EXPATS: EXploiting sPATiotemporal cloud patterns to advance severe Storms process understanding and detection

Claudia Acquistapace, **Paula Bigalke**, Daniele Corradini

Institute for Geophysics and Meteorology, University of Cologne, Cologne (DE)

follow us on socials

@EXPATS-ideas4s

UNIVERSITY OF COLOGNE

Deutscher Wetterdienst
Wetter und Klima aus einer Hand

Italia – Deutschland science-4-service network in weather and climate (IDEA-S4S)

collaborative research framework between **Italy** and **Germany** for advancing **weather** and **climate** research

Italia – Deutschland science-4-service network in weather and climate (IDEA-S4S)

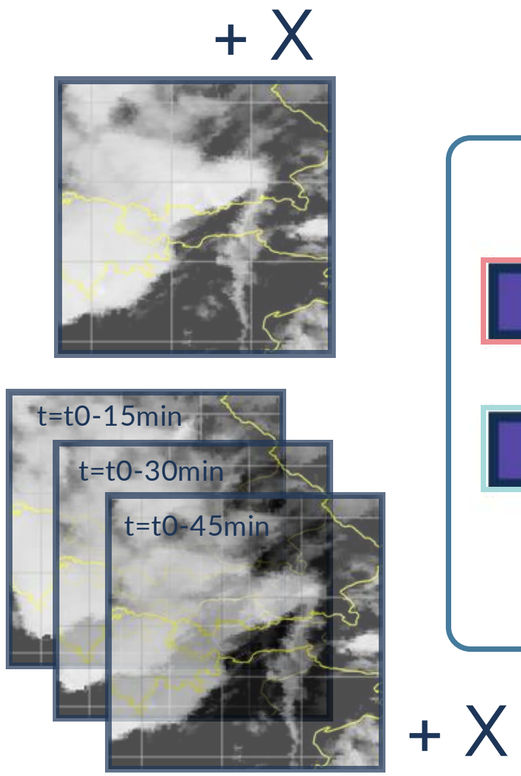
collaborative research framework between **Italy** and **Germany** for advancing **weather** and **climate** research

EXPATS

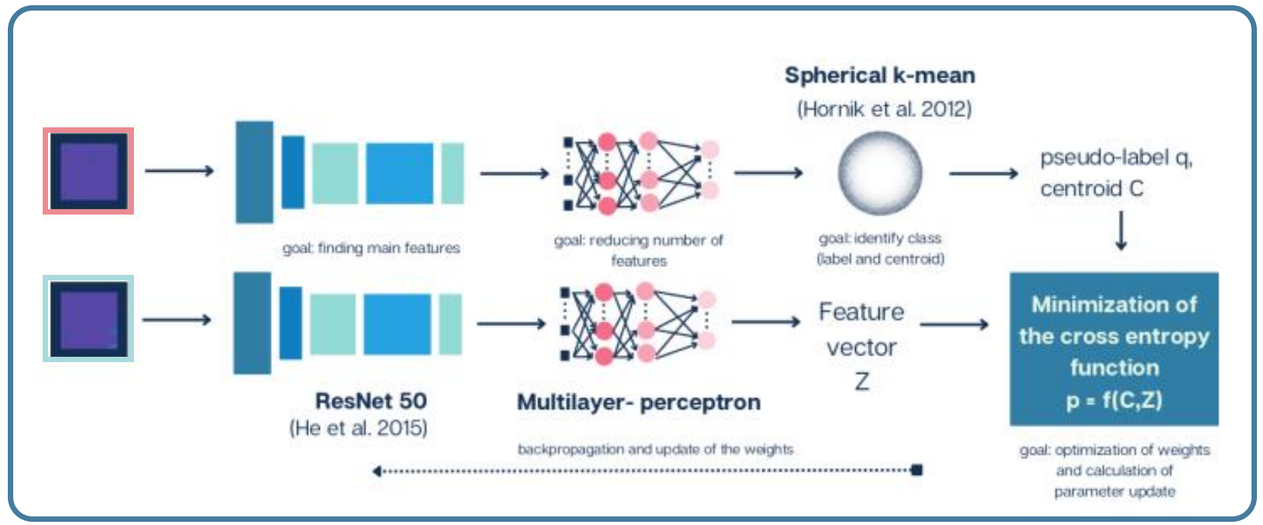
- Research group funded within IDEA-S4S
- Improve understanding of extreme events & hail by making use of large amounts of geostationary satellite imagery & advanced deep learning methods

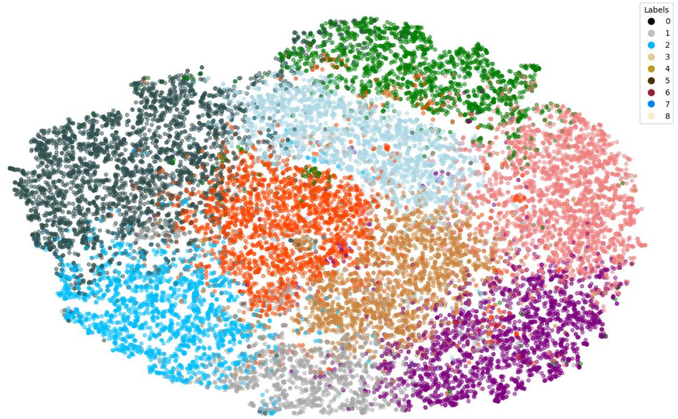
- EXPATS' work is based on EWC infrastructure. Access via national representative DWD.

Our main approach



based on Chatterjee et al. (2023)





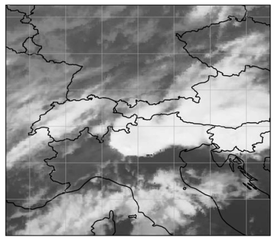
MSG **IR 10.8** micron as images or timeseries + other data sources

Train a self-supervised deep learning model (2 or 3D CNN or transformers)

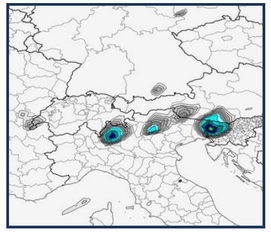
feature space clustered into k classes → analyse + adapt for further applications

Overall work flow

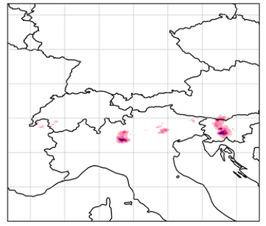
Download different kinds of data locally



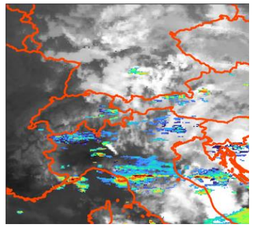
MSG/MTG radiances + CMSAF



MWCC-H hail



EUCLID lightning

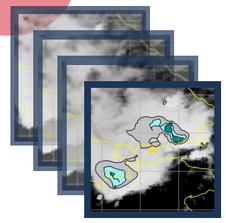


Radar composites

pre-processing:

- parallax correction
- matching of different datasets

Upload to EWC buckets



prepare data

- build time series
- crop to specific subdomain

save on VM

Train deep learning model on GPU

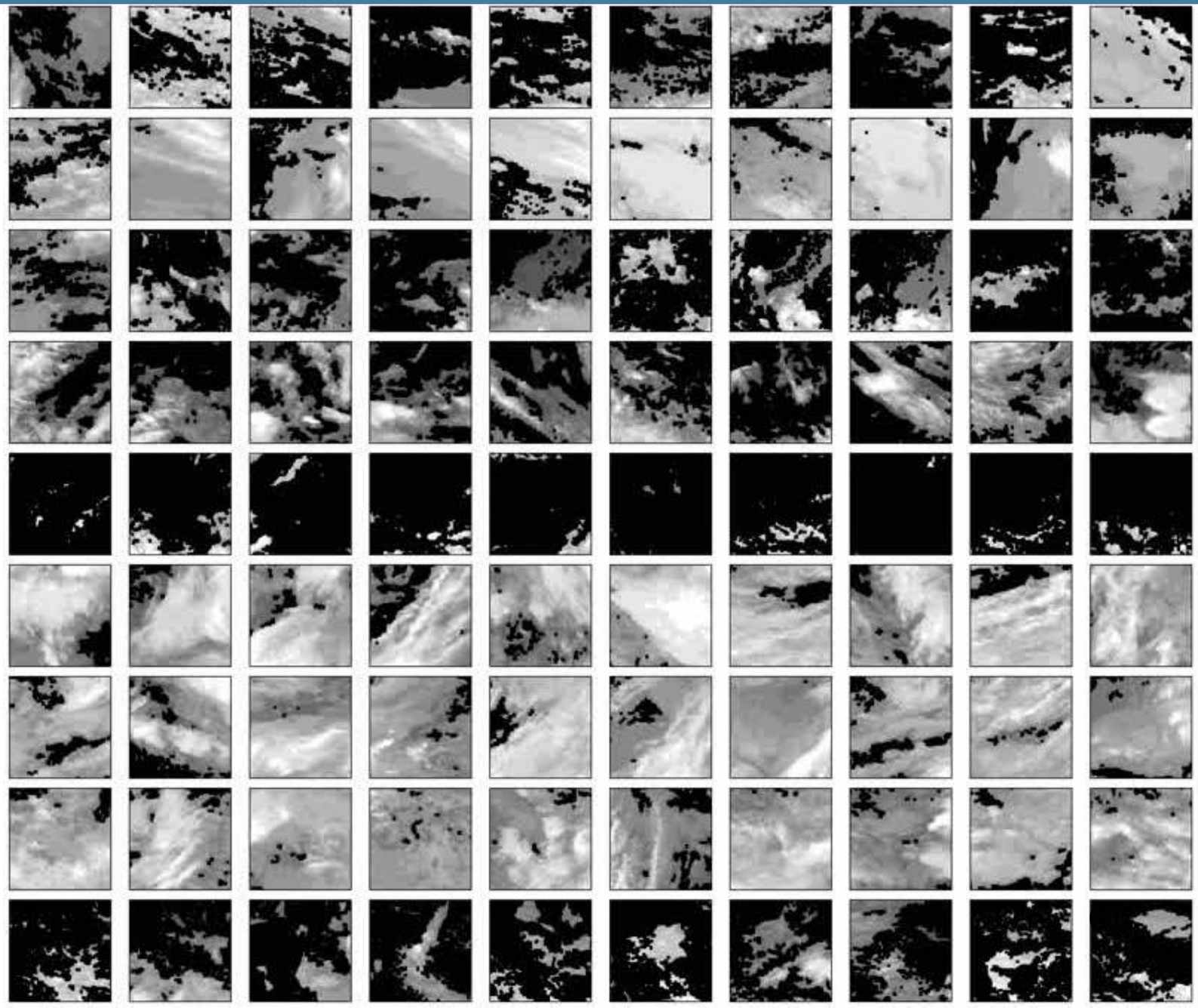
Post-processing and further analysis on VM or local

Main results: video classification works!

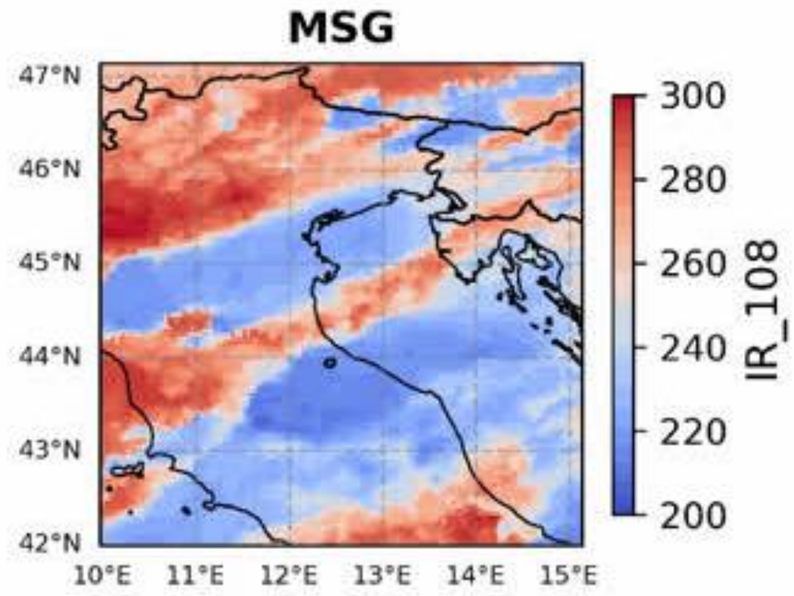
BROKEN: low to high cloud cover, broken clouds structures, low level clouds,

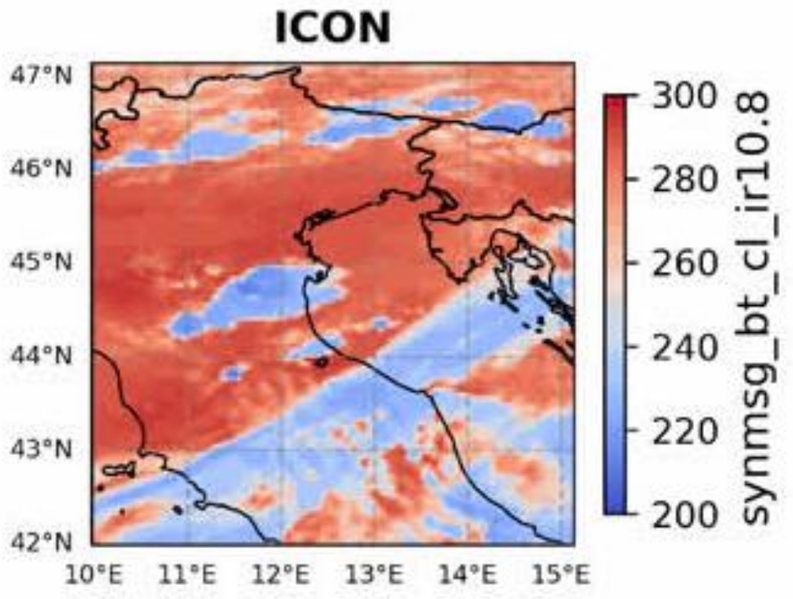
CONVECTIVE: broken clouds with average cloud cover, various cloud top heights, deep cores

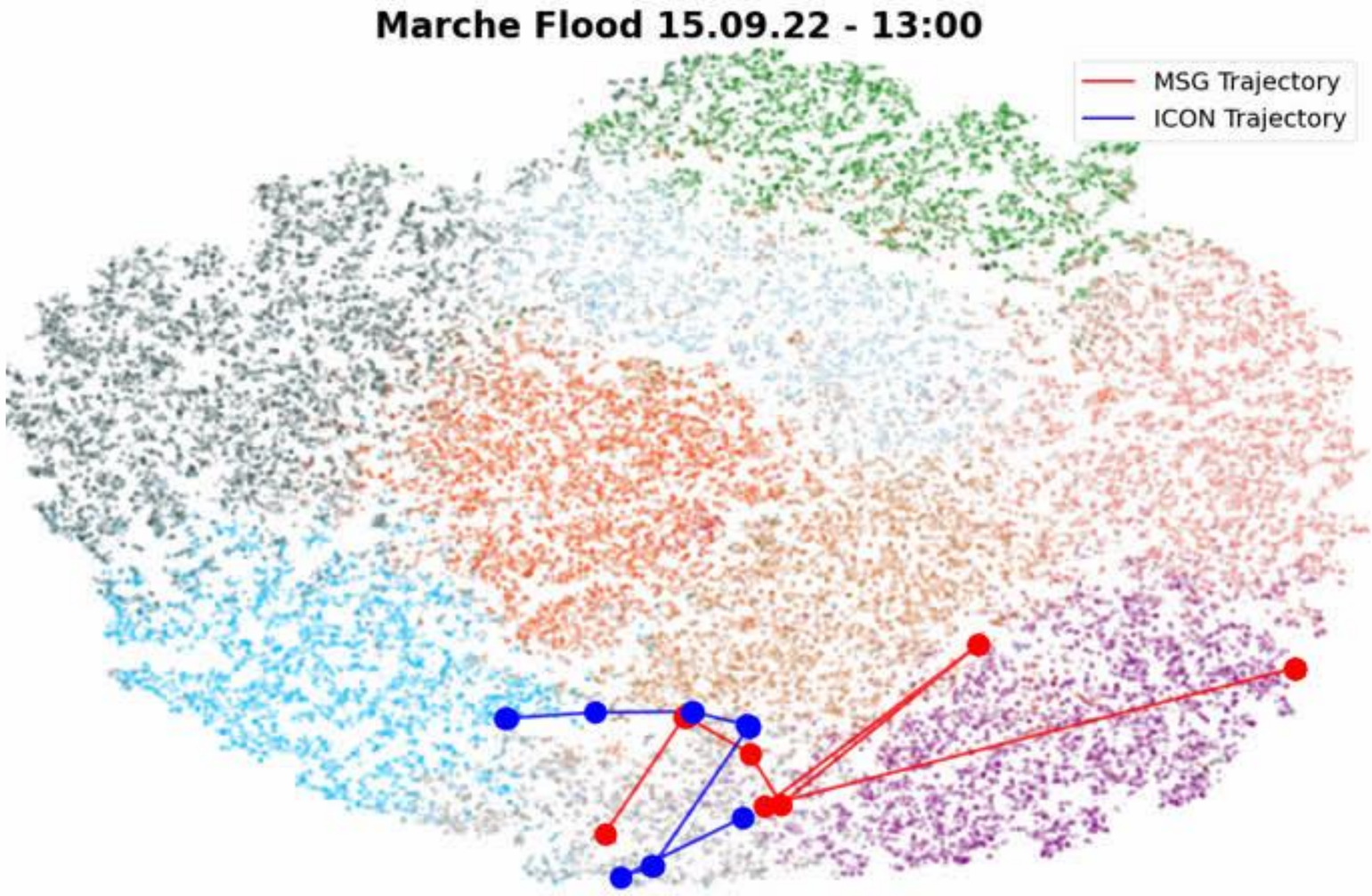
OVERCAST: high and homogeneous cloud cover, multilayer clouds, different motions



Main results: same framework in space for model evaluation

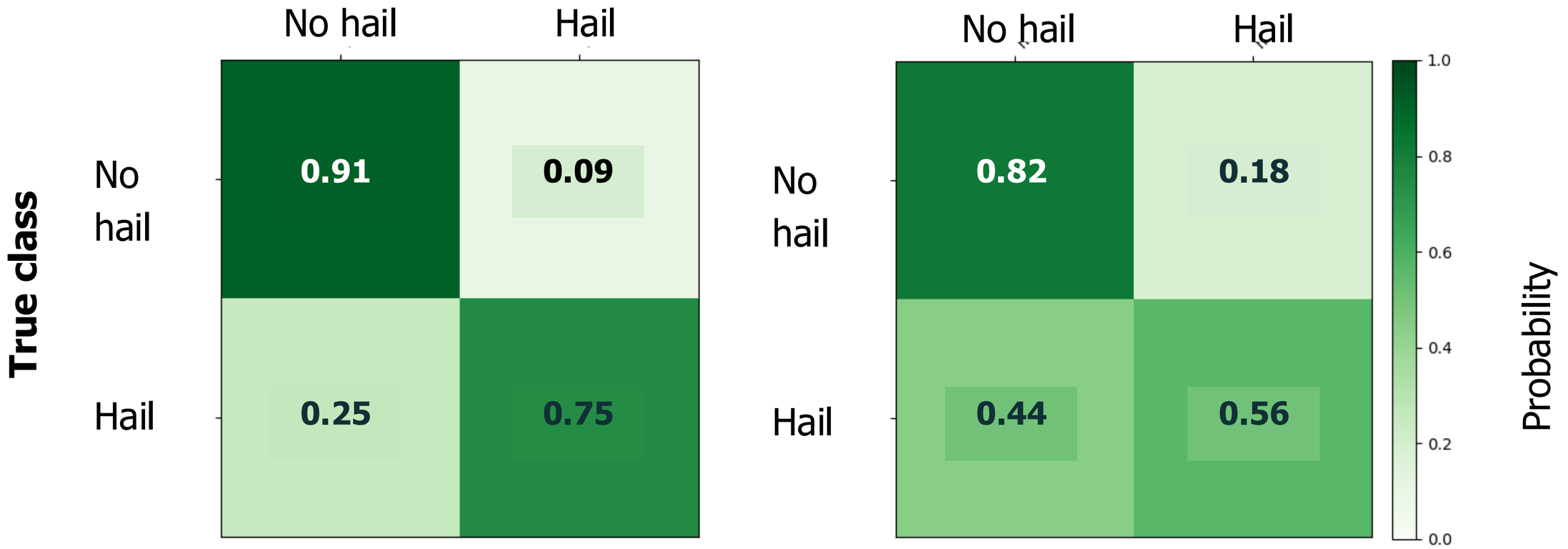






Main results: Work to identify hail patterns in progress

Supervised model performance vs logistic regression



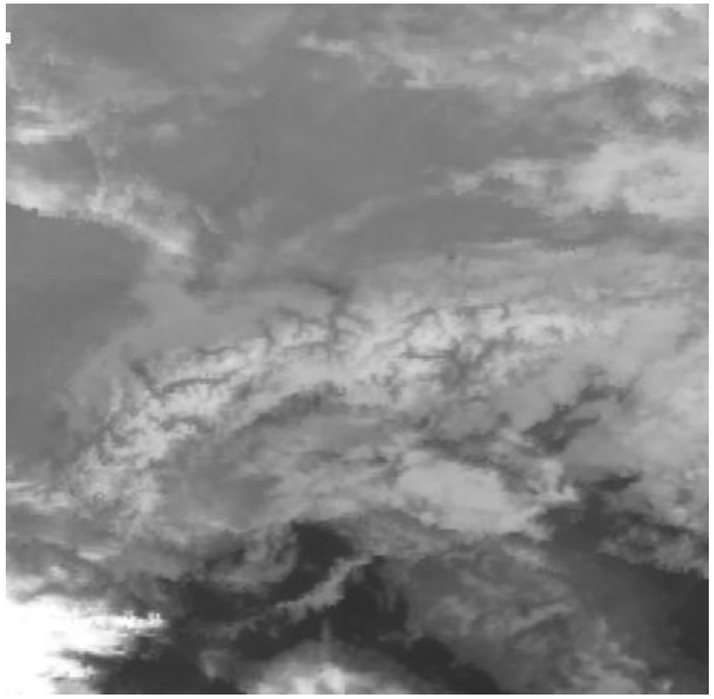
supervised CNN on test dataset (unseen data)

logistic regression (linear fit)

Data preparation and buckets

MSG

- 10.8 µm (future also others 6.2, 3.9)
- 2013-2024, Apr-Sept
- Over Alps (5-16E, 42-52N)
- 0.04° regular grid



Procedure:

- Download from EUMETSAT with data tailor
- Apply parallax correction using CTH
- Crop to our domain
- Upload to EWC data bucket

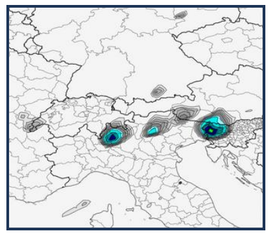
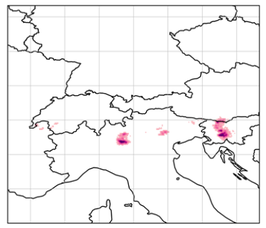
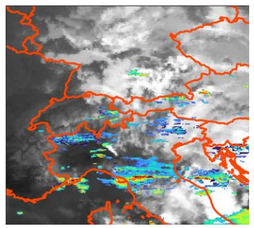
Feedback:

- MSG in available buckets are not corrected for parallax when made available with radar data.
- Some tutorial scripts for downloading EUMETSAT products (as it is available for OCA) would be nice
- Downloading MTG - bit difficult

Data preparation and buckets

Collecting **other data sources** on buckets for same domain (subdomains) and time period

- **Radar** data from DWD and ARPAE
- **EUCLID** lightning data
- Cloud properties derived from **CM-SAF** products
- **ERA 5**
- **MWCC-H hail probability**

Feedback:

- Great option to store in buckets
- Easy access from all VMs

Our suggestions / thoughts:

- Back up of data buckets would be appreciated
- Can we make our data buckets available with associated DOI?

What future for our buckets?

Good reasons to make our buckets (with our training dataset) **publicly accessible**:

- they can be **linked to the publications** in respect of FAIR principles
- they can be **shared with the community** & re-used for training other models, improving verification chances
- they can be **exploited for training students**.

Is this a possibility? What options are on the table?

Additional feedback

- Not too easy to settle in if you’re new to it
- **Onboarding**:
  - useful but also overwhelming at beginning
- **User support**:
  - Great!
  - Direct contact via Mail or chat
  - Quick responses
- **Knowledge base webpage**:
  - Is very appreciated!
  - Sometimes bit hard to find what you’re looking for

Thank you for your attention

Claudia Acquistapace, **Paula Bigalke**, Daniele Corradini

Institute for Geophysics and Meteorology, University of Cologne, Cologne (DE)

follow us on socials

@EXPATS-ideas4s

UNIVERSITY OF COLOGNE

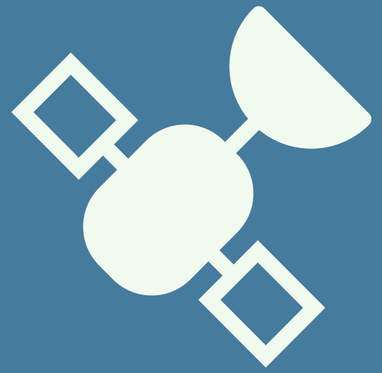
Deutscher Wetterdienst
Wetter und Klima aus einer Hand

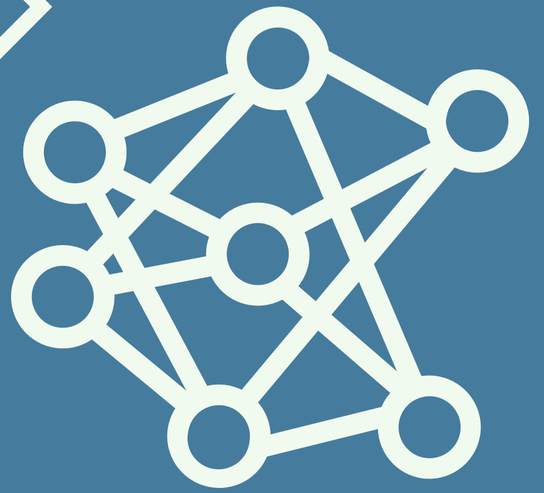
DWD

Back-up Slides

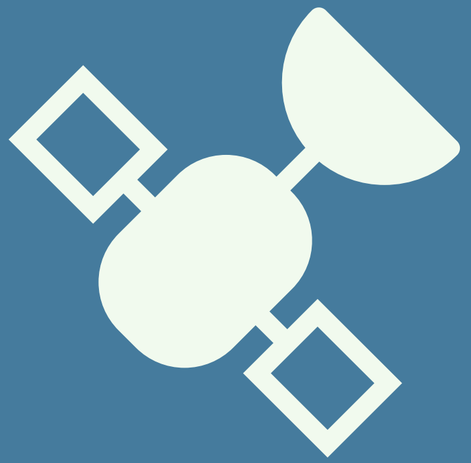
PhD Paula: Investigation of changes for hail storms over the Alps

PhD Daniele: Self-supervised learning using satellite imagery for evaluating cloud structures in ICON-GLORI



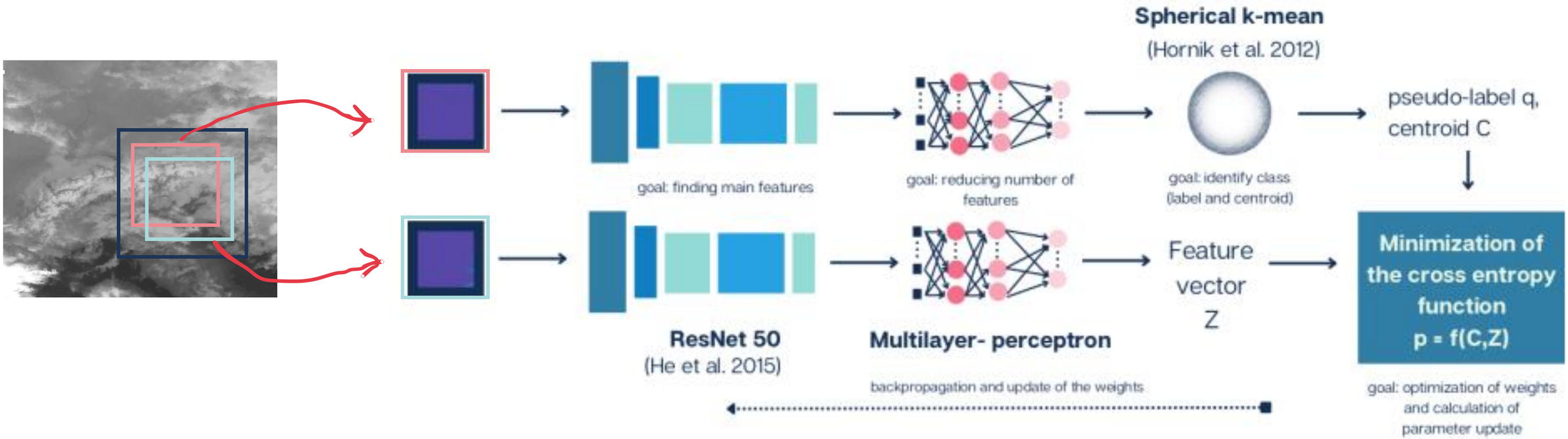


PI research: clustering of cloud spatiotemporal evolution regimes



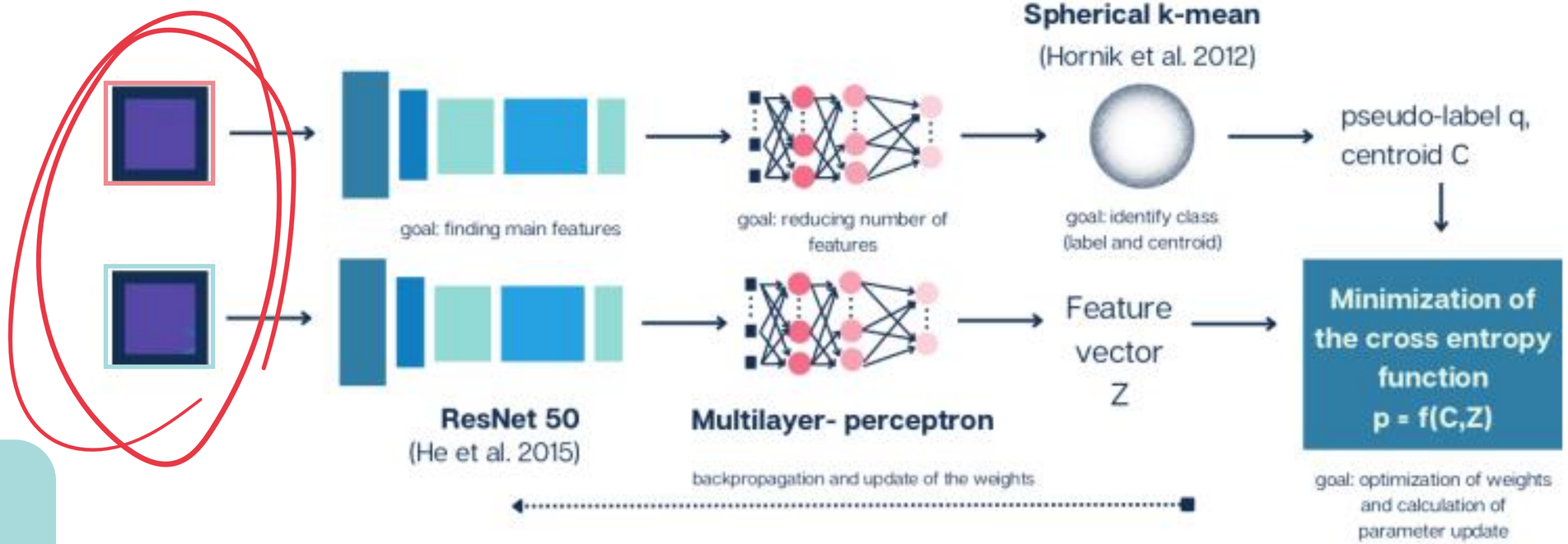
PhD Daniele: Self-supervised learning using satellite imagery for evaluating cloud structures in ICON-GLORI

Evaluating ICON cloud structures using ML



Evaluating ICON cloud structures using ML

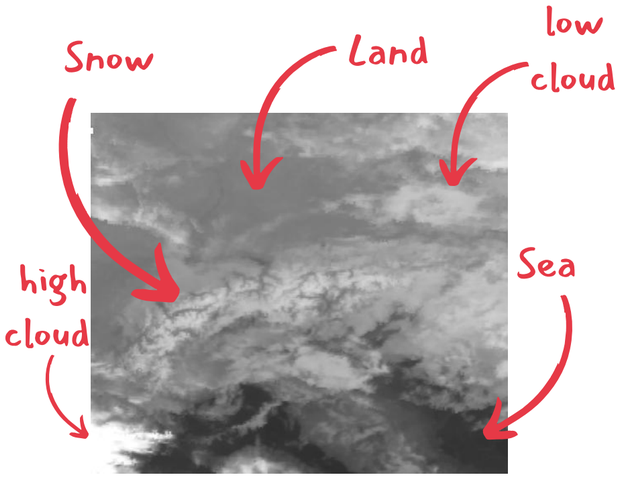
Adapt the self-supervised framework from Chatterjee et al. (2023)

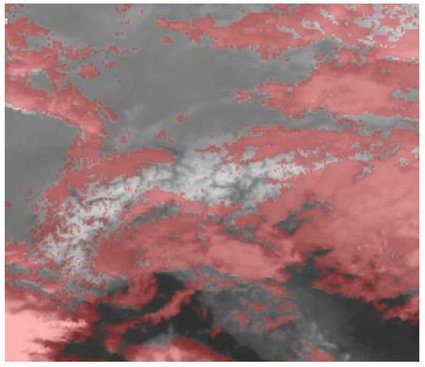


Assess the impact of using **IR channels,** compared to using cloud optical thickness.

Channel **IR 10.8** micron is provided from the model output and data availble also during nightime

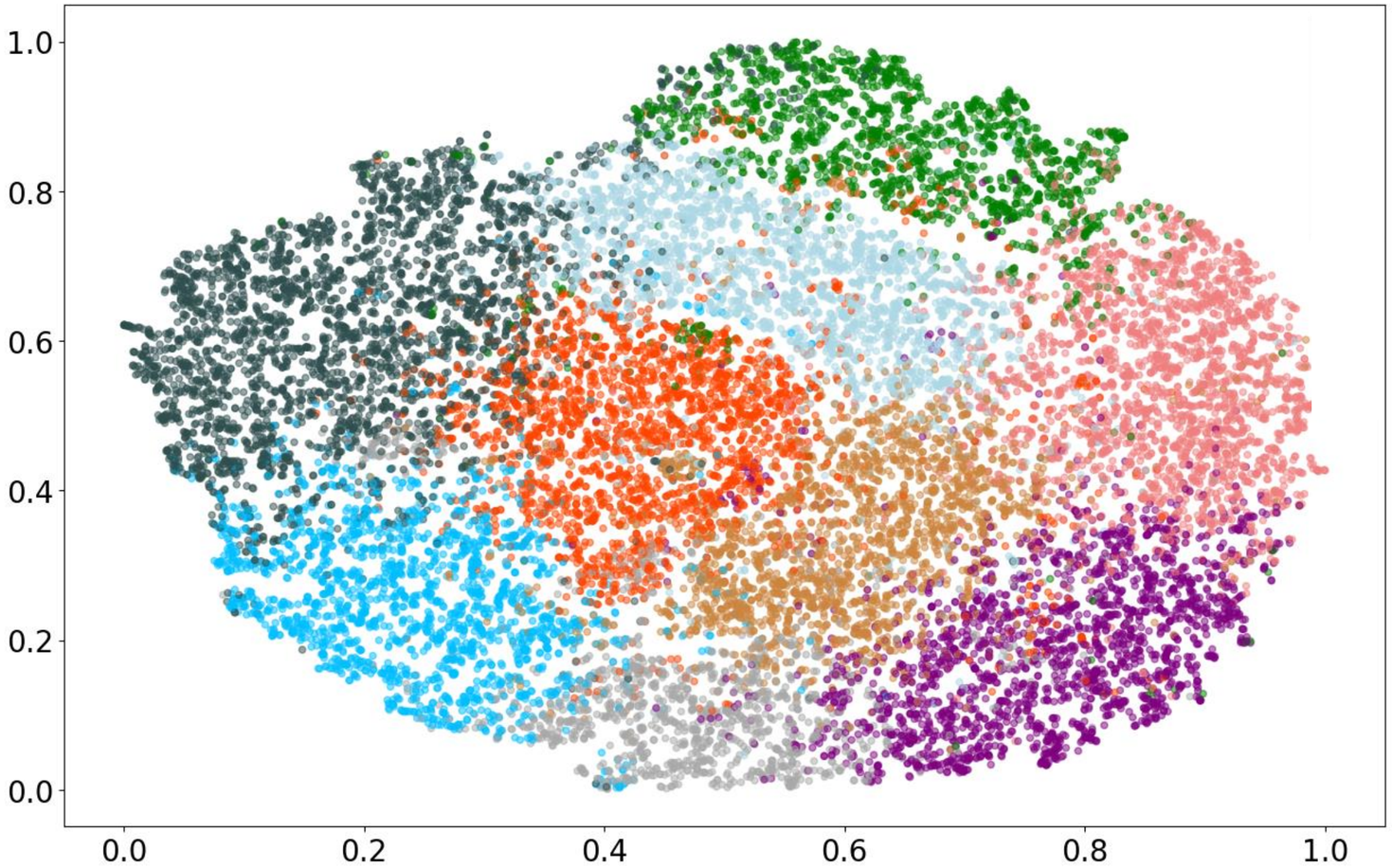
BUT

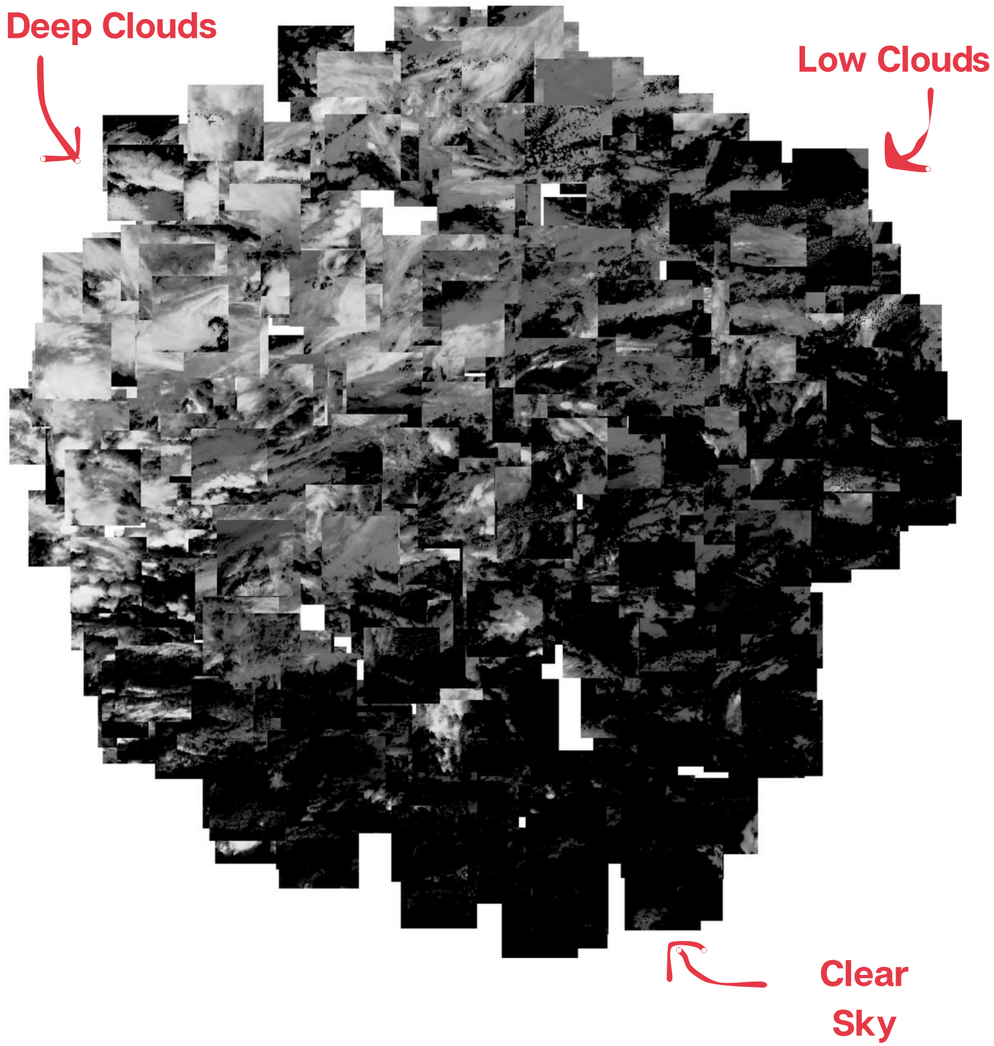




Cloud mask from CMSAF is applied to filter out ground signals

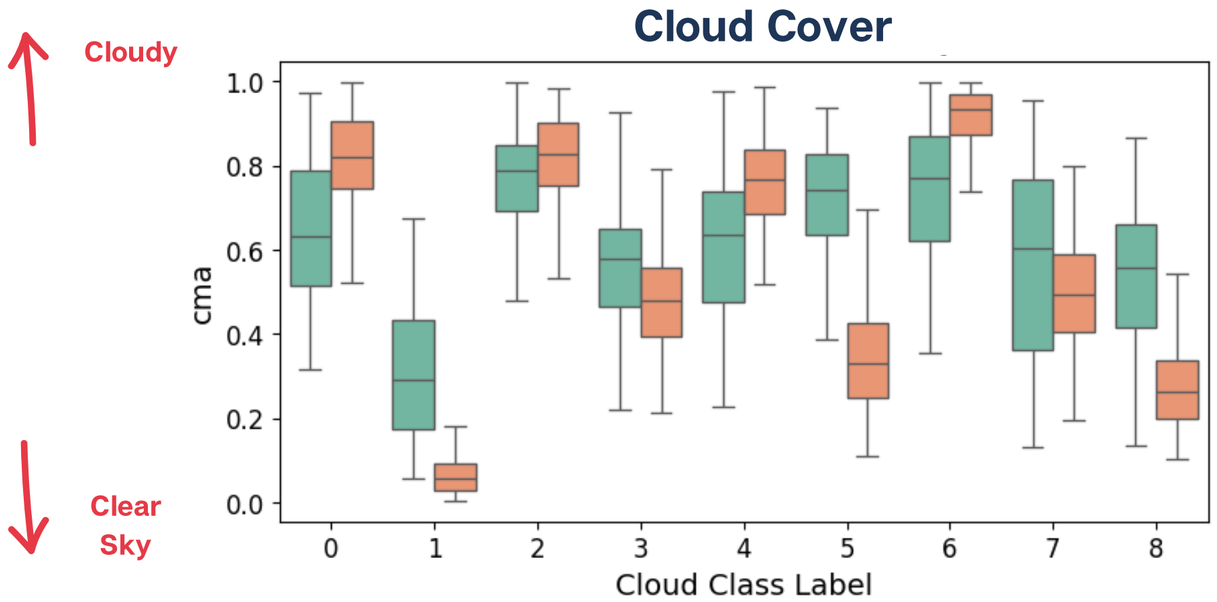
2D Embedding





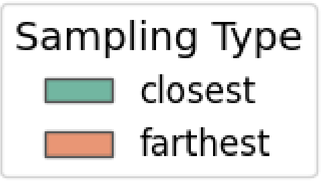
The 128-dimensional feature space is reduced to 2D using **tSNE** for visual inspection.

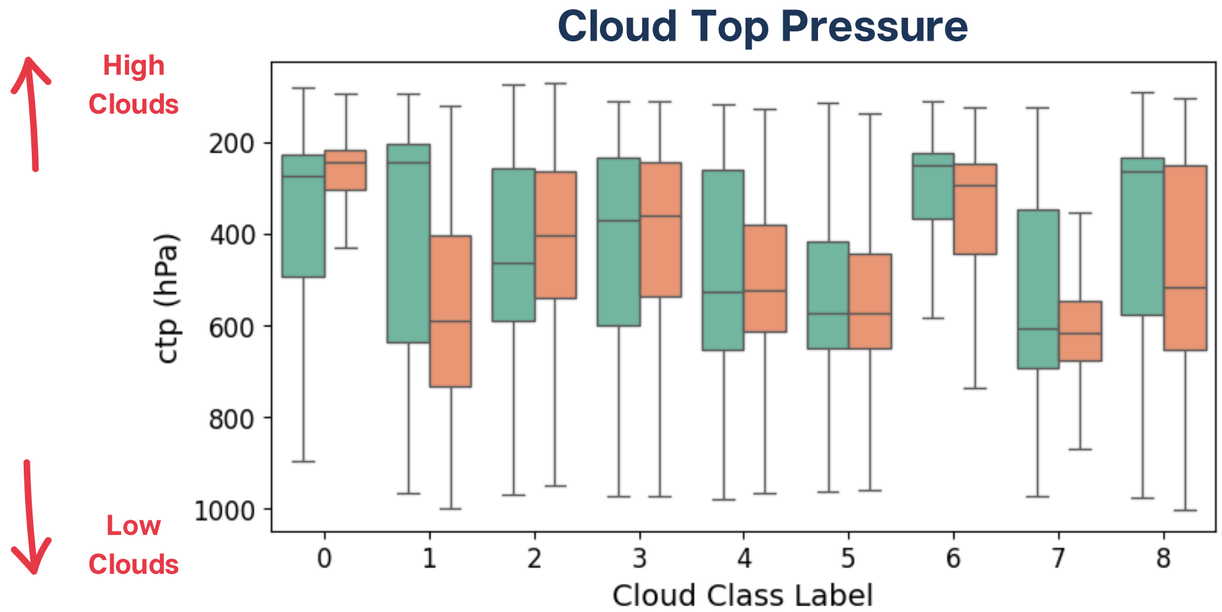
Intra- and inter-class variability

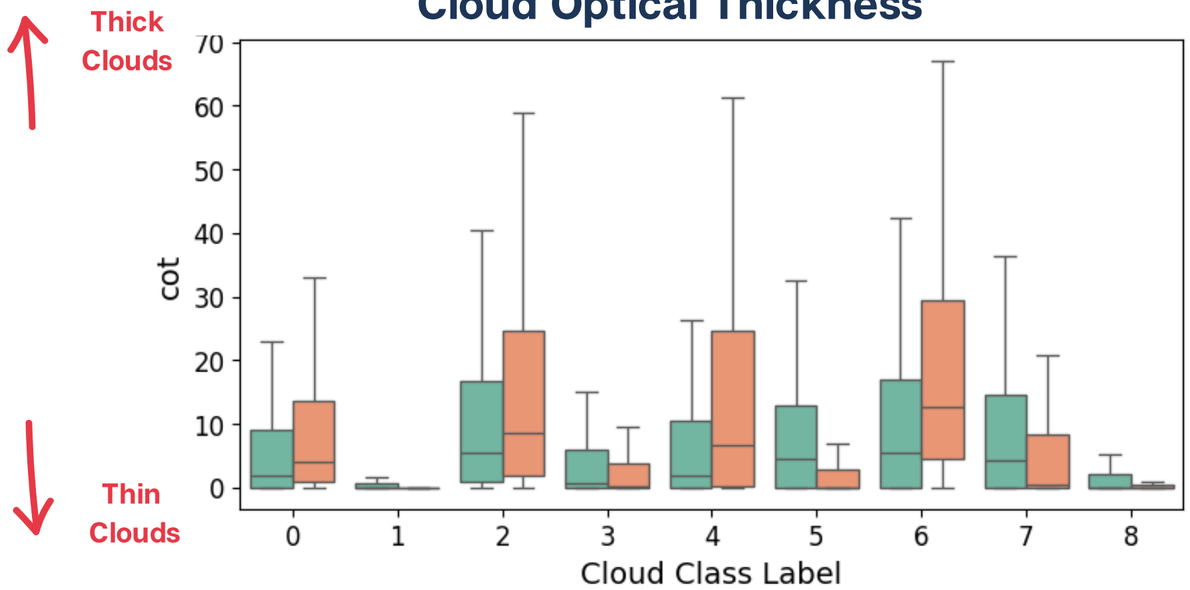


100 samples closest to and farthest from the cluster centroids.

Products from **CMSAF** data.

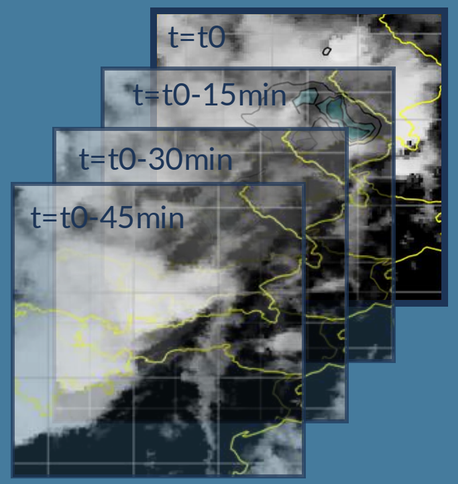






PhD Paula: Investigation of changes for hail storms over the Alps

My approach



large hail

Convolutional Neural Network

e.g. ResNet

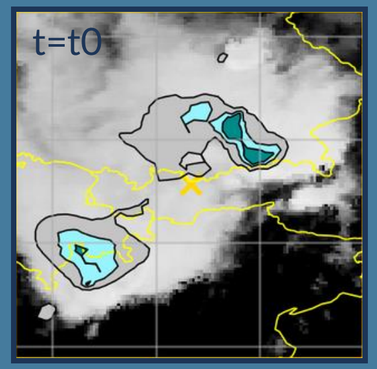
Extracts spatiotemporal features

Representation vector

Hail

No hail

This was HAIL



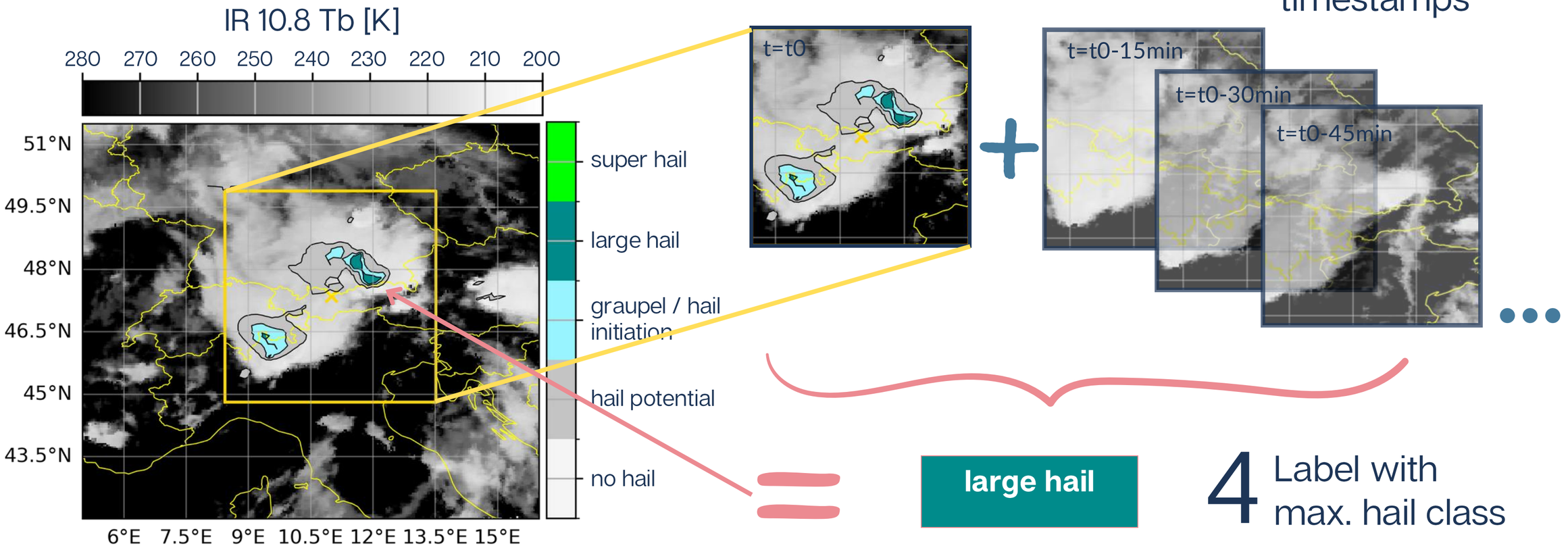
Truth based on PMW overpasses
(Laviola et al., 2020)

Constructing spatiotemporal dataset

1 Match PMW satellite overpass with MSG timestamp

2 Crop over overpass / hail area

3 Extend timeseries by previous MSG timestamps



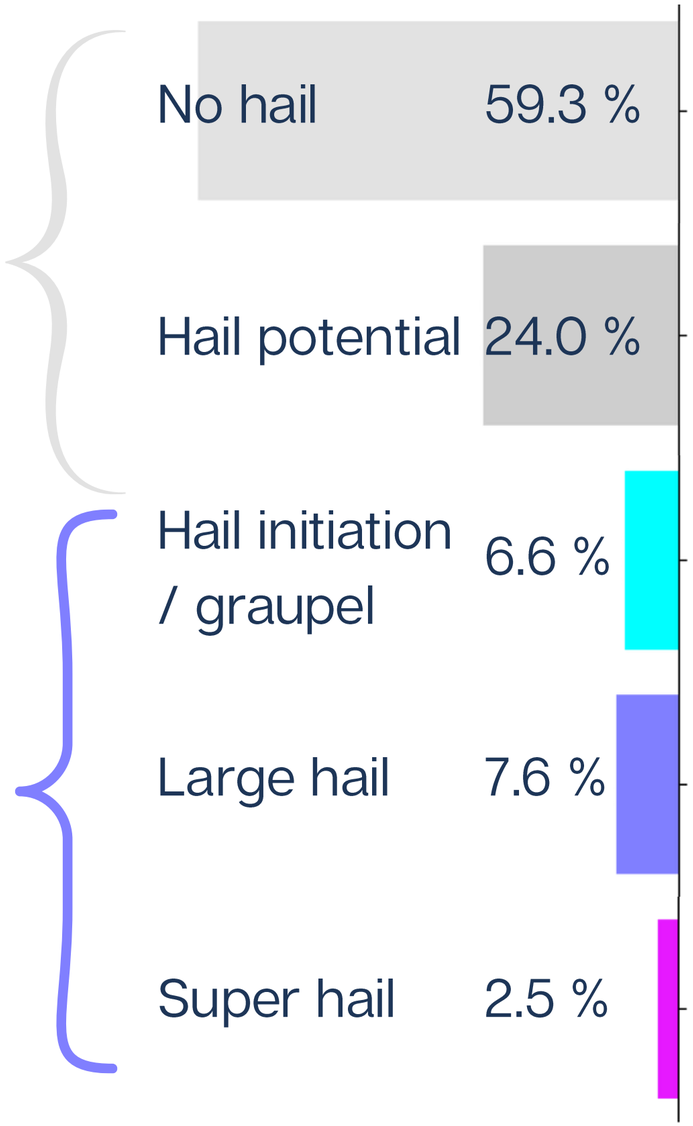
4 Label with max. hail class

Dataset for pre-study

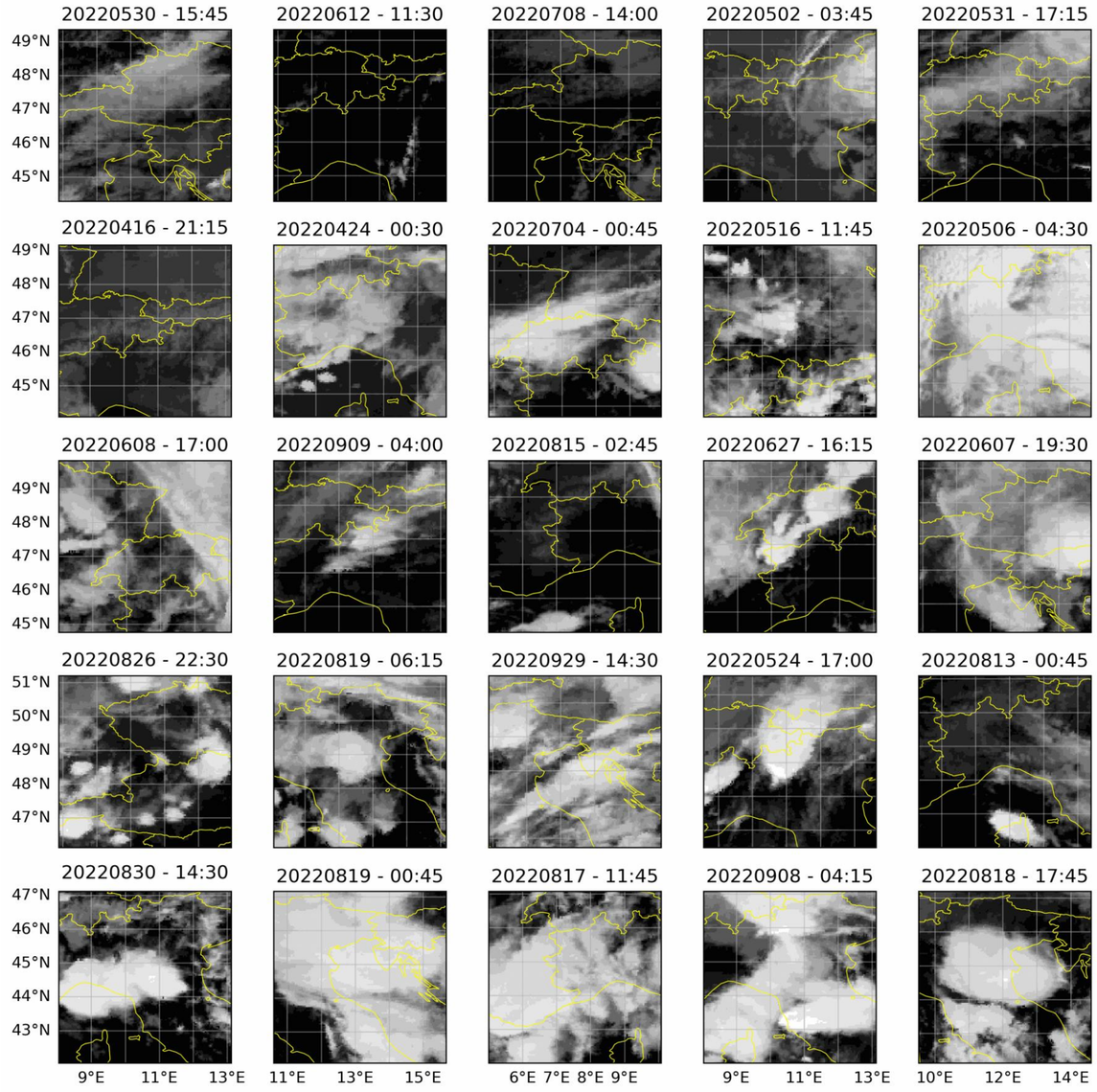
- 2022, Apr-Sep
- Total number of timeseries: 1244

No hail
83.3 %

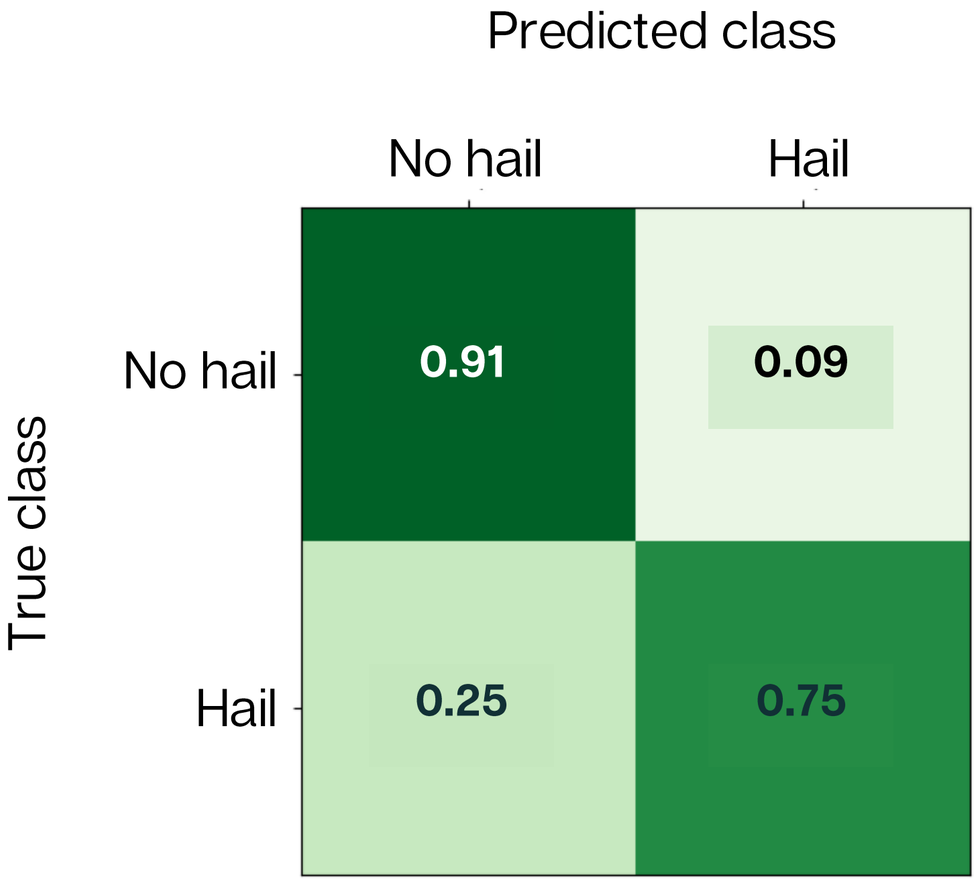
Hail
16.7 %



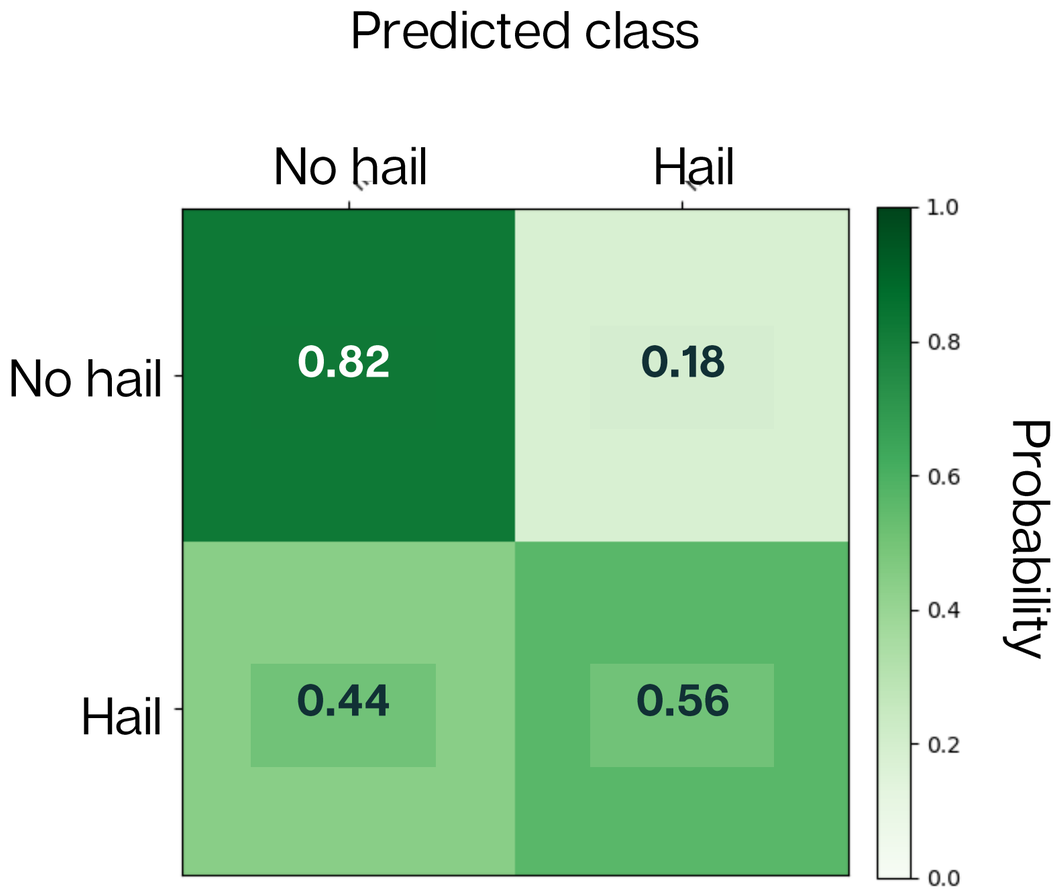
Examples



Model Performance



on test dataset
(unseen data)



logistic regression
(linear fit)

Daniele’s and my thoughts from before

- Summary on our activities (Sat data, SSL) taken from above talks
  - using GPU to train ML models (1 gpu for each VM is not enough!)
  - We might need more also for students project work.
  - Having personal GPUs is good in intense work phase, but maybe when not using gpus all the time is more beneficial to share nodes among different tenants?
- We heavily use buckets to store and share data - this is great! (point to improve, bucket backup!)
- Easily share data VMs belonging to the same provision. Buckets aren’t the best while using data from ML (data are needed locally)
- Buckets already provide with some sat data (MSG) but the processing is left to the users (e.g. parallax correction is not there).
- How to deal with data publication when papers are accepted? FAIR principles