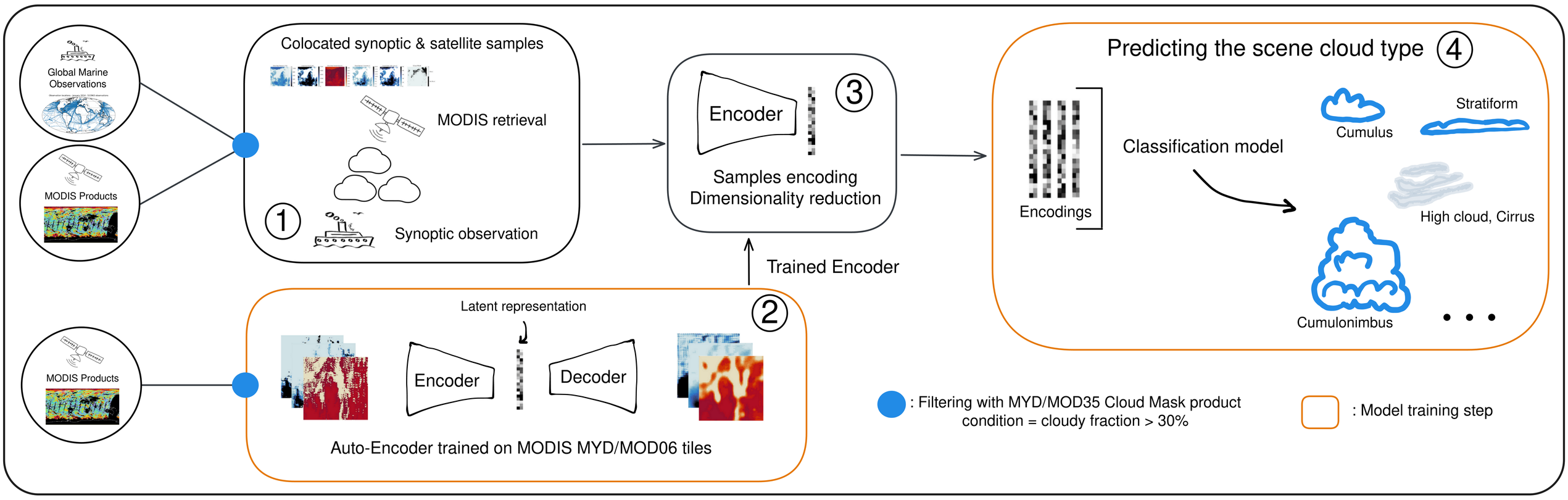
From MODIS cloud properties to cloud types using semi-supervised learning

Julien LENHARDT[1], Johannes QUAAS[1], Dino SEJDINOVIC[2]

[1] Institut für Meteorologie, Universität Leipzig, Leipzig, Germany

[2] Department of Statistics, University of Oxford, Oxford, United Kingdom



→ Clouds play a crucial role in the Earth's energy budget.
→ Different cloud types = very different radiative properties and interact in numerous ways with aerosols.
→ Aim is to classifiy cloud regimes objectively.

Method developed[1] based on colocated satellite retrievals of cloud properties and surface observations of cloud types.
→ square tile (128 x 128 km) of cloud properties around the cloud type retrieval.

Issue of the low number of colocated samples and their high-dimensionality
→ semi-supervised encoding step using an AE[2] to extract meaningful representations from the cloud scenes.

Data

Labels = Global marine meteorological observations[3], Cloud type observations.
Input features = MODIS Atmosphere L2 Cloud Product[4], cloud-top properties, cloud optical and microphysical properties.

Colocated dataset ~ 5 000 samples (year 2008, global)
MODIS tiles dataset ~ 600 000 samples (year 2016, global)

Data product	Description	Variables	Resolution
Global marine meteorological observations	Synoptic observation	Cloud type (30 categories)	Latitude/longitude coordinates 0.1° Hourly/daily observations
MODIS Atmosphere L2 Cloud Product (MOD06)	Cloud-top properties, cloud optical and microphysical properties	CTH (m) COT (a.u.) CWP ($g.m^{-2}$)	1-km resolution Daily overpass
MODIS Atmosphere L2 Cloud Mask Product (MOD35)	Cloud pixel flag	Cloud mask	1-km resolution Daily overpass

Methodology summary

1) Colocate surface synoptic observations and satellite retrievals.
2) Train AE model on MODIS cloud properties tiles.
3) Encode colocated cloud-scene samples using the trained encoder.
4) Train cloud classification model with encodings and cloud type labels.

Acronyms:
AE : Auto-Encoder
CTH : Cloud-top height
COT : Cloud optical thickness
CWP : Cloud water path
CTP : Cloud-top pressure
N_d : Droplet number concentration
LWP : Liquid water path

References:
[1] Lenhardt et al. (in prep)
[2] Hinton et al. (2006)
[3] Met Office (2006)
[4] Platnick et al. (2017)
[5] Kuma et al. (2023)
[6] Garcia et al. (2009, 2010)
[7] Rossow et al. (1991)
[8] Jungclaus et al. (2022)

Next steps and perspectives:
→ Evaluate cloud representation in global simulation outputs from ICON-ESM[8] (in collaboration with Daniel Klocke, Max Planck Institute for Meteorology)

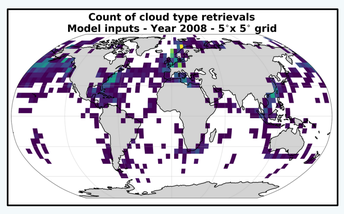
→ Investigate cloud adjustments to aerosol cloud interactions (e.g. N_d – LWP relationship) through the scope of the cloud classification developed here.

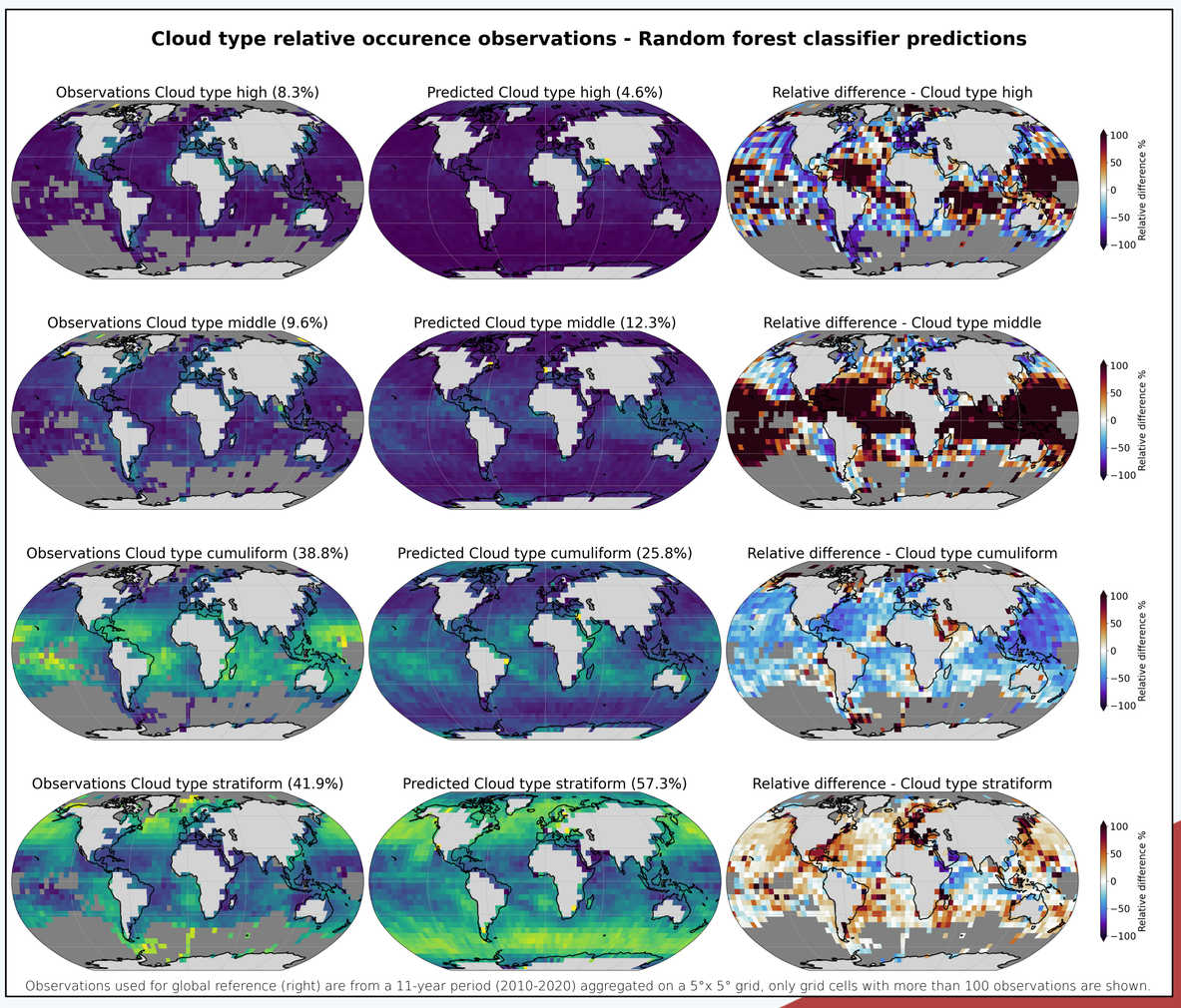
Results

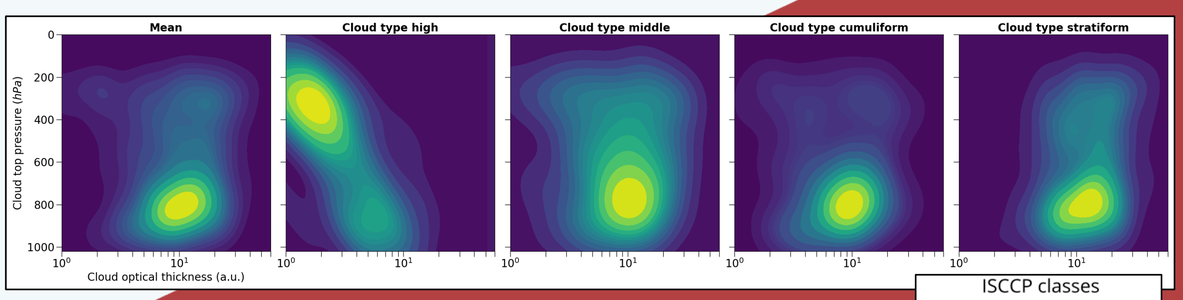
Class labels reworked into either 4 or 10 cloud groups[5].
4 groups → high, middle, cumuliform, stratiform (shown here).
10 groups → cirrus, cirrostratus, cirrocumulus, altostratus, altocumulus, cumulus, stratocumulus, stratus, cumulonimbus and cumulus + stratocumulus.

Still highly unbalanced dataset → over-sampling & IBA[6] of the geometric mean

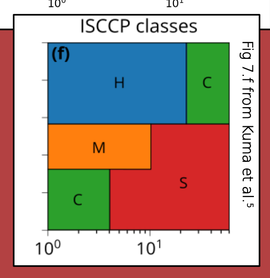
Generalization = one of the main objectives since spatial coverage decreases greatly during the colocation process.







Comparing to the ISCCP COT-CTP classes[7]
→ Separation between cumuli- and stratiform clouds not clear
→ Middle clouds covering a wide range of classes
→ High clouds cluster captured



Contact:
Julien LENHARDT
Institut für Meteorologie
Universität Leipzig
E-Mail : julien.lenhardt@uni-leipzig.de

QR Code to abstract

This project receives funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100.