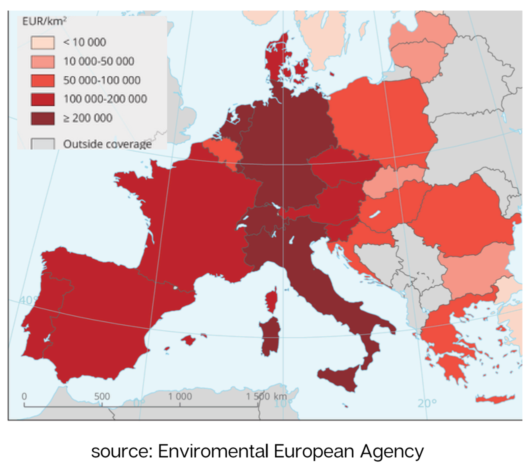
Self-supervised learning using satellite imagery for evaluating cloud structures in ICON-GLORI

Daniele Corradini, Claudia Acquistapace, Paula Bigalke
Institute of Geophysics and Meteorology, University of Cologne, Germany

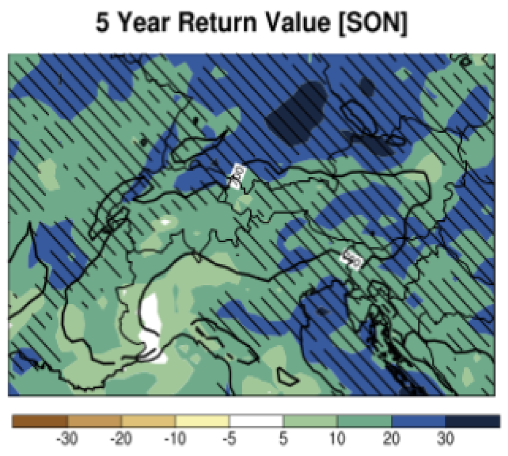
Deutscher Wetterdienst
Wetter und Klima aus einer Hand
DWD
UNIVERSITÄT ZU KÖLN

1. MOTIVATION



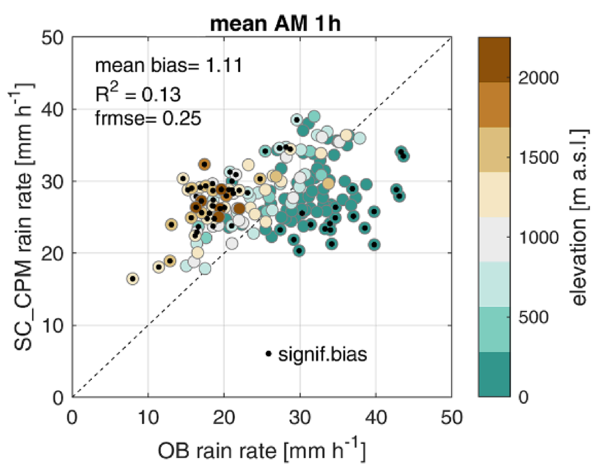
Extreme weather events are a serious **public safety hazard**

Fig. 1: Total economic loss per sq. km caused by weather-related extreme events in Europe (1980-2020) [Daniell et al. (2016)]



Severe **storms** are expected to intensify as **climate change** progresses.

Fig. 2: Projected changes (%) in 5-year return value of 1-day precipitation event during fall season (Gobiet et al., 2014).



Precipitation modelling is biased due to the **orography**.

Fig. 3: Comparison of observed and simulated annual rain rate maxima at 1 h duration (Dallan et al., 2023)

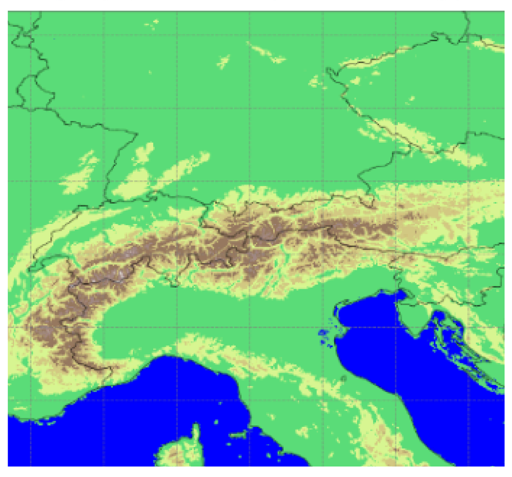
2. GOAL

Develop a **self-supervised** machine learning framework for evaluating **ICON-GLORI** model **cloud** representations, especially of severe thunderstorms, by leveraging the spatial features as seen from the **geostationary satellites MSG.**

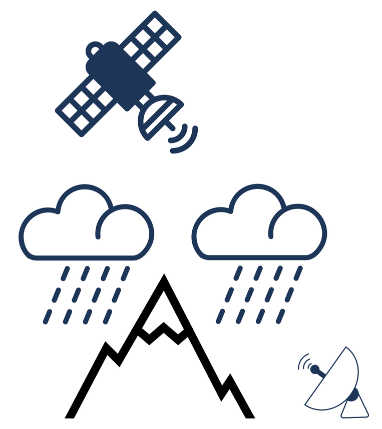
GLORI — Global to Regional ICON Digital Twin

Global-to-regional short-range **high resolution** Digital Twin, configurable, on-demand, based on the prediction capability of the **ICON** earth system model.

STUDY AREA: THE ALPS!



WHY MSG? GOOD COVERAGE AND RESOLUTION



3. MACHINE LEARNING METHOD

Self-supervised learning is a form of **unsupervised** learning that does not require manually created labels but generates **pseudolabels** from the data themselves.

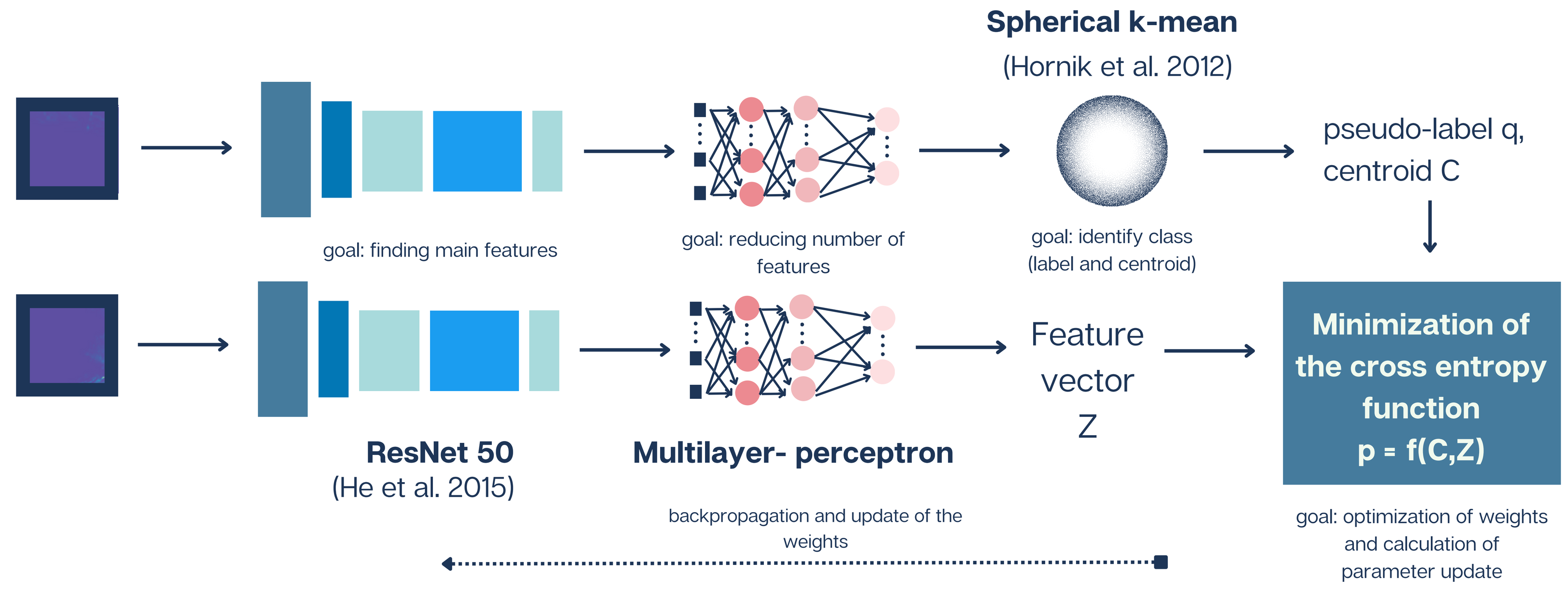
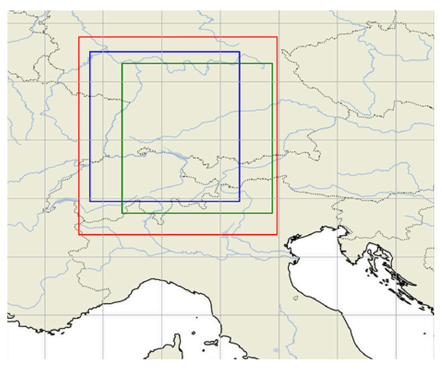


Fig. 4: Architecture of self-supervised ML model adopted in Chatterjee et al., (2023).

4. DATA PREPARATION

We collected a dataset of MSG observations that spans from **2023 to 2013** (April to September). The data were **parallax corrected** and interpolated into a **regular grid** (0.04°x0.04°). We use **Infrared channel 10.8 µm** (IR 10.8), a proxy for cloud top height.

Size: big crops: **128x128 pixels**; subscrops: **96x96 pixels**
96 pixels = 3.84° ~ **300-400 km** (depending on coordinates)
It still needs to be tuned to optimally capture **mesoscale convective systems.**



The ML model processes images, and when converting brightness temperatures to RGB, **color scale normalization** must be optimized. For the first runs, we normalized the colorscale between 200 K and 300 K.

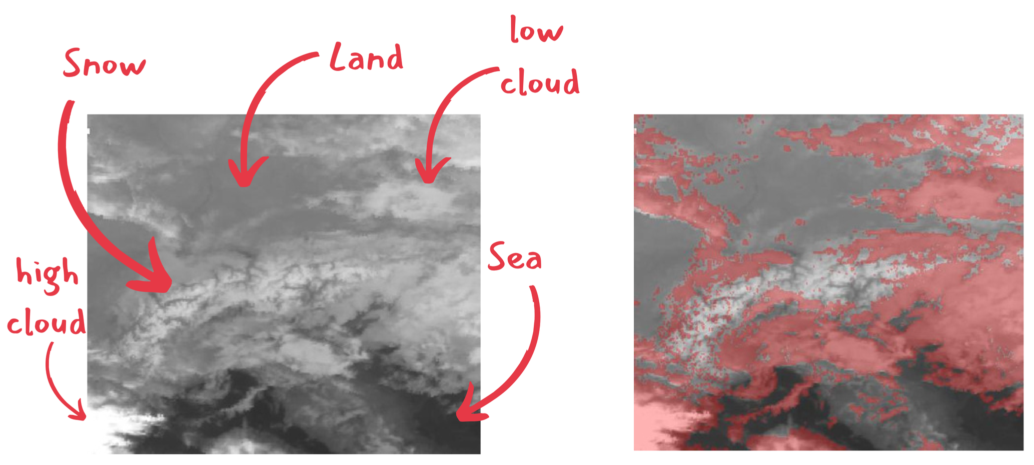


Fig. 5: Left: IR 10.8 channel images highlighting different signals. Right: Same image with CMSAF cloud mask (red) superimposed.

The IR 10.8 channel reflects varying surface emissivities (e.g., snow, land, sea). Using the CMSAF **cloud mask** helps filter out non-cloud signals.

5. CLOUD CLUSTERING

The **feature space** consists of multidimensional vectors of image semantic properties from multi-year satellite crops. To inspect it, dimensionality must be reduced to 2D. Clusters are characterized using physical quantities, like **cloud properties** from CMSAF.

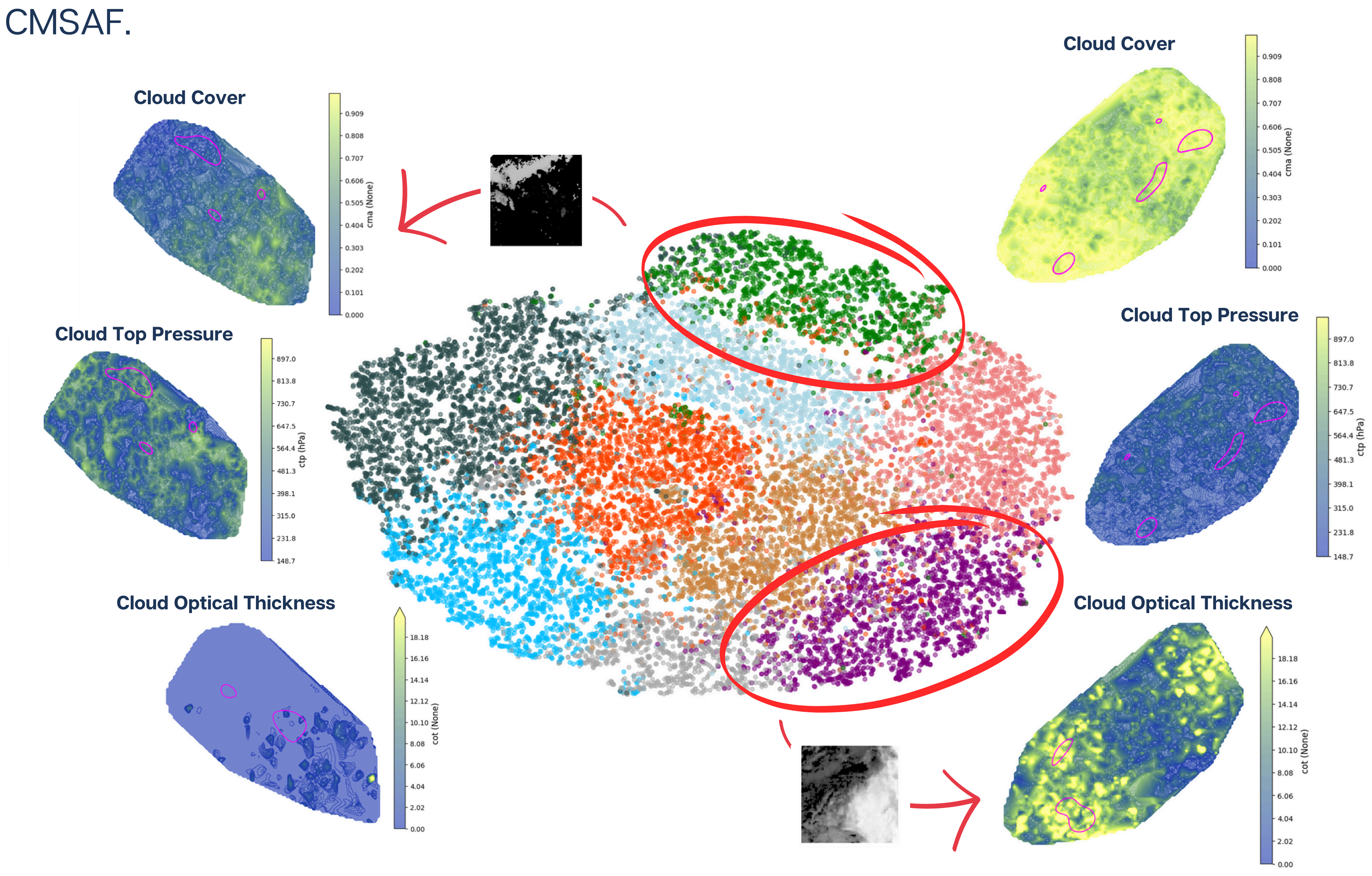
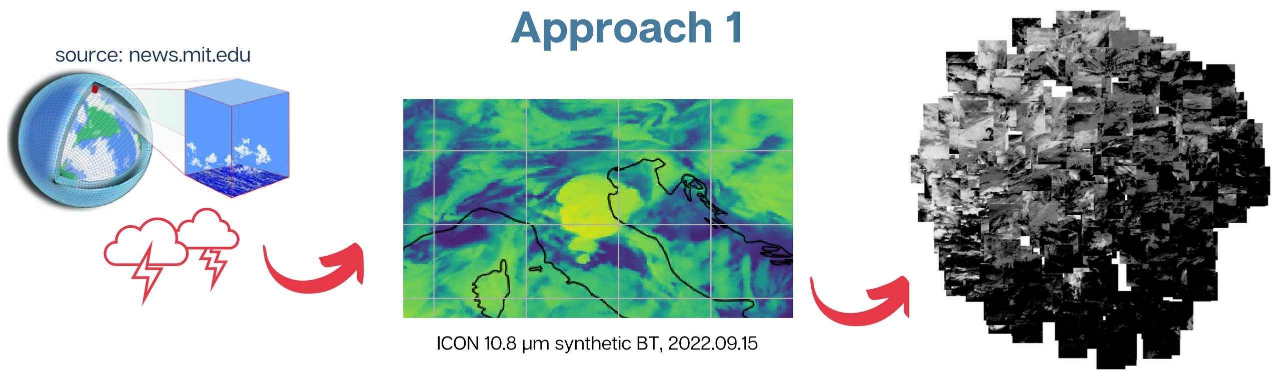


Fig. 6: 2D feature space representation via t-SNE, based on ML model training with IR 10.8 crops (including cloud mask) across 9 classes. Physical characteristics are illustrated using CMSAF products for the green class (left) and purple class (right), with an example crop near the centroid shown over the arrows.

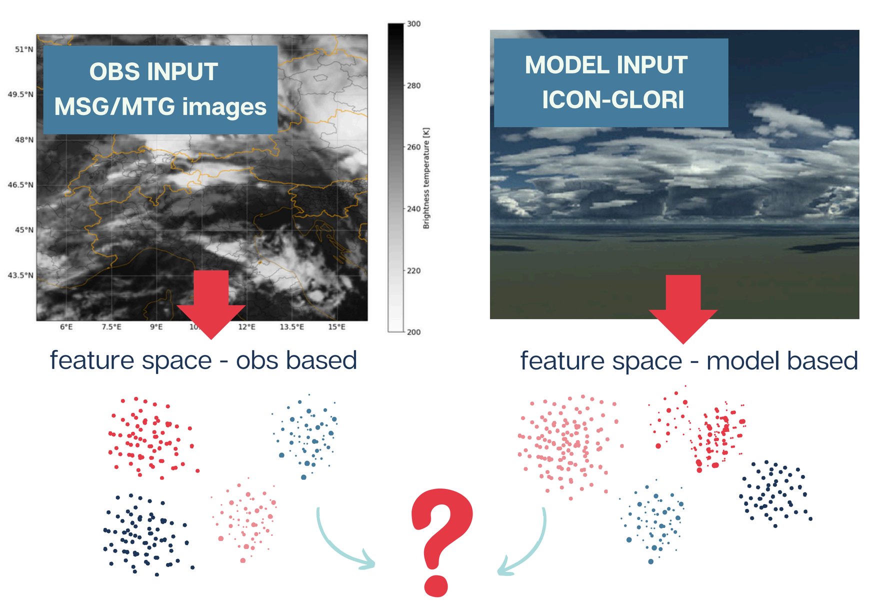
6. MODEL EVALUATION

Approach 1



The ML framework will pinpoint the positions within the observation-based feature space of **relevant case studies**, which IR 10.8 will be modelled by ICON-GLORI. Model effectiveness will be evaluated based on how closely the identified locations align with the actual cloud classes.

Approach 2



Another feature space is built based on **ensemble model** predictions and it is directly compared with the one obtained from satellite observations

7. REFERENCES

- Daniell, J. E., Wenzel, F., & Schaefer, A. (2016). Global economic costs of natural disasters from 1900-2015.
- Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., & Stoffel, M. (2014). 21st-century climate change in the European Alps. Science of The Total Environment, 493, 1138–1151.
- Dallan, E., Marra, F., Fosser, G., Marani, M., Formetta, G., Christoph, S., & Borga, M. (2023). Representing the reverse orographic effect of extreme precipitation. Hydrology and Earth System Sciences, 27, 1133-1153.
- Chatterjee, D., Acquistapace, C., Deneke, H., & Crewell, S. (2023). Cloud systems structure analysis using machine learning. Artificial Intelligence for the Earth Systems.
- Hornik, K., Feinerer, I., Kober, M., & Buchta, C. (2012). Spherical k-Means Clustering. Journal of Statistical Software, 50, 1-22.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition.
- Marsigli, C. and Potthast, R. and the the GLORI Team: The Global-to-Regional ICON Digital Twin, EMS Annual Meeting 2022, Bonn, Germany, 5–9 Sep 2022, EMS2022-690.
- van der Maaten, Laurens & Hinton, Geoffrey. (2008). Viualizing data using t-SNE. Journal of Machine Learning Research. 9. 2579-2605.
- Tenenbaum, Joshua & Silva, Vin & Langford, John. (2000). A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science. 290. 2319-2323.

contact: dcorrad1@uni-koeln.de

follow us
webpage
@expats_ideas4s
EXPATS-IDEAS4S