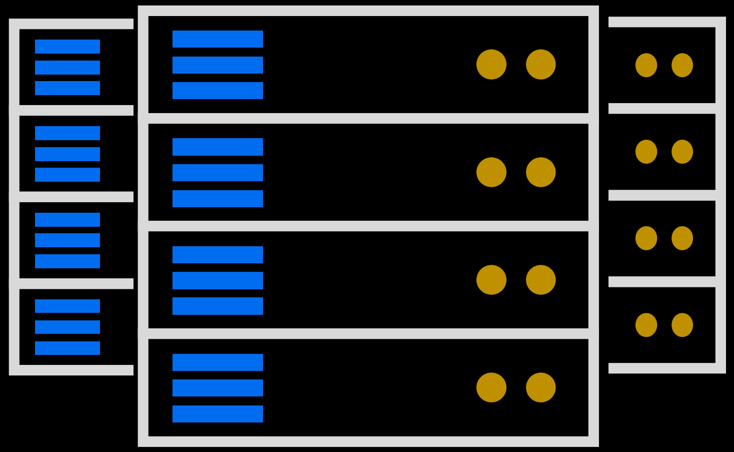
# KI:STE PROJECT WORK PACKAGE 3



- In work package 3 we integrate environmental data required by the AI platform.
- The goal here is design modular interfaces to the AI platform such that open source datasets can be easily integrated and used for machine learning model training.
- This work package is led by the company 52°North, experts in geo-informatics and spatial information infrastructures.

# KI:STE PROJECT WORK PACKAGE 4

- The goal of work package 4 is to design and implement the Earth-AI-platform.
- Milestones include implementing benchmark machine learning frameworks, containerize the data flows and create an easy-to-use user interface.



- The Earth-AI-platform should be flexible and be compatible with cloud computing and high performance computing at JSC.
- This work package is lead by the company Ambrosys.

# KI:STE PROJECT WORK PACKAGE 5

- Work package 5 assures smooth collaboration.
- The KI:STE project is coordinated by the JSC, tasks include:
  - Communication within the project
  - Public outreach  (@KisteProject)
  - Administration
  - Project Management
  - Presenting KI:STE on EGU and hoping for researchers to be interested and new collaboration partners.

Visit our KI:STE website here.

Email Scarlet Stadtler for further details:

s.stadtler@fz-juelich.de

K I : S T E

# MEET THE KI:STE PARTNERS

**AP3**: Benedikt Gräler, Christian Autermann

**AP4**: Markus Abel, @Thomas Seidler

**AP1**: Earth-AI development and evaluation

| Supervisor | Doctoral researcher |
|---|---|
| Julia Kowalski | @Ann-Kathrin Edrich |
| Ribana Roscher | @Timo Stomberg |
| Susanne Crewell | @Dwaipayan Chatterjee |
| Stefan Kollet | @Kaveh Yousefi |
| Martin Schultz | @Ankit Patnala |

**AP2**: We all!

**AP5**: @Scarlet Stadtler

Top Page