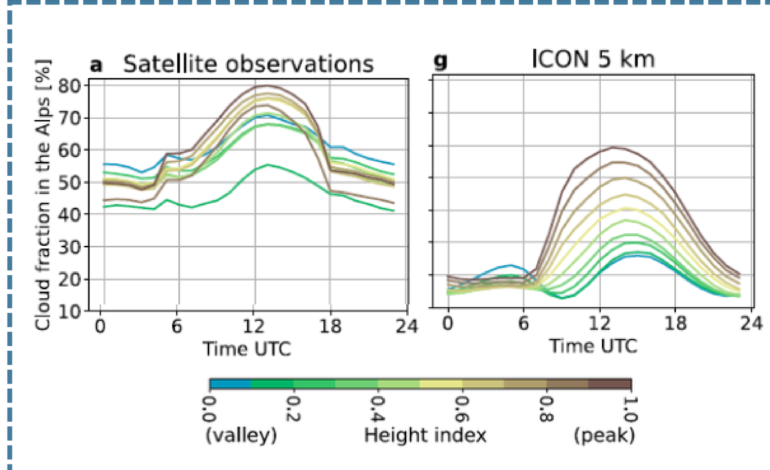
Evaluating Cloud Spatial Structures in 500-m ICON Simulations during the TEAMx Campaign Using Self-Supervised Learning

Daniele Corradini (1), Paula Bigalke (1), Bernhard Pospichal (1), Elsa Cattani (2), Annika Oertel (3), Claudia Acquistapace (1,4)

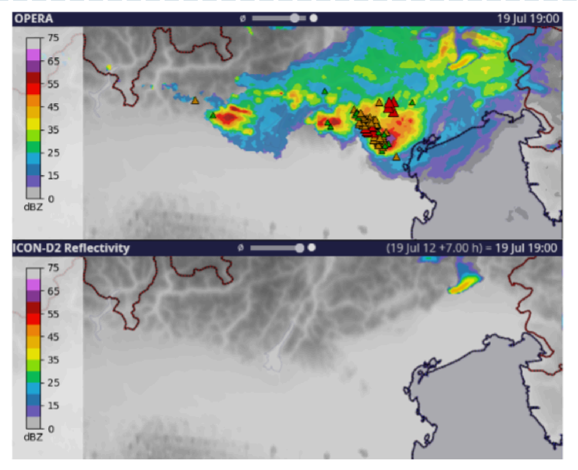
1. (former) University of Cologne, Germany; 2. CNR-ISAC, Bologna, Italy; 3. IMKTRO, Karlsruhe, Germany; 4. (now) Department of Geoscience, University of Padua, Italy.

1. MOTIVATION



Cloud cover is still not well captured by **ICON** over **orography**

Fig. 1: Cloud cover diurnal cycle from satellite observations and ICON 5 km over the Alps [Poujol et al. 2025].



Models still fail to capture precipitation over **orography**

Fig. 2: Severe hailstorm on 19 July 2023 in Northern Italy: OPERA radar reflectivity vs. ICON-D2 forecast. [Fischer et al., 2024]

2. GOAL

Compare cloud structures seen by geostationary **satellite** with the ones simulated by **ICON** model in the Alpine region.

3. DATA

Why Meteosat Second Generation (MSG)?

Good Coverage over the Alps
Spatio-temporal resolution
Long dataset

Why 10.8 µm Infrared channel ?

Night-time availability
Height proxy
Model compatibility

Training set - years: 2013-2020 **~140k samples**

Data Processing:

- **Parallax correction**
- Interpolation into a **regular grid** (0.04°x0.04°)
- Application of a (corrected) **cloud mask**

1 Random Crop every 15 min

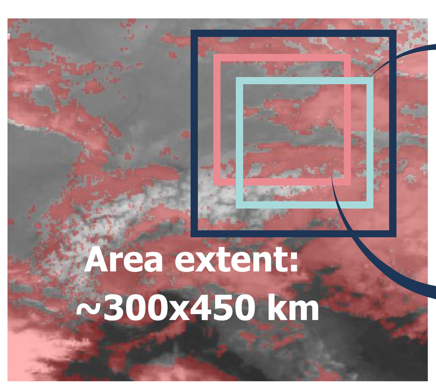


Fig. 3: IR 10.8 channel image with CLAAS3 cloud mask (red) superimposed. Crops are 100x100 pixels.

4. DEEP LEARNING METHOD

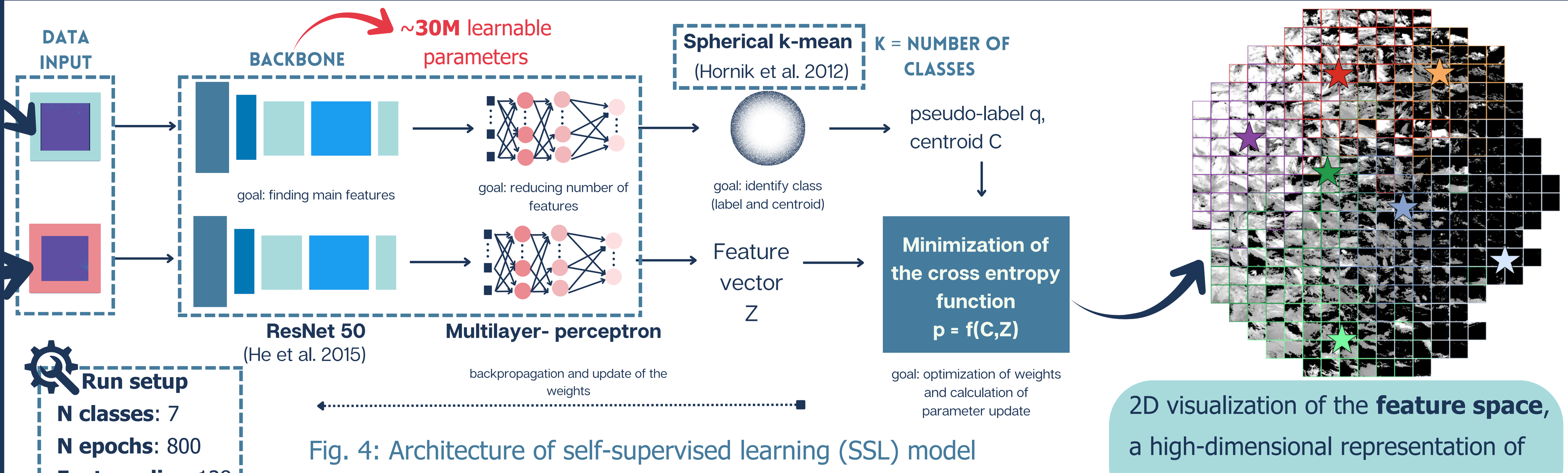


Fig. 4: Architecture of self-supervised learning (SSL) model adopted in Chatterjee et al., (2023).

2D visualization of the **feature space**, a high-dimensional representation of satellite image semantic properties.

5. CHARACTERIZATION

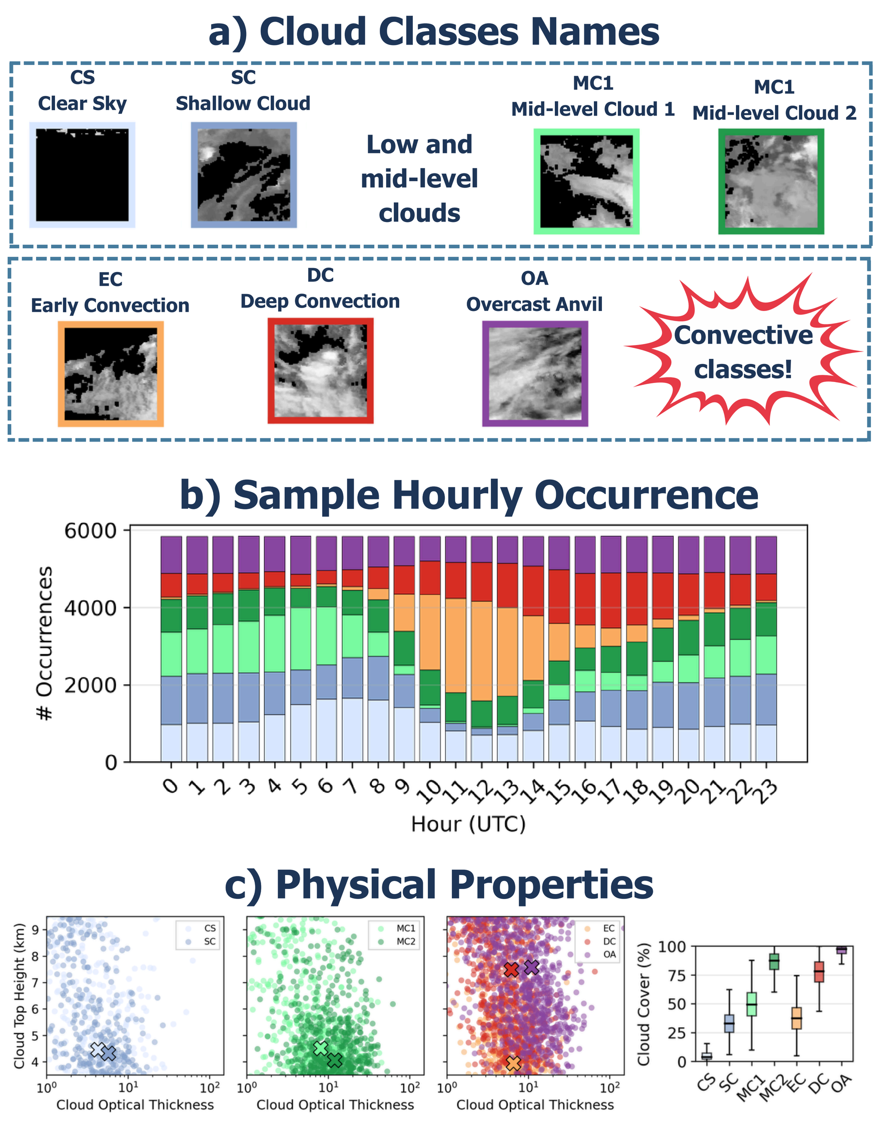
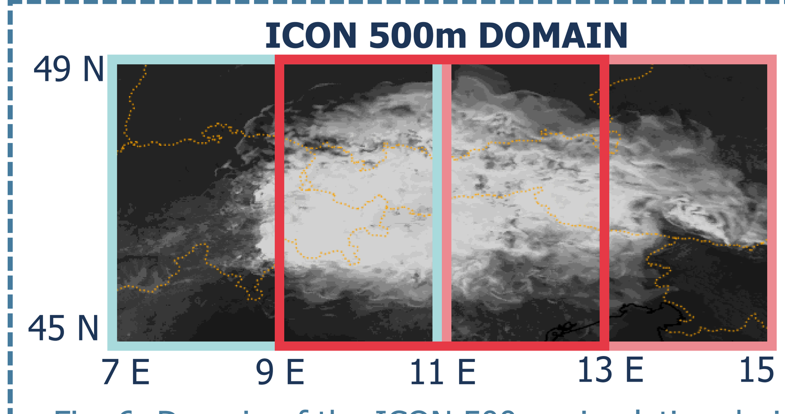


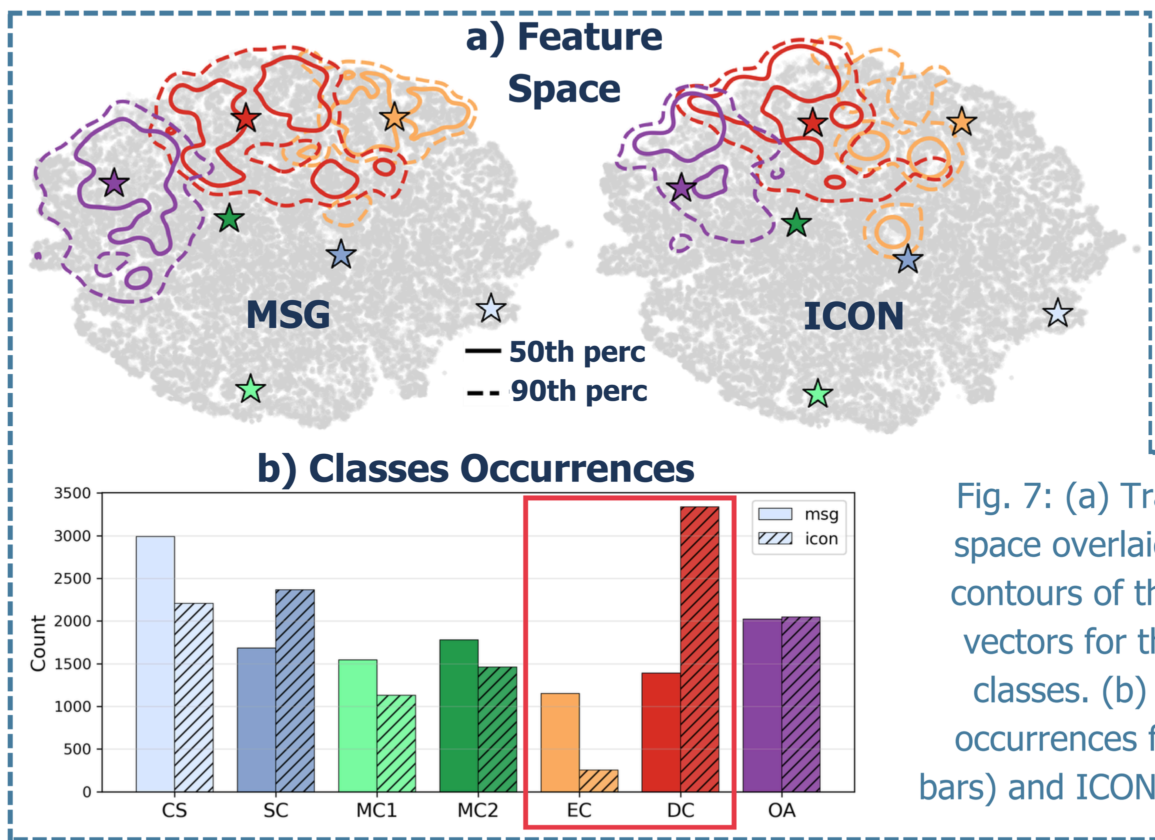
Fig. 5: (a) Class names and acronyms, with examples of image crops closest to the centroid of each classes. (b) Hourly occurrence counts of the classes. (c) Physical characterization of the classes: scatterplot of cloud-top height versus cloud optical thickness using the 1000 nearest samples to each class centroid (mean shown as cross), and boxplots of the cloud-cover distribution. All variables are derived from the CLAAS-3 dataset of the CMSAF.

6. PRELIMINARY ANALYSIS



Apr-Sept 2025
hourly output
ICON init 00
Tot Test Samples
MSG: 12558
ICON: 12810

Fig. 6: Domain of the ICON 500 m simulation during the summer period of the TEAMx campaign, shown with an example of the simulated IR 10.8 µm channel. The domain is divided into three regions: west (blue), centre (red), and east (pink). To maintain consistency with the training dataset, each region covers a 4° × 4° area, with image crops of 100 × 100 pixels. Same crops are derived for both MSG and ICON data.

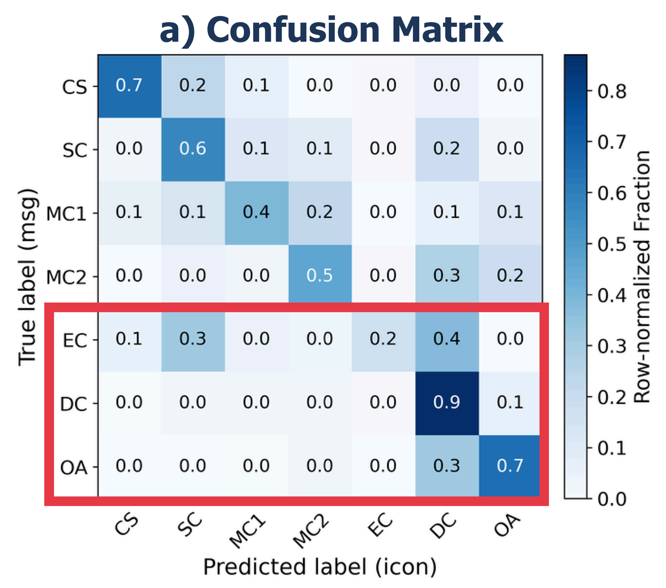


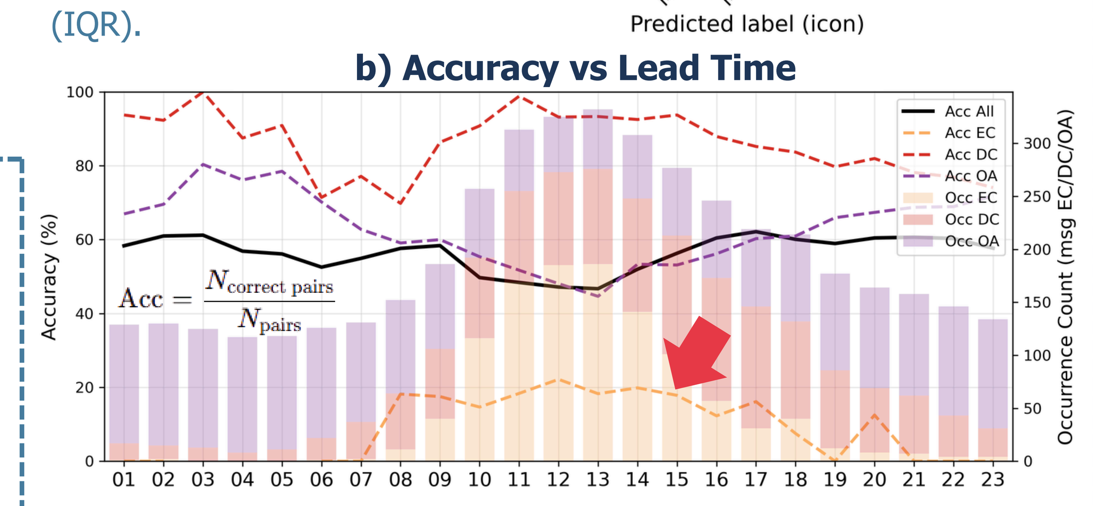
ICON underestimates EC clouds in favor of DC clouds, which is also the class that shows the lowest Accuracy and Distance scores

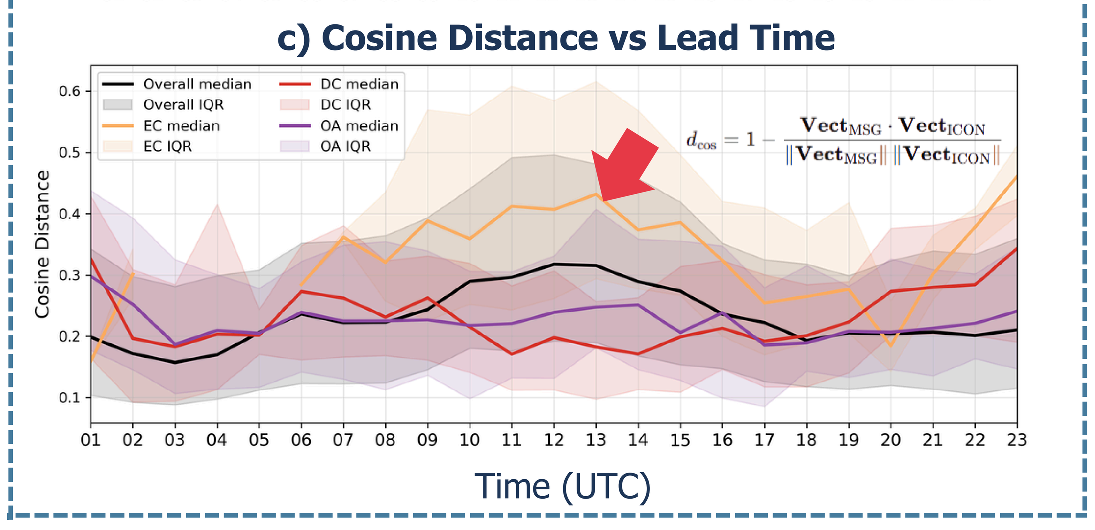
Fig. 7: (a) Training feature space overlaid with density contours of the test feature vectors for the convective classes. (b) Overall class occurrences for MSG (filled bars) and ICON (hatched bars).

ICON Representation Performace

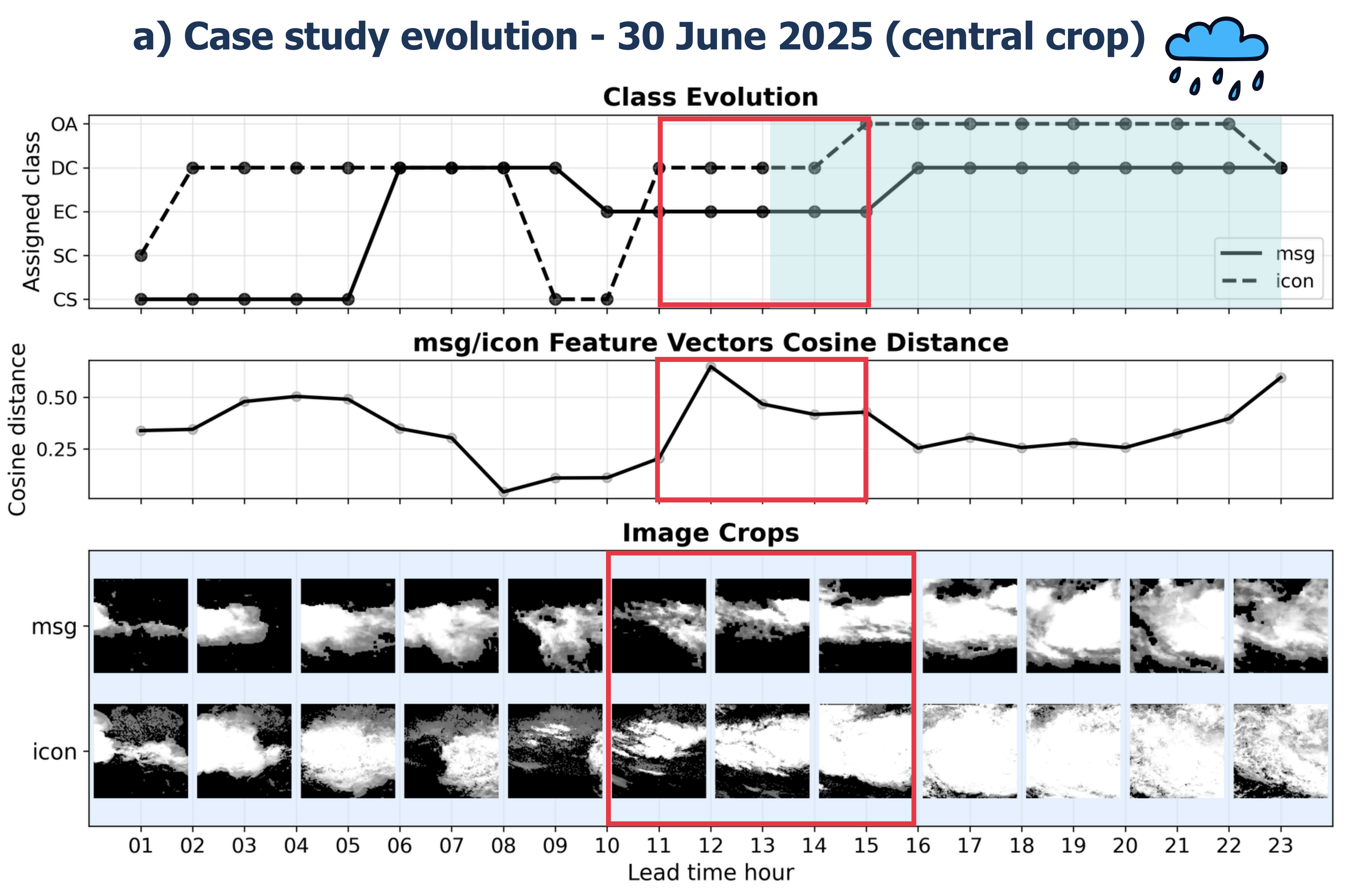
Fig. 8: (a) Confusion matrix for the test set classification, with MSG samples as ground truth and ICON samples as predictions, matched by datetime and domain region. (b) Classification accuracy as a function of forecast lead time for all classes and convective classes; MSG convective class occurrences are shown in the background. (c) Same as (b), but using cosine distance, showing median and interquartile range (IQR).





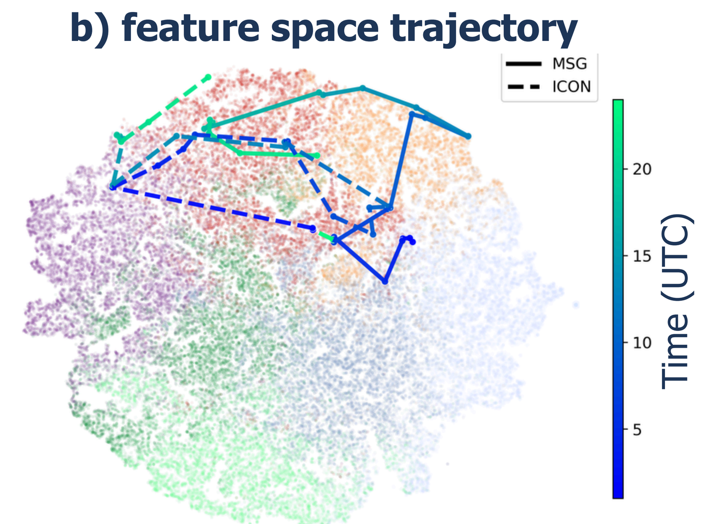


7. CASE STUDY ANALYSIS



CASE STUDY DESCRIPTION

- southwesterly moisture advection
- large CAPE / small CIN
- deep convection with precipitation after 13 UTC
- weak synoptic forcing / thermal winds



During the stage which precedes the start of precipitation, ICON tends to produce bigger cells compared to MSG.

Fig. 9: (a) Case study of the convective classes on 30 June 2025. The evolution over forecast lead time is shown through class assignment for MSG (solid) and ICON (dotted). Shaded area indicates precipitation was recorded (top). Cosine distance between MSG and ICON feature vectors (middle). Examples of image crops every 3 hours (bottom). (b) Same case study as in (a), but the evolution is represented as a trajectory in the 2D feature space, with lead time color-coded.

8. SUMMARY

- **Trained SSL model on 10.8 µm IR to classify clouds; identified three convective classes (EC, DC, OA).**
- **EC class is underrepresented while DC class is overrepresented;**
- **ICON simulations do not correctly represent convection initiation phase. Case study shows early, oversized convective cells.**

Future work:

- **improve SSL framework,**
- **use higher-res satellite data (MTG),**
- **analyze convective transitions,**
- **include TEAMx ground observations.**

dcorrad1@uni-koeln.de

References