Characterization of cloud variability through novel satellite-based observations
Dwaipayan Chatterjee, Hartwig Deneke, Susanne Crewell
Cloud system appearance (structure and distribution) results from complex interplay of physical processes and large-scale environmental conditions → large interest in cloud classification
Traditional approach of using human-defined metrics
Janssen et al., 2019

Unsupervised machine learning
Denby, 2019

Visual inspection
Stevens et al., 2019

How can we classify cloud systems?
Our approach

Previous studies

• only consider tropical clouds over water → expand to central European domain

• make at least some use of human decisions → consider a purely data driven approach

Challenges

• Cloud systems transform continuously → categories are transient

• Likelihood of closeness should be assessed by machine learning

Use a bottom-up approach where the network learns a suitable hidden representation of the data
Guiding Questions

1. How can we use deep learning to identify different cloud systems?

2. How can we physically interpret the identified classes?

3. What are the measures to identify the right number of classes?

4. Can the model generalise well on unseen data?
Overview

1. Data
2. Deep learning architecture
3. Classification
4. Gaining understanding
5. Generalization
6. Conclusions & Outlook
1. Satellite Data

Satellite data

- Enhanced resolution MSG data set (Deneke et al., AMT, 2021)
- Cloud optical depth (COD) and 10 additional retrieval products, e.g. top of the atmosphere reflectivity, downwelling solar radiation
- April to July 2013 + 2015

Meteosat Second Generation (MSG) → Meteosat Third Generation (MTG)

Two different sizes tested
- 128x128 = 256x256 km
- 64 x 64 = 128x128 km

4 random crops at any instant
2. Deep Learning Architecture

Facebook artificial intelligence research (FAIR) (Caron, 2018, 2021)
FAIR Open source Vision Self Supervised Learning (VISSL) library (Priya Goyal, 2021)
2. Deep learning architecture and data pipeline

By using data augmentation such as random crops in our case, we discourage memorization and encourage generalization.
3. Classification result

Deep learning model assigns unique class (coloured frame) to each image
3. Classification

Evolution of understanding by the deep neural network

Epoch = 800
3. Classification

Good convergence achieved from 250 epoch
After 700 epoch clear segregation
3. Classification: Closeness

Top 5 nearest neighbours from Train Data set
Training Domain

Test Image

20130401t0615
20130401t0620
20130402t1630
20130706t0900
20130717t0615
20130401t0605

20130728t0850
20130728t0845
20130728t0855
20130728t0840
20130728t0900
20130526t1520

20130613t1615
20130613t1610
20130613t1620
20130613t1555
20130613t1600
20130613t1625
4. Gaining understanding: Physical significance

Surface downwelling radiation (Wm$^{-2}$)

Top of the atmosphere reflected radiation (Wm$^{-2}$)

Deep convective cloud systems

Centeroid position and closest image of one class

Clear sky
4. Gaining understanding: Number of classes

Centeroid images of 10 classes identified for small (64x64) region

- class1
- class2
- class3
- class4
- class5
- class6
- class7
- class8
- class9
- class10

Frequency of occurrence in COD-CTP space following Oreopoulos et al., 2016

Cloud Top Pressure (CTP) vs. Cloud Optical Depth (COD)

Similar behaviour found in different classes
4. Gaining understanding: Number of classes

N = 9

N = 8
Goal: Identification of highly variable conditions of solar energy

4. Gaining understanding: Targeted Parameter

Surface downwelling radiation (Wm$^{-2}$)

- **5 classes**
- **8 classes**
- **9 classes**
- **10 classes**

Less representation vs. More classes Repetition
Inta Cluster distance of seen and unseen data from Centroids Which are closely related as 0.99 show good consistency in the spread
1. How can we use deep learning to identify different cloud systems?
   DeepClusterV2 based self-supervised learning approach can identify distinct cloud regimes in complex central European environment

2. How can we physically interpret the identified classes?
   Satellite retrieval products support the physical interpretations of identified cloud regime classes

3. What are the measures to identify the right number of classes?
   Depending on purpose of classification different metrics can be used

4. Can the model generalise well on unseen data?
   Generalisation tested with unseen data using centroid embeddings → correlations up to 0.99 and similar distance distributions