CHARACTERIZATION OF CLOUD VARIABILITY THROUGH NOVEL SATELLITE-BASED OBSERVATIONS

INAUGURAL-DISSERTATION



Author: Dwaipayan Chatterjee (Student ID: 12731)
Supervisor: Prof. Dr. Susanne Crewell

Institute for Geophysics and Meteorology
Faculty of Mathematics and Natural Sciences
University of Cologne

Cologne, 2024

Berichterstatter:
Prof. Dr. Susanne Crewell
Prof. Dr. Peer Nowack

Datum der mündlichen Prüfung: 12. September 2024
Die Dissertation ist von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln angenommen worden.

Abstract

Cloud formation and evolution across different spatiotemporal scales present some of the most complex mechanisms in Earth's atmosphere. These complexities lead to various cloud types displaying diverse characteristics regarding their water content and impact on the radiative budget. Our limited capacity to represent the governing physical laws, likely because of too many unknown parameters and insufficient high-resolution observations, hampers the semi-empirical sub-grid representation of clouds in the numerical models. Consequently, clouds emerge as substantial sources of uncertainty in climate sensitivity experiments, diminishing our confidence and constraining our capacity to predict the future trajectory of our climate. Along with this, especially low-level clouds strongly modulate the surface solar irradiance (SSI), which impacts solar photovoltaic production. The relevance of this impact has increased with time as our electric grids have been increasingly infused with renewable energy, thus becoming more dependent on their intermittent nature. This causes sudden stress on microgrids, which causes them to struggle to balance the resulting variations around the standard frequency.

The current state-of-the-art Geostationary Satellites (GS), like Meteosat Third Generation (MTG), offer spatial resolution at 1km, yet this still falls short of resolving the fine-scale nature of many cloud objects, which extend to hundreds of meters or finer. However, by treating the aggregation of these individual cloud objects spanning up hundreds of kilometers as spatial distributions, the resolution likely becomes adequate to characterize their variability. Research increasingly supports the notion that these cloud fields (CF) exhibit distinct cloud radiative effects and form under different large-scale conditions. Consequently, advanced-generation climate models must consider their variability.

This work develops two neural networks (defined together as the framework) based on self-supervision principles to taxonomize the entire daytime visible mesoscale CF spectra from satellite images. Their individual pretext tasks simplify the entire cloud spectrum by focusing on the essential features of the CFs and thus reduce their high dimensional intricate distribution. We use cloud optical depth images retrieved from GS measurements as an input in both of our networks. Exploiting these networks and GS observations over central Europe (land) and north Atlantic trades (NAT, ocean), we demonstrate how different regimes can be distinguished as well as investigate their organizational interconnections. Our investigation includes the spring to autumn period over land and the dry season over NAT which both feature convective activity.

The framework maps the spatial distribution of CFs based on their optical depth characteristics, facilitating the identification of distinct classes over both ocean and land. Their physical uniqueness is validated independently using supplementary satellite-

retrieved products and reanalysis data. This enables the identification of low-level CFs, which are subsequently further distinguished to capture their difficult-to-separate diverse convective stages. Capturing different regimes of the low cloud family allows us to better characterize their particular contribution to SSI temporal fluctuations using ground-based pyranometer measurements.

The exploration of organizational interconnections among low-level shallow clouds takes place within the downstream trades, where a framework systematically arranges them based on their organizational similarities. This method enables the identification of weak break points in the cloud organization continuum, which are interpreted as regime transitions. A novel metric has been introduced to measure the relative changes in CF organizational status during their transition. Further, a simplified representation of the complete CF spectrum, along with its partitioning, facilitates a comparison between state-of-the-art human labels and machine labels, thus assessing their positions in the continuum. Despite these advancements, shortcomings persist, particularly with lower consensus labels being misclassified, leading to an uncertain allocation within the continuum.

Our analysis indicates that the results depend on the spatial scale of the CFs. Assessing the generalization potential of the framework by ingesting the continuous temporal data over Jülich, Germany, gives us the flexibility to explore the ground-based measurements from the Jülich Observatory for Cloud Evolution (JOYCE). We track the temporal evolution of the CF and emphasize the reasoning within the latent space. Our study illustrates how intrinsic cloud properties influence changes in neural embeddings, where we validate the results with ground-based radar and pyranometer observations. The developed framework opens up the possibility of using artificial intelligence framework-driven metrics for model evaluation against observational data.

Zusammenfassung

Die Bildung und Entwicklung von Wolken auf unterschiedlichen räumlichen und zeitlichen Skalen stellen einige der komplexesten Mechanismen in der Atmosphäre der Erde dar. Diese Komplexitäten führen dazu, dass verschiedene Wolkenarten unterschiedliche Eigenschaften hinsichtlich ihres Wassergehalts und ihres Einflusses auf das Strahlungsbudget aufweisen. Unsere begrenzte Fähigkeit, die zugrunde liegenden physikalischen Gesetze aufgrund zu vieler unbekannter Parameter und unzureichender hochauflösender Beobachtungen darzustellen, behindert die semi-empirische Unterteilung von Wolken in den numerischen Modellen. Folglich werden Wolken zu wesentlichen Unsicherheitsquellen in Klimaempfindlichkeits-Experimenten, was unser Vertrauen beeinträchtigt und unsere Fähigkeit zur Vorhersage der zukünftigen Entwicklung unseres Klimas einschränkt. Darüber hinaus modulieren insbesondere tiefe Wolken stark die einfallende Solareinstrahlung (SSI), was die Produktion von Solarstrom beeinflusst. Die Relevanz hiervon hat im Laufe der Zeit zugenommen, da unsere Stromnetze zunehmend mit erneuerbarer Energie versorgt werden und damit stärker von ihrer intermittierenden Natur abhängig sind. Dies führt zu plötzlichem Stress in Mikronetzen, die Schwierigkeiten haben, die resultierenden Schwankungen um die Standardfrequenz auszugleichen.

Die aktuellen Geostationären Satelliten (GS), wie Meteosat Third Generation (MTG), bieten eine räumliche Auflösung von 1 km, was jedoch immer noch nicht ausreicht, um die kleinskalige Natur vieler Wolkenobjekte aufzulösen, die sich über Hunderte von Metern oder feiner erstrecken. Durch die Aggregation dieser einzelnen Wolkenobjekte, die sich über Hunderte von Kilometern erstrecken, als räumliche Verteilungen könnte die Auflösung ausreichen, um deren Variabilität zu charakterisieren. Untersuchungen unterstützen zunehmend die Vorstellung, dass diese Wolkenfelder (CF) unterschiedliche Wolkenstrahlungseffekte aufweisen und unter verschiedenen großskaligen Bedingungen entstehen. Folglich müssen moderne Klimamodelle ihre Variabilität berücksichtigen.

Diese Arbeit entwickelt zwei neuronale Netzwerke (zusammen als Framework definiert), die auf Selbstüberwachungsprinzipien basieren, um das gesamte mesoskalige CF-Spektrum aus Satellitenbildern im sichtbaren Wellenlängenbereich zu taxonomisieren. Ihre individuellen pretext Aufgaben vereinfachen das gesamte Wolken-Spektrum, indem sie sich auf die wesentlichen Merkmale der CF konzentrieren und somit ihre hochdimensionale komplexe Verteilung reduzieren. Wir verwenden die aus GS-Messungen abgeleitete Wolkenoptische Tiefe als Eingabe in beiden unserer Netzwerke. Durch die Nutzung dieser Netzwerke und GS-Beobachtungen über Mitteleuropa (Land) und den Nordatlantik-Passatgebieten (NAT, Ozean) zeigen wir, wie verschiedene Regime unterschieden werden können und untersuchen deren Organisationszustände. Unsere Untersuchung umfasst die Frühjahrs- bis Herbstperiode über Land und die Trockenzeit über NAT, die beide konvektive Aktivitäten aufweisen.

Das Framework kartiert die räumliche Verteilung von CFs basierend auf ihren optischen Tiefecharakteristika und ermöglicht die Identifizierung unterschiedlicher Klassen über Ozean und Land. Ihre physische Einzigartigkeit wird unabhängig anhand zusätzlicher satellitengestützter Produkte und Reanalysedaten validiert. Dies ermöglicht die Identifizierung von niedrigen Wolkenfeldern, die anschließend weiter differenziert werden, um ihre schwer zu trennenden verschiedenen konvektiven Stadien zu erfassen. Das Erfassen verschiedener Regime der niedrigen Wolkenfamilie ermöglicht es uns, ihren spezifischen Beitrag zu den zeitlichen Schwankungen der SSI anhand bodengestützter Pyranometermessungen besser zu charakterisieren.

Die Untersuchung der Organisationszustände von niedrigen Wolken findet entlang der Passatwinde statt, wo das Framework sie systematisch basierend auf ihren organisatorischen Ähnlichkeiten anordnet. Diese Methode ermöglicht die Identifizierung schwacher Breakpoints im Kontinuum der Wolkenorganisation, die als Regimeübergänge interpretiert werden. Es wurde eine neue Metrik eingeführt, um die relativen Veränderungen im organisatorischen Status von CF während ihres Übergangs zu messen. Darüber hinaus erleichtert eine vereinfachte Darstellung des vollständigen CF-Spektrums zusammen mit seiner Partitionierung einen Vergleich zwischen modernen menschlichen Klassifikationen und Maschinenlabels, wodurch ihre Positionen im Kontinuum bewertet werden. Trotz dieser Fortschritte bleiben Mängel bestehen, insbesondere bei geringem Konsens, wenn Bilder fehlerhaft klassifiziert werden und zu einer unsicheren Zuordnung innerhalb des Kontinuums führen.

Unsere Analyse zeigt, dass die Ergebnisse von der räumlichen Skala der CFs abhängen. Die Bewertung des Generalisierungspotenzials des Frameworks durch die Aufnahme der kontinuierlichen zeitlichen Daten über Jülich, Deutschland, gibt uns die Flexibilität, die bodengestützten Messungen vom Jülicher Observatorium für Wolkenentwicklung (JOYCE) zu erkunden. Wir verfolgen die zeitliche Entwicklung des CF und betonen die Argumentation im latenten Raum. Unsere Studie veranschaulicht, wie intrinsische Wolken-Eigenschaften Veränderungen in neuronalen Einbettungen beeinflussen, wobei wir die Ergebnisse mit bodengestützten Radar- und Pyranometerbeobachtungen validieren. Das entwickelte Framework eröffnet die Möglichkeit, metrikengetriebene künstliche Intelligenzframeworks zur Bewertung von Wetter- und Klimamodellen anhand von Beobachtungsdaten zu verwenden.

Contents

1 Introduction **1**
1.1 Motivation . . . 2
1.2 Objectives . . . 5
1.3 Overview of the studies . . . 6
1.3.1 Study 1: Understanding mesoscale cloud system's structure and distribution . . . 6
1.3.2 Study 2: Capturing the diversity of mesoscale cloud organizations 8
1.3.3 Study 3: Towards solar energy application . . . 9

2 Theoretical background **11**
2.1 Clouds at different scales . . . 11
2.2 Cloud property retrievals . . . 12
2.3 Cloud optical depth retrievals . . . 15
2.4 Additional auxiliary datasets . . . 16
2.5 Self-supervision in general . . . 17
2.5.1 Convolutional neural network . . . 19
2.5.2 Vision transformer . . . 20

3 Understanding mesoscale cloud systems **22**
3.1 Introduction . . . 23
3.2 Satellite dataset . . . 27
3.3 Deep learning architecture . . . 29
3.3.1 Definition of the network input . . . 30
3.3.2 Network general architecture . . . 30
3.3.3 Upper branch of the network . . . 31
3.3.4 Lower branch of the network and cost function minimization . . 32
3.3.5 Network learning behavior . . . 33
3.3.6 Optimization for the convergence . . . 33
3.4 Results . . . 35
3.4.1 Analysis of the clustering capability . . . 35
3.4.2 Determination of the optimal number of classes . . . 36
3.4.3 Optimum classification . . . 38
3.4.4 Physical significance of the identified classes . . . 39
3.4.5 Generalization . . . 41
3.4.6 Persistency of cloud regimes and probability of transition . . . 43
3.5 Conclusions and outlook . . . 45
3.6 Appendix 1 . . . 49
3.6.1 Determination of the optimal number of classes for 64 x 64 images 49

4 Capturing mesoscale cloud system's diversity **50**
4.1 Introduction . . . 51
4.2 Satellite dataset . . . 53
4.3 Methods . . . 54
4.4 Results . . . 56
4.4.1 Continuous and discrete representations . . . 56
4.4.2 Machine versus human labels . . . 58
4.5 Transitions . . . 61
4.6 Conclusion . . . 63
4.7 Appendix 2 . . . 65
4.7.1 Domain description . . . 65
4.7.2 Network architectures . . . 65
4.7.3 Determination of optimal cluster number . . . 70
4.7.4 Sensitivity tests . . . 71
4.7.5 Attention maps of N1 . . . 75
4.7.6 Environmental characteristics . . . 76
4.7.7 Transition probability . . . 77

5 Towards solar energy application **79**
5.1 Introduction . . . 80
5.2 Data and method . . . 84
5.3 Results . . . 85
5.3.1 Robustness of the network . . . 85
5.3.2 JOYCE SSI characterization . . . 94
5.4 Better differentiating the low cloud regimes . . . 96
5.4.1 Low cloud continuum . . . 97
5.4.2 Daily cloud evolution . . . 101
5.5 Conclusion . . . 103

6 Conclusion and outlook **105**
6.1 Conclusion . . . 106
6.2 Outlook . . . 108
6.3 Towards model evaluation . . . 111
6.3.1 Introduction . . . 111
6.3.2 Evaluate cloud regimes . . . 112
6.3.3 Evaluate shallow cloud regime transitions . . . 113
6.3.4 Transition probability . . . 114

References **116**

Acknowledgements **142**

Eidesstattliche Versicherung **144**

List of Figures

1.1 MTG satellite's first image of Earth at 11:50 UTC on 18 March 2023 using FCI, highlighting cloud structures with higher resolution 2
2.1 Illustration of atmospheric processes at various spatio-temporal scales . 12
2.2 The spectral response function of two visible and four near-infrared bands from ABI on GOES-16, with a high spectral resolution atmospheric transmittance plot in gray 13
2.3 The spectral response function for the 10 infrared spectral bands from ABI on GOES-16, with a high spectral resolution infrared Earth-emitted spectrum in red 14
2.4 Illustrating the architecture of a convolutional neural network 19
2.5 Illustrating the architecture of a vision transformer 20
3.1 Geographical map of Europe over laid with MSG high-resolution cloud optical depth image 29
3.2 Schematic diagram of deep learning neural network using convolutional neural network and multi-layer perceptrons 31
3.3 Progressive visualization of the dimensionally reduced feature space . . 34
3.4 Closest and farthest COD images belonging to a certain cloud regime . 36
3.5 Correlation patterns among classes for the number of cloud regimes varying from 12 to 7 for 128x128 configuration 37
3.6 Visualization of classification by the deep neural network at k = 7 for 128 x 128 configuration and k = 8 for 64 x 64 configuration 38
3.7 Radiative and physical properties of 1000 randomly selected images from each cloud regime 40
3.8 Joint histogram of COD and CTP for 128 x 128 and 64 x 64 pixel configurations. 42
3.9 Correlation of centroids obtained from 2013 and 2015 datasets 42
3.10 Two example cases for temporal evolution and transition of cloud systems over Juelich 44
3.11 Cloud systems persistency and transition probability over Juelich . . . 45
3.12 Correlation patterns among classes for the number of cloud regimes varying from 12 to 7 for 64x64 configuration 49
4.1 Characteristics of the continuum space visualized with satellite images, and its partition supported by cloud amounts measured at four different vertical heights 57
4.2 Comparison of human and machine labels in the continuum 60
4.3 visualization and quantification of the sugar to flower transition 62
4.4 North Atlantic trades domain description and random cropping schematic 65

4.5 Illustration of the Vision transformer neural architecture (N1) 66
4.6 Dimensionally reduced discrete feature space of N2 and effect of continuum partitioning by only using random global crops of size 96x96 in N2 69
4.7 Results of three different metric scores of distortion, silhouette, and Calinski-Harabasz, shown along with varying cluster numbers 71
4.8 Sensitivity test for obtaining reasonable classes by varying sample sizes 72
4.9 Sensitivity test for obtaining reasonable classes by varying dimensional reduction techniques 73
4.10 Sensitivity test where N1 and N2 are initialized with different random seeds 74
4.11 Selecting different global crop size (65%, 75%, 85%) and training N1. t-SNE is intialized by PCA, and using cosine distance as a unit of distance, we find k =7 as the common consensus among the three. 74
4.12 Visulaization of six self-attention maps from the final head of the teacher ViT while ingesting four cloud systems with different organizations . . 75
4.13 Large scale environmental conditions of human labels and their nearest neighbors selected from the continuum feature space 76
4.14 Matrix representing the probability of transition of each machine-identified cloud regime to another over the Barbados cloud observatory. 77
5.1 Visualization of the dimensionally reduced original N2 feature trained using old and extended data set 86
5.2 Each cloud regime's relative frequency of occurrence from 2013-2019 except 2014 87
5.3 Radiative properties (surface solar irradiance and top the atmosphere reflectivity) of 1000 randomly selected images from each cloud regime for old and extended dataset 88
5.4 Cloud physical properties (cloud optical depth and cloud top pressure) of 1000 randomly selected images from each cloud regime for old and extended dataset 88
5.5 Comparision of the diurnal cycles of the seven cloud regimes between old and extended dataset 91
5.6 Comparision of the probability density estimate of measured transmissivity using a pyranometer at the JOYCE observatory of the seven cloud regimes between old and extended dataset 93
5.7 Spatially averaged and hourly averaged temporal variability of surface solar irradiance (SSI, W/m^2) for the seven cloud regimes derived from the extended dataset 95

5.8 Visualization of four low-level cloud regimes, characterization of their macrophysical properties and their normalized distribution constructed by linear interpolation into the t-SNE plane 97
5.9 Matrix representing the probability of transition of each low cloud regime to another class derived over Jülich 98
5.10 Box plot of solar irradiance relative variability and clearness index ramp rates of the low-level cloud systems 99
5.11 Daily low-level cloud system evolution visualized in the feature space and assessing the feature changes with radar and pyranometer measurements 102
6.1 Illustration of multi-modality combining radar with satellite measurements 109
6.2 Illustration demonstrating the use of pre-trained neural networks to exploit short-term surface solar irradiance (SSI) forecasting 110
6.3 Geographical map of Europe along with MSG high-resolution cloud optical depth divided into 8 subdomains 114
6.4 Matrix representing the probability of occurrence among classes in the next 30 minutes between subdomain 'a' and 'b' 115

List of Tables

3.1 The relative frequency of occurrence and cloud macrophysical characteristics for the seven classes 39
5.1 Characteristics of the seven cloud regimes for the extended dataset followed by old dataset 90
5.2 Comparison of probabilities for ramp's absolute value being >= 0.5 between our study for the four low-level clouds and those reported by Lohmann et al. (2016) for mixed sky conditions 101

Acronyms

ABI	Advanced Baseline Imager in GOES satellite
AI	Artificial Intelligence
CBH	Cloud Base Height
CERES	Clouds and Earth's Radiant Energy System
CI	Clearness Index
CMASK	Cloud Mask
CF	Cloud Fraction
CMSAF	Satellite Application Facility on Climate Monitoring
CNN	Convolutional Neural Network
COD	Cloud Optical Depth
CR	Cloud Regime
CTP	Cloud Top Pressure
CTT	Cloud Top Temperature
DCOMP	Daytime Cloud Optical and Microphysical Properties algorithm
ECMWF	European Centre for Medium-Ranged Weather Forecast
ERA-5	ECMWF 's reanalysis 5th generation
FAIR	Facebook Artificial Intelligence Research
GOES	Geostationary Operational Environmental Satellite
GPU	Graphic Processing Unit
HRV	High-Resolution Visible
IWV	Integrated Water Vapor
LARC	Large Batch Training of Convolutional Networks
LCF	Low Cloud Field
LES	Large Eddy Simulation
MBL	Maritime Boundary Layer
NAT	North Atlantic Trades
NN	Neural Network
MLP	Multi-Layer Perceptron
MODIS	Moderate Resolution Imaging Spectroradiometer
MSG	Meteosat Second Generation
MTG	Meteosat Third Generation
PCA	Principle Component Analysis
RFO	Relative Frequency of Occurrence
RMSE	Root Mean Square Error
RStd	Relative Standard deviation
TOA	Top of the atmosphere
TRS	Top-of-the-atmosphere Reflected Solar radiation
SGD	Stochastic Gradient Descent
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
SSL	Self-Supervised Learning
SDS	Surface Downwelling Solar radiation
SSI	Surface Solar Irradiation
SST	Sea Surface Temperature
UTC	Coordinated Universal Time
VISSL	VIsion library for state-of-the-art Self Supervised Learning
ViT	Vision Transformer

1 Introduction

1.1 Motivation

One of the most intriguing facets when viewing Earth from space is the distribution of clouds (see Fig. 1.1). Clouds considerably influence weather and climate by modulating Earth's energy budget and water cycle. At any point in time, clouds cover about 67% of the surface of the Earth according to satellite-based estimates (M. D. King et al., 2013), making the planet appear white from space. Many individuals adjust their behavior based on the appearance of the cloud field from the ground. For instance, they might apply sunscreen when they see only a few shallow clouds or carry an umbrella upon spotting darker clouds in anticipation of rain.

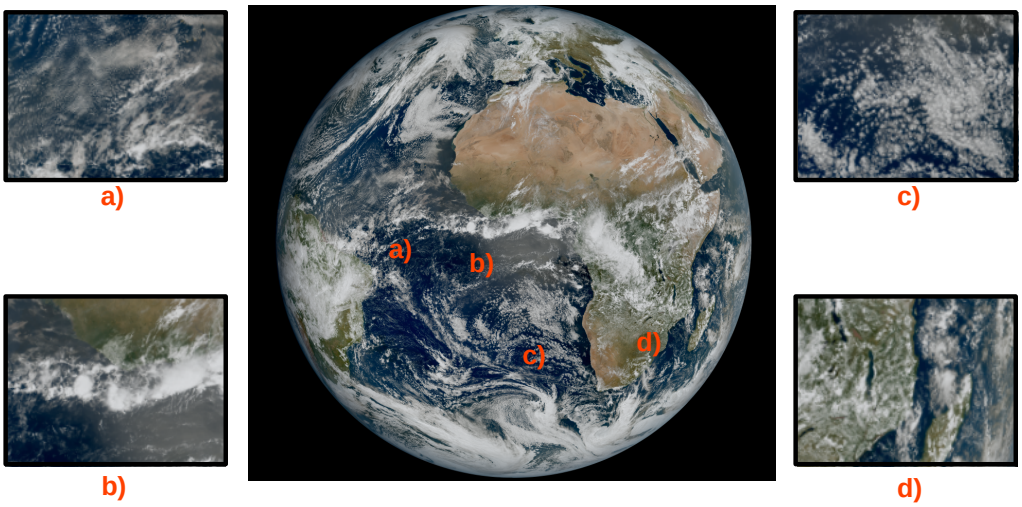


Figure 1.1: Meteosat third generation (MTG) satellite's first image of planet at 11:50 UTC on 18 March 2023 using Flexible Combined Imager (FCI) instrument. Zooming in on the finer structures and distribution of different cloud systems (see a-d) highlights the benefit of higher resolution.

Water vapor primarily redistributes heat and moisture from surplus tropical to shortage arctic regions. Clouds also facilitate the transformation of latent heat into sensible heat and moisture into liquid; thus, embedded within the atmospheric circulation, they make heat and moisture available to higher latitudes (Haar & Suomi, 1971). Furthermore, different cloud types display diverse traits regarding their water content and impact on the radiative budget (Höjgård-Olsen et al., 2020). Depending on factors like altitude and optical thickness, clouds serve a dual role in regulating the temperature of the Earth system (Hartmann et al., 1992; Stephens, 2005). Lower-altitude clouds, such as cumulus, primarily composed of liquid droplets, tend to be warmer and effectively reflect incoming solar radiation (Bony, Dufresne, Treut, et al., 2004). Conversely, higher-altitude clouds, such as cirrus, are colder, contain ice particles, absorb infrared radiation emitted by the Earth (Liou, 1986), and re-emit some of it back into

space and towards the Earth's surface. Consequently, clouds contribute to a complex interplay of warming and cooling effects on Earth's net radiative budget. In response to global warming, the precise dynamics of this balance and how it might evolve remain uncertain (Stephens, 2005; Bony et al., 2006).

The atmospheric part of the Earth system is primarily driven by solar energy absorbed at its surface (Trenberth & Fasullo, 2011; G. Stephens et al., 2012). Albedo and outgoing long wave radiation of the Earth are predominantly determined by cloud characteristics such as cloud type and cloud fraction (Ockert-Bell & Hartmann, 1992). Research has shown that depending on the cloud regime, this bi-variate relation is either linear, exponential, or both (Feingold et al., 2017). However, understanding the process by which broken cloud fields control the surface energy variability is nontrivial (Lane et al., 2002). Especially the shallow cloud fields represent a particular challenge due to their small spatial scale and rapid evolution. (Lamer & Kollias, 2015).

The transition towards green energy, spurred by initiatives like the Green Deal, has led to widespread adoption of renewable sources such as solar power (EU-Commission et al., 2021). For instance, the European Union (EU) aims to deliver over 320 GW of solar photovoltaic (PV) by 2025 and almost 600 Giga Watt (GW) by 2030 (Kougias et al., 2021). However, this shift has increased reliance on the intermittent nature of solar energy, posing challenges for grid stability. One significant factor modulating steep fluctuations in Surface Solar Irradiance (SSI) is the presence of low-level shallow clouds. These clouds, characterized by their fragmented structure, create distinct cloud shadows and sunlight exposure, influencing PV output fluctuations (Gristey et al., 2020; Lohmann et al., 2016). Hence, managing stable grid frequency has become crucial, particularly with high solar energy penetration, necessitating effective control strategies to minimize fluctuations within a narrow range (Anvari et al., 2016).

Hence, various cloud systems exert diverse influences on both climate systems and PV production. Estimating the impacts of these systems relies on two distinct data sources: 1) Numerical weather prediction data or model data and 2) observations.

Climate models are complex computer simulations designed to replicate Earth's past, present, and projected future climates. Within these simulations, researchers carefully represent various components of the Earth's system (i.e., land, atmosphere, ocean, and their coupled interactions), drawing upon understanding the component's physical behavior. Hence, these models serve as invaluable tools for assessing both anthropogenic influences on the Earth's climate and the impacts of natural variability. However, despite their fidelity, uncertainties persist, with cloud systems feedback emerging as a prominent source of variability.

Discrepancies in climate sensitivity among current models often arise from differences in how clouds are represented (see Chapter 8.6.3.2 of the Intergovernmental Panel on Climate Change report (IPCC Core Writing Team et al., 2007)). This is because

even subtle alterations in cloud amount or distribution can disproportionately impact climate compared to changes in greenhouse gas concentrations or aerosol levels (Zelinka et al., 2017). For instance, observations from the Clouds and Earth's Radiant Energy System (CERES) satellite instrument (NASA et al., 2017) indicate that clouds contribute -21 W/m^2 to Earth's net energy budget globally, a value 91% higher than the forcing exerted by a major greenhouse gas like CO_2 (1.66 W/m^2) in absolute terms, see chapter 7 from (Boucher et al., 2013). Thus, accurately representing clouds and understanding their behavior at different time scales in climate models becomes crucial.

Using advanced storm-resolving climate models simulating at 1km resolution (50 times finer than previous generation models), one can capture the finer details of the circulation that makes the storm in the atmosphere or eddies in the ocean (Hohenegger et al., 2023). However, despite these advancements, the small-scale shallow clouds, which occur at a finer scale of 10 meters, remain poorly resolved (Schneider et al., 2019). Also, the heterogeneity of these shallow clouds complicates their representations in the numerical models, limiting the model's ability to produce realistic simulations (Nordling et al., 2024).

Concurrently, at the same time, present and upcoming generation geostationary satellites offer continuous high spatio-temporal resolution observations, further enabling our understanding of rapidly changing dynamic systems in the Earth's atmosphere. For example, the Flexible Combined Imager (Durand et al., 2015) onboard the current Meteosat Third Generation (MTG) has a very high spatial resolution, around 1 km, with a repeat cycle of 2.5 minutes over Europe.

Despite clouds' small individual extents, the advanced geostationary satellite imagery captures unprecedented detail of distinct cloud fields and their continuous spatio-temporal behavior (Schulz et al., 2021a; Vogel et al., 2021) spanning vast areas (a few 100 km) of our planet's atmosphere(Stevens & et al., 2021). For example, during the preparation of the Eurec^{4}a campaign, Stevens et al. (2016) identified four shallow convective regimes and named them sugar, gravel, flower, and fish based on their cloud structure and appearance. These regimes exhibit differences in net cloud radiative feedback (Bony et al., 2020) and are related to different environmental conditions (Schulz et al., 2021b).

However, recent satellite-based studies (Denby, 2020a; Janssens et al., 2021) pointed out that placing four discrete classes (i.e., sugar, gravel, flower, and fish) on the diversity of the observable structure does not account for the intermediate cloud patterns or transient states. To account for the importance of these transition states, Large Eddy Simulation (LES) studies (Narenpitak et al., 2021, 2023; Dauhut et al., 2023) demonstrated the importance of understanding the governing processes guiding these transient stages. If ignored, several crucial mechanisms for climate feedback can be disregarded or overlooked. Additionally, recent research attempting to quantify these

labeled well-organized systems discovers that they only occur approximately 50% of the time over North Atlantic Trades (NAT) (Janssens et al., 2021; Schulz et al., 2021c; Vial et al., 2021), and there are some disagreements between the labeler and the data (Schulz, 2022). Also, these labels, being specific to the NAT regime, cannot be transferred to other climate regimes, e.g., to study cloud structures over Jülich.

From a meteorological standpoint, this dissertation is driven by the challenge of effectively harnessing geostationary observations to accurately depict the structure and distribution of cloud systems in their entirety. Central to this effort is the development of methods aimed at comprehending and capturing crucial features that characterize individual clouds as well as the overall cloud system distribution. These features must possess robustness in identifying distinct cloud systems while smoothly connecting to their transitional states. Simultaneously, from a solar energy perspective, the objective is to explore the potential of these representative features in targeting shallow cloud regimes and synergizing with ground-based instruments to characterize short-term, sharp fluctuations in SSI.

1.2 Objectives

In automated learning for essential features of cloud systems from satellite imagery, we aim to understand their complex organizational interconnections and taxonomize them effectively. Our goal is to explore reliable data sources that have been under-explored so far, such as geostationary satellite measurements and ground-based observations, and develop a scalable approach that is not constrained by the availability of labels. We plan to assess certain learning approaches using self-supervision in neural networks such that during training, the network learns the data properties defined through their pretext tasks, which can potentially guide the network in revealing complex relationships within cloud systems. Although it is challenging to quantify the performance of self-supervised frameworks without a comparable supervised task, this approach presents an opportunity to generate insights into their physical properties and test the networks on additional datasets.

Land and ocean convection each present unique challenges. Cloud convection over land is characterized by rapid heating, variable moisture, and pronounced diurnal cycles, with peak activity typically occurring in the late afternoon and early evening when surface heating is at its maximum. Significant topographical influences also lead to intense but often short-lived convective events. In contrast, oceanic convection tends to be more organized, influenced by the ocean's high heat capacity, abundant moisture, uniform temperature distribution, and large-scale atmospheric circulation patterns, resulting in more sustained and widespread cloud formations. There can be slight increases in convective activity during the nighttime due to radiative cooling at

the top of the atmosphere. These differences are essential to capture and represent, as the developed methodologies must be robustly applicable in both cases.

This dissertation formed a component of the KI:STE project funded by the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection, "AI Strategy for Earth system data," accessible at `https://kiste-project.de/`, and contributed to its overarching goals. KI:STE sought to leverage contemporary artificial intelligence methodologies for pattern recognition and analysis of spatio-temporal data from environmental monitoring. The dissertation aimed to accomplish several objectives within this framework, specifically focusing on cloud systems and their variability.

The first objective involves identifying unique cloud patterns and assessing their spatio-temporal evolution. Secondly, the study investigates whether neural networks (NN) can detect organizational links between various convective stages of cloud systems, aided by developing innovative metrics for quantifying these connections. The study aims to monitor structural changes in mesoscale cloud formations and their correlation with prevailing large-scale environmental conditions by analyzing the temporal progression of these connections among cloud systems. Lastly, the research aims to evaluate fluctuations in SSI variability modulated by shallow clouds for solar energy applications.

Hence, this thesis aligns with the overarching goals of the KI:STE project by offering a comprehensive meteorological examination of cloud systems' structure and distribution across both land and ocean under contemporary climate conditions. We leverage our insights into cloud systems to establish useful connections between low-level clouds and their capability to regulate SSI fluctuations, critical for short-term solar energy fluctuations and microgrid stability.

Addressing the objectives outlined above, the thesis includes three studies for which an overview is given below.

1.3 Overview of the studies

1.3.1 Study 1: Understanding mesoscale cloud system's structure and distribution

The intuition of appearance in mesoscale cloud systems arises from their visual characteristics, which can be measured through their structure and distribution. Here, structure pertains to the intrinsic characteristics of individual cloud entities, while distribution denotes the larger pattern formed by these entities collectively. They result from a complex interplay of physical processes (Williams & Brooks, 2008; Leinonen et al., 2016) and large-scale environmental conditions (Tselioudis et al., 2013; Bony, Dufresne, Treut, et al., 2004; Li et al., 2014). Recent approaches investigating mesoscale

cloud regimes use the joint composites of cloud properties, which completely neglect the spatial structure and organization of the cloud systems within the considered domain (Jakob & Tselioudis, 2003; Oreopoulos et al., 2014; Jin et al., 2016; Tzallas et al., 2022). However, Bony et al. (2020) has shown how these cloud organizations help explain differences in cloud-radiative effects. Furthermore, (Schulz et al., 2021c) have shown that these patterns include various cloud morphologies that arise in correlation with various meteorological settings.

Studies examining the spatial fields of cloud systems (Denby, 2020a; Bony et al., 2020; Janssens et al., 2021; Kurihana, Moyer, & Foster, 2022) often overlook the impact of varying spatial scales of cloud systems on the network's generalization capacity or the meaningfulness of the identified cloud systems. In other words, the question of how different image sizes influence the generation of meaningful cloud classes has not been deeply discussed.

The availability of significant volumes of high-resolution data records from modern geostationary satellite systems motivates us to create a consistent and unbiased cloud system classification scheme. Consequently, our first study aims to address this gap by developing an algorithm based on a modern self-supervision technique that processes large quantities of satellite cloud observations and learns to partition them according to their structure and distributions. To train our algorithm, we utilize the Cloud Optical Depth (COD) parameter derived from second-generation Meteosat satellite data retrieved over central Europe (Deneke et al., 2021), which has been further enhanced to resemble the high resolution of upcoming third-generation Meteosat. In the satellite dataset, COD has the best retrieval accuracy and is physically related to cloud attenuation, the variable that most substantially influences solar energy yields. We let the network identify useful cloud regimes that are as distinct as possible in their joint composite of COD and cloud-top pressure (CTP) physical space (Oreopoulos et al., 2016).

The resulting unique land cloud systems are grouped in different COD textures, and additional retrieval parameters from the satellite dataset, such as cloud top pressure, reflectivity, and surface downwelling solar radiation, help us verify their distinctness. A crucial aspect of the study was to evaluate how the spatial scales of cloud systems impact the data-driven clustering process. This was achieved by testing the clusters of the cloud systems on unseen data, which gave valuable insights into the spatial distribution of mesoscale cloud phenomena.

Furthermore, we examine the temporal variability of the cloud regimes over a particular region (Jülich, Germany) to assess the applicability of the network, which was not trained with any temporal information. We quantified the average persistence of specific cloud systems and the probabilities of their specific next transitions. This quantification is vital for planning solar power plants as power plants can be informed

early of the evolving cloud state. Overall, this study contributed to a deeper understanding of the dynamics of cloud systems and their implications for renewable energy applications.

1.3.2 Study 2: Capturing the diversity of mesoscale cloud organizations

Understanding the intricate dynamics of mesoscale shallow marine cloud organization and their net radiative impact on Earth's energy balance poses a significant challenge (Stevens et al., 2021). This difficulty arises, in part, from the inherent complexity of objectively quantifying the organizational state of the atmosphere.

Previous methodologies studying mesoscale cloud formations over North Atlantic Trades (NAT) have encountered limitations (Vial et al., 2017; Vogel et al., 2021; Schulz et al., 2021b) due to their narrow focus on specific types of distinct cloud organizations (Stevens et al., 2020; Bony et al., 2020; Rasp et al., 2020). In contrast, the new approach presented in this study clears this obstacle by harnessing the complete mesoscale marine cloud spectra over NAT as measured from the geostationary observations.

We devise a framework comprising two NNs where the first one projects the structural likeness of cloud images into a multidimensional feature space. Subsequently, the second NN categorizes these scenes into distinct clusters with physical significance. This strategy broadens the scope of capturing cloud formations and their different convective stages, which opens up the possibility of capturing cloud system transitional states.

What sets our approach apart is simplifying the cloud organization's complexities by efficiently categorizing the continuum into useful interpretable classes. This offers greater adaptability and generalizability than previous metric-based methods (Janssens et al., 2021). Furthermore, it diverges from other unsupervised learning (Denby, 2020b, 2023) frameworks and surpasses previous human-based classifications in coverage (Schulz, 2022). We establish the robustness of our framework through sensitivity studies and comparisons of machine labels with state-of-the-art human labeling.

We deepen our understanding of the mesoscale shallow cloud organization transitions through consistent developments observed within the feature space of our framework, coupled with reanalysis data. Quantifying such transitions illustrates the future potential for representing complex cloud system dynamics and underscores the challenge of studying such transitions.

Consequently, our approach enables statistical analysis over large regions and extended time spans by reducing image complexity. Initial analyses have already revealed the preferred occurrence of certain transitions within the first half-year period. These identified regimes hold promise for future exploration, either in isolation or in tandem with ground-based observations.

1.3.3 Study 3: Towards solar energy application

Shallow cloud systems, such as cumulus clouds, are widespread, spanning vast expanses of continents and trade wind regions across oceans (Warren et al., 1988). Characterized by their small spatial scale, broken cloud formations, and rapidly evolving tendencies (Berg & Kassianov, 2008), they play a crucial role in modulating SSI and its variability (Lane et al., 2002; Perez et al., 2016). This makes them of considerable interest for solar energy applications.

Comprehending the rapid, short-term fluctuations in solar photovoltaic (PV) output, closely correlated with changes in Surface Solar Irradiance (SSI) modulated by such clouds, is crucial (Lave & Kleissl, 2010; Haaren et al., 2014). Exceeding allowable limits in these fluctuations could potentially lead to total power system failure, especially given the limited response capability of transmission system operators (Marcos et al., 2014). This risk is particularly pronounced in scenarios with high PV penetration rates in stand-alone microgrids, such as those found on islands or remote villages.

This variability, encapsulated in SSI, includes predictable solar geometry and cloud/weather influences. Employing a normalized metric is advantageous for dissecting the impact of cloud-related factors, removing solar geometry while preserving cloud-induced variability (Lohmann et al., 2016). The Clear Sky Index (CSI) represents the ratio of SSI to extraterrestrial irradiance, and its variations over a specified time period δt are often referred to as the ramp rate. This time interval can vary widely, ranging from seconds to hours, depending on the concerns of grid operators (Ranaweera et al., 2014; Perez et al., 2015).

With a focus on solar energy application, this chapter revisits the work from Chatterjee, Acquistapace, et al. (2023) and performs a detailed investigation of the different cloud regimes' SSI characteristics using pyranometer measurements from Jülich Observatory for Cloud Evolution (JOYCE). Two shallow cloud regimes from Chatterjee, Acquistapace, et al. (2023) observed the highest spatio-temporal variability in measured SSI. Exploiting the versatility of our NNs, we refine our differentiation approach to better capture the variability within these low cloud classes.

Our focus is to accurately delineate the unique characteristics of low-level clouds across the entire cloud spectrum. For instance, we investigate ramp fluctuations from a few seconds to 15 minutes (Lohmann et al., 2016) important for short-term grid operations and intra-day energy trading. Such ramps are significant for a localized distribution system over JOYCE such as the Living Lab Energy Campus (LLEC) at Forschungszentrum Juelich (FZJ) (Althaus et al., 2022), which may lead to voltage control issues.

Finally, as an outlook, we apply transfer learning to track rapid shallow cloud system evolution (Tomson, 2010) over JOYCE and emphasize reasoning within the

latent space. Our study illustrates how intrinsic cloud properties influence changes in neural embeddings. We assess the results with ground-based radar and pyranometer observations.

2 Theoretical background

2.1 Clouds at different scales

Figure 1.1 illustrates various individual cloud objects at different spatial scales, and looking up from space, their collective arrangement together shows us the cloud fields. These cloud fields have convective clouds of different sizes and are individually at different development phases. The observed diversity of these cloud fields on different spatial scales is large, from a few 100 km to more than 2500 km (see Fig. 2.1). Given the scale variability, this thesis has focused on cloud systems that occur at a spatial scale of 100 km to 1000 km and a time scale from hours to days. They are known as mesoscale cloud systems.

Representing realistic mesoscale cloud fields is a challenge for numerical models. Even though there are global storm-resolving models at a 1-3 km scale, they do not resolve the individual clouds. Large eddy simulations resolve up to 50 to 100 m, but they can cover only a spatial scale of up to 500 km or 1000km. So, the large-scale aspects of mesoscale cloud fields are missing in them. Therefore, mesoscale systems' interaction with the sub-mesoscale and super-mesoscale systems is not adequately understood or represented in numerical models (Prein et al., 2020; Feng et al., 2018; Neumann et al., 2019; Na et al., 2022).

For climate models with a grid size of 50 km or 100 km, their parameterization schemes try to account for the effect of the individual clouds, and therefore, the concept of cloud fields is not embedded in them. Recent research shows that it could be essential to account for their representation, as hinted in the study of (Bony et al., 2020), which shows changing cloud fractions with different cloud fields over NAT.

We focus on utilizing geostationary observations to investigate mesoscale cloud systems. Due to their good quality spatial and temporal coverage and long-term record, they offer the information required to study the role of clouds in the climate system. Additionally, their continuous coverage allows for the resolution of the diurnal cycle of the clouds (Shang et al., 2018) and thus could be used to study cloud evolution. Geostationary satellites primarily support short-range forecasting, and their contribution to high-impact weather forecasting is gaining relevance. Their support for the solar renewable energy industry, such as the nowcasting of solar surface irradiance, is also in demand for their efficient planning and operations (Müller & Pfeifroth, 2022).

We use cloud optical depth (COD) retrievals from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument on board the Meteosat Second Generation (MSG) satellite (Deneke et al., 2021) for coverage over central Europe. It mimics the spatio-temporal capabilities of MTG and is based on the Cloud Physical Properties (CPP) (Roebeling et al., 2006) algorithm developed at the Royal Meteorological Institute of the Netherlands. While North Atlantic trades (NAT) retrievals

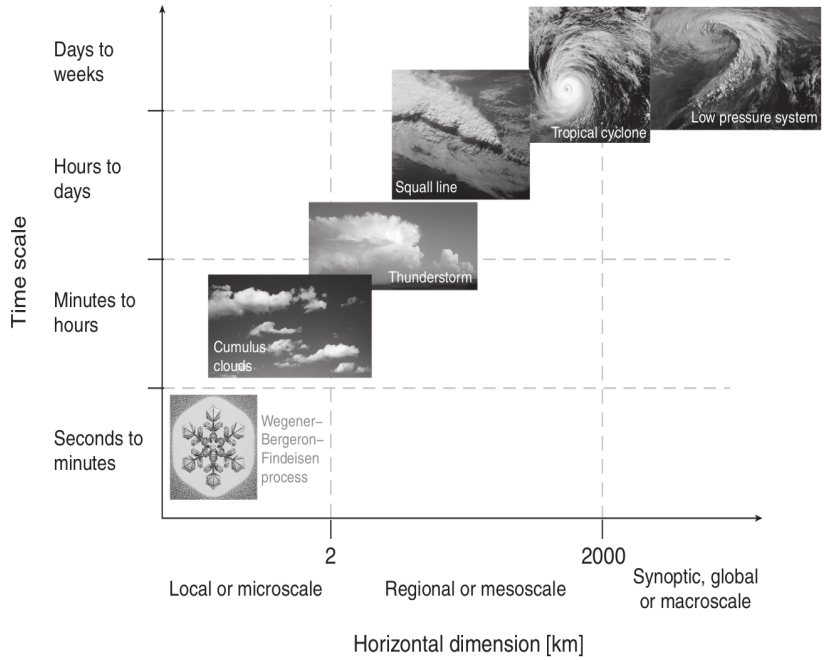


Figure 2.1: This illustration from U. Lohmann et al. (2016), chapter 9, Figure 9.6, shows atmospheric processes at different spatial and temporal scales. The scale separation shown here is taken from Orlanski (1975)

are taken from the Daytime Cloud Optical and Microphysical Properties algorithm (DCOMP) (Walther & Heidinger, 2012) from GOES-16 E Advanced Baseline Imager (ABI) (Schmit et al., 2005).

2.2 Cloud property retrievals

Solar radiation, primarily incident in the visible range of the electromagnetic spectrum, interacts with the different components present in the earth's atmosphere, such as aerosols, cloud droplets, and gasses, and in the process, they reflect, absorb, and transmit the incoming energy, which results in driving different atmospheric processes. For instance, out of 100 units of incident solar radiation, 30 units are reflected back to space: 20 units by clouds, six units by cloud-free air, and four units by the Earth's surface. Due to their large particle sizes, hydrometeors in clouds, such as rain drops and ice particles, are less selective in their scattering properties and reflect visible light equally effectively. Additionally, 19 units of solar radiation are absorbed by the atmosphere, while the remaining 51 units are absorbed by the Earth's surface (Menzel, 2005). The Earth disposes of this absorbed energy through a combination of infrared radiation emissions and sensible and latent heat fluxes.

Incoming solar radiation is concentrated in shorter wavelengths, peaking at about 0.5 μm, while the thermal infrared radiation emitted from Earth's atmosphere has its maximum energy in longer wavelength at about 10 μm (Menzel, 2005). Therefore,

the overlap between solar and infrared spectra is relatively insignificant, reducing the complexities for remote sensing applications.

Several spectral regions are considered useful for cloud remote sensing from satellites. These spectral regions are also called window channels, meaning that radiation is not significantly absorbed by the atmosphere's gases, which allows looking downward at the radiating surface of the earth or clouds. Different satellite missions choose the central wavelengths that, individually or in combination, allow the discrimination from the Earth's surface, such as between different altitudes of cloud types, vegetated and non-vegetated surfaces, and atmospheric aerosol content.

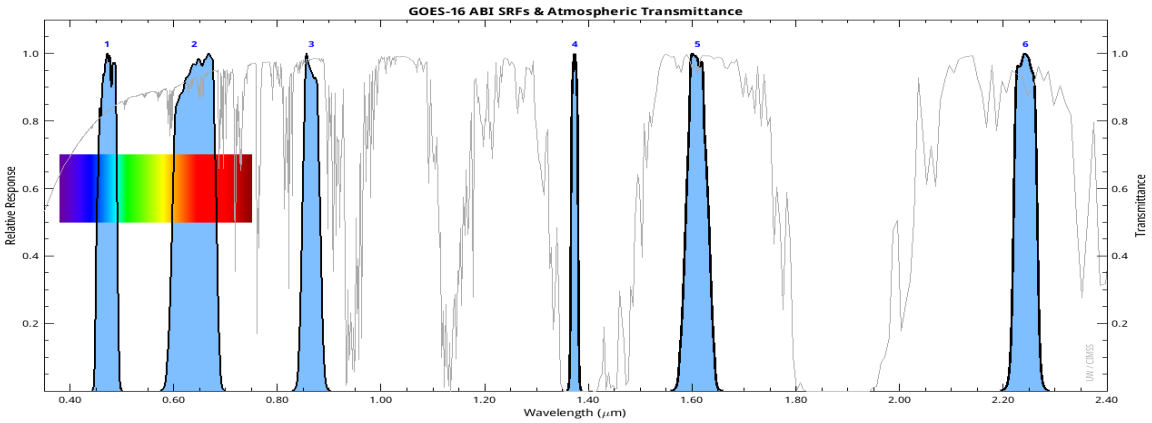


Figure 2.2: The two visible and four near-infrared spectral bands from Advance Baseline Imager (ABI) in the Geostationary Operational Environmental Satellite (GOES) - 16 are plotted as spectral response functions (SRFs). A high spectral resolution atmospheric transmittance plot in gray. This SRF plot is made publicly available at Cooperative Institute for Meteorological Satellite Studies `https://cimss.ssec.wisc.edu/goes/calibration/`

In the visible range (see Fig. 2.2), satellites receive part of the solar radiation reflected by the clouds or the earth's surface. Visible bands (mainly limited to daytime use) are primarily used for daytime cloud mapping and albedo determination. The bands are ideal for identifying visual features such as cloud edges and structures, which helps detect small-scale features such as river fog, fog edges, overshooting tops, or cumulus clouds during the day (Schmetz et al., 2002a). For instance, the 0.47 μm band provides daytime observations of dust, smoke, and clouds where smoke and dust signals are dominating when the sun is low in the sky because they are more effective at forward scattering than backward scattering (Yung, 2003). Also, because of the smaller wavelength, Rayleigh scattering is more prevalent. Meanwhile, higher wavelengths such as 0.64 μm have reduced Rayleigh scattering, and help monitor snow and ice cover, diagnose volcanic ash, and analyze hurricanes (Schmit et al., 2018).

The near-infrared (NIR) channels, going beyond the visible red range, include wavelengths between 0.7 μm and 3 μm and represent a unique portion of the electromagnetic

spectrum. Instead of focusing on an object's infrared emission, researchers examine how objects reflect, transmit, and absorb the Sun's near-infrared radiation. This method is particularly useful for assessing vegetation health and soil composition. NIR gets strongly reflected compared to visible ranges over leaves and canopy of healthy plants. Also, plants with high chlorophyll content reflect more near-infrared energy than unhealthy plants. From a cloud perspective, it can better contrast water against the surrounding features. The NIR channel at 1.37 μm is significant for its strong water vapor absorption, which makes it effective for detecting very thin cirrus clouds during the daytime. Additionally, the NIR band at 1.61 μm helps infer the phase of the cloud top: liquid water clouds appear bright due to their high reflectivity, whereas ice clouds and snow are darker because ice absorbs radiation around 1.61 μm rather than reflecting it.

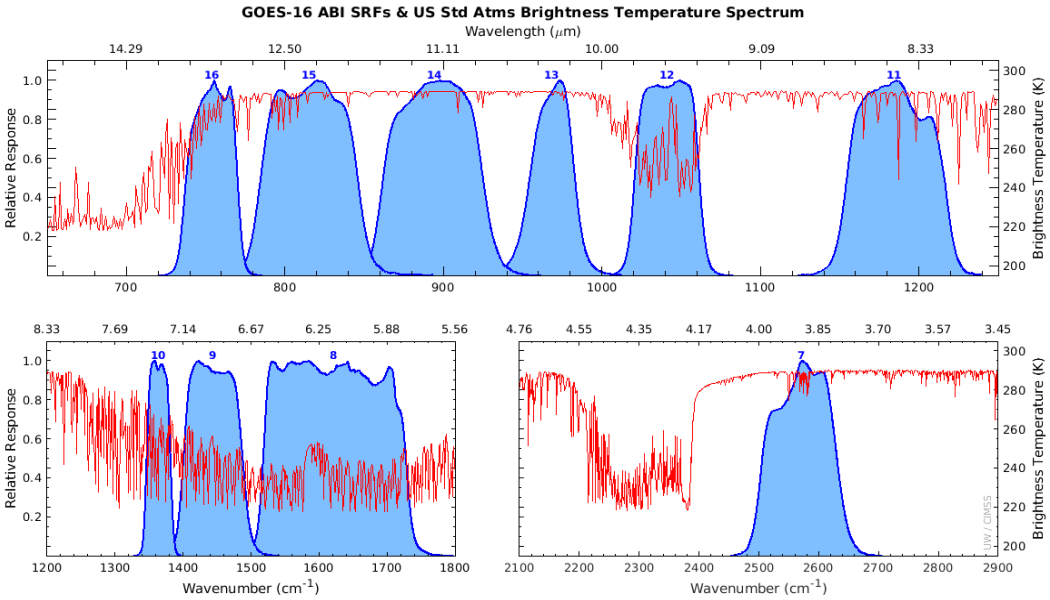


Figure 2.3: Same as Fig. 2.2, but the 10 infrared spectral bands from ABI in GOES - 16. A high spectral resolution infrared Earth-emitted spectrum is plotted in red. This SRF plot is made publicly available at Cooperative Institute for Meteorological Satellite Studies `https://cimss.ssec.wisc.edu/goes/calibration/`

In the infrared (IR) region (see Fig. 2.3), satellites measure brightness temperatures related to thermal radiation emitted by the Earth's surface or clouds, depending on their temperature. Infrared channels are primarily used for cloud mapping and surface temperature determination, both day and night. Infrared absorption bands allow satellites to detect water vapor at different altitudes in the atmosphere. For instance, wavelengths of 6.2 μm, 6.9 μm, and 7.2 μm correspond to atmospheric water vapor's upper and mid-levels. They can also serve dual roles depending on the time of day. For example, the IR 3.9 channel, known as the shortwave window in the infrared region, can detect fog and low-level clouds at night and differentiate between water clouds and

ice surfaces during the day. Additionally, combining infrared channels enhances their utility; for example, the IR 8.7, IR 10.8, and IR 12.0 channels can be used to determine the cloud phase. Thus, integrating different spectral bands allows the generation of higher-level-derived products used for comprehensive measurements of the cloud state, such as their composition, temperature, and height.

2.3 Cloud optical depth retrievals

Cloud optical depth (COD) or cloud optical thickness (COT) physically represents how much light the cloud can intercept. COD along particle size distribution represents the clouds' radiative properties in the solar spectral region (Walther & Heidinger, 2012). COD depends on the cloud's thickness, state (liquid droplets or ice crystals), and the size distribution of the water content. We have used COD in our study for several reasons. It is closely associated with the radiative effects of the clouds. Its usage avoids the influence of solar zenith angle, surface effects, and aerosol effects on the measured radiances. CPP and DCOMP robustly tackle the aleatoric uncertainties from the atmosphere and surface. Specifications about the algorithms and additional choices for data usage are mentioned in Section 3.2 and 4.2 .

Various techniques have emerged in the last three decades for extracting cloud optical thickness and effective radius using satellite radiances. They are based on using the non-absorbing visible and moderately absorbing near-infrared regions of the spectrum (Nakajima & King, 1990; Han et al., 1994; M. King et al., 1997; M. D. King et al., 2004). The primary variations among these techniques mainly arise from different satellites used and the specific visible and near-infrared wavelengths employed. Also, the complexities of the interpolation and iteration scheme employed for extracting cloud physical properties affect the retrievals. The fundamental principle guiding these methods is that cloud optical thickness primarily controls the reflectance of clouds at visible wavelengths and has very little dependence on particle size. Meanwhile, in near-infrared wavelengths, particle size mainly influences cloud reflectance.

In the classical retrieval from Nakajima and King (1990), they employ a visible wavelength (0.75 µm) along with one of the two absorbing near-infrared wavelengths (2.16 or 3.7 µm). This dual use of absorbing near-infrared wavelengths minimizes uncertainty in determining the effective radius, particularly for optically thin clouds. Nakajima and King (1990) products may experience retrieval failures, substantial biases, and uncertainties due to sub-pixel variability and 3D radiative effects (Cho et al., 2015). Additionally, the accuracy of these products is influenced by solar and viewing geometry.

Modern retrievals like CPP use a non-absorbing visible wavelength of 0.6 or 0.8 µm and a near-infrared wavelength of 1.6 or 3.8 µm. The cloud reflectance at 1.6 µm is

suitable for thicker clouds, whereas 3.8 µm is suitable for thinner clouds. However, the uncertainty of retrievals from 3.8 µm is high because of its closeness to the thermal emitted radiances. Also, the solar irradiance at 3.8 µm is low compared to 1.6 µm. Therefore, the CPP algorithm mainly considers 0.6 and 1.6 µm reflectances for COT, particle size, and cloud liquid water path (CLWP) retrievals.

DCOMP builds upon the bi-spectral approach of Nakajima and King (1990) and attempts to synergize the retrieval algorithm, which can be generalized over multiple sensors (such as SEVIRI and ABI) in a physically consistent manner. It can be applied to sensors measuring any of the three-channel pairs (0.6–1.6 µm, 0.6–2.2 µm, and 0.6–3.75 µm). Regarding uncertainty estimates, uncertainty estimates from the inputs are utilized to calculate uncertainty estimates of the retrieved parameters. Finally, DCOMP is also flexible regarding using auxiliary data, such as water vapor profiles from numerical weather models, which differ from standard static climatological estimates in other retrieval algorithms.

2.4 Additional auxiliary datasets

To support our investigation of mesoscale cloud structures and their association with large-scale environmental conditions (integrated water vapor (IWV), horizontal and vertical wind speed, relative humidity), we have used ERA-5 (Hersbach et al., 2020). It is the fifth generation of the atmospheric reanalysis data product from the European Center for Medium-range Weather Forecast (ECMWF). It has an hourly resolution and covers the Earth with a 31km grid. Please note that in our second study over NAT, we utilized ERA-5 to represent the best possible state of the atmosphere as it assimilates vertical atmospheric measurements from the EUREC^{4}A campaign's HALO drop sondes (Stevens et al., 2016).

The Clouds and Earth's Radiant Energy System (CERES, (NASA et al., 2017)) experiment offers satellite observation of Earth's radiation budget and cloud properties. CERES instruments are deployed on various satellites, including Terra (Kelly et al., 2014), Aqua (Salomonson et al., 2002; Barnes et al., 2003), Geostationary, and Tropical Rainfall Measuring Mission (TRMM, (Simpson et al., 1996)). By fusing data from sources like MODIS imagers on Terra or Aqua, 4-D var weather assimilation data (Dimet & Talagrand, 1986; Lewis & Derber, 1985), and measurements from geostationary satellites, CERES delivers comprehensive radiative flux information at different levels—top of the atmosphere, within the atmosphere, and at the surface. Furthermore, it provides insights into associated cloud and aerosol properties.

CERES mainly provides measurements of reflected solar radiation and emitted thermal infrared radiation to space, covering a spectrum from ultraviolet to far-infrared wavelengths. Our work uses the Edition4A SYN1deg data product, which has TOA,

surface, atmospheric radiative fluxes, and cloud properties at 1 x 1 degree regional and hourly resolution. Compared to other product releases, it is processed using consistent atmospheric reanalysis data.

Jülich Observatory for Cloud Evolution (JOYCE, (Löhnert et al., 2014)) station observations in Juelich, Germany, further support our investigation for mesoscale cloud regimes over land. Here, we use the Cloudnet target classification (Illingworth et al., 2007), which characterizes clouds by their geometrical extent and microphysical properties. It uses measurements from a Ka-band cloud radar (JOYRAD-35) combined with microwave radiometer and ceilometer observations having 30 seconds temporal resolution. To characterize the influence of shallow clouds on ground-measured shortwave radiation, we use a pyranometer (CMP21) at a one-minute temporal resolution to measure incoming SSI.

2.5 Self-supervision in general

The surge in deep learning applications within Earth system science, particularly in the earth's atmosphere, has shown a fast pace since 2020. The European Centre for Medium-Range Weather Forecasting (ECMWF) now has its ensemble of AI models called Artificial Intelligence Forecasting System (AIFS) launched in September 2023. This notable success in short to medium-range forecasting primarily involves a regression task of learning future time steps variables exploiting ERA-5 reanalysis products (Lam et al., 2023; Keisler, 2022; Bi et al., 2023).

However, other critical tasks like classification, segmentation, and sparse ground truth prediction have experienced limited advancements. For these tasks, the bottleneck lies in the scarcity of annotated data sets, a challenge influenced by factors such as the high cost of annotations, time consumption, divergent research group interests, and instances where annotations become ambiguous due to varying human interpretations. Research shows that interpreting a satellite image, such as mesoscale distribution of cloud systems, typically demands more time per image than labeling natural objects used in classical computer vision applications. For instance, shallow cloud systems over NAT were labeled at an estimated rate of an inter-quartile range of the 20s to 38s per image (Rasp et al., 2020), whereas image samples from ImageNet were labeled at an average rate of 50 images per minute (Russakovsky et al., 2015).

Deep-learning models are commonly trained through a supervised learning paradigm. The models can map inputs, for instance, a satellite image of a cyclone, to corresponding outputs, such as predicting its intensity (Maskey et al., 2020; Xu et al., 2021; Wimmers et al., 2019). This learning approach has consistently demonstrated its effectiveness in training specialized models that excel at the specific tasks they were trained for. However, apart from the practically difficult task of annotating obser-

vations (Sylolypavan et al., 2023; Rother et al., 2021), the conventional approach of supervised learning limits the feature learning capacity of neural architectures as the final layers strongly fit the distinct learning objectives (Grigg et al., 2021) and hence cannot fully utilize the rich information in the data. The annotations also come with human biases (Misra et al., 2016; Iosifidis & Ntoutsi, 2018), which influence the weights and biases during training.

Satellite datasets often lack curation, resulting in unregulated data distribution and potential imbalances within the sample space. In such cases, training neural network parameters in a supervised manner may worsen what is termed the long-tailed effect. Recent studies propose that self-supervised learning (SSL) methods may offer a more effective alternative, as they can capitalize on the uncategorized nature of the datasets (Liu et al., 2022). By doing so, these methods effectively address the long-tailed effect, facilitating the more efficient learning of robust features.

SSL, at its core, entails actively learning structures in the data through pretext tasks, which are automatically generated from the data itself. Put simply, the data acts as its own supervisor in training the model. This approach to learning draws inspiration from infant learning processes (Orhan & Lake, 2024), where observation, inborn perceptions, and environmental stimuli drive the automatic grasp of knowledge. Further, infants' immediate surroundings facilitate their learning abilities as contextual guides, which enables understanding of fundamental principles without much external guidance (Rani et al., 2023).

Early iterations of SSL introduced pretext tasks such as geometric estimation of image rotation (Dosovitskiy & Brox, 2016; Gidaris et al., 2018), grayscale conversion (Larsson et al., 2017), image inpainting (Pathak et al., 2016), and autoencoding (Donahue et al., 2017), tailored to the dataset's characteristics to produce pseudo-labels. Also, partly the success of SSL builds on the able shoulders of modern-day neural architectures such as convolution neural networks (CNN) (LeCun et al., 2015), Vision transformers (ViT) (Vaswani et al., 2023), and multi-layer perceptions (MLP) (Rumelhart et al., 1986). The design of these architectures introduces inductive biases of feature learning, such as 1) attention mechanisms in ViT, 2) CNN's kernels, or filters, which learn the spatial structures in images, or 3) MLP's effective high-level reasoning through non-linear dimensionality reduction, all together contribute to the success of SSL.

NNs typically comprise node layers, encompassing an input layer, one or more hidden layers, and an output layer. Each node is associated with a weight and threshold and is interconnected with others. Activation occurs when the output of a node exceeds the designated threshold, initiating the transmission of data to the network's subsequent layer. Otherwise, no data is transferred to the next tier of the network. We now describe these architectures in concise detail below. For more details, please refer

to their original papers.

2.5.1 Convolutional neural network

A convolutional neural network (CNN, (LeCun et al., 2015)) is a type of NN specialized in handling data with a grid-like topology, such as images. It operates through three core layers: 1) the convolutional layer, 2) the pooling layer, and 3) the fully connected layer. Different research groups have experimented with various combinations of these layers to suit their specific objectives better. In the convolutional layer, a dot product is performed between two matrices: one representing the restricted area of the receptive field (the size of the region in the input that produces the feature (Araujo et al., 2019)) and the other representing the set of learnable parameters, known as a kernel or filter.

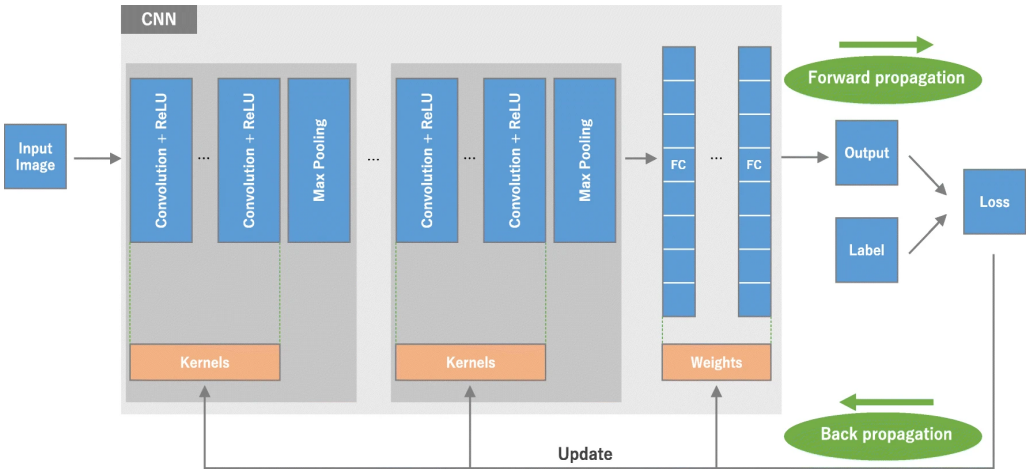


Figure 2.4: Illustrating the architecture of a convolutional neural network (CNN). Its training process entails several key components, such as convolution layers, pooling layers (such as max pooling), and fully connected (FC) layers. During training, learnable parameters, including kernel weights and biases, are adjusted based on the loss value through backpropagation. Adjusting the weights generally employs the gradient descent optimization algorithm. This Figure is taken from Yamashita et al. (2018), their Fig. 1, used under the Creative Commons Attribution 4.0 International License. (`http://creativecommons.org/licenses/by/4.0/`)

As the kernel slides over the image's height and width during the forward pass, it generates a picture representation of the receptive region. This process yields an activation map, which serves as a two-dimensional depiction of the image, showcasing the kernel's response at each spatial location within the image. The pooling layer replaces the output of the network at certain locations by deriving a summary statistic of the nearby outputs, aiding in reducing the spatial size of the representation and thereby decreasing the required computation and weights. In a fully connected layer, neurons maintain complete connectivity with all neurons in both the preceding and succeeding layers, typically computed via matrix multiplication.

CNNs excel in image processing due to their effective utilization of sparse interaction, parameter sharing, and equivariant representation. Sparse interaction is achieved by employing kernels smaller than the input image, thereby minimizing the number of parameters within the neural network. Unlike traditional neural networks where each weight in the matrix is utilized only once, in convolutional networks, weights assigned to one input are shared across other inputs to produce consistent outputs. This parameter sharing enables CNNs to recognize objects even if they appear at different locations within the image, contributing to their robustness and effectiveness in image recognition tasks.

As one layer feeds its output into the next layer, extracted features can hierarchically and progressively become more complex. However, because of many kernel operations deployed to extract and learn the essential features of an image, a CNN is generally successful in learning the spatial hierarchies of visual features in the image. Please refer to the work of Yamashita et al. (2018) for a detailed overview of CNN.

2.5.2 Vision transformer

This thesis centers on comprehending the structure and distribution of mesoscale cloud systems. Structure pertains to the intrinsic characteristics of individual cloud entities, while distribution denotes the larger pattern formed by these entities collectively. In this context, Vision Transformers (ViT, (Vaswani et al., 2017)) offer an advantageous tool as they excel in capturing global dependencies and local contextual information.

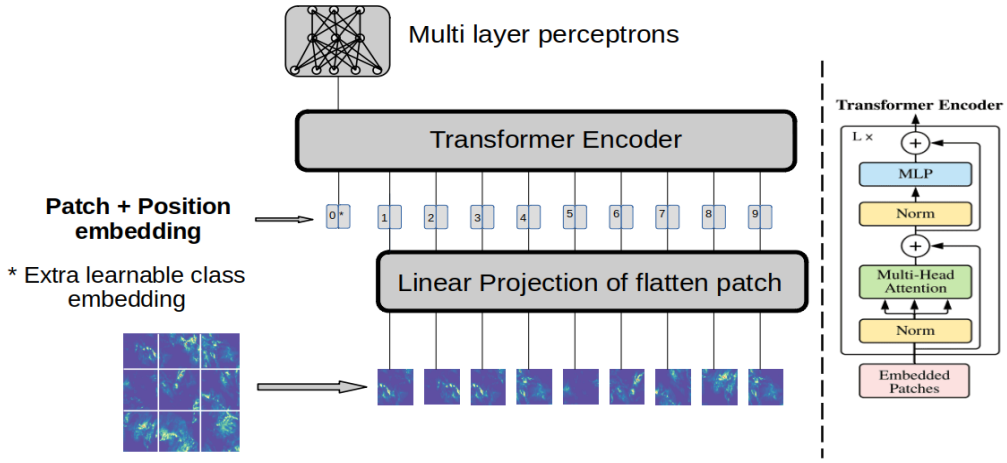


Figure 2.5: Illustration taken from Dosovitskiy et al. (2021) made openly available by Google research and adjusted for demonstrating cloud system. ViT partitions an image into patches of fixed sizes, flattens them into a 1D vector, linearly embeds each flattened patch, incorporates position embeddings, and feeds the resultant sequence of vectors into a conventional Transformer encoder. The illustration of the Transformer encoder was inspired by Vaswani et al. (2017).

ViT stands out for its notable departure from certain inductive biases found in

CNNs (refer to Section 2.5.1). Unlike CNNs, ViT relies heavily on self-attention mechanisms to process images. This approach allows ViT models to capture the intricate relationships among sequenced image patches. This flexibility is particularly advantageous as it demands minimal prior knowledge about the data structure, a departure from the more rigid architectures of CNNs as noted by LeCun et al. (2015). While CNNs are confined by their fixed receptive fields and filter numbers, transformers can attend to the entire sequence, facilitating the understanding of long-range dependencies. ViT employs self-attention to dynamically assess and prioritize various aspects of the input data during the learning process. Essentially, self-attention evaluates the relevance of each image patch within the sequence relative to others. Consequently, when provided with abundant data, ViT's self-attention mechanism outperforms CNNs in addressing various vision-related challenges (Goyal et al., 2022). For a deeper understanding of ViT, readers are encouraged to explore the comprehensive work by Khan et al. (2022).

3 Understanding mesoscale cloud systems

Chatterjee, D., C. Acquistapace, H. Deneke, and S. Crewell, 2023: Understanding Cloud Systems' Structure and Organization Using a Machine's Self-Learning Approach. Artif. Intell. Earth Syst., 2, e220096, https://doi.org/10.1175/AIES-D-22-0096.1.

The content of this chapter has been published under the Creative Commons Attribution 4.0 License. Formatting changes were made to adopt the format of this thesis.

Personal contribution: The idea to study cloud systems over Europe and associate them with solar energy was presented by Prof. Susanne Crewell. The study was designed by myself. Hartwig Deneke provided the satellite dataset. I wrote the manuscript with helpful suggestions from all the co-authors and with the graphic color scheme help of Claudia Acquistapace. Furthermore, three anonymous reviewers gave valuable input, which helped me improve the graphics and structure of the manuscript that I carried out.

Abstract

In this study, we introduce a self-supervised deep neural network approach to classify satellite images into independent classes of cloud systems. The driving question of the work is to understand whether our algorithm can capture cloud variability and identify distinct cloud regimes. Ultimately, we want to achieve generalization such that the algorithm can be applied to unseen data and thus help automatically extract relevant information important to atmospheric science and renewable energy applications from the ever-increasing satellite data stream. We use cloud optical depth (COD) retrieved from post-processed high-resolution Meteosat Second Generation (MSG) satellite data as input for the network. The network's architecture is based on the DeepCluster version 2 and consists of a convolutional neural network and a multilayer perceptron, followed by a k-means algorithm. We explore the network's training capabilities by analyzing the centroids and feature vectors found from progressive minimization of the cross entropy loss function. By making use of additional MSG retrieval products based on multi-channel information, we derive the optimum number of classes to determine independent cloud regimes. We test the network capabilities on COD data from 2013 and find that the trained neural network gives insights into the cloud systems' persistence and transition probability. The generalization on the 2015 data shows good skills of our algorithm with unseen data, but results depend on the spatial scale of cloud systems.

SIGNIFICANCE STATEMENT: This study uses a self-supervised deep neural network to identify distinct cloud systems from cloud optical depth satellite images over central Europe. Satellite-retrieved products support the physical interpretation of the identified cloud classes and help optimize the number of identified classes. The trained neural network gives insights into cloud systems' persistence and transition probability. The generalization capacity of the deep neural network with unseen data is promising but depends on the spatial scale of cloud systems.

3.1 Introduction

Cloud system appearance defined here as structure and distribution results from a complex interplay of physical processes and large-scale environmental conditions. Since the ground-breaking work by Howard (1803), there has been considerable interest in simplifying the complexity of cloud appearance and classifying cloud systems. With the availability of geostationary satellites providing images over large domains with ever-increasing temporal, spatial, and spectral resolution, the need for automatic classification becomes more and more imminent. A prominent pixel-based classification is the International Satellite Cloud Climatology Project (ISCCP) (Rossow & Schiffer, 1999), which defines nine cloud types based on thresholds in cloud optical depth (COD) and cloud top pressure (CTP).

Extending pixel-based cloud classifications to a regional context, ISCCP has also established the concept of cloud regimes - sometimes also referred to as weather states – for analyzing cloudiness at regional scales. This approach was introduced by Jakob and Tselioudis (2003) and considers the joint 2-dimensional histograms of cloud optical thickness and cloud top pressure across a domain as a basis. Representative mean histograms for the different cloud regimes are obtained by k-means clustering (see MacQueen (1967)). The method has subsequently been refined and extended to global scale analyses (Tselioudis et al., 2013), applied to other satellite instruments such as the Moderate Resolution Imaging Radiometer (MODIS) (Oreopoulos et al., 2014), and has been used for climate model evaluation (Williams & Webb, 2008; Jin et al., 2016; Tselioudis et al., 2021). Recently Tzallas et al. (2022) has introduced a cloud regime dataset for Europe based on Meteosat Second Generation satellite observations and the CLAAS2 climate data record (Benas et al., 2017). In a more fitting variety of contexts, our work uses cloud regimes, cloud classes, and cloud patterns interchangeably.

However, one major shortcoming of this approach is that using the joint 2-dimensional histograms of cloud properties completely neglects the spatial structure and organization of clouds within the considered domain. For cloud systems over oceans in the trade-wind regimes, some preliminary indications from recent research by Bony et al. (2020) showed how different mesoscale cloud patterns impact the top-of-atmosphere radiation budget differently and thus the local climate. Therefore, assessing the organization of cloud systems is important to understand regional climate (changes) but also has potential for other applications such as renewable energy.

Various authors also proposed to use of machine learning algorithms to perform the cloud classification task. Visa et al. (1998) suggested an unsupervised approach making use of Self-Organizing Maps (SOM) using five channels from Advanced Very High-Resolution Radiometer (AVHRR) in polar orbit, which were fine-tuned by the Learning Vector Quantization (LVQ). However, in that approach, the uncertainty in the cloud classification was high, and there was ambiguity in the physical understanding of the classification (Tian et al., 1999). Fabel et al. (2022) used a self-supervised approach to classify clouds from ground-based all-sky images and labeled the clouds according to low, medium, and high cloud base height. However, when prescribed to identify specific cloud features, the neural network does not fully utilize the rich representations present in the data as it focuses on a narrow objective. To overcome the issue of a missing "ground truth," Stevens et al. (2020) tried to understand cloud systems through visual inspection from scientific peers labeling satellite images identifying four major patterns. Rasp et al. (2020), through a crowdsourcing effort, asked a group of scientists to label multiple independent and overlapping cloud organizations present

in visible images of roughly 2,000 km domain into these predefined classes. Their study showed that the average participant needed 30 seconds to classify one image, and agreement over cloud patterns was only reached in 37 percent of cases, showing significant disagreement among scientists for the labeling task. Training a neural network with limited classes can also suppress the information on relevant other cloud regimes. Thus, Denby (2020a) used unsupervised machine learning where input image combinations and loss function train the neural network's parameters to learn the similarity between the same images again over the tropical North Atlantic. He showed that the identified classes are physically independent, supported by spectral bands representative of the physical properties. However, the approach involved human decisions in their learning stage, like the triplet selection of cloud images for training purposes. Additionally, in contrast to crowdsourcing and Denby (2020a) approach, (Kurihana, Moyer, Willett, et al., 2022; Kurihana, Moyer, & Foster, 2022) used learning of rotation-invariant dimensionally reduced features through unsupervised auto-encoders to generate a unique 22-year long record of global cloud classification with 42 clusters from Moderate Resolution Imaging Spectroradiometer (MODIS). However, their dataset covers ocean regions only.

In this work, we want to investigate the suitability of a self-supervised deep neural network to identify distinct cloud systems over central Europe. In contrast to the tropical oceans, the central European domain is characterized by more synoptically driven cloud systems, strong land surface type, and topography variability that might hamper our endeavor. We use a resolution-enhanced cloud optical depth (COD) product following Deneke et al. (2021) derived from Meteosat Second Generation (MSG). This can act as a proxy for future improvements in the spatial resolution by Meteosat Third Generation (MTG) observations, launched at the end of 2022. While other approaches directly used RGB images (Denby, 2020a) or infrared brightness temperature images (Schulz et al., 2021a) and thus relied directly on the level-1 satellite observations, we opt for COD and thus a level-2 product as input. The choice of COD is beneficial, as it is a cloud parameter that should be invariant to other influences, as it, e.g., avoids the influence of the diurnal cycle evident in the reflected radiances measured by the satellite. Furthermore, COD has the highest accuracy from all the retrieved parameters in the satellite data set (Deneke et al., 2021) and physically relates to the attenuation of clouds being the quantity that modulates solar energy yields most strongly. However, it needs to be noted that many of the boundary layer clouds driving the variability in solar energy might still not be fully resolved with the spatial resolution of our product.

The main challenge we tackle in the present study is whether deep neural networks can understand different cloud regimes based on the structure and distribution of COD and how this helps explore the cloud systems' physical properties.

We can frame four guiding research questions to detail our work better:

1. How can we use deep learning to capture cloud variability from COD images to identify different cloud regimes?
   Satellite images contain rich visual information, which we transfer into the feature space from which a given number of classes is extracted.

2. How can we physically interpret the identified classes?
   In the traditional computer vision domain, self-supervision is often used as pre-training, followed by fine-tuning with some labels as a downstream task for vision applications, e.g., for medical image classification (Azizi et al., 2021). Without using labels to classify the scenes, we use information obtained by cloud and radiative properties retrieved from the multi-channel MSG images to interpret the physical meaning of the classes.

3. What are the measures to identify the correct number of classes?
   Our principle for selecting the correct number of classes is that the right number of classes should be the best in identifying distinct properties for the classes in the physical parameter space avoiding correlation between classes.

4. Can the model generalize well on unseen data?
   In deep learning, the true capability of a trained model is realized on unseen data, so testing the model on unseen data (i.e.)., another year in our case, will ultimately reveal the network's performance.

One of the inherent challenges of our study is the closeness of the cloud system patterns as they transit from one form to another. Cloud systems are generally chaotic, meaning their transformation from one type to another can be quick and transient. We want the neural network to understand this continuous relationship and to generate an understanding of this connection in the feature space, i.e., the n-dimensional space where the features of the cloud fields are collected.

An additional challenge is that the traditional datasets in computer vision, such as ImageNet (Russakovsky et al., 2015), are prepared in a controlled environment and have a well-balanced class distribution. In contrast, the satellite images are uncurated as their distribution space is not controlled. It is thus very likely that the classes are not well balanced and hence have a distribution sample imbalance among the categories, also known as the long-tailed effect. Under this effect, during training, the major cloud systems present in the satellite measurements drive the decision boundary of the neural network. Recent efforts have shown that Self-supervised learning (SSL) has a good tackle on the imbalanced learning problem (Liu et al., 2022). The learned features are more robust and easy to generalize as the SSL extracts more comprehensive information from the data space.

The present work has the following structure: section 3.2 describes the satellite dataset and our working domain. Section 3.3 provides details on the deep learning architecture, section 3.4 presents our results that aim at answering the questions posed above, and finally, section 3.5 concludes on our results and points at future work. In the following work, cloud regimes, cloud classes, and cloud patterns have been used interchangeably.

3.2 Satellite dataset

Geostationary satellites provide high-spatial and temporal resolution monitoring of the Earth with ever-increasing capabilities and, thus, data volume. For Europe, the Meteosat Second Generation satellites located over the equator and the prime meridian offer full-disk multi-spectral images at 15 min temporal and at 3 x 3 km^2 nadir spatial resolution and in 11 narrow-band spectral channels since its launch in 2002 (Schmetz et al., 2002b). To mimic the capabilities of the upcoming Meteosat Third Generation (MTG) mission, we use a Meteosat Second Generation (MSG) cloud dataset, which features an improved spatial resolution (Deneke et al., 2021). Herein, the Meteosat - 9, Spinning Enhanced Visible and InfraRed Imager (SEVIRI) is used, which is positioned at 9.6°E as backup satellite and provides the so-called rapid scan service at 5 min repeat cycle. Note that within the operational Meteosat program, Meteosat - 11 has taken over the rapid scanning at 9.6° E, and Meteosat - 9 has drifted over to 45.5° E. Channels 1 - 3 (λ = 0.635, 0.810, 1.640 μm) are used in combination with the broadband high resolution visible (HRV) channel (λ = 0.4 - 1.1 μm) to retrieve and downscale COD and effective cloud radius (r_e) from the reflected solar radiation observations. The retrieval is based on the Cloud Physical Properties algorithm developed at the Royal Meteorological Institute of the Netherlands and applied within the context of the Satellite Application Facility on Climate Monitoring (CMSAF) for the generation of the CLAAS-2 cloud data record (Roebeling et al., 2006). Together with COD and r_e, other cloud parameters are retrieved, such as liquid water path (W_L) and droplet number concentration (N_D). The data set also includes cloud top pressure (CTP) derived from thermal infrared channels via radiative transfer calculations, but this dataset does not benefit from the resolution enhancement. All products have 2 km x 1 km spatial and 5 min temporal resolution over central Europe.

We decided to use COD as input for our self-supervised approach because (i) the use of COD instead of reflectance(s) avoids the consideration of the solar zenith angle and surface characteristics, which have a strong influence on the reflectance, (ii) COD is the most accurate retrieval product (Deneke et al., 2021) at the enhanced resolution, and (iii) COD is related to surface radiation and thus important for solar energy application which we aim to target at a later stage. Nevertheless, it is important to understand the magnitude and impact of retrieval

uncertainties. The bi-spectral retrieval method used for estimating the COD relies on the principles described by Nakajima and King (1990) and is highly non-linear. A variety of uncertainty sources affect the overall accuracy, as is discussed in great detail in Grosvenor et al. (2018). Specifically, retrieval results are affected by uncertainties of the instrument itself and its calibration, atmospheric correction, and assumptions on surface albedo. Errors in COD become excessively large outside a value range of about 5 - 50 but are expected to lay below 8 - 10% within this range (see Platnick et al. (2017)'s Figure 14). This estimate is valid for warm clouds as observed by the MODIS instrument. Given the efforts directed to validation and instrumental calibration within the context of the Satellite Application Facility on Climate Monitoring (see Benas et al. (2017); Meirink et al. (2013)), a similar accuracy is expected for the MSG-based retrievals of COD used here. It must, however, be noted that this uncertainty estimate does not take into account the effects of unresolved cloud variability, retrieval uncertainties arising for mixed-phase and ice clouds, as well as 3-dimensional radiative effects. Given the focus of the present paper to classify cloud scenes into characteristic regimes based on the COD, there is hope that such uncertainties are regime-dependent and implicitly taken into account by the classification.

We focus on a central European domain (Fig. 3.1), which exhibits a wide variety of clouds associated with synoptic events and local forcings. While we intentionally apply a machine-learning (ML) based cloud classification over land surfaces for the first time, we exclude the Alps' topography, where the reflective snow can complicate cloud retrievals. The study period includes spring and summer, i.e., from April to July, to capture more variability in the cloud conditions, including locally triggered low clouds, which pose a challenge to numerical modeling and solar energy production due to their high variability. We avoid dusk and dawn by limiting the period to 6 am to 6 pm UTC. The year 2013 is used for training purposes, and the year 2015 to test the general applicability of the approach.

Ad-hoc, it is unknown which spatial scale is most meaningful for classification. Here we use images of 128 x 128 pixel size corresponding roughly to 260 km x 135 km, where cloud systems distributions up to the mesoscale are well represented. This choice is consistent with previous studies that used domain sizes around 250 km x 250 km (Denby, 2020a) or 1 x 1 to 2.5 x 2.5 degrees (Jakob & Tselioudis, 2003; Oreopoulos et al., 2016). We also test a smaller image size of 64 x 64 pixels, roughly 130 km x 68 km, to study the influence of image size. At any instant, we crop four random images from our parent central European domain (Fig. 3.1). Therefore, with 144 parent images per day for about 120 days in 2013, we obtained 68,000 images per year. We keep 10,000 images unseen from the model for evaluation purposes. Finally, we also extract data over the specific location of Juelich to investigate the temporal variability and the transition from

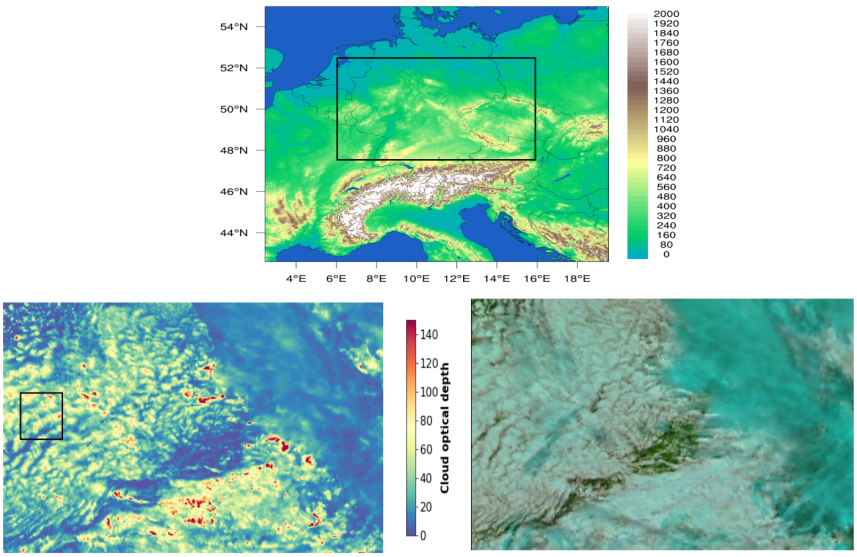


Figure 3.1: Top: Geographical map of Europe shown along with topography, thick black lines are coast, the thin black lines are political boundaries, and the black box shows the central European domain. Bottom left: MSG high-resolution cloud optical depth of the central European domain on May 22, 2013, at 13:30 UTC. Four images from 128 x 128 and 64 x 64 configurations are randomly cropped out at every time step. The boundary box inside the COD field represents a 128 x 128 area over Juelich, Germany; cropped at every time step is not used during training but is specifically used for the transfer learning task. Bottom right: The central European domain's natural color RGB image at standard resolution is shown for better context. Note that cyan colors in the RGB indicate ice.

one class to another. The data over Juelich is not used during training and is specifically used for the transfer learning task. Also, data from 2015 are left out in training to investigate the generalization.

3.3 Deep learning architecture

Recent years have seen enormous progress in representation learning by the deep learning community (Caron et al., 2020; van den Oord et al., 2017; He et al., 2020). They facilitate the intelligent use of uncurated images with modest computational requirements. In our case, we want the neural network to learn distinctiveness in the distribution pattern of the cloud systems. The network should learn cloud systems' global and local patterns and distribution properties from satellite images. As cloud systems change continuously, the neural network should learn the features of the satellite images such that it can discriminate at an instantaneous level.

To set up this architecture, we make use of the software package DeepCluster version 2 (named DCv2 in the following) from Facebook Artificial Intelligence Research (FAIR) (Caron et al., 2019, 2020). The open-source VISSL computer vision library (Goyal et al., 2022) was used to adapt the DCv2 neural network to

our requirements. Based on sensitivity tests, we set the number of epochs always to 800, and depending on the available GPU type, batch size is set to 750 or 800.

For the classification task, we opted for a self-supervised approach because, during the representation learning phase, it learns more comprehensively and freely from the data distribution. In our self-supervised method, the network learns directly from the satellite COD observations from an initial dataset of 58,000 images from 2013. The network ingests data in batches, i.e., small groups of images, and for each epoch, the entire 58,000 image dataset makes the complete pass in 72 batches. For each batch, the network adjusts its parameters through stochastic gradient descent to achieve the minimization of the cross-entropy loss function. After 800 epochs (see above), the parameters are optimized. The network can now work with a dataset of COD images the model has not been trained on, and it will assign new images into k categories (or cloud classes). The user decides the number of categories k for clustering clouds; for this study, we varied k between 4 and 12. Our experiment is guided by the well-established ISCCP classification with nine cloud classes and the clear sky category. Thus, we started with k = 10. To investigate the sensitivity of cluster numbers, we run the experiment also with a higher class number, such as k = 12, and reduce them until k=4. The optimum number is investigated in section 3.4.3.

3.3.1 Definition of the network input

N satellite images of COD built the input training data set $X = \{x_1, x_2, x_3....., x_N\}$ of the deep learning architecture illustrated in Fig. 3.2. For robust and effective global mapping, image features are learned separately from two slightly differing sub-images. The physical reason behind this approach is that, since physical laws guide the appearance of the cloud systems, the neural network should know that the two random crops almost represent the same system. Thus, for creating two sub-images, each satellite image x_i is transformed using random cropping to obtain two global views x_{it} of the original image (x_{it1} and x_{it2} in Figure 3.2). Each x_{it} covers 0.75 part of x_i.

3.3.2 Network general architecture

The neural network's task is to learn visual features from each satellite image. A function f describes the learning process via a nonlinear mapping $f(x_i) = z_j$ with $i = 1...N, j = 1....M$ that maps the image x_{it} into the array of features Z $=\{z_1, z_2, z_3....., z_M\}$, where M is the dimension of the feature array Z and in this study, the chosen dimension M is equal to 128. The randomly cropped image x_{it} has dimensions 96 x 96 (0.75 of 128 x 128) or 48 x 48 (0.75 of 64 x 64). The function f represents the transformations performed by the network's architecture components: a convolutional neural network (CNN) as the backbone and a multi-layer perceptron (MLP) as the head. The CNN can identify spatial

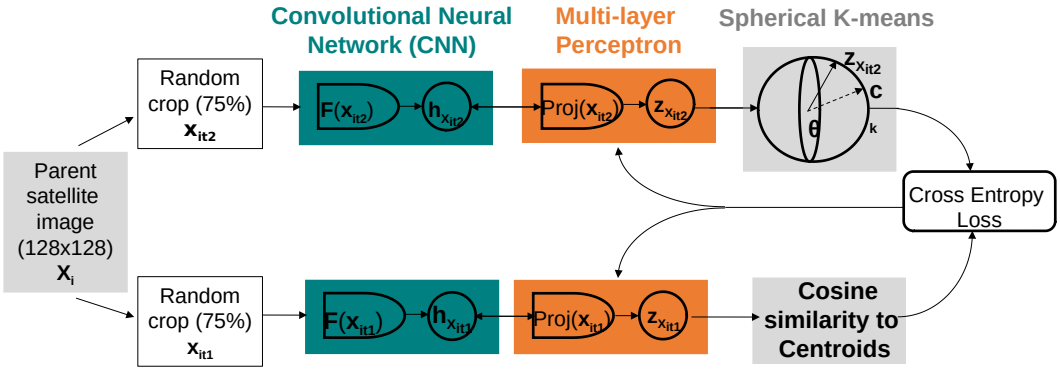


Figure 3.2: Schematic diagram of the deep learning network adapted in this work from VISSL library. From left to right: X_{it} is the random crop from X that is split into two similar subsamples $x_{it1/2}$. F_x represents the convolutional neural network (CNN) and h_x its output. For the CNN, the ResNet-50, a residual neural network architecture (He et al., 2015) 50 layers deep with 23.521 million parameters is used in our setup to learn essential features h_x from the high dimensional satellite images. Then, Proj(h_x) represents a multi-layer perceptron (MLP). The MLP attached to the CNN output, comprises three fully connected layers; the first two have 2048 neurons and the last layer with 128 neurons. To activate the neurons, the rectified linear activation function (ReLU) (Fukushima, 1975) is used. The feature vector z_x represents the output from MLP. The convolution of the functions F_x and Proj(h_x) is the function f, representing the non-linear transformations performed by the entire network's architecture. The regularized architecture limits the latent space's information capacity and learns the cloud systems' representation, minimizing the multinomial cross-entropy loss.

dependencies in an image by learning relevant filters. Its output kernel values are then mapped by the MLP nonlinearly into the feature array Z (Fig. 3.2 for more details). Figure 3.2 represents the network's architecture, and the two branches graphically indicate the paths that the global views x_{it} follow in the network. It is important to note that the parameters shared by CNN and MLP of the upper and lower branches are the same.

3.3.3 Upper branch of the network

In the upper branch (Fig. 3.2), applying the function f to the image x_{it2}, we obtain the feature vector Z $\{z_1, z_2, z_3....., z_M\}$ that then enters the Spherical k-means algorithm (Hornik et al., 2012), that clusters the data in the following way. The algorithm is initialized with random centroids $C = \{c_1, c_2....c_k\}$ of dimension M. The number of centroids provided to the Spherical k-means corresponds to the number of classes in output k, decided by the user. Each centroid C has a label q_l

with $l = 1....k$. The algorithm identifies the closest centroid to the feature vector Z through cosine similarity and assigns Z a pseudo-label q_s corresponding to the selected closest centroid C_s. q_s is a pseudo label because it is not the definitive label for the feature vector we will obtain at the end of the iterations as it changes with epochs to reduce the cross entropy loss. Thus, through the upper branch, the network assigns a preliminary label q_s corresponding to a centroid C_s to the feature vector Z obtained from the image x_{it2}.

3.3.4 Lower branch of the network and cost function minimization

In the lower branch, the function f is applied to the image x_{it1}, similar to the image x_{it2} that went in the upper branch. The function f outputs a feature vector Z, which is first compared with the centroids of upper branch C by calculating a normalized dot scalar product. Then, Z is provided as input to the cross entropy loss function, together with the output of the upper branch C_s, q_s. The network now minimizes the multinomial logistic loss function L given by:

$$L(Z, q_s, C_s) = -\sum_{l=1}^{k} q_l log(p_l) \tag{1}$$

Where p_l is the normalized exponential of the cosine score for the class l divided by the sum of the cosine scores for all the classes and represents the probability distribution of the cosine score:

$$p_l = \frac{\exp(\frac{1}{\zeta} Z^T C_l)}{\sum_{m=0}^{k'} \exp(\frac{1}{\zeta} Z^T C_m)} \tag{2}$$

Note that the top and bottom branch between the Spherical K-means block and the cosine similarity is somehow connected as the centroids come from the k-means of the top branch. In equation 2, ζ is a normalization factor called temperature that controls the sharpening of the p_l, and $Z^T C_l$ represents the scalar product between the vectors Z and C_l. We have taken it as ζ from the benchmark configuration in VISSL. When the feature vectors of the two branches capture similar information about the original image, the loss becomes lower and higher when they diverge. That's how the network branches are encouraged to focus on the image characteristics, which progressively makes the feature vectors similar. The loss function minimization happens progressively layer by layer, derivating the loss function L with respect to Z parameters and adjusting parameter values in each layer by backpropagation. At the end of the minimization, we obtain a configuration θ of the parameters of the network that will be ready for a new iteration (batch) over the two branches. To investigate the suitability of the DCv2 model for cloud classification, we changed the loss code to obtain the distance of the feature vectors from the centroids, the features of the centroids, and the class

category or the pseudo labels of each image sample. Furthermore, we made some changes in the nearest neighbor test algorithm in the VISSL library to find the nearest neighbors from the centroids.

3.3.5 Network learning behavior

To better understand how the computational model works internally and learns progressively from subsequent epochs, we extract the features at different levels of training to visualize and better comprehend the progress in the clustering around the k classes provided as input. Figure 3.3 shows the ability of a self-supervised approach to extract the rich representations in the dataset. In our case, we do not have labels to measure the accuracy of the scene classification task. Therefore we assess the quality of the learning algorithm by looking at the dimensionally reduced feature vector.

To represent the M-dimensional feature vector Z, in this case, we used the t-distributed Stochastic Neighbor Embedding (tSNE) algorithm (van der Maaten & Hinton, 2008a), which tries to preserve the relative position of points while mapping it to lower dimensional space. Here, we map the 128 dimensions into two, and for the visualization, in Figure 3.3, we prescribe the number of classes to ten (k=10).

The neural network ingests the data at the first epoch and performs spherical clustering for the first time using the embeddings (synonym for feature vector Z). The model, in this way, generates the first idea of a few different class distributions (for instance, pink and yellow dots in Fig. 3.3a). At epoch 25, spirals of data points are clustering together (Fig. 3.3b). To interpret self-supervised learning using an energy-based regularized framework (LeCun & Huang, 2005), an energy function, the cross-entropy loss here, assigns low energies to training samples. It uses a regularizer, spherical k means here, and the loss function pushing the energy down along the data points that minimizes the volume of the low energy regions. In such context, Ranzato et al. (2007) showed that compressing the information to represent the data can help reduce the training points' energy and maximize the unobserved points' energy. We observe a spiraling behavior supporting the energy-based model theory. At epoch 250 (Fig. 3.3c), the model achieves convergence, and near epoch 800 (Fig. 3.3d), a clear segregation among the classes appears with maybe some outliers (like in yellow dots of Fig. 3.3d). The good performance of the network displayed convinced us to stop the training at 800 epochs.

3.3.6 Optimization for the convergence

We used stochastic gradient descent (SGD) (Ketkar, 2017) to optimize the training process with momentum as compared to gradient descent which uses the average gradient of all training samples. SGD allows a slightly different loss curve

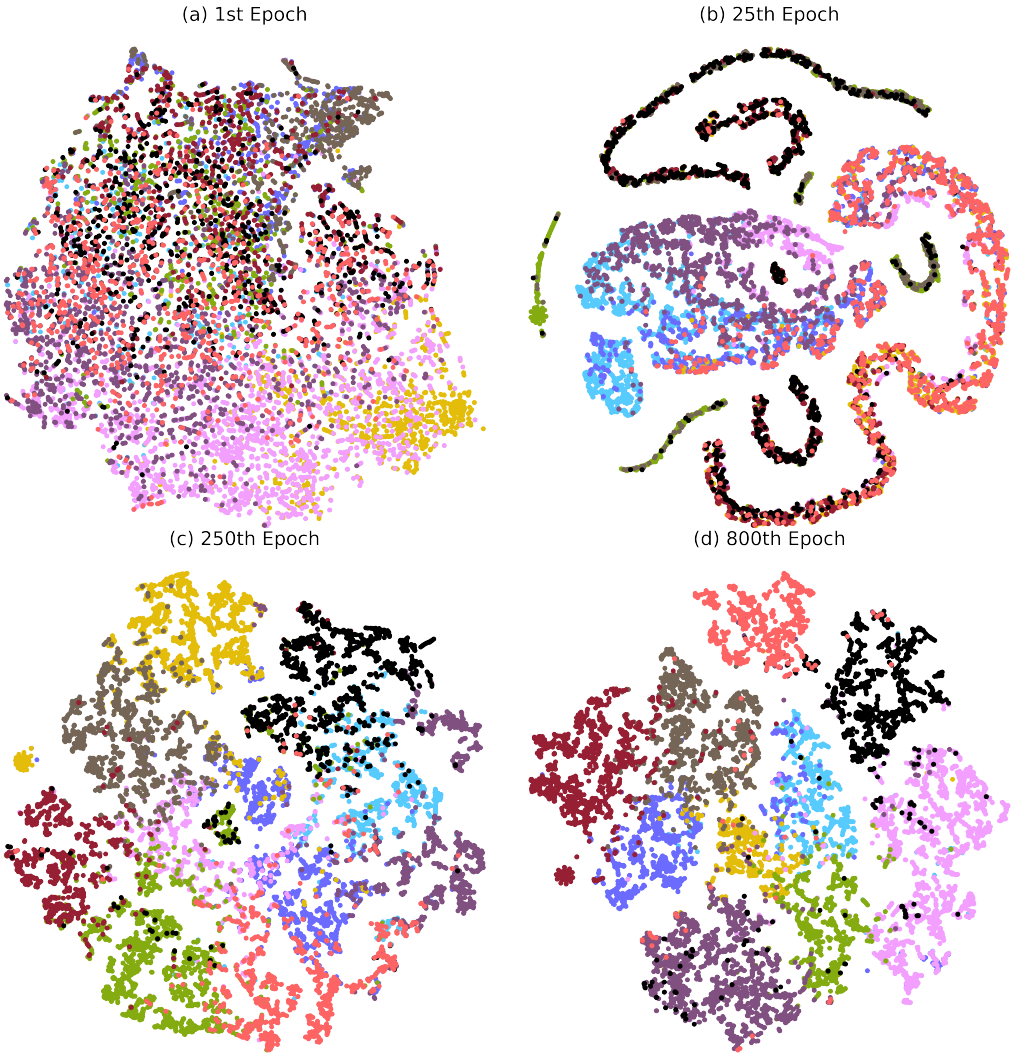


Figure 3.3: Progressive visualization of the dimensionally reduced feature space for ten classes(color coded) and sub-samples size of 15000 at a) 1 epoch, b) 25 epoch, c) 250 epoch, d) 800 epoch by the neural network. The perplexity used for t-SNE runs is 30.0, and the epsilon derived from auto-configuration is 1150.

with a randomly selected mini-batch, and this helps to reduce the computations enormously. Using momentum in the SGD lets the past and the current gradients guide the overall direction of the loss curve. In addition, DCv2 uses a learning rate scheduler that modifies the learning rate based on the progression in the training process. We use the learning rate scheduler of linear warm-up plus cosine with the linear start value of 0.3, end value of 4.8 and cosine start value of 4.8 end value of 0. VISSL supports mixed-precision training from NVIDIA's Apex library to reduce the model memory requirement and better training speed. We have used mixed precision to save GPU memory and thus allow larger batch-size training. Therefore, NVIDIA's Large Batch Training of Convolutional Networks (LARC) and SGD are used to improve large batch size training convergence. To avoid collapse, DCv2 applies batch normalization for stability apart from the cluster-

ing constraint. The task of the batch normalization layer is to take outputs from the last layer and normalize them before passing them as input to the next layer. For multi-GPU training, synchronized batch normalization normalizes the input within the mini-batch because it helps share the statistics across all the devices. For conversion to global batch normalization, VISSL supports NVIDIA's Apex SyncBatchNorm. We have used four NVIDIA V100 GPU'S for training purposes. For training the neural network for 800 epochs on 4 V100 GPUs it took 16 hours or 64 core hours in total.

3.4 Results

3.4.1 Analysis of the clustering capability

Once the network has sorted the individual images into different classes, we want to check the consistency within a single class. Thus, we quantify how close to each other the evolving cloud patterns are in the feature space. We can check how images from the same cluster that are spatially close or far in the feature space are visually similar to gain confidence in the quality of the representation of the different cloud patterns in the feature space. For this analysis, for each centroid, we select the closest image to a centroid (called centroid in the following) over Central Europe (Fig. 3.4) and compare them with their two closest and their two farthest images. The seven centroids from Figure 3.4 show distinct COD patterns; however, the physical separation of identified classes and their interpretation will become clearer in the later sections. Centroid 2 is associated mainly with clear sky conditions, and centroid 3 has optically thin clouds. Centroids 4 - 6 reveal extended cloud fields with a mean cloud fraction higher than 94% though their distribution differs, with centroid 6 having the highest COD values representing intensive convective cells. The images closest to the centroids resemble their behavior, while the ones farthest away sometime deviate to a larger extent. That indicates the distance to a centroid can be used to measure the classification's quality. It is worth noting that Figure 3.4 visually demonstrates the distinctiveness and separation of the features learned within a single class from the cloud systems found in other classes. Additionally, to quantitatively assess the quality of our embeddings in seven clusters, we calculated the silhouette score (Rousseeuw, 1987a) for our high-dimensional and complex dataset, yielding a score of 0.54. This score ranges between -1 to +1, and low or negative values indicate that the clustering configuration may have too many or too few clusters. Thus, the mean score of 0.54 for all the embeddings shows that clusters are considerably compact and non-overlapping, supporting our qualitative assessment. Therefore, scores like silhouette can also be used to find the optimal number of clusters, as discussed in section 3.4.2. Further analogies in the cloud patterns will be investigated in section 3.4.4, where we aim at a physical interpretation.

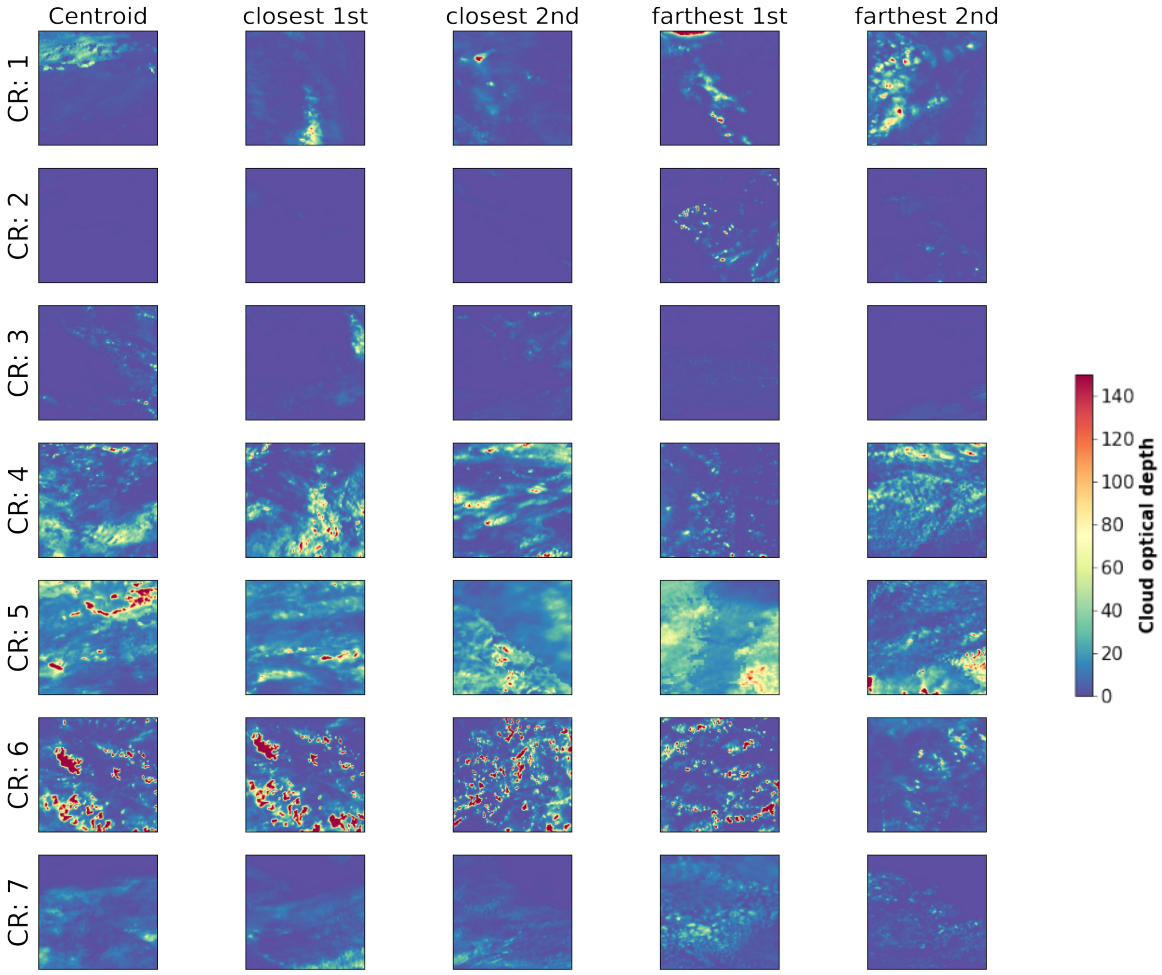


Figure 3.4: Each row shows five 128 x 128 COD images belonging to a certain cloud regime (CR) over central Europe. From left to right: centroid, closest 1^{st}, closest 2^{nd}, 2^{nd} farthest, farthest are shown along with the color scale for COD reference purposes.

3.4.2 Determination of the optimal number of classes

As stated in section 3.3, users provide the number of classes k utilized by the network to cluster the data. We want to understand, for each configuration, what is the optimum number of classes. This number should be as low as possible, but at the same time, it should separate the radiative and physical properties of different clouds. We use for this purpose the spatial distribution of COD and CTP as retrieved from MSG. In this way, we follow Oreopoulos et al. (2016) who distinguished cloud regimes based on COD - CTP using k-means clustering methods, starting from a core ensemble of six cloud regimes and then verifying if each regime can be split into sub-regimes. To split a regime into sub-regimes, they evaluate the differences in the spatial pattern correlation coefficients between their centroids (mean joint histograms) to some threshold values.

Here, we use a bottom-up approach, starting from varying amounts of cloud regimes (from 12 to lower cluster numbers). We check the correlation in the joint histograms of COD and CTP and also look at sensible distributions of the

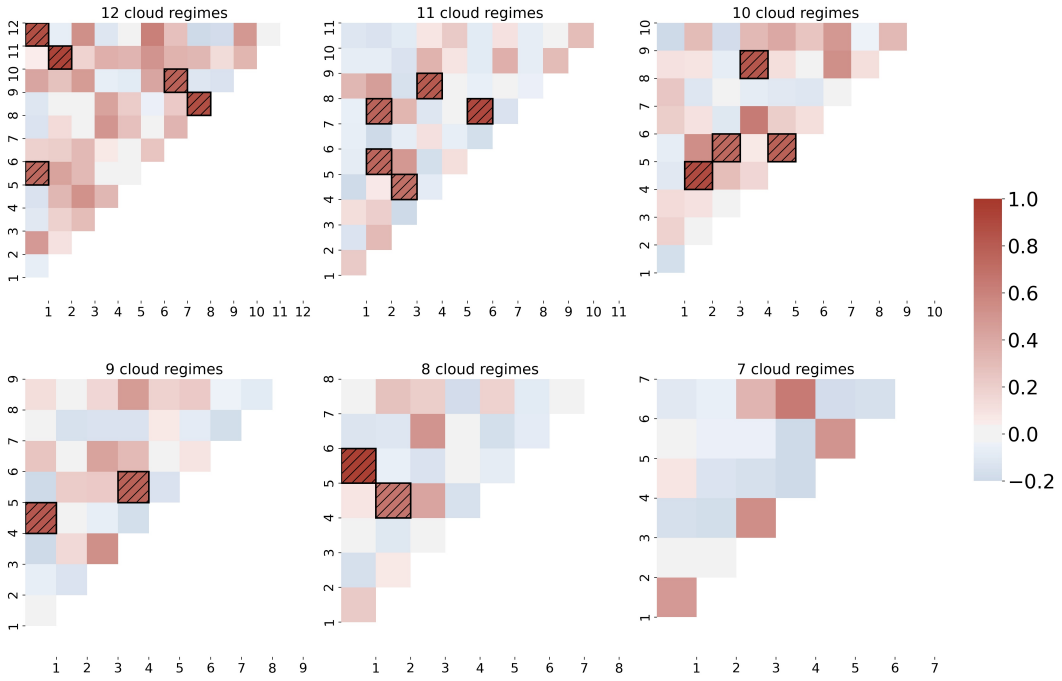


Figure 3.5: Correlation patterns among classes for the number of cloud regimes varying from 12 to 7 for 128 x 128 configuration. Red squares with hatches indicate the classes with a correlation of higher than 0.65 in COD-CTP space. Note that for smaller numbers of cloud regimes, no significant correlation appears. A similar figure for the 64 × 64 configuration is shown in appendix 3.6

mean cloud fraction and the relative frequency of occurrence (shown in Fig. 3.6) over the central European domain (more details on the joint histograms can be found in section 3.4.4). Based on this, we define the following thresholds for the correlations: 0.65 for the 128 x 128 and 0.7 for the 64 x 64 sample images. A correlation between two clusters higher than a threshold value would indicate that clusters are not sufficiently independent. In fixing the threshold, we got inspired by Oreopoulos et al. (2016), who selected a threshold of 0.6, but in this work, the thresholds are slightly more restrictive. Additionally, for 64 x 64, since the domain is comparatively smaller than the 128 x 128 sample images, it has more variability in terms of cloud system pattern and distribution. Here we restrict with a relatively higher threshold of 0.7.

For the 128 x 128 configuration, Figure 3.5 shows that the configuration with seven classes is the first one in which the correlation among the classes is never larger than the prescribed threshold. The number of classes whose correlation is higher than the threshold (red squares in Fig. 3.5) gets reduced if we decrease the number of cloud regimes from 12 to 7. The network can distinguish seven classes, thus the largest number of classes, without producing some superposition of the cloud radiative properties. In conclusion, we find the optimal number of classes to be 7 for the 128 x 128 configuration and 8 for the 64 x 64 configuration.

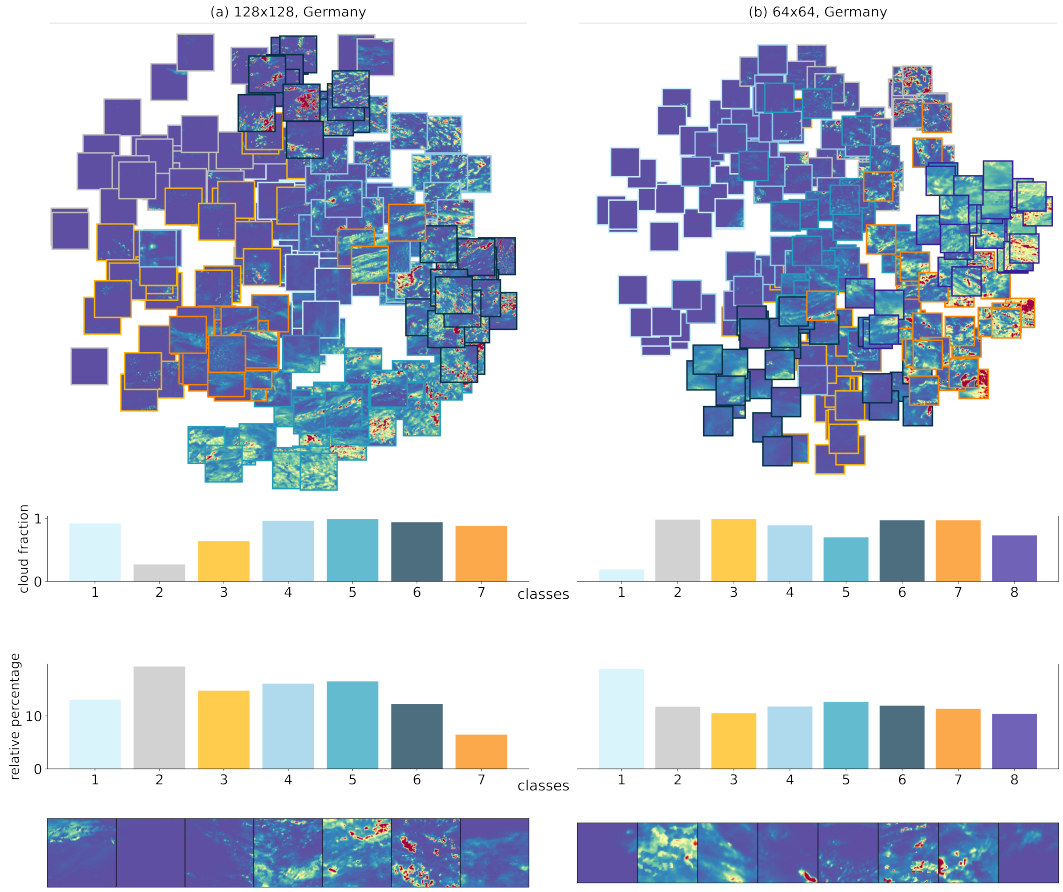


Figure 3.6: Classification by the deep neural network at a) k = 7 for 128 x 128 configuration and b) k = 8 for 64 x 64 configuration. Thirty random samples were selected over central Europe and visualized for each class. Each image is assigned a unique class through a colored frame. The bar charts in the lower part of a) and b) represent the number of images in relative percentage and cloud fraction in each regime. To better associate the class number with the cloud regime, the centroid image is shown as well.

3.4.3 Optimum classification

Now we want to investigate how the clustered satellite images arrange themselves in the representation space. Our goal is to see if, giving k classes (as determined in section 3.4.2) as a clustering constraint, the features from the satellite images are strong enough to be captured by the deep neural network for a qualitative classification of distinct types of clouds. Since visualizing the feature vector Z is impossible because it has 128 dimensions, we reduce the dimension of the feature array to a 2D space using the tSNE algorithm by van der Maaten and Hinton (2008a) and visualize the classes in a 2-dimensional space in Figure 3.6.

Figure 3.6 shows the representation learning capability of the neural architecture using the self-supervised algorithm. When looking at the 2-dimensional reduced feature space, distinct cloud patterns occupy distant areas. As already noted in section 3.4.1, class 2 relates to clear sky conditions and is located clearly on the edge of the 2D space (Fig. 3.6a left). In contrast, on the opposite edge

(Fig. 3.6a right), we find the classes with the strongest convective cloudiness (classes 5 and 6) as identified by small areas of enhanced COD. Similar distinct behavior can be found in other regions of the 2D space and will be discussed in more detail in the next section. Comparing the configurations with 128 x 128 and 64 x 64 pixels, one can initially notice that the latter captures more variability in the cloud systems because smaller images can detail more information on cloud structures. Section 3.4.4 will investigate this in more detail.

Table 3.1: The relative frequency of occurrence (RFO), Mean and standard deviation (std) of cloud mask (CMASK), cloud optical depth (COD), cloud top height (CTH), cloud top pressure (CTP), cloud top temperature (CTT), shortwave downwelling radiation (SDS), TOA reflected solar radiation (TRS), for all the classes identified by the network in the 128 x 128 configuration.

Class	1	2	3	4	5	6	7
RFO	13.2	19.6	15	16.3	16.8	12.4	6.6
CMASK mean (%)	92.2	27.3	64.5	96.5	99.1	94.8	88.2
CMASK std (%)	10.7	21.6	17.6	5.5	3.6	8.5	12
COD mean	5.3	1.2	3.3	15.3	24.8	23.4	7.4
COD std	4.6	2.7	3.3	6.9	11.5	13.9	4.4
CTH mean (m)	8212	5152	3198	3866	6001	5642	3529
CTH std (m)	1734	3606	2226	1625	2160	2245	2259
CTP mean (hPa)	374	613	736	657	498	530	697
CTP std (hPa)	101	245	164	118	142	146	157
CTT mean (k)	236	260	271	264	248	251	265
CTT std (k)	10	24	15	11	14	13	14
SDS mean (Wm^{-2})	442	546	582	408	253	194	488
SDS std (Wm^{-2})	183	232	196	143	104	123	171
TRS mean (Wm^{-2})	298	155	224	401	530	284	328
TRS std (Wm^{-2})	183	232	196	143	104	123	171

3.4.4 Physical significance of the identified classes

In this section, we analyze the physical properties of the identified cloud classes. The goal is to verify the independency of the identified classes in the physical spaces given by a combination of radiative and cloud properties. An overview of different characteristics in different classes is given in Table 3.1. We first use the surface downwelling solar radiation (SDS), and the top of the atmosphere (TOA) reflected solar radiation (TRS) (Fig. 3.7a). Note that these products have been derived with radiative transfer calculations (see section 3.2). Moreover, we analyze CTP and COD (Fig. 3.7b). As in Denby (2020a), we randomly select 1000 images for each of the seven optimal classes of the 128 x 128 configuration. Then, we derive their mean position in the space of the SDS-TRS (Fig. 3.7a) and similarly for COD and CTP (Fig. 3.7b).

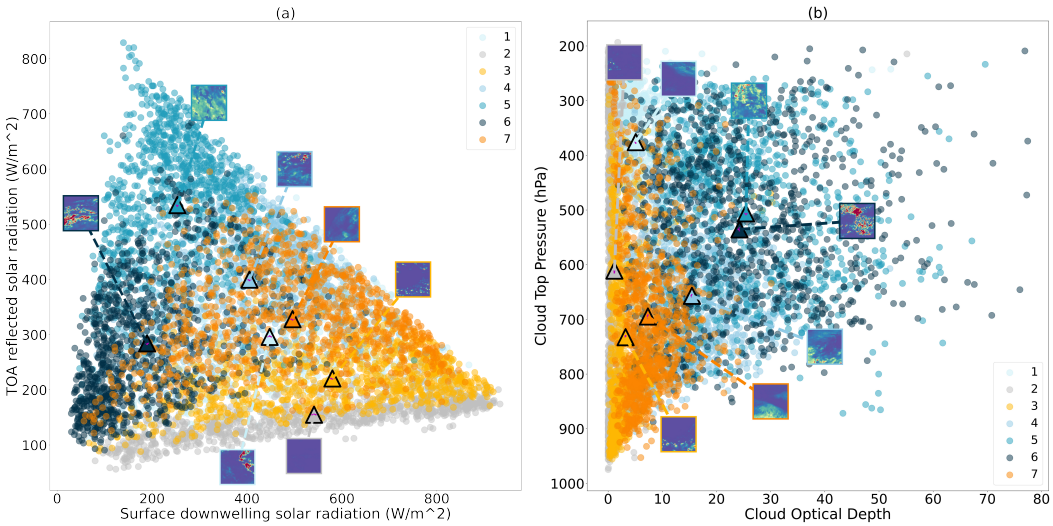


Figure 3.7: a) Mean TOA reflected solar radiation (TRS) versus mean surface downwelling solar radiation (SDS) given by dots of 1000 randomly selected images of the 128 x 128 configuration with color indicating the class they belong to. Triangles indicate the mean position of the classes. The images associated with the triangles are the closest to the mean position of the 1000 images. The crosses in the triangles represent the error bars of each cluster group for the radiative properties on x and y. b) Same as a), but in the 2d space given by COD (abscissa) and CTP (ordinate).

The collocation of the classes in the radiative and cloud properties space permits the identification of the types of clouds the network distinguishes (Table 3.1, Fig. 3.7). Class 1 (light blue dots) probably corresponds to cirrus clouds with a COD mean value smaller than ten and cloud top pressure around 400 hPa, approximately 7500 m. Cirrus show a large variability in albedo and surface downwelling solar radiation that depends on the natural variability of cirrus clouds: they can be thicker or semitransparent and contain larger or smaller ice crystals, profoundly impacting their radiative properties. The cloud-free class 2 has the lowest cloud fraction (27%), with most solar radiation reaching the surface. Due to the lower albedo of the surface compared to clouds, less radiation is reflected back to space. Note that SDS and TRS experience a strong diurnal cycle, and only at noon SDS values > 700 Wm^{-2} are reached. Class 3 and 7 correspond to relatively low-level clouds (orange-shaded dots in Fig. 3.7) with a medium cloud fraction of 65 and 88%, respectively. These clouds have relatively small cloud optical depth values, with their mean values displaying cloud top altitude below 600 hPa (around 3,000 m). Note that frequently low-level clouds are topped by overlaying (semi-transparent) ice clouds (as shown in Fig. 3.1), which explains the relatively high standard deviation in CTH of more than two kilometers. However, there are differences between class 3 and class 7: class 3 has only half of the COD as class 7, but its relative standard deviation is much higher, indicating high variability, which is demanding for solar energy production. Class 7, possibly associable with deeper low-level clouds, displays larger

COD values and smaller surface downwelling solar radiation than the shallower low-level clouds of class 3.

Classes 4, 5, and 6 (blue-shaded dots in Fig. 3.7) all describe overcast situations with cloud fractions of more than 94%. For these classes, average COD values are larger than 10, and they span a broad region of CTP values above 800 hPa, indicating various stages and intensities of convective development. There are more interesting different features among the classes. Class 5 and 6 have very similar mean values of COD and CTP around 25 and 500 hPa, which especially differ from the mean COD of 15 and the CTP of 650 hPa of class 4. These differences indicate that classes 5 and 6 correspond to deeper and thicker clouds than class 4. However, class 6 has the highest relative standard deviation of COD (55%) of the three classes, which aligns with the impression of the strongest convective activity as discussed with respect to Figure 3.6. Still, there is a distinct difference, as class 5 reflects twice as much radiation into space as class 6. Thus, despite of similar cloud fraction, the network is able to identify different structural features within these three "overcast" classes.

Finally, we investigate the joint histograms of cloud optical depth and cloud top pressure (Fig. 3.8) for distinctness. Clearly, each class occupies its characteristic areas in the 2-dimensional space; however, some overlap exists: From the three overcast classes, both classes 5 and 6 include high clouds though class 6 is more limited to the highest tops and has a broader spectrum of COD. The third overcast class (class 4) also includes lower clouds and thus slightly overlaps with the low-cloud class 7, but the occurrence of high COD values (23 and above) becomes more prominent for class 4 at the mid-cloud level. The overlap between classes 2 and 3, clear sky and shallow low-level clouds, respectively, can be explained by the rapidly changing nature of low-level clouds, which can easily form or dissolve and show large variability in space.

For the fine resolution of 64 x 64 also, no strong overlap exists but it appears that the high occurrence is bound to smaller areas in the COD-CTP space except for classes 6 and 8, likely due to the stronger separation into eight classes. In fact, this might be helpful to identify low-level clouds, which seemed to be more confined in class 5 as in the 128 x 128 configuration. Overall, we find that the classes identified and the properties represented through these classes are exclusive to a large extent though some overlaps occur. Therefore using satellite-retrieved products supports the physical interpretation of the identified cloud classes. Note that a detailed investigation of the different classes' characteristics is beyond this study's scope.

3.4.5 Generalization

This section will test how the model behaves when ingested with satellite data from a different year. The research question we want to answer is whether the

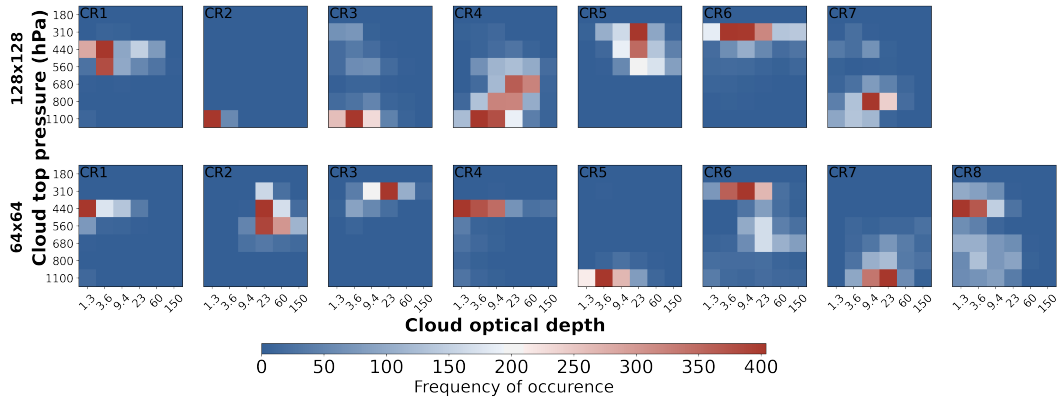


Figure 3.8: Joint histogram of COD (abscissa) and CTP (ordinate) for 128 x 128 (top) and 64 x 64 pixel (bottom) configurations.

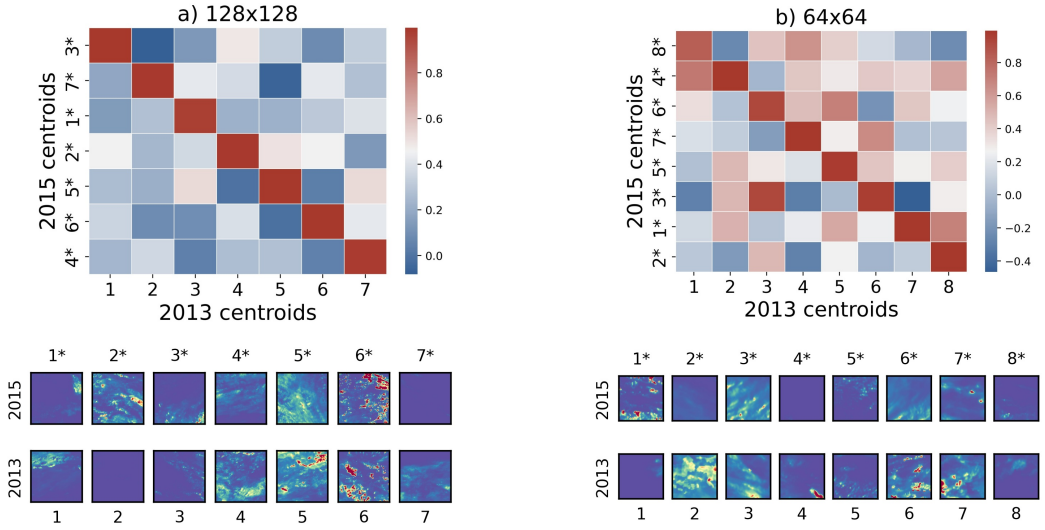


Figure 3.9: Correlation of centroids obtained from 2013 and 2015 datasets for a) 128 x 128 configuration and b) 64 x 64 configurations. Below each joint histogram are the images closest to the centroid from 2015 (indicated with *) and 2013, respectively, for each class. Note that the class number is resorted as indicated by * according to matching in the correlation of centroids.

model can identify the same centroids detected with the data from 2013 using data from 2015. The goal is thus to discover how the centroids will arrange themselves to represent the cloud regimes of unseen cloud system distributions. The only information the network retains from the run on 2013 data are the values of the internal parameters of the neural network (CNN and MLP). We also classify the 2015 data set into the same number of optimal classes as discussed in section 3.4.3.

To answer this question, we then correlate the 2013 and 2015 centroid embeddings and compare them around other respective centroids for both configurations (Fig. 3.9). This type of analysis should reveal the inter-cluster spread in both years. The 128 x 128 configuration generally performs better on unseen data

than the 64 x 64 configuration. In the 128 x 128 configuration, all centroids from unseen (2015) data coincide with the centroids from seen (2013) data (Fig. 3.9a). The highest off-diagonal column-wise centroid correlation is 0.53, which suggests that in a 128 x 128 configuration, the newly identified classes from the 2015 data set fit nicely with the pre-trained network from 2013.

In the 64 x 64 configuration, it is more challenging for the network to classify scenes similar in both years. Here the off-diagonal column-wise centroid correlations are much higher, with values up to 0.93. This especially holds for two classes: 1) a mixture of deep convections along with a spread out of low conviction (class 3) and 2) varying clear sky conditions (class 1). For example, both centroids of classes 3 and 6 from 2015 display a correlation higher than 0.9 (both main and off-diagonal) with class 3 of 2013, while other centroids get discriminated more clearly (Fig. 3.9b). These classes have almost the same type of mixed convection, but rather than being in one class as in 2013; they belong to two classes in 2015. Furthermore, from 2015 class 8 represents very low convection, whereas class 4 groups almost clear sky cases. Both these classes correlate for 0.84 (main diagonal) and 0.74 (off-diagonal) with class 1 of 2013, representing an almost clear sky. Similar to the mixed convection case, rather than being in one class of almost clear sky type, it is present in two different cloud classes in 2015.

While the generalization works well for 128 x 128 configuration, problems appear for the smaller scale. We think this might be related to two probable reasons. 1) The spatial scale of cloud systems, i.e., the scale at which cloud systems occur over a region. For the network to properly represent the clouds, such a scale should be well represented in the input dimensions (128 x 128 representing well over central Europe). Still, in the 64 x 64 case, the scale captured might be small to understand the structure and distribution of cloud systems, which is especially challenging when scales include a mixture of clear sky, convective, and stratiform regimes. 2) We get more classes of variability on a smaller scale and thus likely need more training data. However, letting the network run on 2015 data for a few more epochs might also improve its performance and decrease the centroid correlations in complex cloud systems.

3.4.6 Persistency of cloud regimes and probability of transition

Finally, we move towards applications and use the network's ability to identify cloud regimes over a fixed location (Juelich, Germany, in our case). We can study the temporal development of specific cloud regimes and exploit the 5-minute temporal resolution of the dataset to observe the transformation of cloud systems with time at a higher resolution. Note that the network trained with random crop settings over central Europe has not previously ingested the temporal continuous data over Juelich. In other words, the neural network has no idea of the temporal

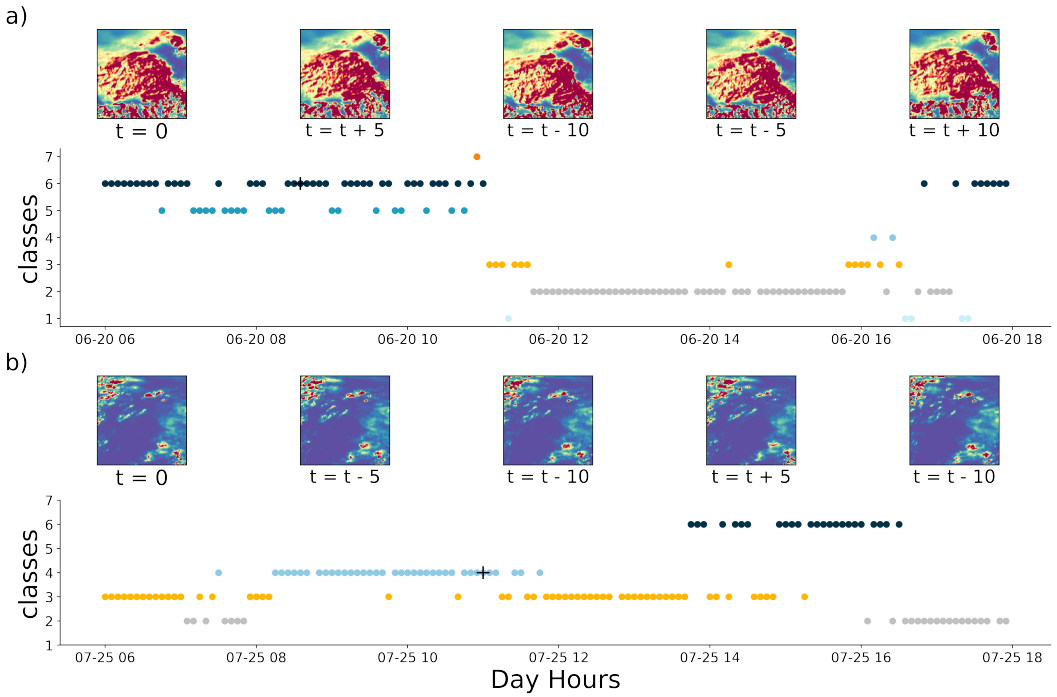


Figure 3.10: Temporal evolution and transition of two example days for the cloud systems over Juelich a) on the 20 June 2013 as identified by the neural network. Here t = 0 represents a COD image that is given as a test case (identified with the vertical line + symbol), and the neural network is asked for the four nearest images sampling over Juelich. The time stamps represent the order of their closeness from the test image (t = 0) in the feature space. b) same, for 25 July 2013.

cohesivity of the satellite images. This experiment also gives us insights into the application of transfer learning for the neural network to study cloud systems over a fixed geospatial location. Looking at images from Juelich (see Fig. 3.1) for an entire day, we can investigate how one class transitions towards another in time. The goal is to understand how close classes should be in the feature space to make a transition and derive a transition probability from one cloud regime to another. Figure 3.10 illustrates the temporal evolution of cloud regimes for two selected days. In this way, the computational model can provide information on the persistency of the cloud classes over a specific region. Furthermore, it also reveals that sometimes a mixed regime appears, as can be seen for 20 June 2013 (Fig. 3.10a) in the morning when the alteration between two deep convections (class 5 and 6 is observed). The network also shows how deep convection decreases, the lower level cloud appears (class 3), and the clear sky conditions prevail in the afternoon. Figure 10 also illustrates how similar the four closest images (in terms of cosine distance) look to a randomly selected test image for both days.

More specifically, we can detect how long a specific cloud regime stays on average over the location or what is the most probable cloud regime it can evolve into. In the 128 x 128 configuration, the clear sky regime (class 2) displays the longest persistency over the site (Fig. 3.11a)). It is also the cloud regime that

most probably evolves into the shallow low-level type (class 3) (Fig. 3.11b)). Similarly, classes 4, 5, and 6, representing different stages of convective cloud structures and distributions, display a persistency of around 75 minutes and often evolve into one another. Class 7 is a class that stays in one fixed location for the shortest time, an average of 30 minutes. These clouds are located near the boundary layer and could represent deeper low-level clouds. They often evolve into class 5, i.e., grow deeper to a full convective cloud, or with slightly less probability, can also evolve into class 3 (shallower low-level clouds), maybe after a precipitation event. Note that the low persistence time results from the occurrence of frequent jumps between two classes (see Fig. 3.11) during the transition time and that temporal smoothing might be beneficial for the analysis of the long-term time series.

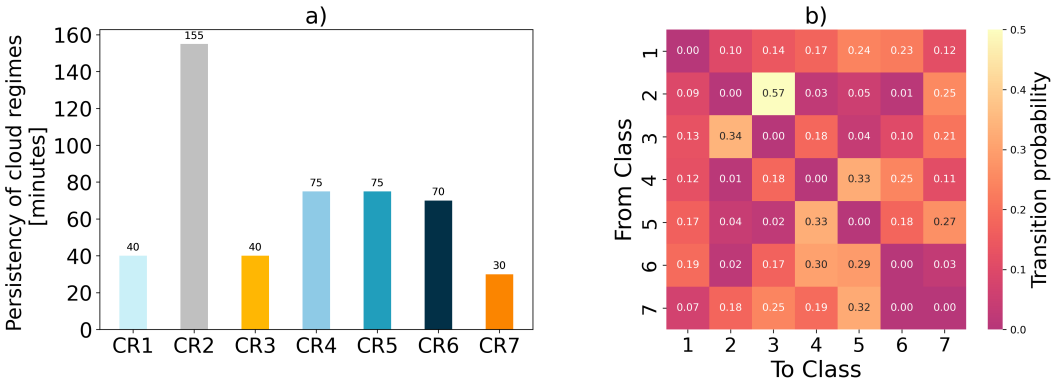


Figure 3.11: a) Cloud systems persistency over Juelich, Germany, given by the average interval of occurrence of cloud regimes over the site. b) Matrix representing the probability of transition of each cloud regime to another class derived over Juelich, Germany. We used one year of data for both analyses.

3.5 Conclusions and outlook

This work uses a deep neural network architecture and a spherical k means algorithm to detect cloud regimes from cloud optical depth images over Europe for the first time. Classifying cloud regimes is a long-standing problem in Earth system sciences. Central Europe's mid-latitude geographical location diurnally gives rise to majorly two cloud families: 1. a mix of stratiform and convective clouds for night time and 2. convective type in the day time. Keeping our long-term vision on solar energy applications, we wanted first to disentangle them and focus on convective cloud systems. We used COD measurements, available in the daytime only, for summer periods. In this way, we were able to investigate a wide variety of convective and stratiform clouds. In the future, with many more years of data, we intend to analyze the diurnal behavior of the different classes. We propose this approach to exploit the greater freedom that the deep neural network offers in understanding the complexity of the cloud patterns. We then

verify the results by analyzing the cloud physical and radiative properties of the cloud regimes obtained from the network. The cloud optical depth values and all the other radiative properties are retrieved as a post-processed satellite product by Deneke et al. (2021) to mimic the Meteosat Third Generation (MTG). In this way, our approach can also be seen as an opportunity to reduce future satellite missions' complexity and data volume. We investigated two different configurations of 128 x 128 and 64 x 64 pixels to study the effect of scale. At first, we tried to understand how close classes are in the feature space to better capture how cloud regimes can transform into one another. A high similarity is found among the closest and farthest images to a randomly selected one from each class. We then determined the optimal number of classes representing the cloud pattern variability. The cloud regime number that adequately represents independent and mutually exclusive cloud patterns is seven for the 128 x 128 configuration and eight for the 64 x 64 configuration. Adopting those number of classes, in fact, the cloud regimes appear visually distinct and positioned in different areas of the feature space.

Exploiting cloud and radiative satellite products, namely cloud optical depth, cloud top pressure, surface downwelling, and TOA reflected solar radiation, we associated each cloud regime with specific physical properties for the 128 x 128 configuration. Class 1 represents cirrus clouds, while class 2 groups clear sky cases; classes 3 and 7 represent two types of low-level clouds, namely shallower and deeper. Finally, classes 4, 5, and 6 comprehend various convective clouds.

To test the applicability of the network, which did not get any temporal information during training, we investigate temporal variability of the cloud regimes over a specific location (Juelich, Germany). In the 128 x 128 configuration, we identified the most persistent cloud regime as the clear sky (155 minutes), followed by convective clouds (70 - 75 minutes). Low-level clouds, as well as cirrus clouds, last around 30 - 40 minutes. These results can help better understand each cloud class's diurnal cycle over the area of interest and are relevant to describe the highly variable conditions for solar energy production.

Finally, we also tested how the network was clustering data on a new unseen dataset by running it with data from 2015. We found that all centroids from unseen data coincide with those from 2013 in the 128 x 128 configuration. In the 64 x 64 configuration, the network had difficulties discriminating the same centroids, possibly also due to the smaller dataset available. However, the generally good agreement between 2013 and 2015 shows that the data distribution between the two years has not changed much. Nevertheless, this highlights the importance of choosing the right image size according to the problem at hand.

This study followed a purely data-driven approach and did not investigate the properties of detected cloud regimes in detail. However, in the future, we aim to investigate which classes correspond to situations most determining for

production and distribution of solar energy. Herein also, the transition probability between different classes offers the potential for short-term prediction capability. For this application, additional training information will also be explored.

Despite using a resolution-enhanced product, geostationary measurements are limited in the detection of boundary layer clouds because of their lower spatial resolution, the question that still can be further explored is whether the lower image resolution offered by geostationary satellites eliminates valuable information for the classification of boundary layer clouds. Sabottke and Spieler (2020) find a case-to-case dependence of image resolutions in radiography. Therefore, determining the optimal satellite resolution for identifying different cloud regimes by machine learning applications remains an open problem and will depend on the application. COD retrieval is also prone to uncertainties (see Sec. 3.2) Tan et al. (2015) find that cloud regimes are less susceptible to minor, random, or statistically insignificant changes in cloud properties. The highest uncertainties occur for COD retrievals at large values outside the range of 50. However, at these high values, downward solar radiation is already heavily extinguished so that the application for solar energy, which we aim to pursue next, will not be affected strongly.

This work shows the potential inductive bias of convolutional architecture to handle complex patterns during the first few initial epochs of self-supervised training. At each epoch, the representation learning of the feature improves. However, future work can focus on seeing whether the network can also learn some hidden properties of the satellite images, such as the geo-localization of cloud systems or environmental parameters associated with a given cloud pattern. The network in the feature space might be able to reveal the specific combination of environmental parameters leading to convection.

In this work, one of the main difficulties connected with the training of the DCv2 on satellite images is the absence of control over the distribution of images. For example, specific cloud types like shallow clouds could be overrepresented within the datasets compared to deep convection. In this case, the model could saturate, and it would not be able to extract more information when fed with additional images. Moreover, by design, DCv2 favors a well-balanced class dataset; therefore, the network's performance is expected to decrease when the algorithm works with satellite images because naturally occurring cloud systems do not conserve class balance properties. One solution to attempt in future work could be to attach a weight (1D Tensor) to the cross-entropy loss function that contains the prior knowledge about the class distribution of the dataset.

Finally, satellite images have more factors of variation than typical benchmark computer vision data sets such as ImageNet. For example, clouds usually spread over the image, generating a different distribution from a usual computer vision data set, which is normally highly object-oriented and centered. Next, it is also

likely that two or more very different cloud patterns may be present in an image. Thus have multiple concepts in a single image. However, the spherical k-means algorithm uses the dominant concept (pattern and distribution) present in the images and attributes the whole image to a single label. We thus think it would be beneficial if we could extract multi concepts from the features and use them for training. Present-day transformers pre-trained using self-supervision and different pre-text tasks (Atito et al., 2021; Caron, Touvron, Misra, Jégou, et al., 2021) have shown good promise in understanding the contextual relation of semantics present in the image. Their suitability will be investigated in the future.

Acknowledgments. Dwaipayan Chatterjee's research has been supported by Federal Ministry for the Environment, Nature Conservation, Nuclear Safety, and Consumer Protection (Grant agreement: 67KI2043, https://kiste-project.de/). This work was performed as part of the Helmholtz School for Data Science in Life, Earth, and Energy (HDS-LEE). Dwaipayan would also like to thank Quentin Duval and others from Facebook's artificial intelligence research for maintaining the VISSL computer vision library. Dwaipayan would like to thank the three anonymous reviewers for taking the time to provide constructive comments and for making the manuscript better. Claudia Acquistapace's research has been supported by the Deutsche Forschungsgemeinschaft (Individual proposal, project number 437320342), project "Precipitation life cycle in trade wind cumuli" (`https://gepris.dfg.de/gepris/projekt/437320342`, last access: 23 December 2021).

Data availability statement. The data and code to produce this work can be accessed at DOI:10.5281/zenodo.7437949 or via web link https://zenodo.org/record/7437949

3.6 Appendix 1

3.6.1 Determination of the optimal number of classes for 64 x 64 images

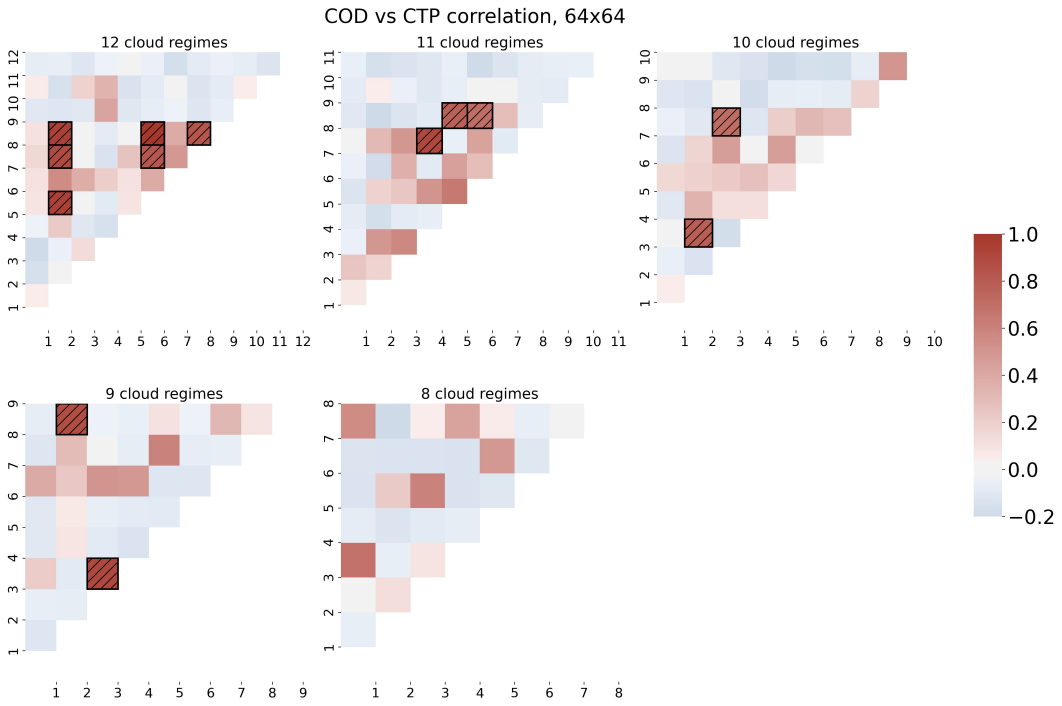


Figure 3.12: Correlation patterns among classes for the number of cloud regimes varying from 12 to 7 for 64 x 64 configuration. Red squares with hatches indicate the classes with a correlation of higher than 0.70 in COD-CTP space.

4 Capturing mesoscale cloud system's diversity

Chatterjee D., S. Schnitt, P. Bigalke, C. Acquistapace, S. Crewell, 2024: Capturing the Diversity of Mesoscale Trade Wind Cumuli Using Complementary Approaches From Self-Supervised Deep Learning. Geophysical Research Letters, 51, e2024GL108889. https://doi.org/10.1029/2024GL108889

The content of this chapter has been published under the Creative Commons Attribution 4.0 License. Formatting changes were made to adopt the format of this thesis.

Personal contribution: Susanne and Sabrina originally proposed investigating the trade wind cloud organizations. I designed the study. Sabrina helped discuss the cloud field continuum results and downloaded the ERA-5 data. Claudia provided scripts for the quick plotting of ERA-5. Paula collected the center of latitude, longitude, and timestamps from the human-labeled dataset. I wrote the manuscript with helpful suggestions from Susanne. Furthermore, I made all the graphics except Figure 1.b and Figure 1 in the supplementary. I am thankful to Andi Walther for providing the GOES-16E data. Furthermore, two anonymous reviewers gave valuable input, which helped me improve the graphics and structure of the methodology section that I carried out.

Abstract

At mesoscale, trade wind clouds organize with various spatial arrangements, shaping their effect on Earth's energy budget. Representing their fine-scale dynamics even at 1 km scale climate simulations remains challenging. However, geostationary satellites (GS) offer high-resolution cloud observation for gaining insights into trade wind cumuli from long-term records. To capture the observed organizational variability, this work proposes an integrated framework using a continuous followed by discrete self-supervised deep learning approach, which exploits cloud optical depth from GS measurements. We aim to simplify the entire mesoscale cloud spectrum by reducing the image complexity in the feature space and meaningfully partitioning it into seven classes whose connection to environmental conditions is illustrated with reanalysis data. Our framework facilitates comparing human-labeled mesoscale classes with machine-identified ones, addressing uncertainties in both methods. It advances previous methods by exploring transitions between regimes, a challenge for physical simulations, and illustrates a case study of sugar-to-flower transitions.

Plain Language Summary Clouds are a fundamental player affecting our planet's energy balance, making their accurate representation crucial in climate models. One open question is how they organize on a scale of a few 100 km (mesoscale) in the tropical northern Atlantic region east of Barbados. Satellite observations can help to categorize these clouds, but previous methods had limitations in capturing the full range of cloud arrangements and transitions between different cloud forms. We have introduced a novel approach that utilizes machine learning and geostationary satellite data to address this issue. Our machine learning model autonomously learns to recognize various cloud patterns and distributions. We conducted a comparative analysis between the categories generated by the machine and those identified by human experts to understand the strengths and weaknesses of both methods. Additionally, we explore a case study where clouds undergo a transformation, changing from a structure resembling sugar to one resembling flowers. This particular transformation was found difficult to capture with physical simulation before. The clear signatures of the transition identified by our machine learning approach can help to better understand cloud evolution, which is crucial for improving climate models and predicting how cloud behavior may change in a changing climate.

4.1 Introduction

Shallow convective clouds, though individually small (measuring in tens of meters), cover large areas of the tropical oceans, forming distinct cloud fields that

span hundreds of km. They are vital in regulating the Earth's energy balance, exerting a net cooling effect by reflecting more sunlight than retaining outgoing long-wave radiation (Bony, Dufresne, Treut, et al., 2004). However, the representation of these clouds, even in the advanced 1km scale climate simulations, is insufficient (Schneider et al., 2019). This contributes to a significant inter-model spread in predicted cloud feedback and climate sensitivity (Bony & Dufresne, 2005; Nuijens & Siebesma, 2019). To address this challenge, Bony et al. (2017) proposed the EUREC^{4}A field campaign, organized in January-February 2020, around the Barbados region of the North Atlantic Trades (NAT) (Stevens et al., 2021). This initiative aimed to enhance our understanding of shallow cloud dynamics by leveraging a diverse set of observations and thus possibly improving their representation in numerical models.

During the preparation of the campaign Stevens et al. (2020) identified four shallow convective organization regimes (*Sugar, Gravel, Flower, Fish*) (SGFF), with frequent occurrence on meso-β (20 to 200 km) and meso-α (200 to 2,000 km) spatial scale. These regimes exhibit differences in net cloud radiative feedback (Bony et al., 2020) and are related to different environmental conditions (Schulz et al., 2021c). Of specific interest are transitions between different organizations, e.g., from sugar to flower, which has been studied in Large-Eddy-Simulation (LES) to understand the governing processes and prove to be difficult (Narenpitak et al., 2021; Dauhut et al., 2023).

Yet, imposing four distinct classes on the diversity of the observed organization does not cover the intermediate cloud patterns or transient states, as highlighted by LES studies. Hence, some processes critical for climate feedback may be ignored or neglected. Furthermore, recent studies trying to quantify these labeled well-organized systems find that they occur only around 50% over NAT (Janssens et al., 2021; Schulz et al., 2021c; Vial et al., 2021) and some ambiguities in agreement from the labeler's side exist (Schulz, 2022).

Denby (2020b) and Janssens et al. (2021) argue for a continuum of cloud organization where Denby (2020b) employs an unsupervised neural network for grouping similar cloud structures and demonstrate its effectiveness via hierarchical clustering (HC) and associated radiative properties. However, their training approach involved a possibility of false negative sampling (Huynh et al., 2022), where the negative pair's distant tile (taken from a random location on a different day) does not necessarily guarantee a dissimilarity in their cloud system's structure and distribution.

Further, employing high-dimensional features in HC has performance and scalability issues (Du, 2023; Gilpin et al., 2013). Janssens et al. (2021) assumes a linear combination of traditional cloud metrics for describing the cloud systems. Utilizing these metric scores and a k-means algorithm, they attempted to partition their metric space into seven arbitrary clusters, as finding meaningful cloud

regimes (CRs) seemed non-trivial.

The overarching goal of our study is to develop a simplified approach to describe cloud organization from high-resolution images. In this way, it should open up new pathways to exploit the information content of existing comprehensive satellite data records. Our first objective is to develop a streamlined representation that captures the entire cloud spectrum's organizational relationships, which we call a continuum. Second, we target the four somewhat arbitrary classes from Stevens et al. (2020) and delve deeper into finding useful CRs from an interpretable continuum. We approach our objectives by developing a two-step self-supervised deep learning approach (Section 4.3) applied on GOES – 16 E cloud optical depth (COD) images (Section 4.2). Section 4.4.1 delves deeper into the representations and their characteristics, highlighting the differences to Denby (2023)'s approach. Our work demonstrates that the presence of derived partitions facilitates a comparison of human labels with these partitions (Section 4.4.2). Finally, in Section 4.5, we illustrate how the partitioning of the continuum supported by environmental data allows us to monitor when a particular cloud system transitions to another.

4.2 Satellite dataset

We use COD retrieved from GOES-16 E Advanced Baseline Imager (Schmit et al., 2005) using the daytime cloud optical and microphysical properties algorithm (DCOMP) (Walther & Heidinger, 2012) at 2 km horizontal resolution and 10 – 15 minutes temporal resolution. Our domain in NAT (5 - 20° N and 40 – 60° W) is similar to domains used in past studies (Bony et al., 2020; Schulz et al., 2021c). The regional climate defines December to May as dry and June to November as wet seasons (Stevens et al., 2016). We consider November to April 2017 - 2021 as our study period. November is added to the typical dry period because we want to see how stronger convective events influence our approach.

We chose COD because it is closely related to the cloud radiative effect and mitigates solar and surface influences. The uncertainty associated with COD retrieval remains below 10% for all ranges in water clouds (see Figure 4 in Walther and Heidinger (2012)). Note that some fine-scale cloud systems, such as sugar and gravel (meso-β scale), their individual cloud cells might not be fully resolved with the spatial resolution of this product. However, since our study focuses on the organizational aspects of shallow convection clouds (spanning hundreds of km), we expect the resolution limit to have a limited impact on our study.

Representation learning, also known as feature learning, is a specialized field within machine learning that focuses on extracting meaningful features of a given dataset. To better represent the mesoscale cloud distributions, we use six images per timestamp, including an additional fixed image over the Barbados domain (see appendix 4.7.1). Although they might overlap in some instances, random

cropping aims to get mesoscale distributions as diverse as possible without human interference. Note that the Barbados domain enables comparison with ground-based measurements in future studies. To have an adequate spatial scale of typical occurring cloud fields over NAT (as discussed in Section 4.1), we use 256 x 256 pixels (roughly 512 square km) as also found in Muller and Held (2012). We exclude crops affected by glint or poor retrieval quality using the respective data flags. Time stamps are limited to 9 am - 3 pm Barbados local time to avoid sun glinting. We use land class data to filter out images with convection over land, specifically over the northeast of the South American continent. Finally, to mitigate uncertainties at high COD from DCOMP retrieval, COD values above a threshold of 50, already indicating deep clouds, are clipped to 50. This results in a sample size of 51,000 satellite images.

For further analysis, we make use of hourly ERA-5 (Hersbach et al., 2020) large-scale environmental parameters (integrated water vapor (IWV), horizontal and vertical wind speed, relative humidity) and cloud fraction at a spatial resolution of 0.25°. Hourly cloud amount for four vertical ranges (surface-700 hPa, 700 hPa-500 hPa, 500 hPa-300 hPa, 300 hPa-tropopause) is used from the Clouds and Earth's Radiant Energy System fourth edition (CERES, Edition - 4A) (Wielicki et al., 1996), characterized by a spatial resolution of 1°.

4.3 Methods

The workflow is as follows: a) A neural network (N1) ingests satellite images to continuously sort cloud organizations based on visual similarity, yielding the feature vector 'Z' (384 dimensions) for each image. b) Z is reduced to a 2-dimensional (2D) space for visualizing a continuous arrangement of images with respect to their cloud structures (continuum). c) The optimal number 'K' of meaningful clusters or CRs is derived from the 2D representation, d) A second neural network (N2) of similar architecture as N1 but constrained by 'K' classes ingests the satellite images to finally assign each image to a discrete class.

a) N1 identifies the structural similarities in the cloud systems and maps the learned visual features into the 384-dimensional feature space Z. To learn similar embeddings of semantically similar mesoscale structure and distributions, every epoch, we opt for two random global crops with a 0.75 fraction (192 x 192 pixels) within a single parent satellite image. Each crop is processed in a separate branch (called student and teacher) by a Vision Transformer (ViT), which has a sequence of self-attention (Vaswani et al., 2023) and feed-forward layers (Bebis & Georgiopoulos, 1994). Note that both branches have the same general architecture, but the parameters (weights and biases) learned during training are slightly different. As the largely overlapping global-crop pair has very similar cloud structures, the network learns their essential features and puts the pair closer to each other in the high-dimensional feature space. A cross-entropy loss function min-

imizes the difference in the output of both branches and yields the 384-element feature vector Z. More details on the implementation are given in appendix 4.7.2. This way of training, unlike Denby (2020b), eliminates the need for a negative pair and avoids linearized assumptions like in Janssens et al. (2021).

b) Z includes the continuously sorted representation of cloud organization. We reduce its 384 dimensions to two dimensions using the well-established t-distributed Stochastic Neighbor Embedding or t-SNE algorithm (van der Maaten & Hinton, 2008b). It preserves relative local positions by using cosine distance in affinity computation and tries to retain global structure by initializing with principal components for mapping to a two-dimensional space. This proves helpful because high-dimensional data when directly applied to cluster analysis, face challenges like the curse of dimensionality (Aggarwal, Hinneburg, & Keim, 2001), where increased dimensions make distances between data points less meaningful. Also, the presence of noise and outliers can distort clusters, hindering the algorithm's ability to identify distinct clusters (Steinbach, Ertöz, & Kumar, 2004).

c) After obtaining the continuously sorted 2D representation of cloud systems (see Fig. 4.1.a), we intend to find optimal boundary conditions within the sorted order to derive distinct clusters (CRs). Selecting a meaningful and interpretable number of clusters is crucial to avoid over-fitting, where excessive clusters can capture noise, and also under-fitting, where too few clusters can miss significant patterns in the data. On this 2D representation space, we apply a set of three statistical approaches, namely metric scores of distortion, silhouette (Rousseeuw, 1987b), and Calinski-Harabasz (Caliński & Harabasz, 1974) to identify the number of optimal classes into which the given features could be clustered. Schubert (2023) suggests taking a collective inference from these three methods to best fit the spherical k-means clustering algorithm used during the training of N2. Appendix 4.7.3 illustrates how the three metrics point to an optimal clustering of the continuum into seven CRs. Note that the choice of seven classes is robust, as illustrated by several sensitivity tests (shown in appendix 4.7.4), such as the dimensionality-reduction technique, size of the dataset, initial weights of the network, and different global crop sizes.

d) N2 from Chatterjee, Acquistapace, et al. (2023) is similar to N1 concerning the two branches: When the feature vectors of both branches capture similar information from the global crops, the loss decreases; conversely, it increases when they diverge. However, before the cross-entropy loss is computed in each branch, a spherical k-means clustering is applied. Herein, the feature vector from the upper branch gets assigned a class (target label) based on proximity to its nearest centroid while the lower branch feature vector tries to reduce its cosine distance with the allocated centroid (predicted proxy) to reduce the loss. In this way, the network learns to allocate both global crops to the same class. After obtaining the label of each satellite image, we transfer the assigned class to the continuum space,

which proves helpful because N1 has learned the sorting arrangement of keeping similar cloud systems closer. Therefore, it helps to visualize how each cluster with distinct characteristics can form a separate local region. Additionally, the N2 feature space is i) more sparse than N1 (see appendix 4.7.2 for explanation) and ii) arranged by closeness to the centroids, which, unlike N1, may not be ideal for representing smooth transitions of cloud systems. Note that there are further differences between N1 and N2, e.g., image augmentation, which are detailed in S2.

4.4 Results

4.4.1 Continuous and discrete representations

We now analyze the diversity of cloud systems included in the satellite data record within their continuous and discrete representations. Both are visualized in 2D continuum space using the t-SNE algorithm (Section 4.3). The organization state captured in the satellite images changes smoothly and different cloud organizations can be identified in different areas of the continuum (Fig. 4.1.a). Going anticlockwise from the top, arch-shaped cloud systems lie in the top-left, followed by flower-type distributions on the left side of the continuum. Close to the flowers in the bottom-left are the flowers spreading out into stratocumulus. Note that physically simulating the transition is challenging as modeling studies struggle to capture the stratocumulus to cumulus transition (Sarkar et al., 2020), although they lie adjacent in the continuum.

The bottom part of the feature space contains long bony skeletons, i.e., fish-type cloud systems, and the bottom-right corner shows an extended part of fish-type cloud organizations delineated by unusually large cloud-free regions. The top-right region of the continuum is a collection of deep convective cells. These primarily occur in the month of November. Arc-shaped cloud systems appear on the left and top-left of the continuum. Vogel et al. (2021) suggest that the horizontal structure of mesoscale arcs is intrinsically linked to gravel, flowers, and fish. In sequence, Figure 4.1a shows a continuous link in the spatial arrangement of cloud systems rather than the distinct classes. This demonstrates the good performance of our continuous approach, which is further supported by the analysis of attention maps in appendix 4.7.5. Note that any newly taken satellite image can be placed into this continuum using the weights of N1, allowing a quick assessment of its organizational status. Also, similar trajectories of subsequent images can be tracked within the continuum space.

After training N2, each of the images can be attributed to one of the seven classes (refer to Section 4.3), revealing distinct spaces within the continuum (Fig. 4.1.b). To assess how well the seven classes separate, they are evaluated using cloud amounts at four different height levels from CERES data. This analysis,

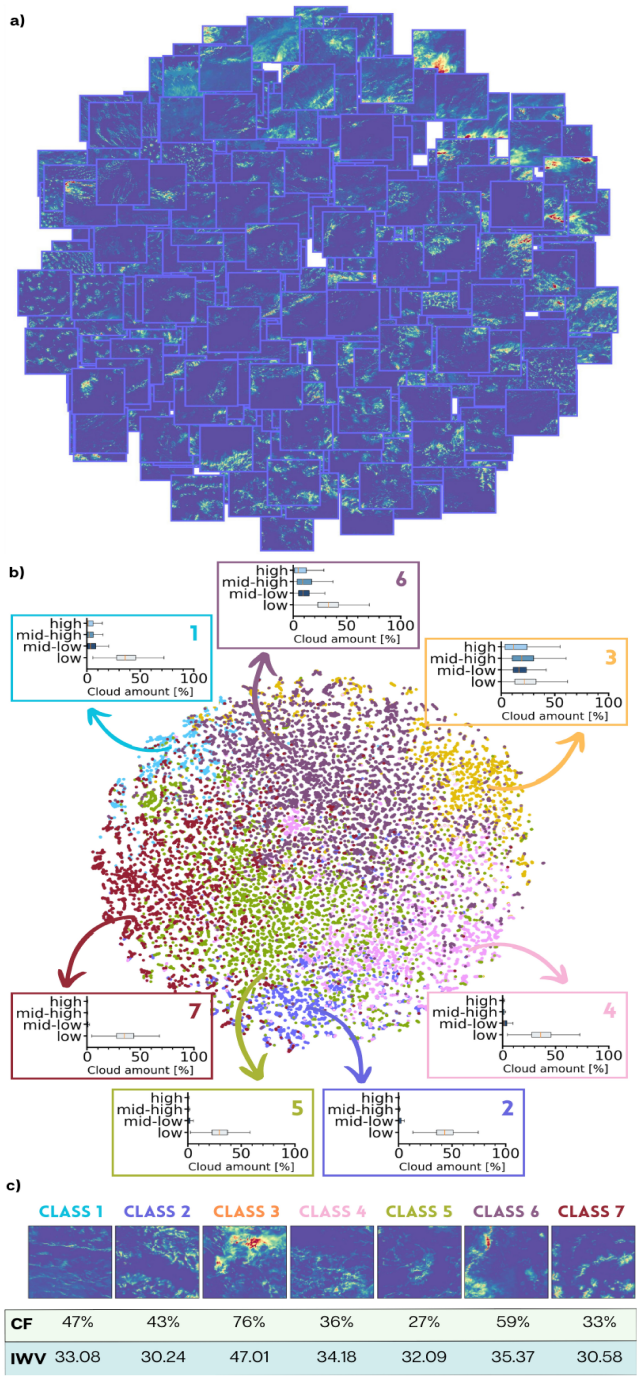


Figure 4.1: a) Visualization of four hundred randomly selected satellite images arranged in the continuum space. b) Same as a), but now, instead of an image, the discrete class determined by N2 is shown (colored). For each class, statistics on low, mid-low, mid-high, and high cloud amount (%) obtained from the CERES hourly data set are provided. c) Centroid COD images belonging to seven clusters as identified by the discrete neural network (N2). The table shows per class the average of cloud fraction (CF, %) from the GOES retrieval and integrated water vapor (IWV, kgm^{-2}) from ERA-5.

on the one hand, reflects how each class differs from the others, and on the other hand, it reasons for the underlying closeness of each class with neighbor classes

in the continuum. The difference between the seven clusters is especially evident when looking at their centroid images (Fig. 4.1.c).

Deep convective class three has by far the highest cloud fraction of 76% and a third more water vapor (47.0 kgm^{-2}) than all other classes (mean = 32.5 kgm^{-2}). We use IWV as a fingerprint for the origin of air masses and intend to test it later to investigate the connection between CR and air mass origin. Figure 4.1.b already shows that class 3, which by far has the highest IWV, is also related to the deepest convection. Neighboring class six includes less frequent higher-level clouds and has a reduced CF of 59% compared to class three. All other classes are dominated by low-level clouds with lower than 50% CF. Classes one and four (neighbor to class six) still have some mid to high-level cloud amounts (below 10%). Class one can be interpreted as representing arch-shaped cloud systems, and four resembles the fish class with a more open sky (also shown by reduction in CF).

Classes two, five, and seven, being close in the continuum, have similar cloud vertical distributions and IWV ranging from 30 to 32 kgm^{-2}; however, their organization is very different, as illustrated by the centroids (Fig. 4.1.c) and mean CFs (43%, 27%, and 33%, respectively). Class two primarily comprises shallow cloud cover, corresponding to cloud systems resembling fish-type formations. Class five has the lowest CF and is an intermediary class type between classes two and seven. Finally, class seven has a presence of low cloud amounts and negligible mid to higher cloud amounts, which visually resembles flower-type cloud distributions. Therefore, discretizing the continuum helps us visually find three main classes (one, two, and seven) frequently resembling features identified by humans, i.e., sugar, fish, and flower, respectively. However, it also shows the remaining diversity and their characteristics in a cohesive approach. Note that in contrast to the challenges faced by Denby (2023) or Janssens et al. (2021) in isolating meaningful clusters, our N1 + N2 framework excels in simplifying the cloud organization complexities by efficiently categorizing the continuum into seven interpretable classes.

4.4.2 Machine versus human labels

While we checked for visual correspondence and class-wise characteristics in Section 4.4.1, our framework now creates the opportunity to quantify how human labels compare to the machine's seven clusters. For this, we use the dataset from Schulz (2022), which is a 1km x 1km resolution manually labeled dataset for the NAT region and EUREC^{4}A time period (47 days). Approximately 50 scientists generated the dataset by identifying mesoscale patterns (SGFF) and marking variable-sized rectangles around homogeneous organization states. Overlapping rectangles allowed a single grid point to be labeled with multiple patterns by a scientist. Individual uncertainty is expressed through each pattern's classification

mask (c_m) (Schulz, 2022). For example, if a grid point is within both gravel and sugar rectangles, the c_m would be 0.5 for both and 0 for the other two patterns. Mutual agreement among scientists for each pattern at a grid point is determined by averaging c_m values, ranging from 0 to 100%.

We hypothesize patterns with higher agreement are most likely attributed to their meaningful partitions within the continuum (as discussed in Section 4.4.1). For each time-stamp where at least one of the four patterns was identified within our domain, we select a 256 x 256-pixel satellite image centered over the area of highest human agreement. In this way, we ensure the best possible intercomparison. This leaves us with 52 samples of human-labeled satellite images (fish: 19.3%, gravel: 26.9%, flower: 28.8%, sugar: 25.0%). Note that even with the highest consensus criteria, there's still diversity in agreement. The inter-quartile agreement range is 35%, while the minimum and maximum agreements show consensus levels of 7% and 91%, respectively.

The framework classifies 40% flower-labeled cloud systems in class seven (see the hit rate for each class in Fig. 4.2.a) while sugar-labeled cloud systems are 31% classified in class one and 20% in class four. Gravel has a total of 44% representation in classes one and five, whereas fish annotated labels are allocated 30% in class two and 20% each in classes four and five. Further, examining example images visually (Fig. 4.2.a), it becomes apparent that images with lower human agreement notably diverge from the established definitions (provided in Stevens et al. (2020)) of SGFF cloud structures, in contrast to images with high human agreement.

Within the continuum (Fig. 4.2.b), flowers detected with high probability mostly occur in areas of class seven, which was already well reflected in the centroids. Following a similar agreement is sugar (street-type cloud systems), which can be found in areas of class one. However, 38% of sugar samples, with a low agreement, lie in classes four and five, which are extended fish and flower type classes (Section 4.4.1). Note that even though these samples reside in those regions of the feature space, their confidence is less than 25%. Similarly, in the gravel pattern, 21% samples belong to class six and exhibit minimal human confidence. In contrast, the rest from the gravel class are positioned between classes one and seven, suggesting that gravel cloud cell sizes fall between sugar and flower. Rightly, no human-labeled samples are found in class three, which predominantly comprise deep convective cells. Finally, the fish class exhibits relatively higher confidence in human labels, aligning well with the feature space characteristics, and lies in class two (fish) and four (extended fish-type cloud structures with large cloud-free regions). Hence, cloud systems characterized by higher agreement among human observers are situated within the designated regions, while those with lesser consensus are positioned within the ambiguous regions of the continuum.

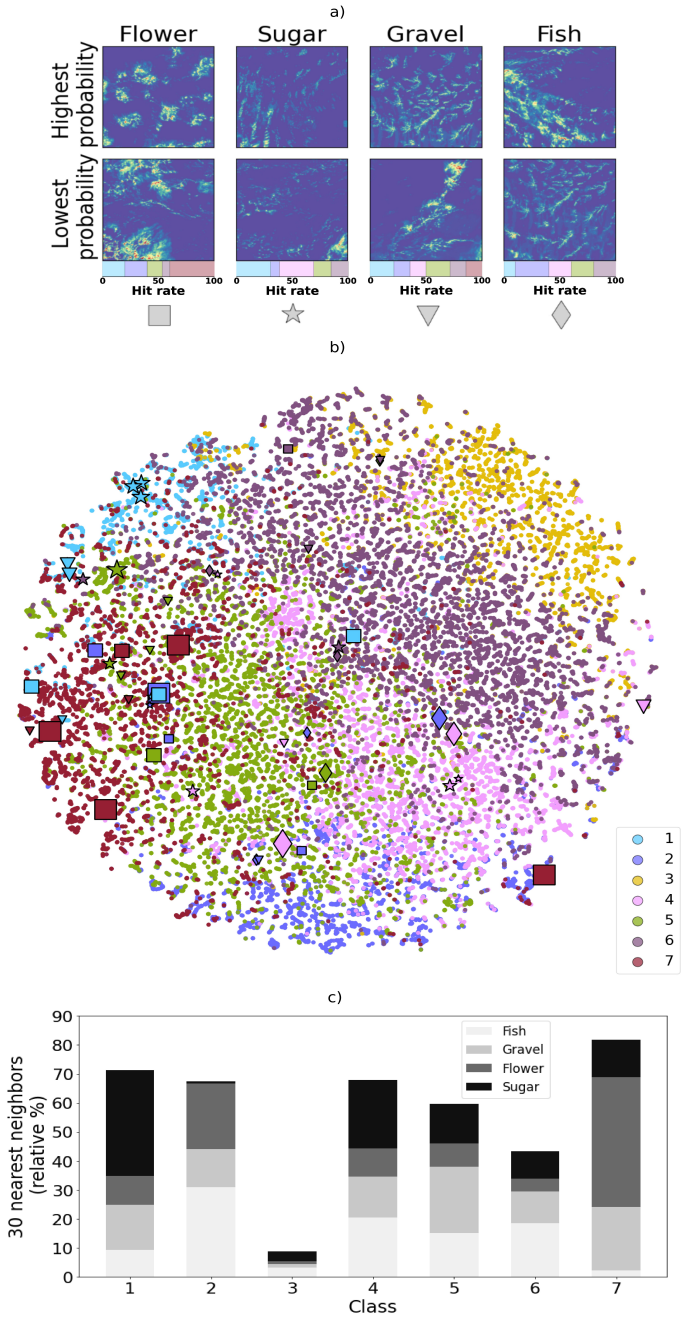


Figure 4.2: a) To enhance visualization and reference for human labels, each column displays 256 x 256 COD images of a specific class, with the highest and lowest human agreement shown in two rows. Below, the images in each column show the hit rate, representing the N2-predicted class for each human label. b) Continuum space colored with different classes (1-7) in the background, along with Human labels (fish, sugar, flower, gravel) in the foreground. Ascending symbol sizes with low (0-0.25), mid-low (0.25-0.50), mid-high (0.50-0.75), and high (0.75-1.00) agreement are shown. c) Relative occurrence of 30 nearest neighbors to human-labeled fish, gravel, flower, and sugar along the seven machine-labeled classes.

To compensate for the limited number of human label samples, we analyze the 30 nearest satellite images to each human label as identified by N1 (Fig.

4.2.c). The majority of neighbors in human-identified fish-type cloud systems (more than 50%) belong to machine-identified classes two and four. The gravel regime includes members of all classes, with notable contributions from classes one, five, and seven, which exhibit cloud cell characteristics similar to gravel systems. The variability in the spread can be linked to the limited representation of gravel glass in Schulz (2022)'s dataset, as gravel occurrences were sporadic during the EUREC^{4}A campaign. Additionally, 75% of gravel labels in our subsamples had agreement levels below 0.25. In contrast, the flower regime mainly belongs to class seven (46 %), further aligning with the high confidence of human labels. Regarding sugar-type cloud systems, 37 % of the neighbors fall into class one, while those with low human agreement are scattered across the remaining classes. Therefore, we find that machine-labeled classes of the 30 nearest neighbors encompass the human-labeled ones, especially for sugar, flower, and fish, but not so clearly for gravel.

Further, in appendix 4.7.6, using ERA-5 large-scale environmental variables and cloud physical properties, we demonstrate that the neighbors and the human crops share a similar, homogeneous distribution of physical properties. Therefore, this analysis, for the first time, shows how to exploit the labels and physical properties of the semantically similar nearest-neighbors of any cloud system of interest further to enhance our understanding of the connection between organizations.

4.5 Transitions

To showcase an application that highlights the intelligible partitioning of the continuum, we explore the "sugar" to "flower" (S2F) cloud system transition on February 2, 2020. Using LES, Narenpitak et al. (2021) showed a strengthening of large-scale upward wind motion and an increase in total water path and optical depth as the transformation develops towards the flower. Here, we look at how the transition in COD is represented in the feature space. For example, where do the representations of transitions lie in the feature space? How smooth is the transition in the feature space?

Covering the temporal developments, 47 COD images were collected (after applying quality filter checks (see Section 4.2)), centered at 12.5° N, 50° W. They cover the time from 10:50 to 19:20 UTC, with a gap between 17:00 to 18:00 UTC likely caused by local sun glint. We ingested the available samples into the trained framework and collected their features (from N1) and machine labels (from N2).

Sugar systems comprise small and shallow clouds with a large spread of individual cloud cells in a domain, as evident in the beginning (10:50, Fig. 4.3.a). In contrast, flower systems appear in multiple deeper aggregates surrounded by large dry areas and are detected first in the southeast cover at 16:50 before becoming dominated at 19:20 over the full domain. In general, the transition features lie at

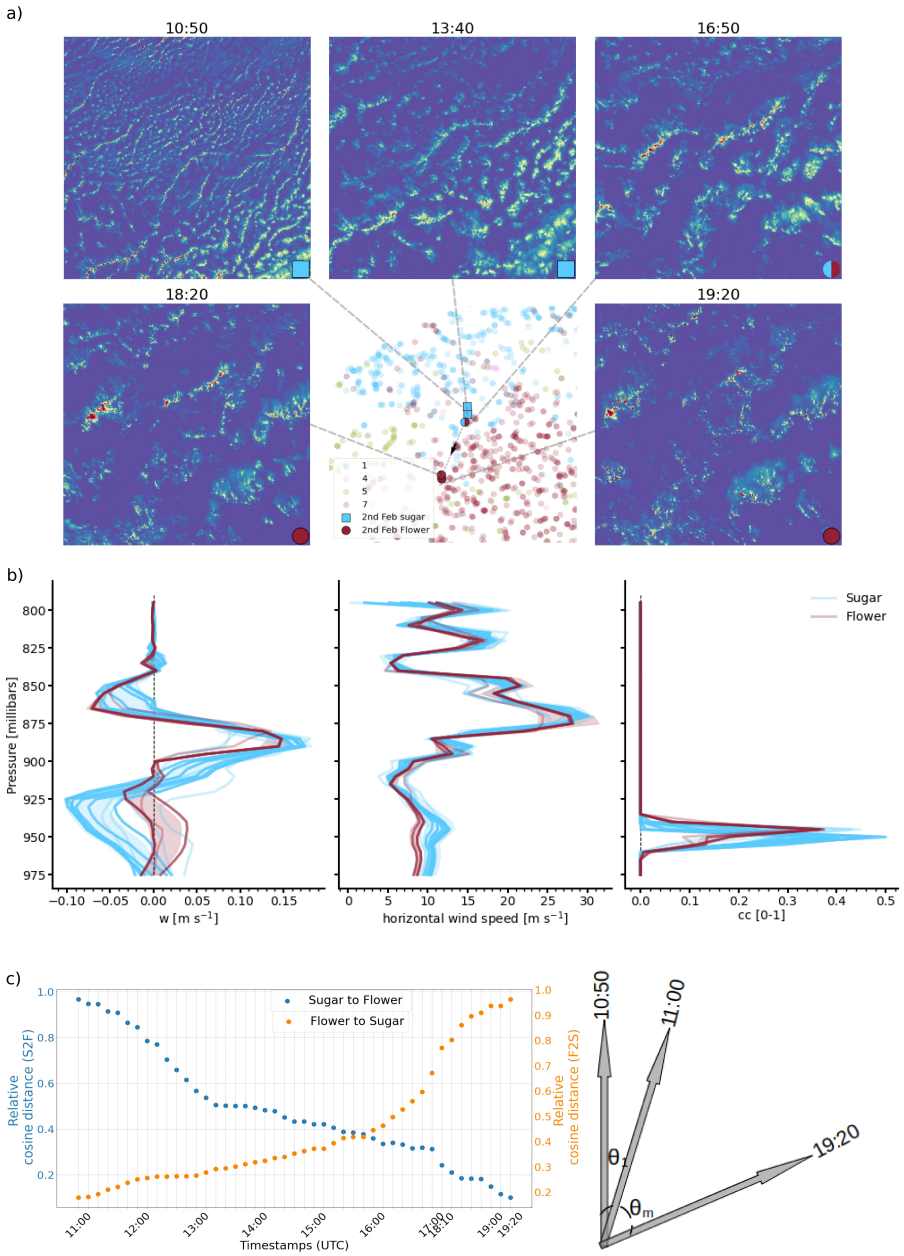


Figure 4.3: a) Five COD images covering the transition period between sugar and flower on the second of February 2020. Their position in the continuum is indicated in the center of the bottom row. b) Individual and standard deviation profiles of 1) vertical, 2) horizontal wind speed describing the atmospheric dynamics, and 3) cloud cover showing changes in mesoscale structure of the transition samples. c) Illustration of temporal transition development inside the feature space: cosine distance of the first daytime image feature obtained at 10:50 UTC compared with the cloud system evolution features for the rest of the day (blue). The last obtained image at 19:20 UTC towards the first image (orange) and θ_m represents the increasing cosine distance.

the border of well-defined clusters one ('sugar') and cluster seven ('flower') (Fig. 4.3.a), and the framework is able to capture their intermediary nature as they are neither perfect sugar nor flower type. We use wind speed (vertical and horizontal)

to represent changes in atmospheric dynamics and changes in cloud cover to account for the changes in mesoscale structure from the ERA-5 product. A gradual increase in vertical velocity is observed as the system transitions from S2F, and consequently, the surface wind speed gradually reduces its strength (Fig. 4.3.b). In addition, as expected, cloud fraction profiles show a gradual decrease as the transition progresses with time.

Sugar-type mesoscale organizations typically occur during the daytime with shallow boundary layers, while flowers occur at night with deeper boundary layers (Vial et al., 2021). We use the cosine distance between features, a unique quantifiable distance metric derived from N1, to show the gradual development of the S2F transition inside the feature space (Fig. 4.3.c). The transformation appears smooth initially, with relatively more significant changes occurring later (post-18:00 UTC) as the system approaches the flower state. We link the relatively high changes in cosine distance during flower stages, as opposed to initial sugar stages, to the progression of convective developments. It becomes more accelerated as the system approaches the well-defined flower state.

Therefore, the framework reveals unbiased relative changes from the point of interest (in space or time) solely based on changes captured in high-dimensional feature space. Also, unlike previous works of Denby (2020b) or Janssens et al. (2021), the intelligible partitioning of the continuum allows us to see when a particular system transitions to another. Appendix 4.7.7 provides insights into the transition probability of one class transforming to another over the Barbados domain.

4.6 Conclusion

In this work, we develop a two-step self-supervised learning framework to study shallow convective organization properties and their transitions. By analyzing organization in a continuous approach without imposing predefined classes, we include all occurring patterns and transitional states in our analysis. Moreover, the approach shows that mesoscale cloud organizations in NAT can be partitioned into seven reasonable CRs for the time period considered. Exploiting the cloud amount at different vertical levels from CERES measurements, we show how the classes are interlinked with each other within the continuous space and thus capture the variability of tropical clouds in more detail.

We compare human-labeled cloud systems (Schulz, 2022) with the machine-identified partitions and underscore challenges in human-labeling of cloud organizations. Cloud systems with higher agreement among humans lie in the "correct" region of the feature space, while the ones with less consensus are in the "wrong" regions of the feature space. Also, the potential and interpretability of the continuum space become more evident when examining the classification and physical properties between human labels and their nearest neighbors.

Two of the seven CRs are strongly related to sugar and flower. Representing the S2F transition case study (Narenpitak et al., 2021) for February 2, 2020, in the continuum illustrates the capability to identify and represent the observed transformations smoothly in their clearly interpretable regions. We evaluate the transition's large-scale environmental parameters and observe a gradual increase in vertical wind speed and a gradual decrease in cloud amount. Finally, we showcase the use of cosine distance metric in capturing clear signatures of the S2F transition, which can help better understand cloud evolution. This is crucial for improving climate models and predicting how cloud behavior may change in a changing climate.

One of the limitations of this study is the use of only daytime cloud retrievals; hence, the organizations' nocturnal nature cannot be captured. Future studies will use infrared satellite measurements for 24-hour coverage. We aim to fine-tune our framework with ground-based observations of the EUREC^{4}A campaign and extend our analysis to a climate scale. The developed workflow could be a testing ground for investigating the newly adjusted subgrid parameterization effects in high-resolution global digital twins (Hoffmann et al., 2023) for mesoscale cloud systems or atmospheric processes at different scales.

Open Research CERES, Edition-4A is available at (NASA et al., 2017), and ERA-5 reanalyses data (Hersbach et al., 2023) is available from the Copernicus climate change services. GOES-16 data has been accessed from the National Oceanic and Atmospheric Administration (NOAA), Climate Data Records (CDR) facility Oceanic and (NOAA) (2024b). Here, the COD retrieved using DCOMP algorithm (Walther & Heidinger, 2012) from GOES-16 measurements is available at Oceanic and (NOAA) (2024a). The code to produce this work and pre-trained weights of N1 and N2 can be accessed at Chatterjee, Schnitt, Bigalke, Acquistapace, and Crewell (2023).

Acknowledgments Dwaipayan Chatterjee's research was supported by the Federal Ministry for Environment, Nature Conservation, Nuclear Safety, and Consumer Protection. Claudia Acquistapace's (CA) research was funded by Deutsche Forschungsgemeinschaft (DFG). CA also acknowledges funding from Federal Ministry for Digital and Transport (BMDV).

4.7 Appendix 2

4.7.1 Domain description

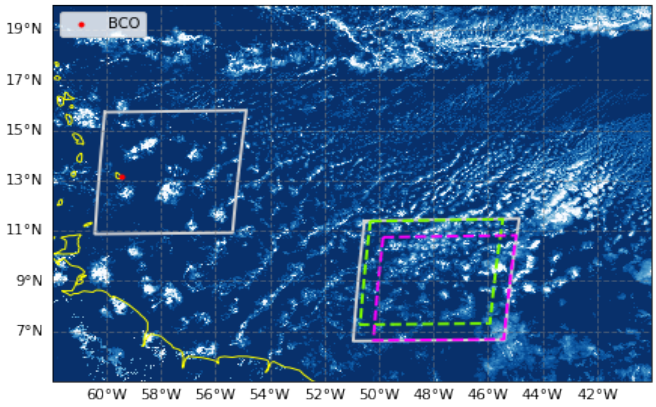


Figure 4.4: GOES's COD image on February 2, 2020, at 13:00 UTC with coastal boundaries (thick yellow) and Barbados Cloud Observatory (red dot). One (out of five) random and a fixed (Barbados domain) 256 x 256-pixel crop over EUREC^{4}A domain are shown. During the learning process, each crop is twice randomly sub-cropped (pink and green dashed lines) by the network, leading to a spatial dimension of 75% (192 x 192 pixels) of the original crop. The Barbados domain enables comparison with ground-based measurements in future studies.

4.7.2 Network architectures

Continuous network (N1)

Definition of the network input N satellite images of COD built the input training data set $X = \{x_1, x_2, x_3....., x_N\}$ of the deep learning architecture illustrated in Figure 4.5. The only intuitive augmentation we opt for here is global random cropping for learning continuous representations. For random cropping, we opt for two global crops (x_1, x_2) with a random 0.75 fraction (192 x 192 pixels) of the parent satellite image to focus on the global distribution of the cloud system. Figure 4.5 shows two random crops (teacher t and student s) are fed into different branches of the network. Therefore, it becomes challenging for one side of the network to know what part of the parent satellite image the other is being fed with; therefore, during the learning process, it focuses on the critical semantics of global cloud distribution.

General network architecture The neural network's task is to learn visual features from each satellite image. A function g represents the transformations performed by the network's vision Transformer (ViT) as $g(x_i) = h_j$ with $i = 1...N, j = 1....M$ that maps the image x_i into the array of features h $=\{h_1, h_2, h_3....., h_M\}$, where M is the output dimension of ViT feature arrays. The

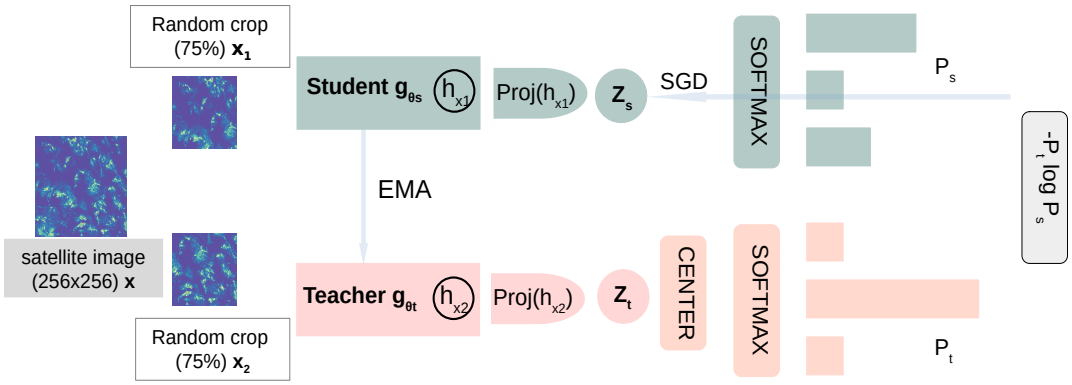


Figure 4.5: This work adopts a deep learning architecture from Caron, Touvron, Misra, Jégou, et al. (2021), where x_1 and x_2 are 75% random crops of the parent satellite image x. The student and teacher vision transformers ($g_{\theta s/t}$) have the same number of trainable parameters (weights and biases) θ. The feature output h_{xi} from g_θ subsequently connects to $Proj(h_{xi})$, a 3-layer multilayer perceptron. Softmax (Bridle, 1989) normalizes MLP's raw activation ($Z_{s/t}$), and centering maintains teacher activations (Z_t) near batch mean properties. P_s and P_t represent normalized student distribution of Z_s and centered and normalized distribution of teacher activation Z_t. Back propagation in student network optimizes its parameters through stochastic gradient descent (SGD), minimizing cross-entropy between P_t and P_s. Teacher parameters ($g_{\theta t}$) are exponential moving averages (EMA) of students ($g_{\theta s}$), aligning the networks.

selected dimension of M is equal to 384, which means the information contained in the 192 x 192 satellite observation space is being non-linearly dimensionally reduced to 384 vector space. ViT is a sequence of self-attention (Vaswani et al., 2023), and feed-forward layers paralleled with skip connections. The mechanism of ViT (Dosovitskiy et al., 2021) takes non-overlapping contiguous image patches of resolution NxN pixels, where N=16 for this work, along with their positional encoding as an input. Without the positional encoding, the output feature vector from ViT is invariant to the arrangement of these NxN patches. Meanwhile, with positional encoding, it learns the relative position of the objects in the image. Therefore, the model learns the relationship between the patches, and thus, it learns how a particular arrangement of cloud distribution usually occurs. The ViT architecture can identify long-range spatial dependencies (Khan et al., 2022) by learning relevant information in the image. The activation function used in the ViT is Gaussian error linear units (Hendrycks & Gimpel, 2023) (GELU), as the GELU function behaves smoother than other activation functions when values are closer to zero and thus is more effective at learning complex patterns in

the data.

Further, h_j is non-linearly projected to Z_l with $l = 1....L$ using a three-layer multilayer perception (MLP) (Rumelhart et al., 1986) activated by GELU followed by l_2 normalization and a linear layer. Here Z $=\{Z_1, Z_2, Z_3....., Z_L\}$ is the final output dimension of the pipeline. The feature space dimensions are decided based on input dimensions, the complexity of information context, and neural network complexity. Caron, Touvron, Misra, Jégou, et al. (2021) suggest that if the training dataset size is much less than 1.3 million, then the final dimensions of Z_l should be reduced compared to the default dimensions L = 65536. We tested rather two different values (L = 128 and L = 8192) and found a much better suitability of the larger value from visual inspection. Our aim here is not to find the optimal feature vector size but a functional size that can optimize the network and smoothly converge the training. Therefore, the optimal dimension size of the dimensionally reduced images in self-supervised learning is not the focus of this work. Figure 4.5 shows two different branches in the network: student and teacher. The point to note here is that they have the same general architecture and pipeline, but the parameters (weights and biases) learned during training are different.

Upper branch of the network (Student) The upper branch of the network, represented in Figure 4.5, by the student transformer g_s and further projected by MLP, ingests one random augmented global crop of the parent satellite image and outputs feature vector Z_s. Softmax normalizes Z_s such that all values are between 0 and 1, and the integration over all L = 8192 elements of its probability distribution yields 1. This probability distribution (as shown in equation 3) of the feature vector Z_s is input to the cross entropy loss function described later. The soft-max probability for an input x_i of the student network can be described as

$$p_s^{(i)} = \frac{\exp(\frac{1}{\zeta}{}_s Z_s^{(i)})}{\sum_{m=0}^{k} \exp(\frac{1}{\zeta}{}_s Z_s^{(m)})} \tag{3}$$

where ζ_s is the temperature parameter for the student network and is set to 0.1. The ζ parameter controls the sharpening of the probability distribution. A higher value of ζ implies smoothed probability.

Lower branch of the network (Teacher) The lower branch of the network represented in Figure 4.5 by the teacher transformer applies function g_t to the other remaining global crop of the parent satellite image, and the MLP projects outputs feature vector Z_t. Unlike P_s, before normalizing P_t individually with soft-max, vector Z_t is centered around the mean properties of all images in a batch. A batch refers to the number of samples propagating through the neural

network before updating the model parameters. Centering is done to prevent any feature from dominating, as the mean will be somewhere in the middle of the batch sample properties. While applying the temperature ζ_t parameter for the teacher, it is kept lower (ζ_t= 0.05) to sharpen the probability of Z_t artificially. Therefore, the feature vector Z_t of the teacher branch is centered and sharpened before it becomes input for the loss function.

Cross entropy loss of the network When the feature vectors of the two branches capture similar information from the global crops of the satellite parent image, the loss becomes lower and vice-versa. That's how the network branches are encouraged to focus on the common image characteristics, progressively making the feature vectors similar.

$$\min_{\theta_s} \sum_{x \in (x_1, x_2)} P_t(x) log(P_s(x)) \tag{4}$$

This is achieved through the cross-entropy loss function applied on the centered and sharpened probability distribution of the teacher branch P_t and smoothened distribution of the student branch P_s. As shown in equation 4, the loss function minimizes θ_s, i.e., the student network's parameters (weights and biases). Teacher network parameters or P_t guide the student network during the training phase, as discussed in Subsection 4.7.2.

Optimization for convergence The loss function minimization happens progressively layer by layer, derivating the loss function with respect to θ_s parameters and adjusting parameter values in each layer by backpropagation. At the end of the minimization, we obtain a configuration of parameters for the student network that will be ready for the next iteration with a new batch of images. Stochastic gradient descent (Bottou, 2012) (SGD) is only applied to the student network parameters θ_s, and the teacher parameters θ_t are built through past iterations of the student network (Caron, Touvron, Misra, Jégou, et al., 2021). As shown in equation 5, θ_t is the exponential moving average (EMA) of θ_s with λ following a cosine scheduled from 0.996 to 1 during training.

$$\theta_t = \lambda\theta_t + (1 - \lambda)\theta_s \tag{5}$$

During optimization, a collapse can occur regardless of the input provided to the model; the output becomes constant or is predominantly influenced by a single dimension. In other words, the model's predictions across different dimensions or features become uniform, leading to zero ideal loss value. Therefore, centering and sharpening introduced in the upper and lower branches and EMA in the optimization of Subsection 4.7.2 are the easiest acceptable ways to prevent collapsing in the described teacher-student framework.

Training and libraries To set up this architecture, we use the software package DINO from Facebook Artificial Intelligence Research (FAIR) (Caron, Touvron, Misra, Jégou, et al., 2021) based on PyTorch. The open-source VISSL computer vision library (Goyal et al., 2021) adapted the DINO neural network to our requirements. Based on sensitivity tests on training loss, visualization of dimensionally reduced feature space, and ablation study of the original network on longer training showing improving performance, we train the model up to 800 epochs. Training the neural network for 800 epochs on 4 V100 GPUs took 16.5 hours or 66 core hours.

Discrete network (N2) We briefly describe the functional mechanism of the discrete neural network (N2) and its learning scheme. Refer to Section 3 from (Chatterjee, Acquistapace, et al., 2023) for a detailed network description. The data loading nature of N2 remains the same as of N1 (Subsection 4.7.2 of S2). The general architecture has a pipeline similar to the continuous approach setup, with the image processing backbone here being a convolutional residual network with 50 layers of depth (ResNet-50, (He et al., 2015)), followed by a projection head of MLP with ReLU activations (Fukushima, 1975) and a linear layer.

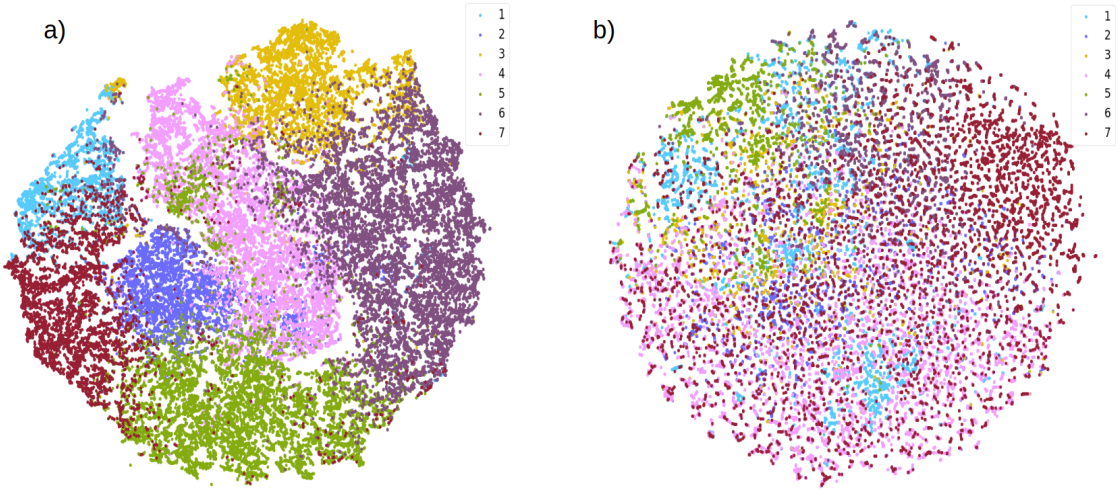


Figure 4.6: a) Sparse 2D feature space obtained from N2 by applying the tSNE algorithm on z_x features of 51,000 satellite images. The perplexity and epsilon derived from auto-configuration for t-SNE runs are 30 and 1150, respectively. b) Same as Figure 1.b in the main text, this uses direct clustering on satellite images with N2, overlaying labels on the continuous feature space from N1 for comparison.

Therefore, similar to Figure 4.5, for the upper branch, the features obtained at the end of the pipeline (like Z_s in the continuous approach) are clustered using spherical k-means (where k=7), and features are allocated a pseudo-label (T) according to their closest centroid. Further, the features obtained from the lower branch are compared with the calculated upper branch centroids using cosine distance (D_T). Finally, T from the upper branch and D_T from the lower branch

are inputs of the cross-entropy loss function as discussed in cross-entropy loss of Subsection 4.7.2 and are progressively minimized during training. We call the labels pseudo-labels during the training stage as they can change to minimize the loss function better. Finally, at the end of the training, we collect the labels for each satellite image and further evaluate their separation using auxiliary datasets.

Data augmentation in N2 The main input difference from N1 is that additional augmented image versions (random flipping and noise addition by random Gaussian blur) of global random crops (see Fig. 4.6.b) are included. Augmentations try to provide auxiliary support to the network's generalizability and better capture the differences in diversity of the shallow cloud systems (Nie, Zamzam, & Brandt, 2021; Paletta et al., 2023)

Supplementary Figure 4.6.b visually demonstrates the impact of clustering when utilizing only two global random crops as input for N2. The assessment of clustering quality, considering both separation and locality, reveals an inferior performance compared to the utilization of augmentations (see Fig. 1.b). This difference can be likely attributed to the inherently more challenging nature of N2's task, which involves identifying distinct partitions within the shallow cumulus family by delving into finer details instead of the sorting task undertaken by N1.

4.7.3 Determination of optimal cluster number

We apply the following metrics to two-dimensionally reduced representations (using tSNE) on h_j from N1 to identify the best optimal cluster:

Distortion metric: The distortion metric considers the cluster's tightness by computing the sum of squared distances (SSD) from each point to its assigned center, which tends to decrease toward 0 as we increase the number of clusters (K). This shows an exponential shape leveling off such that the shape of the curve results in an elbow, but the optimal cluster or the point of inflection represents the point where adding additional clusters stops adding useful information. Also, adding clusters beyond the inflection point also makes the clusters harder to separate; thus, we start to observe diminishing returns by increasing k. The elbow blue line curve in figure 4.7 shows k = 7 as the sweet spot of optimal clustering.

Silhouette metric: Apart from taking cluster closeness into account, this metric also considers distances between points of one cluster and the nearest other cluster center. This means that in order to have a good silhouette score, clusters generally need to be tighter and farther apart from each other. If the Silhouette coefficient for each point is close to 0, it means that the point is between

two clusters; if it is close to -1, then that point is in the wrong cluster, and if it is close to +1, it is in the correct cluster. The average silhouette coefficient calculated for all 51,000 samples shows two local maxima at values of 0.37 (k=3) and 0.36 (k=7), as shown in Figure 4.7. Note that the values are not close to one, meaning the cluster doesn't lie very far from each other, further suggesting the continuous nature of cloud organizations.

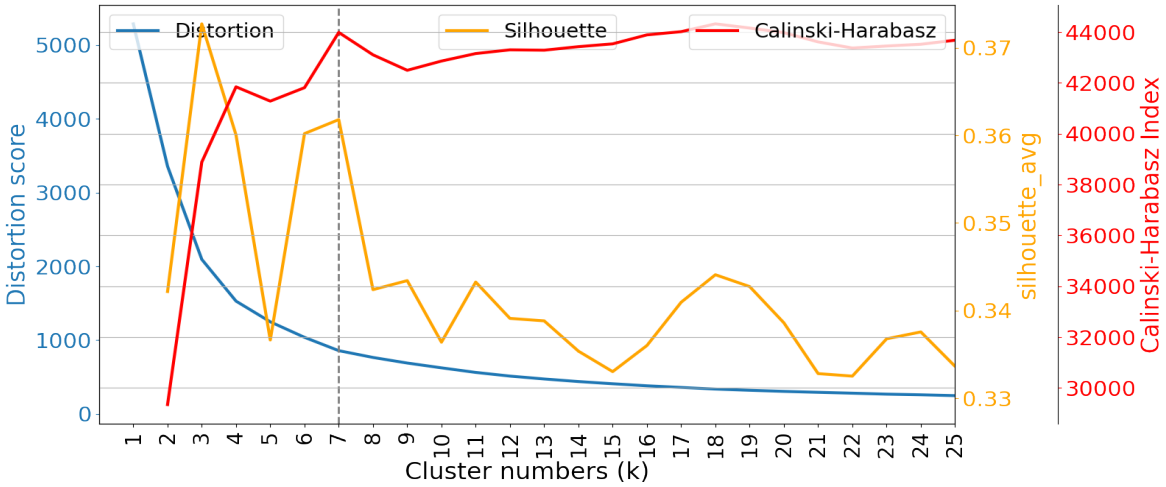


Figure 4.7: Results of three different metric scores of distortion, silhouette, and Calinski-Harabasz, shown along with varying cluster numbers along the abscissa. The vertical-dashed line is drawn at cluster 7, which shows the chosen inflection point for the optimal cluster.

Calinski-Harabasz metric: In comparison, the Calinski-Harabasz metric assesses the separation and compactness of the clusters. It denotes the ratio of the sum of inter-cluster dispersion and the sum of intra-cluster dispersion for all clusters. A good clustering result has a high Calinski-Harabasz Index value. The maximum lies at cluster 7, having a score of 43000.

In summary, the two metrics directs towards k=7, and the difference between the two maxima (k=3 and 7) in silhouette is insignificant. Therefore, we take the common agreement of k=7 as the optimal cluster number and train N2 (Section 4.7.2 of S2) from scratch using 7 clusters.

4.7.4 Sensitivity tests

Here, we show that the choice of seven classes has passed several sensitivity tests, such as the dimensionality-reduction technique, size of the dataset, initial weights of the network, and different global crop sizes.

Different data sample size

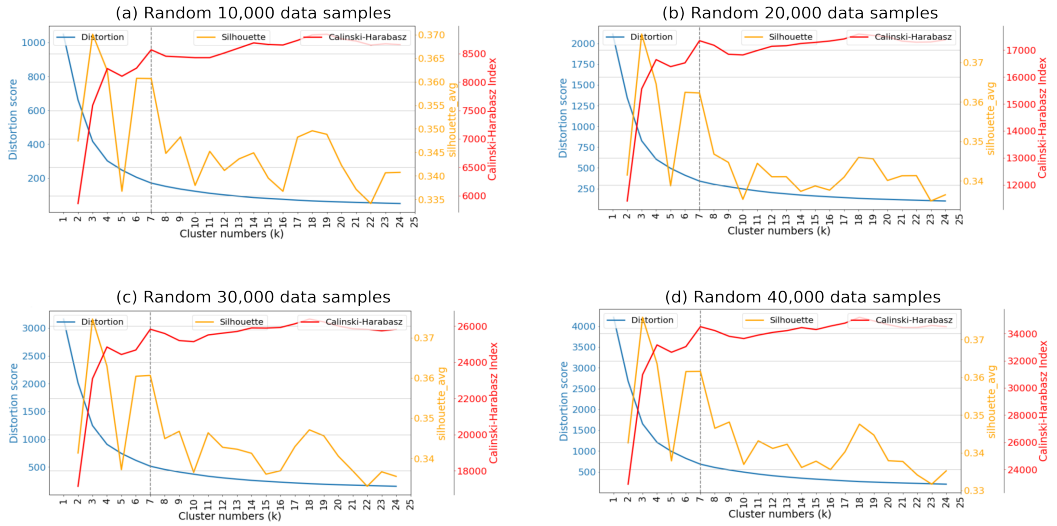


Figure 4.8: t-SNE initialized by PCA and using cosine distance as a unit of distance while constructing the two-dimensional space, on a sample size of a) 10,000, b) 20,000, c) 30,000 d) 40,000 data points.

Dimensionality reduction techniques Manifold extraction algorithm is a generic term used for nonlinear dimensional reduction, or we can call them generalized PCA, which is sensitive to nonlinear structures in the data.

Denby (2020b) train their neural network using 'Euclidean distance' as a metric in the loss function. However, for dimensionality reduction, use Isometric feature mapping (Isomap, (Tenenbaum et al., 2000)), an extension of kernel PCA. Isomap uses 'geodesic distance' as a measure of distance while reducing the dimensions. The geodesic distance looks for the shortest curve in the high dimensional space. Therefore, there are some limitations here:

1. Euclidean distance and high dimensions: First, we would like to point out that Euclidean distance breaks down in high dimensions (Aggarwal et al., 2001). Euclidean distance is sensitive to sparse data distribution in high dimensions. The direction becomes more critical since we normalize our feature vectors (the magnitude becomes one). Therefore, cosine distance (used in N1 and N2) is far more suitable for training.

2. Global versus local structures in the data: Second, we would like to mention that the Isomap algorithm considers maintaining the global pairwise distance (Gao et al., 2021). In other words, it neglects the local structure but only considers the global one.

Overcoming the limitations: t-SNE preserves the neighboring local distances better.

1. We inject the global structure into our initialization of tSNE through PCA, which dictates which regions of the 2D space the points will appear.

2. Second, while reducing the dimensions, we keep cosine distance as a distance measurement criterion in t-SNE.



Figure 4.9: Different dimensionality reduction techniques applied on our high dimensional feature space for cluster number optimization, which is constructed using cosine distance (as unit of measuring distance between two feature vectors). See the main text for an explanation of the distortion, Silhouette, and Calinski-Harabasz algorithms. We can see the consensus between t-SNE and MDS, while Isomap and LLE could not come to a logical conclusion.

For our high dimensional feature space, we demonstrate optimal cluster number sensitivities to different dimensionality reduction techniques in Figure 4.9. In addition to t-SNE and Isomap, we also show multi-dimensional scaling (MDS) and local linear embedding (LLE) methods.

1. MDS, distance-wise, is more geometrically aware as it tries to preserve the inner products between the feature vectors.
2. LLE pays attention to only the local structure of the data. LLE also assumes that the high dimensional data is locally linear and a sample can be represented linearly by several samples in the neighborhood.

Different initial weights of the network

Different overlap threshold Therefore, Figures 4.8, 4.9, 4.10, 4.11 mainly argues about four things:

1. While reducing dimensions, it is important to consider how the original high-dimensional space was constructed. What distance was used while computing affinities? Then, use the same distance while constructing the lower dimensions.

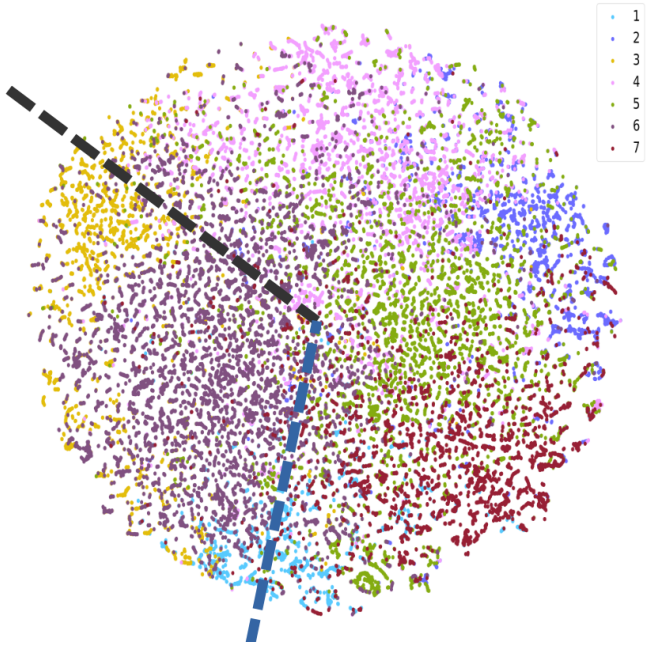


Figure 4.10: N1 and N2 initialized with different random seeds. As compared to main Figure 4.1.b, we find only the orientation of the low dimensional projection to change where the black dashed line is the original location of class 1 and the blue dashed line is with a different random seed. The rest (sorting order and classification) remains the same.

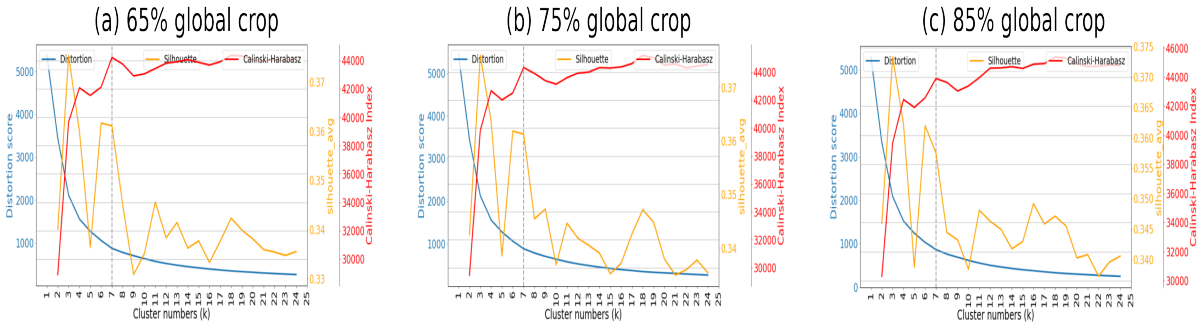


Figure 4.11: Selecting different global crop size (65%, 75%, 85%) and training N1. t-SNE is intialized by PCA, and using cosine distance as a unit of distance, we find k =7 as the common consensus among the three.

2. It is important to consider a dimensionality reduction technique that considers both the local and global structure of the high dimensional data.

3. Varying number of samples from (10,000 to 50,000) still shows k = 7 as the most expressive optimal cluster number

4. Varying the Global crops from the 65^{th} to 85^{th} fraction shows crop size might not be a strong influencing factor in finding the optimal cluster number.

4.7.5 Attention maps of N1

From a human perspective, cloud system distributions may appear to be relatively chaotic and noisy, and while trying to decide their visual characteristics, we may pay attention to some or all of the following: the organizational semantics of convective organization, the semantics of the clear sky regions, deep convective cell distributions, open and closed cells, and shallow convection distributions.

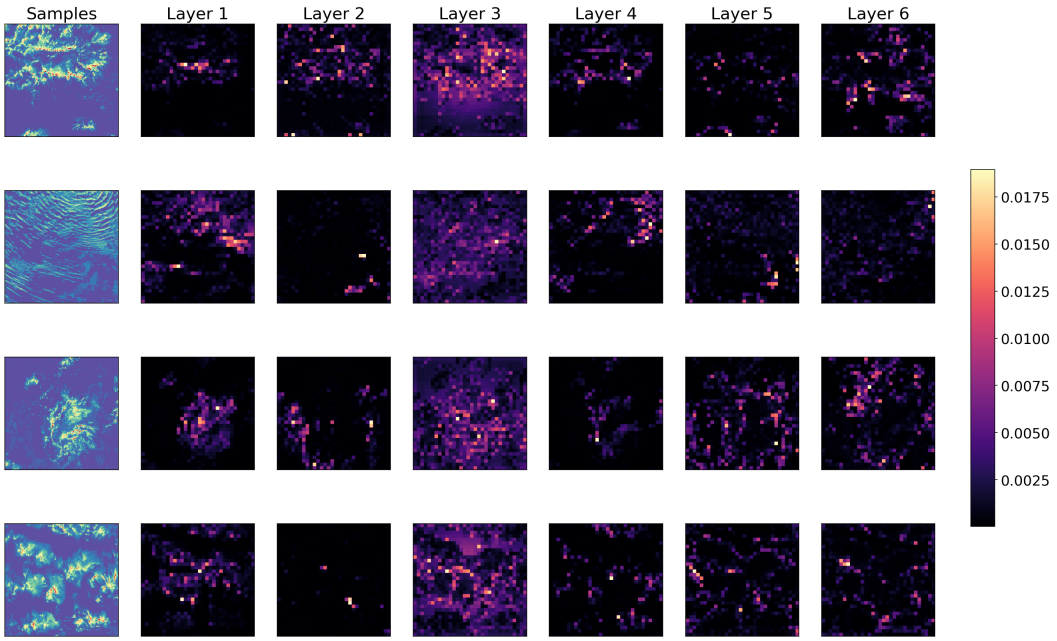


Figure 4.12: Four cloud systems with different organizations are selected as examples. Their respective self-attention maps from the final head of the teacher ViT (Figure 4.5) are shown for the 6 layers of the self-attention head. The color bar indicates the range of the Gaussian error linear units (GELU) activation function for the activation maps. Higher values indicate more important features. All experiments are run with a default of six self-attention heads.

To better understand N1's decision and to build trust in the network's performance, it is crucial to see what the trained N1 architecture has learned to pay attention to when deciding the features of cloud system distributions.

Given a satellite image, the activation space in a neural network allows us to visualize whether a neuron should be activated, indicating what part of the image is important for the network. The self-attention layers in ViT try to decompose the input samples and learn relatively independent features. Thus, this experiment aims to see whether the activation space reveals the abstract patterns that we, as humans, can make sense of while deciding the feature's importance. In this setup, we use a single satellite image sample and pass it through the trained model, freezing the weights. The granularity (N x N), or the number of pixels in a single patch, is controlled by the patch size, which is 8 x 8 pixels in this experiment. This is just for convenience and does not change the result compared to the 16x16 setting used for the main experiment.

Figure 4.12 shows that layer 1 activates at the dominant convective cells and deactivates at thin spread-out convection while layer 2 activates the thin spread convection. Layer 3 seems to try to learn and activate the clear sky features. In contrast to layer one, layer 4 activates the rest of the prominent convections. Like layer 2, layer 5 tries to look at the rest of the thin-spread convection. Layer 6 is uncertain and is not obvious to our eyes, and it may somehow try to deactivate for all the clear sky regions in the majority of cases and look for boundary semantics in the satellite image. Examining other example cases shows the same consistency, and therefore, it can be concluded that although the cloud system distributions are different, each attention map has learned to pay attention to relatively different, consistent, sensible semantics of the cloud systems distribution and further indicates that we can trust the embedding space of the network.

4.7.6 Environmental characteristics

Figure 2.c in Section 4.2 of the main manuscript showed the occurrences of 30 nearest neighbors of human-labeled satellite images (mentioned as human crops below) with machine-identified seven classes. Here, we aim to assess their existing environmental conditions.

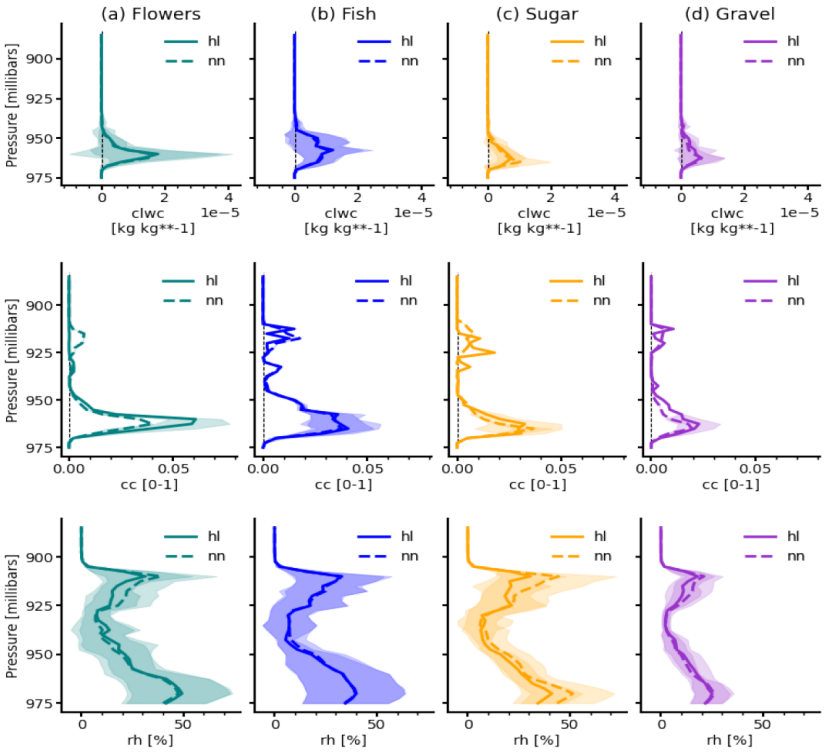


Figure 4.13: Comparison of 52 human labels (hl) environmental conditions with their nearest 30 neighbors (nn) using ERA-5. The top to bottom rows shows weighted-average and standard deviation profiles of cloud water content (clwc, kg kg $^{-1}$), cloud cover (cc), and relative humidity (rh, %) with the exception of cc variability shown in the interquartile range.

This complementary experiment can further help to trust human crops' relative positions in the feature space. If the human crops and the neighbors have a similar homogenous distribution of their physical properties, this implies that the human crops are in the consistent region of the feature space. Here, we take the ERA-5 vertical profile of cloud water content, cloud cover, and relative humidity (Fig. 4.13) to compare the weighted averaged vertical profiles between human labels and their 30 nearest neighbors. When calculating these properties for human-labeled scenes, we weigh them with the level of agreement. In this way, the contribution of well-agreed organizations will contribute more than less agreed cloud organizations. We observe that there is hardly any difference in the vertical profiles except for the relative humidity of sugar and cloud cover for flowers. This may be due to quantitatively using 30 times more data.

4.7.7 Transition probability

Based on our collected satellite imagery derived over Barbados cloud observatory (BCO, see Figure S1 (Domain)), we performed a transition probability analysis for all machine-identified cloud regimes. Here, we exploit the 10-minute temporal resolution of the data set to observe the transformation of cloud systems with time. Note that some of the temporal consecutive satellite imagery could be missing because of the pre-processing step.

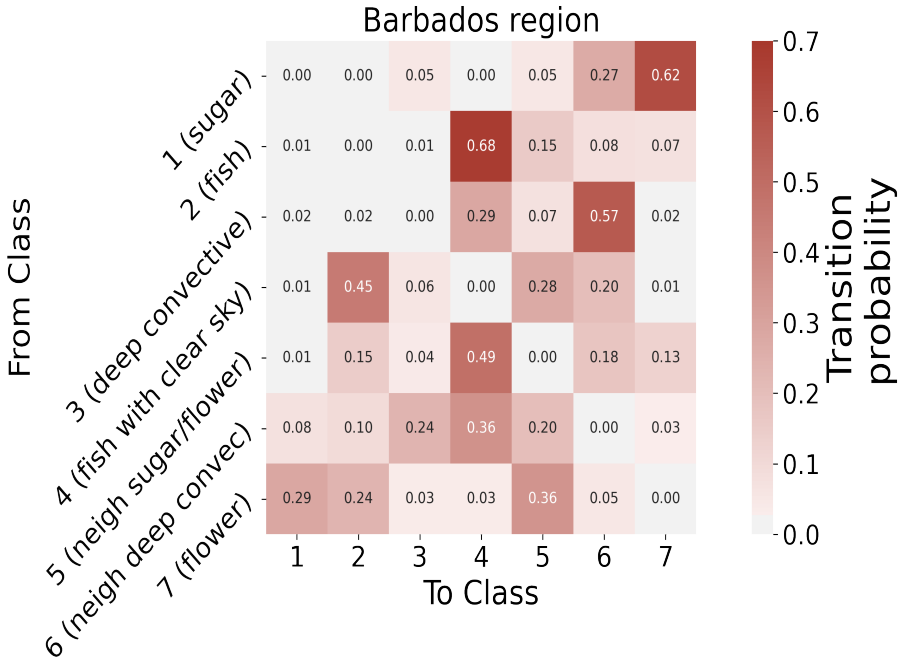


Figure 4.14: Matrix representing the probability of transition of each machine-identified cloud regime to another over the Barbados cloud observatory. The total number of samples for this analysis is 5,470.

The goal is to understand how close classes should be in the feature space to make a transition and derive a transition probability from one cloud regime to

another. We observe that sugar has the highest possibility (62%) of transforming to flower-type cloud systems (Figure reviewer-2:3). A mature developed fish could become a fish with more open sky with 68% probability (possibly due to cold pool downdraft effect). The deep convective systems follow their neighboring class in the continuum, i.e., class six. Fish with more open sky is followed by mature fish-type structures. Class five, which has cloud cells typically found in sugar, gravel, and flower, has 49% chance of a follow-up by fish with more open sky. Finally, the flower class has 36% chance of a transition to class 5 or 30% chance to sugar type distributions.

5 Towards solar energy application

Abstract

Ubiquitous low-level cloud systems play a crucial role in solar energy applications by significantly influencing Surface Solar Irradiance (SSI) and its spatio-temporal availability. The short-term variabilities caused by these clouds impact solar power plant dispatch and the stability of electricity grids, stemming from the fragmented nature and rapid evolution of small spatial-scale clouds. This study employs cloud optical depth (COD) retrieval from high-resolution Meteosat Second Generation (MSG) satellite imagery and a self-supervised neural network to isolate low cloud fields. By integrating ground-based pyranometer measurements with low-level cloud regimes, we investigate the clearness index (CI) over time, quantifying their impact on SSI fluctuations ranging from seconds to thirty minutes. Our findings indicate that low-level clouds are more likely to cause severe fluctuations in SSI compared to other cloud types. Improved differentiation of low-level cloud fields enhances understanding of their radiative and macrophysical characteristics. Additionally, our approach advances prior research by assessing the diurnal evolution of rapidly developing low-level cloud fields using a machine learning algorithm. Validation with ground-based radar and pyranometer observations demonstrates how inherent cloud properties cause changes in neural embeddings of the neural network. Progressing towards explainable artificial intelligence, preliminary results suggest that our ML algorithm applied to satellite data can infer rapid changes in the mesoscale horizontal distribution of low-level cloud systems.

5.1 Introduction

The success of policy initiatives like the Green Deal (EU-Commission et al., 2021) has driven a significant shift towards green energy, marked by the widespread adoption of renewable sources like solar power. For example, the German government has set photo voltaic installation targets of 215 GWp (Giga-Watt peak) by 2030 and 400 GWp by 2040, respectively (Kougias et al., 2021). However, this transition has intensified the reliance on the intermittent nature of solar energy, which is primarily induced by clouds, posing challenges to maintaining grid stability. For example, with solar energy's high contribution, managing stable grid frequency has become crucial, requiring effective control strategies to minimize fluctuations within a narrow range (± 10 mHz) around the standard frequency of 50 Hz (Anvari et al., 2016).

Low-level Cloud Fields (LCF) are one of the main atmospheric drivers causing steep fluctuations in the Surface Solar Irradiance (SSI) time series (Lamer & Kollias, 2015). They contain mostly liquid water and thus have high opacity, and

deeper clouds can develop from them under unstable states of the atmosphere. These clouds, often characterized by their broken structure, create distinct cloud shadows and sunlight exposure (Gristey et al., 2020; Lohmann et al., 2016), demonstrating a distinct bimodal SSI distribution, which drives the fluctuations in the PV output. Also, because of their small spatial scale and rapid evolution, understanding the processes by which the broken cloud fields control SSI is non-trivial (Lane et al., 2002). Hence, by assessing the spatio-temporal dynamics of these low clouds and quantifying their effects on SSI, we can gain insights into how fast microgrid controllers should respond to maintain stable frequency in microgrids (Uriarte et al., 2015; Perez et al., 2016; Krismanto et al., 2018).

SSI variability is a function of predictable solar geometry and cloud/weather influences, where rapid fluctuations in power output are mainly caused by the latter. Employing the transmissivity or the Clearness Index (CI), a normalized metric representing the ratio of SSI to the top of the atmosphere extraterrestrial irradiance is advantageous for discerning the impact of cloud-related factors. CI removes the solar geometry effect while preserving the cloud-induced variability (Lohmann et al., 2016). The change in CI over a specified time interval, δt, is defined as ramp rate and describes changes in solar photovoltaic (PV) power output (Lave & Kleissl, 2010; Haaren et al., 2014). Critical operation ramp rates depend on power grid scales and PV capacity.

Considering different time intervals δt to calculate ramps can provide insights into their varying impacts on solar power. The spatio-time scales can range from seconds and meters up to several hours and hundreds of kilometers, depending on the concerns of grid operators (Ranaweera et al., 2014; Perez et al., 2015). Generally, short-term fluctuations have received less attention in research, partly due to the challenge of accessing high-frequency power data, such as 1 Hz. Also, the common assumption is that rapid fluctuations tend to average across geographic regions. However, with the integration of renewables into the decentralized grids, the short-term temporal imbalances have gained attention (Marcos et al., 2014). Past studies have investigated the short-term effects of these fluctuations occurring in 15 or 60-min intervals (Iversen & Pinson, 2016; Ordoudis et al., 2016; Sengupta et al., 2015) because these scales are needed to maintain the standard frequency of the grid (Hasenbalg et al., 2020). Also, decisions made at short intervals are driven by trading strategies in the energy market. As such, if the error in solar power forecasting is reduced, then electricity markets can trade with more certainty. (see `www.epexspot.com/en/product-info/Intraday/germany`, (Foley et al., 2012)).

Investigating short-term solar irradiance forecasting based on cloud regimes information (McCandless et al., 2016) or temperature forecasts linked to weather regimes (Greybush et al., 2008) have demonstrated superior performance compared to approaches that do not consider specific regimes (Pedro & Coimbra,

2012; Marquez & Coimbra, 2011). For example, McCandless et al. (2016)'s research highlights how the regime-dependent system makes a more accurate deterministic forecast and better captures the expected irradiance variability than a globally trained model, which doesn't consider cloud regimes. This is likely because the cloud systems belonging to a particular regime should have comparable microphysical and macrophysical properties (see Fig. 3.7 from Chapter 3), and, therefore, the models can better learn their individual data distribution separately. Thus, a regime-dependent approach, on the one hand, can differentiate among the cloud field properties, which will likely map the SSI variability better. Second, cloud field distributions can infer information about existing meteorological environments, which could be an auxiliary variable for solar energy forecasting (AlSkaif et al., 2020; Carrera et al., 2020).

Chatterjee, Acquistapace, et al. (2023) (see their Table 3.1) identified two (out of the seven) reasonable LCFs (class 3 and 7) over central Europe, which mainly featured overcast conditions. These two classes had a relative frequency of occurrence of 17 and 7 %, their average surface downwelling solar radiation (the satellite retrieved equivalent of ground-measured SSI) was 582 and 488 W/m^2, and average cloud top height was 3.1 and 3.5 km. While broadly containing the entire LCFs in two classes may be sufficient for some purposes like model assessment of cloud regimes (Williams & Webb, 2009), however, to capture the distinct SSI bimodal characteristics and their short-term rapid developments (Gristey et al., 2020; Sedlar et al., 2021), delineating them into more specific low cloud types becomes crucial for solar energy applications.

The Present and upcoming generation of geostationary satellites offer continuous high spatio-temporal resolution observations, further enabling our understanding of rapidly changing dynamic systems in the Earth's atmosphere. For example, the Flexible Combined Imager (Durand et al., 2015) onboard the launched Meteosat Third Generation (MTG) has a very high spatial resolution, around 1 km, with a repeat cycle of 2.5 minutes over Europe. We hypothesize that the LCFs should be uniquely recognizable with these advanced geostationary observations.

Using geostationary satellite data would enable short-term assessments over large domains, which is currently only pursued from ground-based measurements. Lohmann et al. (2016) assume a fixed sub-hourly scale of 15 minutes as the persistent time scale for all cloud systems evolving over Juelich, Germany. Using this time window, they attempt to account for the temporal variability in ground-based pyranometer network observations during a field experiment (Bomidi, 2020) under different sky conditions. They define different cloud conditions as i) clear sky, ii) overcast (large stratus-type cloud systems), and iii) mixed sky conditions (heterogeneous cloud fields with smaller length scales than those in overcast conditions). However, average persistence time intervals of different cloud regimes

over Juelich can vary by 80 % (see Fig. 3.11 from Chapter 3). Therefore, their analysis most likely includes overlapping effects of other cloud regimes. Also, investigating SSI variability with pyranometers only captures the local scale variability and does not directly account for the effect of large-scale cloud fields and their local impacts, which is important for solar energy applications. More specifically, they could not separately segregate the crucial LCFs to quantify the associated SSI local ramps and their variability precisely.

Our work is also partially inspired by the concept of Living Lab Energy Campus (LLEC) at Forschungszentrum Juelich (FZJ) (Althaus et al., 2022) where they installed a PV power plant with a total capacity of 1.1 MWp (Mega Watt peak) at their former reactor site and another 1.5 MWp on building rooftops. Thus, stabilizing their microgrid (Rohden et al., 2012) by integrating all PV systems on the campus requires high-quality short-term (0 - 30 mins) solar energy forecasts. The LLEC experiment allows us to assess the higher ramps caused by LCFs in short intervals. In this aspect, the high spatio-temporal resolution observation coverage of the mesoscale LLFs on top of Juelich from geostationary satellites can help us better infer the associated short-term variability. In addition, Jülich Observatory for Cloud Evolution (JOYCE, (Löhnert et al., 2014)) offers ground truth for validating our method.

Apart from being regulated by the atmospheric state, LCF evolution over land is a complex process influenced by land surface heterogeneity and highly complex topography. This affects the development of their convective states and triggers their interaction with neighboring clouds in the atmosphere (Rieck et al., 2014). Representing LCF dynamics through geostationary satellite observation involves capturing their evolving convective stages and the transition states over time. As a transfer learning task, we devise a framework (as explained in Section 4.3 of Chapter 4) that follows the manifold of LCF diurnal evolution over Juelich in low dimensional feature space. This enables capturing changes in their mesoscale horizontal appearance. Evaluating the evolving manifold also shows how changes in different dimensions of the embeddings are associated with the cloud field's intrinsic physical properties. The changes are analysed through ground-based radar reflectivity and satellite retrievals.

The overarching goal of this study is to exploit high-resolution satellite observations and pyranometer measurements to investigate fluctuations in SSI associated with LCFs. Our framework of the developed neural networks N1 (Section 4.7.2 of Appendix 4.7) and N2 (Section 3.3 of Chapter 3) help us to differentiate the LCFs from the rest of the cloud field spectra. We extend our work over the central European domain using a one-order magnitude more extensive data set than Chatterjee, Acquistapace, et al. (2023). As a proof of concept, section 5.3.1 analyzes the robustness of the earlier identified seven classes found by Chatterjee, Acquistapace, et al. (2023) with the extended dataset. Section 5.3.2 investigates

the SSI variability associated with the seven classes measured through a ground-based pyranometer. Here, the implications of the different classes are analyzed by matching them with a ground-based pyranometer. Section 5.4 investigates the benefits of better differentiating the low cloud systems for capturing the variability of SSI ramps. Finally, as an outlook, we deploy a transfer learning task, where the framework (as defined in Section 4.3 of chapter 4) tracks the diurnal evolution of the low cloud regimes as captured by the NN to demonstrate swift changes in the cloud field horizontal structures, measured through a ground-based pyranometer. We investigate changes in different dimensions of the neural embeddings to associate them with the cloud field's intrinsic physical properties.

5.2 Data and method

Images of COD retrieved from enhanced Meteosat Second Generation (MSG) (also explained in Chapter 3, Sec. 3.2) already built the base for our first study (Chatterjee, Acquistapace, et al., 2023), where we considered only the spring and summer months (April to July) in 2013 and 2015. Here, we expand the satellite data from April to September 2013 until 2019 [1], which marks a significant extension of 82% compared to our first study. Extending the study period to encompass spring, summer, and partly autumn months allows for a more comprehensive assessment of cloud conditions. This broader time frame enables the inclusion of locally triggered low clouds during various convective situations, which pose challenges to numerical modeling and solar energy production owing to their significant variability. At each timestamp throughout this extended time frame, our methodology involves the collection of five random crops of 128 x 128 pixels over central Europe (6 – 18°E, 47.5 – 53°N). We capture samples from 06:00 to 17:00 every 5 minutes to avoid low solar zenith angles, which affect retrieval quality. Ensuring data quality like in Chatterjee, Acquistapace, et al. (2023), and with about 132 COD images per day, we obtain 660K image samples. Along with COD, other satellite-retrieved cloud properties (Deneke et al., 2021) are available, such as cloud top height (CTH), cloud mask (CMASK) or cloud fraction, cloud top pressure (CTP), top of the atmosphere reflected solar radiation (TRS), and Short wave Downwelling Solar radiation (SDS) which are utilized for the physical interpretation of the cloud systems.

For parts of the analysis, we explore the cloud system evolution over a central site, i.e., Juelich, Germany (50.9224°N, 6.3639°E). Centering over Juelich aims to exploit the ground-based JOYCE station observations. Here, we use the CloudNet target classification (Illingworth et al., 2007), which characterizes clouds by their geometrical extent and microphysical properties. CloudNet classifies 11 classes: ice, drizzle, cloud droplets, aerosol, and insects, to name a few. It uses measurements from a cloud radar combined with microwave radiometer and

[1] except 2014

ceilometer observations merging to a temporal resolution of 30s and a vertical resolution of 30m extending up to 12 km vertically.

We also use a pyranometer (CMP21) at a one-minute temporal resolution to measure incoming SSI. The clear sky index for the location of 50.9083 ° N, 6.4125 ° E is determined based on the work of Perez et al. (2002); Ineichen and Perez (2002) and top of the atmosphere (TOA) extraterrestrial irradiance is taken from PV LIB package (Andrews et al., 2014). Note that the location of the pyranometer at Forschungszentrum Jülich, Germany, via Google Maps differs by 20m horizontally and vertically from the reference point given by Löhnert et al. (2014)(depending on the part of the building and roof/ground).

We use the COD images to train our self-supervised neural networks N1 and N2, as outlined in Section 4.7.2 and Section 3.3, respectively. N1 addresses the task of finding a suitable embedding, hence sorting cloud organizations in a continuous fashion based on intrinsic similarities, and the second neural network (N2) segments the retrieved representations into a fixed number of classes. This process yields a 384-dimensional continuum feature space of cloud field spectra from N1 and a 128-dimensional feature space partitioned into 7 clusters from N2 (see Fig. 4.4.1.a). Chatterjee, Acquistapace, et al. (2023) showed that k = 7 is optimal for the old dataset. However, later on in Section 5.4, we repeat N2 with k = 20 to obtain a more detailed differentiation of the cloud fields.

For dimensionality reduction of the high dimensional feature space (explained in 4.7.4), we use the t-SNE manifold extraction algorithm (van der Maaten & Hinton, 2008b). We use cosine distance as a distance measurement criterion and initialize tSNE through Principle Component Analysis (PCA).

5.3 Results

5.3.1 Robustness of the network

Once the network (N2) has sorted the individual images into seven optimal classes, we want to investigate the robustness of the newly identified classes with the older ones. We examine the consistency regarding feature representation, physical space evaluation, and their diurnal cycle.

Feature space-based qualitative comparison Following the training of N2 using the extended dataset, we examine the newly derived cloud regimes and compare them to those from the old dataset using t-SNE (Fig. 5.1).

Using cosine distance metrics, the arrangement of the classes in the 2D projection (Fig. 5.1b) slightly changes as compared to the old setup (Fig. 5.1a), but the features of cloud systems remain consistent, lying in different areas of the 2D feature space. We still find the classes that are more contrasting in their COD distribution, such as clear sky conditions (class 2) and deep convective sys-

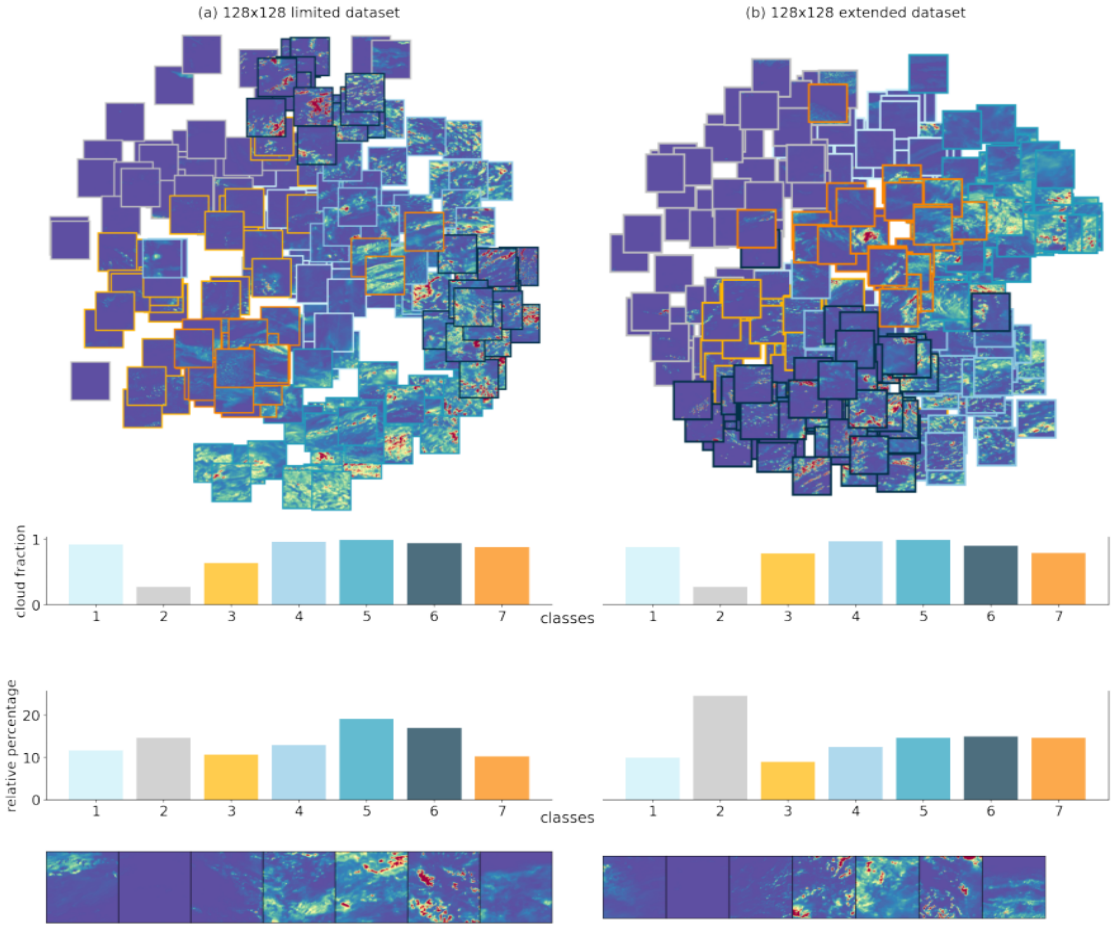


Figure 5.1: (a) Visualization of the dimensionally reduced original N2 feature space, trained exclusively using 2013 data. We display forty random samples of 128 x 128 image configurations from the old dataset within each cluster category (as originally depicted in Fig. 3.6). Each image is bordered by a distinct color corresponding to its assigned class. (b) A similar visualization of the dimensionally reduced but newly trained N2 feature space using the extended dataset. Bar charts located beneath panels (a) and (b) illustrate the relative occurrence (in %) and cloud fraction of each regime. Additionally, we provide the centroid images to facilitate association with the cloud regime.

tems (class 5 and 6) lying on the edge of the 2D projection as in Figure 5.1a. Meanwhile, intermediary systems, such as low-level cloud systems likely having overcast conditions (classes 3 and 4), appear in the middle.

Examination of the cloud fraction (CF) associated with these identified regimes shows a similarity to previous patterns, confirming their qualitative likeness in nature. Also, the centroid's COD images from each cloud regime illustrate their uniqueness and look similar to the centroid images from 2013. Clear sky appears to be the most frequent class in the extended dataset compared to the deep convective class 5 from 2013. Additionally, the occurrence of Class 7 has increased by 50% since 2013 (Fig. 5.1). However, the analysis of cloud regimes' frequency of occurrence using the extended data record indicates the inter-annual variability

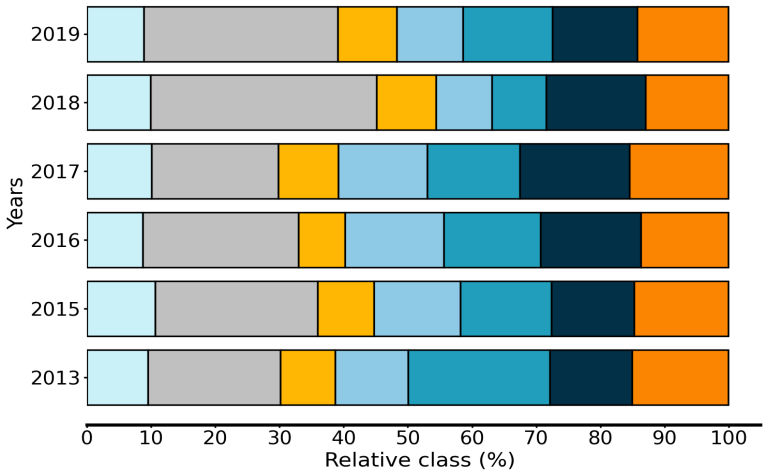


Figure 5.2: Each cloud regime's relative frequency of occurrence from 2013-2019 except 2014

of the classes(Fig. 5.2).

A consistent inverse relation emerges where an increase/ decrease in clear sky conditions (class 2) seems to decrease/increase deep convective system (class 5) occurrences. This is likely because each year is characterized by different characteristics and environmental conditions. The drivers or controlling factors of such inter-annual variability in cloud regimes are to be investigated at a later stage (not in this thesis).

Physical consistency We evaluate the radiative and cloud physical properties of the earlier identified seven cloud regimes with the newly derived ones. The approach involves co-locating randomly picked 1000 images (as in Denby (2020a); Chatterjee, Acquistapace, et al. (2023)) from each of the seven optimal classes and subsequently determining their mean position within the physical space. For example, the Short wave Downwelling Solar radiation (SDS) (note SDS is the satellite equivalent of ground-measured SSI) and top-of-the-atmosphere reflected solar radiation (TRS) illustrate the radiative space (Fig. 5.3) of the cloud regimes. Similar to the well-known International Satellite Cloud Climatology Project (ISCCP) classification, the cloud top pressure (CTP) and COD show the vertical distribution space of the cloud regimes (Fig. 5.4).

The radiative and cloud physical property space shows similar behavior for the old and extended datasets (Fig. 5.3 and 5.4). Class 1 or "cirrus-type" cloud systems demonstrate large variability in albedo and surface downwelling solar radiation. Class 2, or the cloud-free class, has the least COD and allows most solar radiation to reach the surface. Because the surface absorbs most radiation, this class stands out with the lowest TRS, while SSI varies with the solar zenith angle. Among the identified cloud regimes, class 3 has the lowest level clouds (CTP highest) with a mean height of around 780 hPa and allows the highest

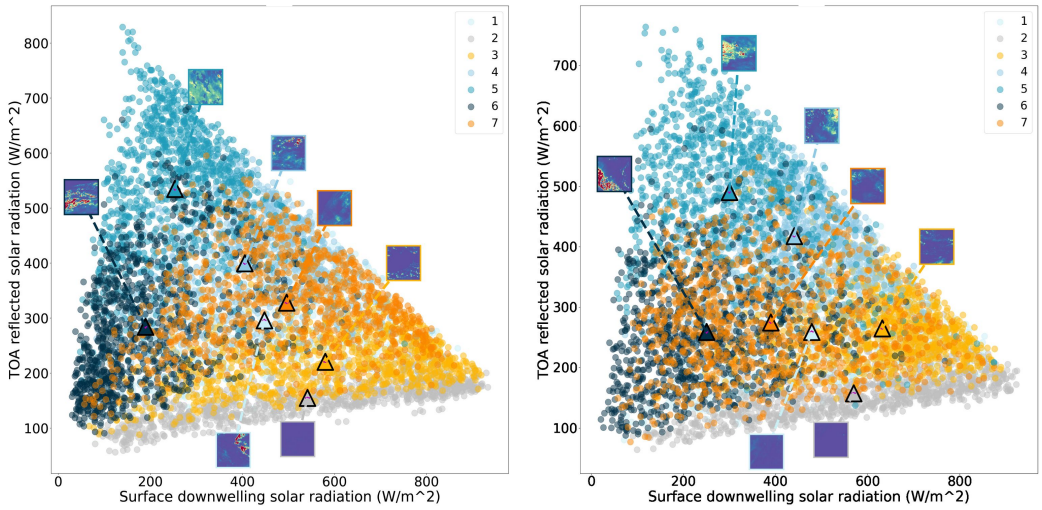


Figure 5.3: Mean TOA reflected solar radiation (TRS) versus mean surface downwelling solar radiation (SDS) given by dots of 1000 randomly selected images of the 128 x 128 configuration. On the left for the older dataset and the right for the extended dataset. Colors indicate the class to which they belong. Triangles indicate the mean position of the classes. The images associated with the triangles are the closest to the mean position of the 1000 images.

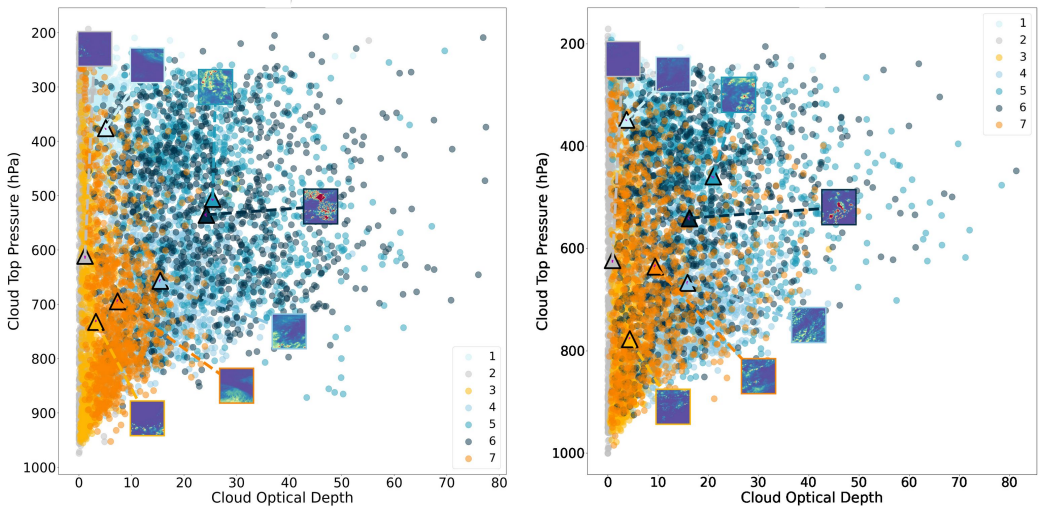


Figure 5.4: Same as Figure 5.3 but for COD (abscissa) and cloud top pressure (ordinate).

solar radiation to reach the ground (see Table. 5.1 for more details). Note that the SSI distribution for class three slightly shifts to a higher SSI compared to what was found earlier with the old data.

Unlike before, where class 7 was the second lowest-lying cloud class in the

older dataset, now class 4, with mean CTH at 3.7 km, contains the second lowest-lying cloud regime among the seven classes (see Table 5.1 or Fig. 5.4). Class 4 demonstrates the same high cloud fraction with a mean value of 97% and COD values with a mean distribution of 15.5. Classes 5 and 6 represent different stages of deeper, thicker clouds with a mean COD of 21 and 17 respectively. However, like the shorter data set, it had a similar cloud fraction, and class 5 reflects twice as much as class 6.

Class 7 stands out as the only class with cloud macrophysical properties slightly contrasting with the older class 7 (refer to the old cloud regime properties in Table 3.1). Its mean cloud height distribution has shifted from 700 to 635 hPa, accompanied by a slight 15% increase in mean COD and a 10% reduction in cloud cover. This implicates more mature convective cells and results in a 20% decrease in SSI reaching the ground and an 18% decrease in reflection into space.

Consequently, the systematic comparison reveals that most identified classes occupy similar positions in the 2D feature space and share co-located physical and radiative properties.

Table 5.1: Characteristics of the seven cloud regimes for the extended dataset followed by old dataset (same as Table 3.1) for comparison: The relative frequency of occurrence (RFO), Mean and standard deviation (std) of cloud mask (CMASK), cloud optical depth (COD), cloud top height (CTH), cloud top pressure (CTP), cloud top temperature (CTT), shortwave downwelling radiation (SDS), TOA reflected solar radiation (TRS).

Class	1	2	3	4	5	6	7
RFO	9.8|13.2	24.4|19.6	8.9|15	12.4|16.3	14.6|16.8	15|12.4	14.6|6.6
CMASK mean (%)	87.8|92.2	26.8|27.3	78.1|64.5	97.1|96.5	98.6|99.1	89.8|94.8	78.5|88.2
CMASK std (%)	9.1|10.7	8.3|21.6	7.7|17.6	7.3|5.5	3|3.6	8.2|8.5	8.6|12
COD mean	3.6|5.3	0.9|1.2	4.3|3.3	15.5|15.3	20.6|24.8	16.8|23.4	8.5|7.4
COD std	3.1|4.6	1.8|2.7	3.8|3.3	6.9|6.9	6.2|11.5	10.9|13.9	5.2|4.4
CTH mean (m)	8759|8212	5013|5152	2675|3198	3760|3866	6769|6001	5752|5642	4354|3529
CTH std (m)	1578|1734	3699|3606	2552|2226	1370|1625	2354|2160	2463|2245	2480|2259
CTP mean (hPa)	349|374	622|613	778|736	668|657	456|498	536|530	634|697
CTP std (hPa)	40|101	49|245	39|164	26|118	25|142	33|146	34|157
CTT mean (k)	233|236	260|260	274|271	264|264	246|248	253|251	262|265
CTT std (k)	9|10	22|24	16|15	12|11	13|14	14|13	12|14
SDS mean (Wm^{-2})	483|442	571|546	532|582	442|408	302|253	247|194	400|488
SDS std (Wm^{-2})	172|183	202|232	185|196	123|143	98|104	116|123	205|171
TRS mean (Wm^{-2})	261|298	156|155	265|224	412|401	491|530	263|284	267|328
TRS std (Wm^{-2})	171|183	242|232	176|196	123|143	114|104	125|123	158|171

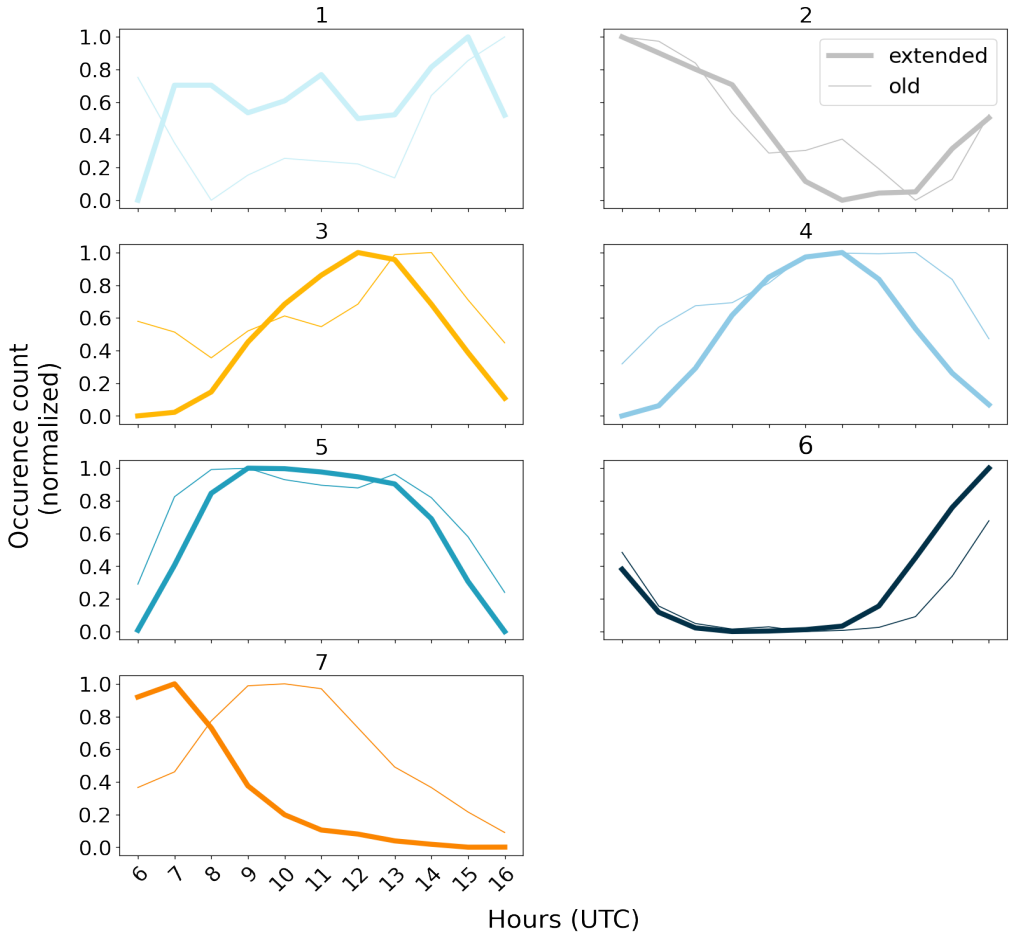


Figure 5.5: The normalized diurnal cycles of the occurrence of cloud regimes 1 through 7 over Juelich, Germany, are depicted here (1-hour averages). Each frame corresponds to a specific cloud regime, distinguished by line color [1-7] as in Fig. 5.1, illustrating its diurnal pattern. The thicker line portrays the daily cycle derived from the extended dataset (2013 - 2019), while the thinner line represents the one for 2013.

Diurnal cycle This section compares the diurnal cycle between the newly derived and the older identified seven classes not shown by Chatterjee, Acquistapace, et al. (2023). The phase of a cloud system's diurnal characteristics represents their unique signature, while amplitude may vary from region to region (Bergman & Salby, 1996) also found in Eastman and Warren (2014). Daytime surface heating and nighttime radiative cooling are two main drivers of the diurnal cycle of low clouds, while higher-level clouds such as cirrus frequently connect to the synoptic systems and largely remain unaffected by the effects of the surface. Understanding these diurnal variations is crucial for evaluating weather and climate models and probing cloud parameterization schemes. Note that COD, as a derived quantity from passive measurements, may see the low-level cloud diurnality masked by high-level clouds.

It's evident that class 1 (Cirrus) exhibits minimal periodicity in their occurrence patterns, and class 2 (clear sky) does not peak during mid-day (Fig. 5.5),

as it shows a slight dip in the afternoon when typical convective cloudiness is typically more pronounced (Ge et al., 2021). However, the remaining classes demonstrate distinct diurnal effects, likely in response to surface heating, manifested by varying phase shifts and different occurrence amplitudes in diurnality. Class 3 peaks typically around noon, with higher occurrence values skewed towards late noon. Conversely, classes 4 and 5 demonstrate a more symmetric occurrence distribution, with class 4 notably exhibiting larger amplitudes.

Class 6, characterized by intense convective activity, tends to appear later in the day. Notably, compared to the patterns observed in the older data set, Class 7 showcases a more pronounced phase shift toward the early morning hours than the old data set. It likely indicates stratus-type cloud systems or early morning fog as also robustly established in Eastman and Warren (2014) (see their Fig. 10). In general, the daily cycle appears more distinct, consistent, and smoother in the newer data set than in the older one indicating its strong statistical robustness. Hence, this detailed comparison provides valuable insights into the evolving diurnal behavior of these cloud classes, shedding light on potential shifts over the day. The diurnal behavior could also be connected to the weather types as, for instance, it is also closely related to the diurnal precipitation cycle (Nesbitt et al., 2008; Zhao et al., 2017).

Measured short wave characteristics As a last test for the robustness of the seven classes, we compare the old and new data sets using SSI measured at the surface with a ground-based pyranometer. The interaction between clouds and radiation depends on a variety of physical attributes of the clouds, such as their vertical and horizontal arrangement, cloud phase, i.e., ice, liquid, or mixed, and the size of cloud particles. The seven cloud regimes identified through COD distribution may not represent entirely distinct cloud systems but composite features derived from multi-layer cloud systems (Mace & Wrenn, 2013). However, we can improve our understanding of cloud-radiative interactions by delineating the seven cloud regime's associated surface radiative fluxes.

Considering this, we move towards exploring cloud regimes based on surface radiation measurement using a pyranometer installed at JOYCE. It is important for solar energy applications to assess how to get from spatial class to their local impacts. We hypothesize that cloud systems belonging to a particular class should have comparable microphysical and macrophysical properties, and, therefore, depending on the cloud regime, it interacts with the incoming short-wave radiation uniquely. To reduce the effect of the diurnal cycle on surface radiation, we normalize pyranometer measurement with TOA extraterrestrial irradiance. Hence, we calculate the CI or transmissivity of pyranometer measurements and assign them to the seven cloud regimes. In this way, we combine ground measurements representing a localized view spanning a few kilometers with satellite data, encompassing a much larger scale—approximately two orders of magnitude

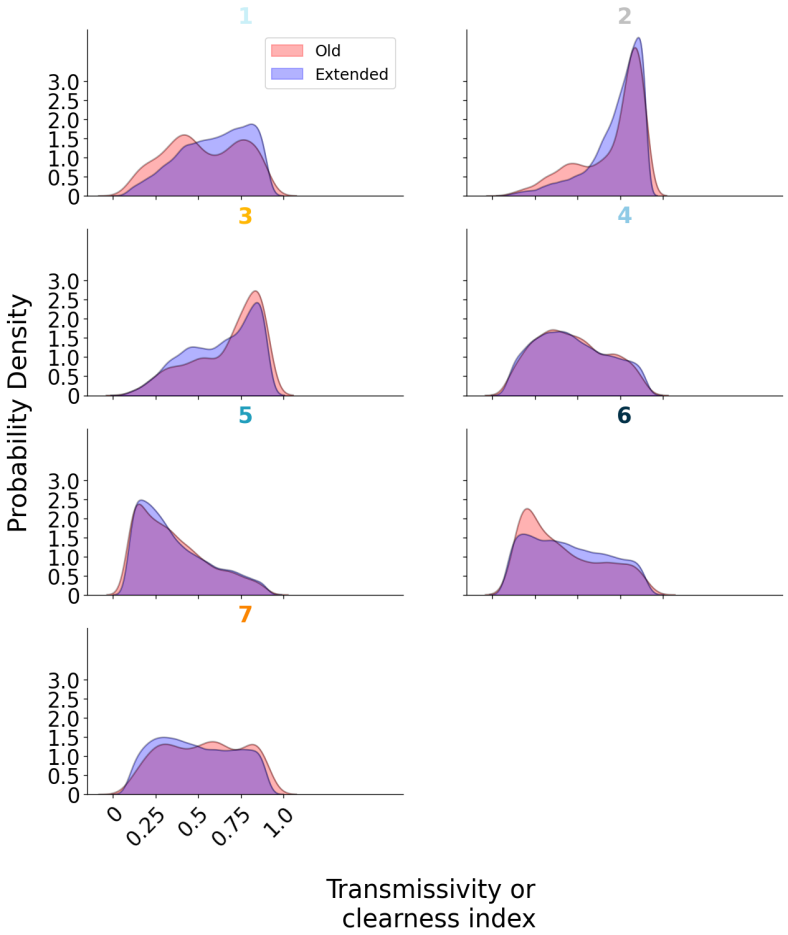


Figure 5.6: The probability density estimate of measured transmissivity using a pyranometer at the JOYCE observatory in Juelich, Germany. The time stamps are considered from 07:00 to 16:00 UTC, as the early morning sunrise and late sunset time stamps are excluded. It is categorized according to seven cloud regimes. Each frame illustrates the density estimates of both the older and extended datasets, distinguished by the colors representing the area under the curves.

greater.

The pyranometer measurement at JOYCE is at a 5-second temporal resolution. To match the satellite time stamp every 5 minutes, we consider radiation values from the pyranometer $\pm$ 5 minutes averaged around the satellite time stamp as its SSI representative.

In the newly derived cloud regimes (Fig. 5.6), class 1 exhibits a stronger unimodal distribution in transmissivity compared with the older and shorter dataset. The new regime shows a higher occurrence of the higher transmissivity ranges. Class 2, being the clear sky and having the highest density at transmissivity close to one, allows maximum measured radiation to reach the ground. Class 3 shows skewness in the higher transmissivity ranges, while class 4 shows the bulk shift in lower ranges. Classes 5 and 6, belonging to different stages of deeper convection, demonstrate the attenuation of solar radiation by more frequent low transmitted values. Class 7 shows a nearly uniform solar energy attenuation between 0.2 and

0.8, likely to be a type of stratiform cloud system, such as stratus or altostratus clouds (as also hypothesized from their diurnal behavior in Fig. 5.5), as their uniform and extensive coverage leads to the relatively consistent transmissivity of solar energy through the cloud layer.

N2 effectively characterizes the diversity inherent in the structure and distribution of mesoscale cloud regimes. The newly identified seven cloud regimes exhibit physical and radiative properties similar to their predecessors (Fig. 5.3 and 5.4). Notably, mesoscale cloud systems with more contrasting features tend to reside on the periphery of the 2D feature space (Fig. 5.1). This observation likely reflects a unique attribute of the latent space, where distinct cloud structures and distributions occupy positions at the extremes, while intermediary systems predominantly populate the middle of the 2D representation.

Expanding the investigation with a larger data set reveals differences in occurrence frequency, particularly notable in clear sky conditions. However, the overall cloud fraction remains consistent, except for class seven (stratus cloud systems). Regimes likely influenced by surface heating exhibit a robust daily cycle, and their interaction with solar radiation displays various degrees of solar energy attenuation across the measured range of the pyranometer. Class seven demonstrates modified responses regarding diurnal patterns and radiative properties, while classes 1 and 3 show some interchange. Note that the diurnal and radiative distributions appear more pronounced and smoother than in the older dataset, underscoring the notion that training with increased data yields improved delineation of boundaries within the mesoscale cloud system distribution.

5.3.2 JOYCE SSI characterization

Our ultimate goal is to link satellite cloud regimes to characteristics relevant to solar energy applications on the ground. In our previous analysis (Section 5.3.1), we used retrieved quantities, e.g., SDS, TRS, CTP, which, however, suffered from the missing vertical resolution of the passive measurements (Rosa & Silva, 2021). Now we look at the pyranometer measurements and asses the cloud regimes' local spatio-temporal impacts.

The influence of cloud regimes on such local variability is crucial for managing the mismatches between supply and demand, which, if unmanaged, can cause power outages or complete grid failures. (Perez et al., 2015; Ranaweera et al., 2014). Also, clearly establishing the relationship between cloud regimes and their influence on local energy production can help assess future ramp scenarios using climate projections (Poddar et al., 2023).

Among the seven identified cloud regimes (Fig. 5.7), Class 3 stands out by causing the highest spatially averaged solar radiation to reach the surface (derived from pyranometer), as depicted in Figure 5.7a. Its class-wise average radiation has a median value of 700 W/m^2, with an inter-quartile variability of 200 W/m^2

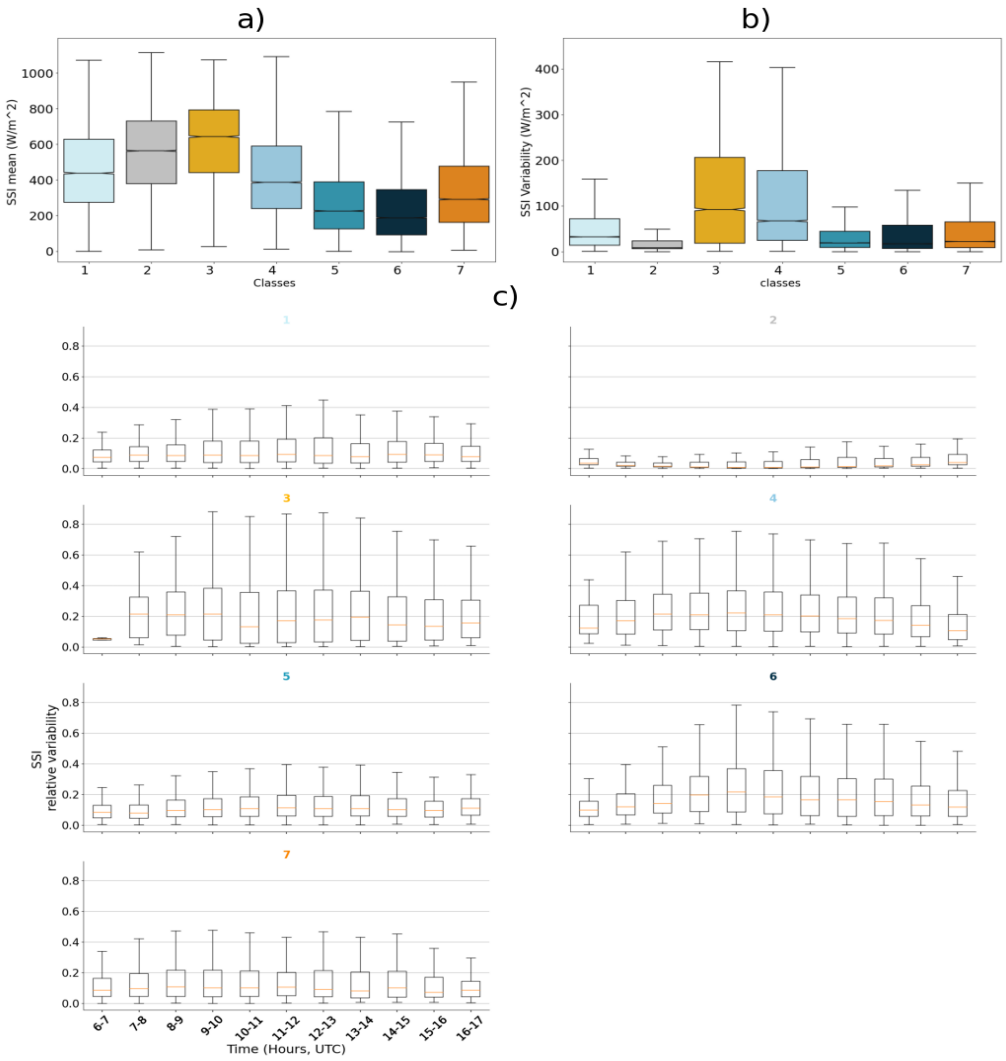


Figure 5.7: Box plots indicating the median (central line) and interquartile (IQR) range of 25th and 75th percentile (box edges) and whiskers (1.5 times the IQR from the lower and upper quartiles) for the seven cloud regimes (n = 129,715) observed by pyranometer at the JOYCE Observatory in Juelich, Germany, spanning from 2013 to 2019 (excluding 2014), illustrating: a) the spatial average distribution of surface solar irradiance (SSI, W/m^2), b) its standard deviation (W/m^2), considered from 06:00 to 17:00 UTC and c) the hourly temporal variability of relative standard deviation (rstd) of SSI between 06:00 to 17:00 UTC for each class.

(derived using 5 minutes average), as shown in Figure 5.7b. Out of the two low-lying cloud systems identified in Section 5.3.1 (also see Table 5.1), class 3 was identified as the lowest altitude cloud regime, exhibiting a peak occurrence post-noon (Fig. 5.5). Following closely, Class 4 has a median SSI of 600 W/m^2 and demonstrates significant variability, with an interquartile range of 160 W/m^2 (20% less than Class 3). It represents the second lowest altitude cloud system, with a symmetric occurrence distribution around noon (Fig. 5.5). In contrast, the remaining CRs do not exhibit such high spatial variability associated with them.

In terms of temporal variability, the diurnal patterns of measured surface radiation (Figure 5.7c) exhibit the highest relative standard deviation (rstd) for classes 3 and 4. A higher rstd indicates greater variability at the surface around the mean SSI, while a lower rstd suggests more stability. The rstd of SSI and its variation with time is crucial for photovoltaic (PV) generation as it considers the occurrence cycle of cloud regimes and can suggest when a certain cloud regime potentially impacts the system performance more than other times when the uncertainty is lower.

On the one hand, class 3 demonstrates a larger and consistently wide interquartile spread throughout the day. On the other hand, class 4, with a 20% higher cloud fraction (CF) than class 3 and a broader range of cloud optical depth (COD) values (refer to Figure 5.3b), exhibits a comparatively reduced temporal uncertainty when compared to class 3. Note the behavior of class 6 (deep convective class), which, even in relative terms, has a string diurnal cycle.

In Section 5.3.1, we explored how the seven distinct cloud regimes exhibit unique cloud properties, leading to varying levels of solar energy attenuation (Figure 5.6). Our focus narrowed down to two specific low-lying cloud regimes, class 3 and 4, which significantly contribute to the observed spatio-temporal variability in measured solar surface irradiance (SSI) (Figure 5.7). It's worth noting that although Cirrus (class 1) and Stratus (class 7) cloud regimes allow considerable incoming solar radiation, their spatio-temporal variability is notably less pronounced compared to classes 3 and 4. For example, the interquartile variability in measured solar irradiance by the pyranometer is 75% lesser for Cirrus and 78% less for Stratus, respectively, compared to the broader variability observed in class 3.

5.4 Better differentiating the low cloud regimes

The low-level cloud regimes (3 and 4) derived from N2 are characterized by significant spatiotemporal variability in solar irradiation and are primarily based on cloud optical depth (COD) distribution—a metric derived from passive measurements. It's essential to recognize that these regimes may not solely encapsulate distinct cloud system properties (Chen & Genio, 2009). Instead, they likely comprise composite features stemming from multi-layer cloud systems, potentially resulting in variations within the regime in terms of radiation (Jakob, Tselioudis, & Hume, 2005) and precipitation (Jakob & Schumacher, 2008). Consequently, there's a need to better distinguish these challenging-to-identify low-level clouds from the overcast systems across the entire spectrum of cloud formations.

Therefore, we further differentiate these low cloud systems from the extended dataset to address the objective of capturing this variability more effectively. We demonstrate the flexibility of N2 in achieving this differentiation among the various low-cloud regimes. We employ N2 to partition our data into 20 classes

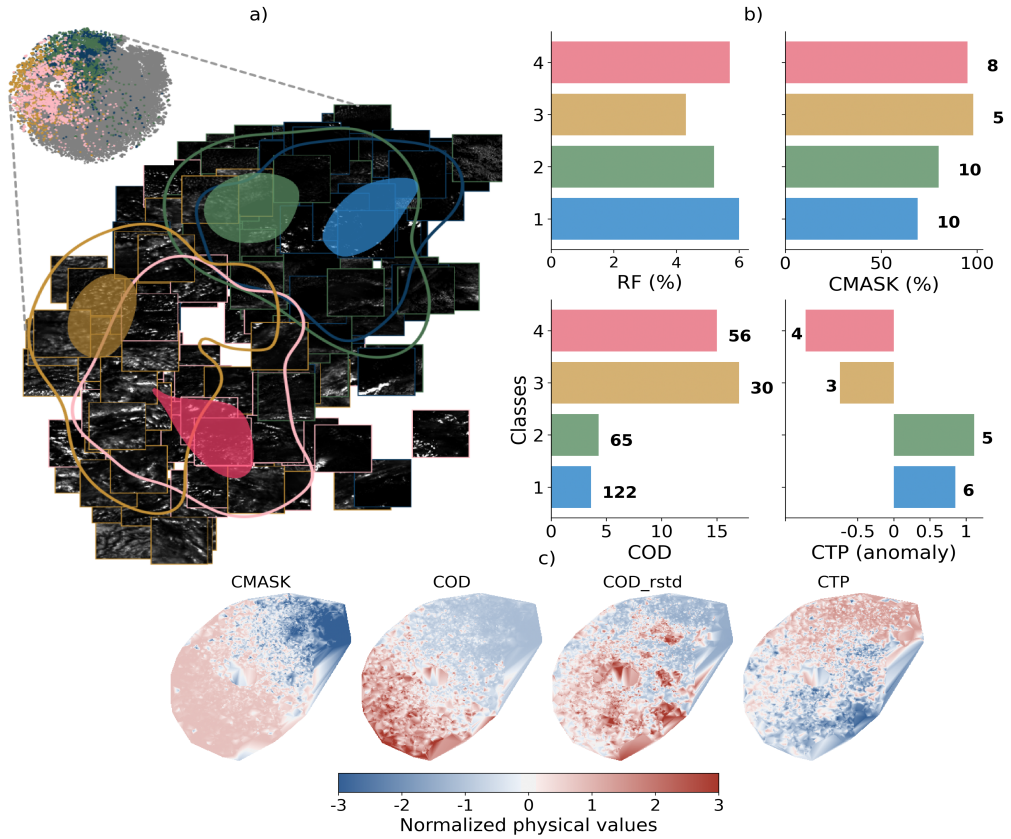


Figure 5.8: a) t-SNE representation of N1 feature space, zoomed into the regions of low clouds, where each random image is assigned a unique color frame. Additionally, the class distribution from N2 is visualized through 40% (spread) and 90% (peak) density contour. b) Satellite image characteristics of the relative frequency of occurrence (RF), cloud mask (CMASK), cloud optical depth (COD), and normalized cloud-top pressure (CTP), followed by the relative standard deviation (rstd in %) of the four different low-level classes. c) Normalized distribution of CMASK, COD, COD relative standard deviation (COD_rstd), and CTP is constructed by linear interpolation into the t-SNE plane.

instead of 7. This choice of partitioning is empirical, driven by the need to force the network to pay attention to finer details and complexities within the COD distribution of our data set.

5.4.1 Low cloud continuum

Our framework learns the COD field distribution of the entire cloud spectrum from the extended dataset and identifies four (out of 20 classes) as low-level cloud fields (Fig. 5.8a) positioned consecutively next to each other within the N1 feature space (continuum). These four classes contain 85% of the data points from classes 3 and 4, the highly variable cloud regimes apparent in measured SSI (Section 5.3.2).

Within the four low-level cloud regimes (I - IV), Classes I and II are more adjacent in the continuum (representation of N1) than III and IV, as demonstrated

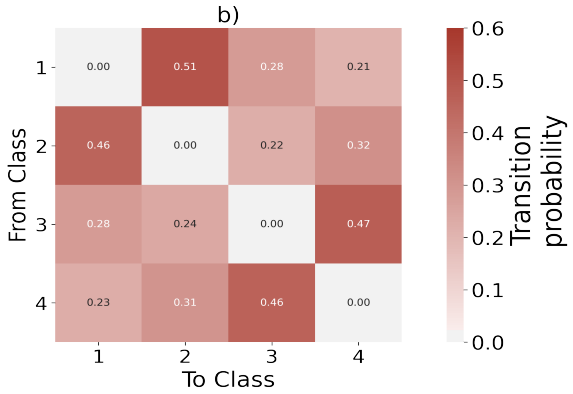


Figure 5.9: Matrix representing the probability of transition of each low cloud regime to another class derived over Jülich from the extended dataset.

by their 40% density contour overlap (Fig. 5.8a). Also, the transition probabilities between Classes I-II and III-IV are high (Fig. 5.9), explaining their closeness in the N1 Continuum. However, their differences become more clearly highlighted by the 90% density contour. An overview of their key physical characteristics in (Fig. 5.8b) leads to the hypothesis that these four classes likely represent the boundary layer (low-level) clouds belonging to different stages of convection.

Class I corresponds to the cloud fields, with the weakest shallow convection having the lowest mean COD and cloud fraction (Fig. 5.8b). It has the highest satellite-derived rstd of COD (122%), suggesting a strong convective activity phase. The neighboring class II displays relatively higher cloud amount and cloud top pressure, and both I and II lie above 750 hPa (below 3 km) on average. Class III shows more developed deeper shallow cloud fields, which are already widespread with the highest mean and lowest rstd of CMASK and COD.

Finally, class IV has a high rstd in COD (56%) (Fig. 5.8b) and the lowest CTP (highest cloud top), indicating deeper clouds, which implies further enhanced convective activity. Note that overlaying semi-transparent ice clouds frequently top these low-level clouds. However, in this case, it seems that the framework can likely isolate only the low-level cloud regimes, as demonstrated by the low standard deviation in CTP.

When spatially averaged cloud macro-physical properties are arranged in the t-SNE space (Figure Fig. 5.8c), a quasi-linear distribution appears for CF, COD, and CTP, which indicates a good functionality of N1. This does not hold for the rstd of COD. A high rstd of COD indicates the presence of both initial and mature cloud cells within a cloud field or multi-layer clouds. Therefore, the non-linearly distributed rstd shows the added complexity of these shallow clouds and their interaction with solar radiation.

Now, we return our view to the ground-based perspective and aim to explain the reason for the observed differences in the four low-level cloud regimes. Class III exhibits the least rstd in measured SSI (Fig. 5.10a), while classes II, III,

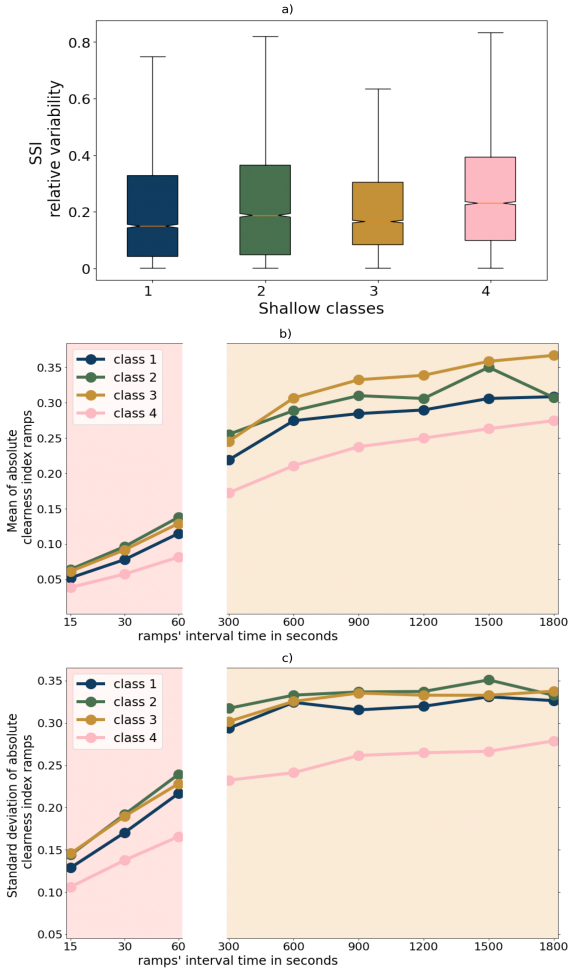


Figure 5.10: a) Box plot depicting four shallow cloud regimes (classes I, II, III, IV) (n = 25,300) observed at the JOYCE Observatory in Juelich, Germany, spanning from 2013 to 2019 (excluding 2014), showcasing the spatially averaged relative standard deviation (rstd) of surface solar irradiance (SSI). b) The mean ($\overline{CI}$) and the standard deviation (σ^{CI}) of the ramps in s^{-1} for each possible combination of time intervals ranging i) from intra-minute fluctuations and ii) five minutes to half an hour for the four shallow cloud regimes (Subplots b and c: courtesy N. Raabe).

and IV display a comparatively larger interquartile spread in measured SSI. This difference is likely due to the deeper formations and widespread nature of clouds in class III, facilitating a more uniform penetration of SSI.

The change in CI or transmissivity of solar energy through the cloud layers reaching the ground surface over a specific time period is known as ramp rates. Ramp rates across different time orders are considered useful (from a few seconds up to several minutes) as they pose different load management challenges and imply different solutions, such as from single distribution systems to large decentralized plants. Therefore, the relevant timescale for evaluating variability depends primarily on the footprint of the solar generation considered (Perez et al., 2016). This also helps bring into perspective the high spatio-temporal variability introduced by the identified LCFs on SSI measured by the pyranometer.

Differentiating the space-time context for mitigating solar energy variability measures, Perez et al. (2016) puts them into four categories depending on PV installation capacity and their spatial extent. They find that plants extending up to a few tens of meters, like small and medium PV installations, ramps calculation up to a few seconds, are relevant (Ranaweera et al., 2014). For large PV plants that extend from a hundred meters to a few km, minute-scale ramps are relevant as they affect the voltage transmission system (PREPA — Puerto Rico Energy Authority, 2013). For spatial scales such as substations extending up to 5 to 20 km, 10 minutes, and higher ramps are important. While for large cities and dense transmission networks extending more than 50 km, 20 - 30-minute fluctuations concern the grid operators.

To characterize the variability in the CI, we employ its two statistical measures—the sample arithmetic mean ($\overline{CI}$, (in s^{-1})) and sample standard deviation (σ^{CI}, (in s^{-1})) — (Fig. 5.10b,c). Quantifying the $\overline{CI}$ and σ^{CI} within different limited timeframes δt across the time series measured by a pyranometer over JOYCE observatory provides an intuitive understanding of the prevailing ramps one can expect once a specific shallow cloud regime persists. We assess ramp rates primarily influencing a) intra-minute fluctuations and b) from five minutes to half an hour. In this way, we consider small-scale PV installations to large decentralized PV power plants.

The $\overline{CI}$ ramp values at an intra-minute scale across the four LCFs show a gradual linear increase with longer time intervals (Fig. 5.10b), while its standard deviation (σ^{CI}) has a comparatively steeper slope (Fig. 5.10c). This is likely because this scale captures quick changes in solar irradiance due to rapidly evolving transient cloud conditions. Conversely, at time scales of 5 minutes and higher, the $\overline{CI}$ continues to increase, though with a gentler slope than the intra-minute developments, while the σ^{CI} possibly reaches a saturation point in ramp rates at later stages. For long-time series, when CI is normally distributed, ramp rates should approach 0.34. Subsequently, the differences among the four LCFs become more apparent as the $\overline{CI}$ ramp values gradually become more distinct at higher time intervals.

Class III demonstrates the highest $\overline{CI}$ ramps at time intervals beyond 5 minutes. This is likely due to its mature cloud objects containing more liquid water, which reflects sunlight, and the gaps between these clouds, which allow sunlight transmission. In contrast, Class IV exhibits the lowest $\overline{CI}$ ramps, possibly due to its layered patchiness that reduces solar energy penetration, resulting in fewer differences in ramps.

Turbulent cloudy conditions, such as those in Class II, show the highest σ^{CI} variability. Classes I and III exhibit intermediate variability in ramps across the considered time range. Meanwhile, Class IV consistently displays the lowest $\overline{CI}$ and σ^{CI} ramp values across all δt intervals.

Table 5.2: Comparison of probabilities for the absolute value of ramps from single pyranometer sensor being >= 0.5 between our study ('C') and (Lohmann et al., 2016) ('L'). Our study provides four probability values corresponding to the four low-level cloud regimes (CR). In contrast, 1^{st} row shows the same probabilities but for mixed sky conditions as reported by Lohmann et al. (2016) in their Fig. 5.

CR	L, 10s δt	C, 5s δt	C, 15s δt	L, 60s δt	C, 60s δt
	0.0044			0.016	
I		0.0033	0.024		0.081
II		0.0048	0.032		0.108
III		0.0068	0.033		0.095
IV		0.0028	0.016		0.045

Lohmann et al. (2016) analyzed the spatio-temporal characteristics of CI for distances ranging from tens of meters to about ten kilometers as part of the HD(CP)^2 Observational Prototype Experiment (HOPE). They utilized 10Hz time resolution data from 99 pyranometers between April 2 and July 24, 2013, and 50 pyranometers between September 3 and October 14, 2013. In their assessment, mixed sky conditions are defined as those that are neither clear nor overcast, characterized primarily by a higher standard deviation of CI. When comparing ramp values from single sensors for their mixed sky conditions (as reported in Fig. 5), we find a higher frequency of large ramps (>0.5) in our study (Table 5.2). This difference can be attributed to the distinct scenarios: our study focuses solely on highly variable low cloud systems, whereas their analysis encompasses all possible mixed sky conditions. Additionally, they used 6 months of pyranometer data, while we used the extended dataset, making our assessment more robust.

5.4.2 Daily cloud evolution

In this section, we investigate the evolution of low-level cloud systems where the rstd in COD is highest (see Fig. 5.8b). Representing cloud system dynamics through satellite observation involves capturing their evolving convective stages and the transition states over time. Changes in cloud system evolution and advection as captured in their horizontal appearance, shown here in Figure 5.11, also provide an opportunity to evaluate whether numerical models could capture timings associated with the crucial stages of cloud system evolution or how well the life cycle is represented in the numerical models.

Here, we use test (unseen) satellite data centered over Juelich to track the diurnal development inside the low-dimensional feature space (t-SNE). Our hypothesis is that intrinsic cloud physical properties drive changes in different dimensions of neural embeddings. To better relate embeddings with the variability of COD present in these cloud fields, we first look at 50 days with the most low-level convective occurrence over Juelich and then exemplarily focus on one

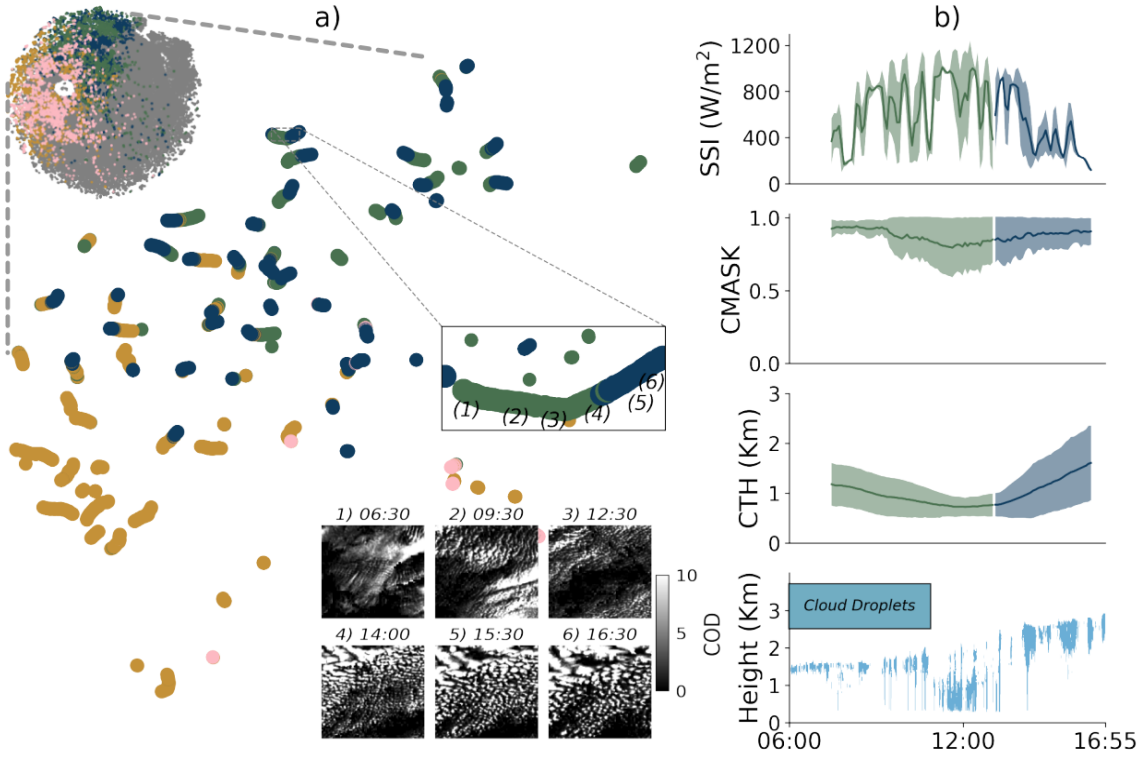


Figure 5.11: a) Fifty days with long-lasting shallow convection over Juelich, Germany, in the t-SNE coordinate. Highlighted is a case study on 12-05-2019 where the cloud system developed from class II to I. Six COD satellite images occurring at different times of the day demonstrate the evolution of the cloud system. b) From top to bottom: pyranometer measurement of SSI (W/m^2), CMASK (%), and CTH (hPa) derived from satellite retrievals and cloud vertical profile from Cloudnet.

specific day covering the initial phase of convection evolving in the shallow space continuum. We input the test samples into the trained framework, gathering their attributed features (N1) and machine labels from N2.

The t-SNE distribution of 50 shallow cumulus days illustrates a continuous development appearing as a 'worm' in the 2D space. Some of them are oriented more vertically, while our example day (Figure 5.10a) is horizontally aligned with slight changes in the vertical. Within this day, class II transitions to class I. We can see the rapid change in the SSI measured by the ground-based pyranometer over a short period (Figure 5.10b). This illustrates the high fluctuation in solar energy imposed by these low-level cloud systems. Note that the variability observed for class II is slightly more than for class I. Although precipitation is a local phenomenon, it is apparent that the dip in the vertical dimension of the 2D space is mainly caused by rain around noon, as captured by the backscatter of a cloud radar operated next to the pyranometer. A cloud radar transmits shorter wavelengths than weather radars (Ka- or W-band; frequency 35 or 95 GHz; wavelength 8 or 3 mm, respectively) and hence it is more sensitive to smaller particles, such as cloud droplets and tiny ice particles. The drop in the vertical dimension

of the 2D space also leads to a drop in CTH and CF, verified by the satellite retrievals. Clearly, the framework captures intrinsic characteristics of cloud field evolution based on COD distribution captured in high-dimensional feature space.

5.5 Conclusion

We present a scalable framework based on satellite images of COD distribution to better classify low-level cloud systems. Extending our work from Chatterjee, Acquistapace, et al. (2023), we utilize a significantly larger dataset to compare the cloud radiative and physical properties. We robustly establish the diurnal cycle and SSI properties of the seven classes and assess their SSI spatio-temporal characteristics using a ground-based pyranometer. The diurnal and radiative distributions appear more pronounced and smoother than in the previous dataset, highlighting the advantage of training with the extended dataset, which yields improved delineation of boundaries within the mesoscale cloud regimes.

Additionally, our analysis reveals that two low-level cloud regimes (classes 3 and 4) from the seven classes identified by Chatterjee, Acquistapace, et al. (2023) exhibit the highest spatiotemporal SSI variability (Section 5.3.2). For solar energy applications, these low clouds are further differentiated into four distinct shallow cloud regimes (classes I, II, III, IV). The structure and distribution of these cloud systems lie next to each other in the continuum, and the relative standard deviation (rstd) in the COD distribution space appears to have a non-linear, complex relationship with the 2D similarity manifold of the N1 continuum.

The analysis of SSI ramps up to 30-minute intervals underscores the importance of refining our understanding of shallow cloud systems. Low-level cloud systems exhibit higher ramp values compared to mostly mixed-sky conditions (neither clear nor overcast) as analyzed in Lohmann et al. (2016). Our multi-year dataset study demonstrates that shallow cloud systems consistently present a higher likelihood of severe fluctuations across different time intervals. This finding is particularly relevant for applications such as grid stability maintenance, where anticipating worst-case scenarios involving strong short-term PV fluctuations is crucial.

We further interpret the learned representations of our neural network, revealing process-wise meaningful insights. By monitoring an exemplary day, we observe consistent transitions in the diurnal evolution of convective stages of low-level cloud regimes within the framework's latent space. Integrating these observations with ground-based measurements enhances our understanding of these transitions and opens up possibilities for extending our comparison with high-resolution advanced numerical simulations.

Our approach is currently limited to 2D t-SNE representations and does not explore all dimensions of the latent space. Effective compression of the high-dimensional space can be achieved by analyzing the singular value decomposition

of the low-dimensional latent space. This will allow us to identify the most important dimensions of the embedding space and likely enhance our understanding of how shallow cloud transitions evolve, as it can better assess the hidden nonlinearities of the embedding space. Further research opportunities include investigating the transition probabilities of these shallow cloud systems, integrating with SSI forecasts, and verifying km-scale climate models.

6 Conclusion and outlook

6.1 Conclusion

This dissertation, in its first two chapters (Chapter 3 and 4), introduces two deep neural networks (N1 Section 4.3 and N2 Section 3.3) designed using self-supervision principles to explore the spatio-temporal variability of mesoscale cloud systems. Initially, we focused on discerning distinct cloud regimes across the mid-latitudes in central Europe. Here N2 learns to put each satellite image into one of the 'K' classes as it progressively improves its feature space's clustering. Subsequently, our investigation extended to understand the organizational linkages among marine shallow clouds across their complete spectrum over the tropical northern Atlantic region. N1 was designed such that it can sort cloud organizations based on their visual similarity and arrange them next to each other in the feature space. By combining these two approaches, we found greater flexibility in comprehending the different difficult-to-separate convective stages of shallow cloud systems over land and assessed their impact on modulating the rapid variability of SSI, crucial for solar energy applications.

Both the NNs used COD available in the daytime to investigate convective activities over the periods ranging from spring to autumn over land and the dry season (November to April) over the tropics. The MSG satellite product used over land retrieves the COD values and other radiative and cloud properties, which are post-processed to mimic the MTG (Deneke et al., 2021). Meanwhile, the maritime clouds observations are taken from publically available GOES -16E data.

To quantify the variability of the cloud regimes based on their structure and distribution, we adopt for obtaining reasonable clusters into which the cloud system spectra could be partitioned. The first study derives these useful regimes from the COD-CTP physical space. The classical, well-accepted cloud genera of ISSCP classification (Rossow & Schiffer, 1999) partially inspired this approach. While doing the partitions, we tried to conserve the independence of the regimes identified in the COD-CTP space. Parallely, to understand the scale effect of mesoscale cloud systems on the clustering and generalization capacity of NNs, this study investigated two different image configurations: 128 x 128 and 64 x 64 pixels. Conserving the non-overlapping nature in the COD-CTP composite space, we found seven mutually exclusive cloud regimes in the 128 x 128 and eight in the 64 x 64-pixel configurations.

The independent nature of these regimes is evident in their radiative and macro-physical space, as retrieved from satellite data. These spaces also aids in characterizing the identified cloud regimes effectively. In the 128 x 128 configuration, class 1 denotes cirrus clouds, while class 2 encompasses clear-sky cases. Classes 3 and 7 correspond to two distinct types of low-level clouds, with differing depths. Meanwhile, classes 4, 5, and 6 comprise various convective clouds.

We observed that the neural network generalizes better with unseen data in

the 128 x 128 configuration, possibly due to the larger image dimensions compared to the 64 x 64 domain. The 64 x 64 domain likely contains greater variability in cloud system patterns and training solely on one year of data might not suffice to capture this diversity.

For solar energy application, we investigated the temporal variability of cloud regimes over a fixed location in Juelich, Germany, and tested the trained NN performance on the persistency of cloud regimes and derived their tendency of transition to other cloud regimes. Later in our third study (Chapter 5), we used the transition probabilities to argue for the closeness of certain shallow cloud regimes in the continuum feature space because they had a higher probability of transition to each other.

In the second study (Chapter 4), we opted for a scale representation of 256 x 256 pixels to explore mesoscale shallow marine cloud organizations. This choice struck a balance between capturing sufficient detail for broad mesoscale structures like fish and maintaining computational efficiency. Smaller scales were inadequate in representing these structures adequately, while larger scales, such as 400 x 400, were deemed computationally expensive.

Most of the cloud systems in our data belonged to the shallow marine cloud family (about 86%) and varied primarily in their organizational aspects. Here the ISSCP analogy of COD-CTP space didn't prove to be very helpful because the CTP didn't vary much. Therefore, we opt for a data-driven way of using the features from N1 to partition the maritime shallow clouds. We used the features as they were found to be more aware of the organizational aspects of the cloud systems.

This approach showed that mesoscale cloud organizations in NAT can be partitioned into seven reasonable cloud regimes for the period considered. We used cloud amount at four vertical levels of the atmosphere (from CERES) and integrated water vapor amount (from ERA-5) to demonstrate the interlink and separateness of the identified cloud regimes in the continuum.

The human labels derived from Schulz (2022) served as a benchmark for classifying mesoscale cloud organizations across the NAT region. Their data set allowed us to compare these human labels with the machine-identified classes. Our analysis indicates that future studies should prioritize labels with higher human confidence, as lower-confidence labels exhibited inconsistent classification and ambiguous positions in the continuum. Furthermore, the potential and interpretability of the continuum space become more apparent when examining the classification and physical properties between human labels and their nearest neighbors.

We demonstrated a case study of tracking sugar-to-flower transitions in the continuum where we show how capturing the continuous transition of cloud regimes benefits from the arrangement setting of the continuum space. The

transient changes in the mesoscale distribution were quantifiable using a unique distance metric, which is inherently embedded in N1. We assess the large-scale environmental parameters of the transition using ERA-5 and note a gradual increase in vertical wind speed alongside a gradual decrease in cloud cover.

The third study Chapter 5, targets solar energy applications and tactfully uses the framework (N1 + N2) to taxonomize the shallow cloud systems over central Europe. First, we perform an intercomparison between earlier identified seven cloud regimes and newly identified regimes derived using an 11 times extended dataset. We did not find significant changes in the cloud's macrophysical, radiative, or diurnal characteristics. This suggests that for a 128 x 128 mesoscale cloud representation, one year of data was enough to capture their diversity.

The analysis of pyranometer measurements located at JOYCE, suggests that among the seven functional classes, classes 3 and 4, representing low-level clouds, exhibit greater spatiotemporal variability in SSI. Leveraging the framework's capabilities, we further distinguish these regimes into four distinct shallow cloud regimes (classes I, II, III, IV), which are closely situated to each other in the N1 continuum feature space. Class I and II represented the initial convective stages, while III and IV represented the later convective stages with more maturely developed cloud cells.

Better differentiation of shallow cloud systems proves advantageous, particularly evident in ramp calculations using the CSI index. Here, we consistently observe that the ramps in SSI induced by these shallow systems exhibit a higher likelihood of severe fluctuations compared to mixed sky conditions (as found in Lohmann et al. (2016)) across various short time intervals.

In our final step, we conduct a case study tracking the rapid daytime development of mesoscale shallow cloud fields over Juelich within the feature space. By correlating these developments with ground-based radar and pyranometer measurements, we extract meaningful insights. For instance, we observe rainfall around noon over Juelich, which leads to changes in the vertical dimension of the 2D embedding space. Thus, we demonstrate the connection between physical drivers influencing changes in the mesoscale distribution of shallow clouds and corresponding alterations in the embedding space.

6.2 Outlook

In this thesis, the focus was on exploring the essential elements of cloud systems, representing their variability and transitional states, and evaluating their impact on SSI modulation. To achieve this, two neural networks were employed, each addressing distinct aspects of categorizing cloud system variability within their feature space. As a result, the study offers a comprehensive meteorological insight into cloud systems and their organizational connections analyzed within the realm of artificial intelligence.

Some possible future directions of research are mentioned below, with an elaborate explanation of the model evaluation of clouds at the end.

Evaluate long-term satellite data record Our strategy aims to streamline the analysis of extensive satellite data over extended periods, offering simplified measures and reducing image complexity within the feature space. With the emergence of N2, we can now effectively classify previously unseen test data, followed by N1's capability to allocate its position in the continuum.

Multi modality, combining radar with geostationary measurements The current research mainly explores the capabilities of geostationary satellite observations, which come with inherent limitations such as inaccurate representation of multi-layer clouds.

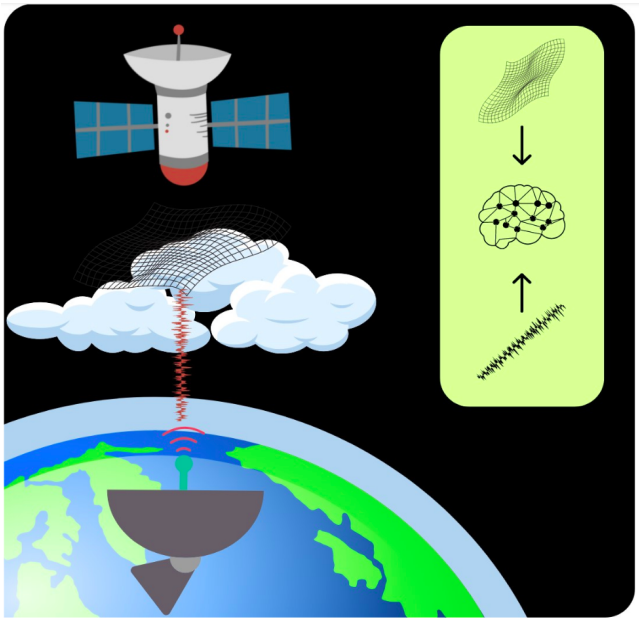


Figure 6.1: Illustration of multi-modality combining radar with satellite measurements. The elements shown are taken from freely available icons in Canva `https://www.canva.com/`. Graphic assistance: courtesy N. Risse

However, based on research outcomes from visual and audio signals (Morgado, Vasconcelos, & Misra, 2020; Morgado, Li, & Vasconcelos, 2020; Morgado et al., 2021), I believe that self-supervision can achieve better representations of cloud structures through multi-modality (see Fig. 6.1). For example, information on the vertical structure can be resolved through radars. The preliminary direction would be to tell the network that satellite and radar information represent the same physical system, i.e., clouds.

Exploit infrared measurements for capturing complete transitions This study's constraint lies in its reliance solely on daytime cloud retrievals, thus

overlooking the nocturnal activities of the cloud organizations. This limitation also centralizes the challenge in model evaluation. Future investigations will incorporate infrared satellite measurements to achieve comprehensive 24-hour coverage. This can assist us in tracking the complete transitions of cloud systems in the feature space.

Pre-trained SSL models and short-term forecasting of SSI When training neural networks with SSL principles, they acquire knowledge about the underlying semantic features or structural details of the data. Subsequently, this acquired knowledge can be leveraged for other tasks, such as predicting SSI.

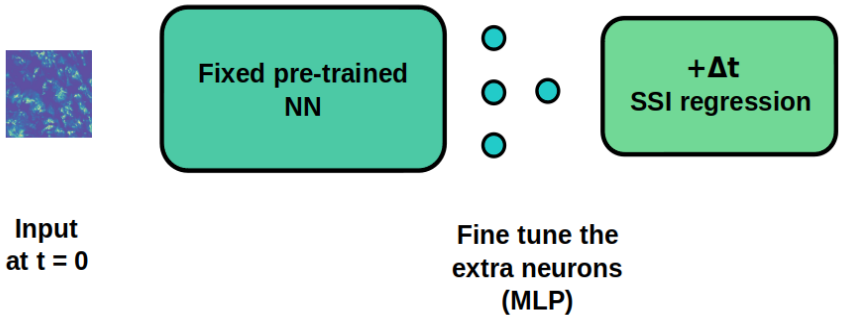


Figure 6.2: Illustration demonstrating the use of pre-trained neural networks (NN), which could be N1 or N2, to exploit short-term surface solar irradiance (SSI) forecasting at Δ t time ahead. Multi-layer perception (MLP) denotes the extra layer of trainable neurons arranged in a feed-forward setting.

An effective approach involves adding additional layers of randomly initialized neurons after the pre-trained last layer in a feed-forward setup (see Fig. 6.2). With limited labeled data, this method enables learning the functional mapping between the output SSI value and input satellite images.

Most important 'n' feature dimensions Our approach currently lacks exploration of all feature space dimensions; therefore, one way to proceed would be to assess the most important 'n' dimensions of the high-dimensional feature. This can be addressed by analyzing the singular value decomposition (SVD) of the embedding space. We can then assess which cloud properties and environmental conditions drive a change in those important dimensions.

6.3 Towards model evaluation

6.3.1 Introduction

Ensuring precise cloud simulation within general circulation models (GCMs) holds significant importance across various timescales (Williams & Bodas-Salcedo, 2017). GCMs predict climate sensitivity, a quantity representing the global mean temperature response to doubling CO_2, consistently within the range of 1.5 to 4.5 K (Houghton et al., 1996, 2001), however, in the 6th assessment report (AR6) of the Intergovernmental Panel on Climate Change (IPCC), this uncertainty has been reduced to 2.5–4 K. This variability in global mean temperature across simulations is mainly attributed to the cloud representations and their radiative feedbacks (Williams et al., 2006; Zelinka et al., 2020). Cloud Radiative Forcing or CRF is defined as clouds' impact on Earth's radiation budget calculated at the top of the atmosphere (Charlock & Ramanathan, 1985).

Crucial in modulating these feedbacks are several cloud parameters such as cloud fraction, top height, and cloud optical depth (COD), which play pivotal roles in either amplifying or mitigating the warming effect resulting from increased CO_2 levels. These parameters are determined by factors such as atmospheric dynamics, convection, mass flux, turbulence, and cloud microphysics. Many of these factors are inadequately represented at the subgrid-scale in today's GCMs (Ryan & et al., 2000). Consequently, metrics for model evaluation against observational data are highly relevant and routinely employed to asses their performance (Murphy et al., 2004; Reichler & Kim, 2008; Bony et al., 1997; Bony, Dufresne, Treut, et al., 2004; Tselioudis et al., 2000).

Traditional evaluation of clouds in GCMs has been based on comparing global averages or maps of cloud variables (Weare & Groups, 1996) such as cloud amount or their radiative forcing (Pincus et al., 2008; Gleckler et al., 2008). However, these approaches do not consider different cloud types' temporal frequency and radiative properties, leading to a disproportionate simulation of errors in the longwave and shortwave radiative effects (Webb et al., 2001).

Recent methodologies mainly explore two main directions:

a) Categorizing cloud information using large-scale variables such as vertical velocity (Bony, Dufresne, Le Treut, et al., 2004), and various measures of lower tropospheric stability (Klein & Hartmann, 1993; Williams et al., 2006). These approaches facilitate linking cloud properties with the dynamic processes influencing them and thus reduce the chances of a model over-fitting due to inherent bias. For instance, Bony, Dufresne, Le Treut, et al. (2004) compiled monthly mean CRF from three GCM simulations and Earth Radiation Budget Experiment (ERBE) observations, categorized by the large-scale dynamics of 500 hPa vertical velocity. However, their approach relies on reliable atmospheric data crucial for cloud formation and also necessitates prior knowledge of meteorological processes associated with each cloud regime (CR) (Jakob & Tselioudis, 2003).

b) Alternatively, another approach involves evaluating joint histograms of cloud optical depth and cloud top pressure (CTP) for cloud cover (Tselioudis et al., 2000; Jakob & Tselioudis, 2003; Gordon et al., 2005; Williams & Webb, 2009). This method compares observations and model outputs simulated in radiative transfer models (RTM) such as Observational Simulator Package (OSP, (Bodas-Salcedo et al., 2011)) or International Satellite Cloud Climatology Project (ISCCP, (Rossow & Schiffer, 1999)) simulators `http://gcss-dime.giss.nasa.gov/simulator.html`. This enables direct comparisons between model-generated quantities and those retrieved from observational platforms like satellites and radars. The possible shortcoming of this approach is that using the composite representation of cloud properties completely neglects the spatial structure and organization of clouds within the considered domain.

A major portion of focus in this thesis, spanning Chapters 3 and 4, primarily delved into the representation of mesoscale cloud system's structure and distribution using geostationary satellite observations. The developed neural networks mainly learn to extract key features of the cloud systems in their embedding space. Embedding space is useful to assess the connection between different cloud systems as it arranges the entire cloud system spectra along a continuum. The neural networks demonstrated refined differentiation of shallow cloud systems, showing benefits for short-term solar energy purposes (Chapter 5).

Additionally, our work sheds light on quantifying spatio-temporal changes in mesoscale cloud organizations during their transitions from one CR to another (Section 4.7.7). Furthermore, assessing mesoscale cloud systems with similar structures and distributions from the embedding space provides valuable insights into the prevailing large-scale environmental conditions (Section 4.13). Consequently, this thesis suggests three potential directions involving integrating neural networks to evaluate cloud representation in models against observational data.

6.3.2 Evaluate cloud regimes

The robust development of reasonable and simplified CRs from long-term satellite image records allows us to evaluate and improve cloud representation in numerical models (Williams et al., 2006; Norris & Weaver, 2001). As an initial step in the evaluation process, we would feed the consistent cloud fields from the simulations using a forward simulator such as the "Radiative Transfer for Television Infrared Observation Satellite Operational Vertical Sounder" (RTTOV) (Saunders et al., 2018) to emulate geostationary satellite measured radiances. RTTOV is a fast radiative transfer model for passive visible, infrared, and microwave nadir-viewing instruments, which also enables rapid calculations of cloud optical depths for each instrument. Preferably, the numerical model simulation chosen could be a global Km-scale simulation (Rackow et al., 2024), aligning well with the resolution constraints of geostationary satellite measurements.

For simplicity, we would start our analysis covering central Europe (Section 5.2), ingest the emulated images into our trained networks N1 and N2 (details given in Sections 4.7.2 and 4.3), and analyze the continuous (N1) and discrete (N2) feature space (refer to Fig. 5.1, 4.1). As in Chatterjee, Acquistapace, et al. (2023), N2 will partition the continuum into seven classes, and we can assess the distinctness of the classes by looking at their frequency of occurrence, and macrophysical composites, like in Table 5.2 and Figure 3.8. We can investigate whether their physical space exhibits similar cloud radiative and macrophysical distributions, as found in the observational space (Fig. 5.3, 5.4).

Additionally, we can possibly examine whether the N1 continuum space displays comparable distributions of environmental conditions, such as Integrated Water Vapor (IWV) (such as Fig. 4.1.b) and relative humidity. This approach will enable us to assess whether the latest generation models can accurately represent the cloud fields' mesoscale variability and horizontal structure from model simulations into the seven clusters.

Note that while our methodology focuses on utilizing COD distributions to examine CRs, this limits the data to daytime only. While we are missing a large part of the day, however, studies by Bony et al. (2006) and Williams et al. (2006) indicate that a significant portion of the error in cloud radiative forcing (CRF) simulated by the GCMs is associated with the shortwave spectrum. Therefore, the CRs derived from COD distribution in our work could serve as a valuable constraint for improving the performance of GCMs.

6.3.3 Evaluate shallow cloud regime transitions

The intricate interplay among low cloud systems and their transitions significantly influences atmospheric circulation in tropical and subtropical regions, potentially shaping the fundamental physics of climate dynamics (Ma et al., 1996; Larson et al., 1999). Many GCMs exhibit conflicting cloud feedback responses in climate sensitivity experiments, particularly regarding boundary layer clouds (Bony, Dufresne, Treut, et al., 2004; Bony et al., 2006; Stephens, 2005). Williams and Webb (2008) further underscores this complexity, attributing 65% of CRF variability among GCMs to two primary factors: i) overly reflective simulations of low clouds, notably stratocumulus regimes, and ii) transitions between these regimes and cumulus formations. Consequently, effectively simulating transitions within shallow cloud systems poses a significant challenge for GCMs (Sarkar et al., 2020).

Our framework development (Section 4.3) allows for evaluating the continuous transition of marine stratocumulus decks as they break up into scattered, randomly distributed cumulus convection. This framework enables us to assess their spatiotemporal progression in the feature space and quantify the relative changes in mesoscale organization. Leveraging the capability of N2, our framework can

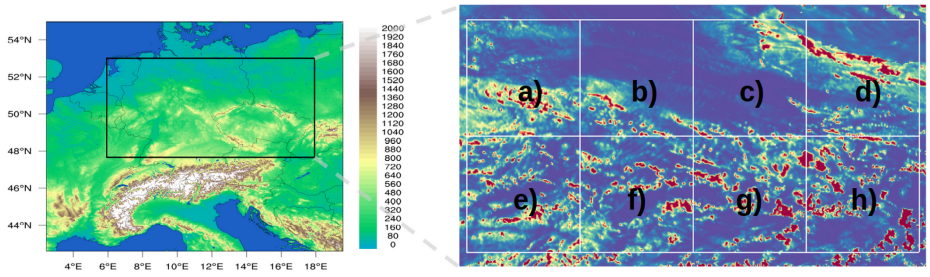


Figure 6.3: left: Geographical map of Europe shown along with topography, thick black lines are coast, the thin black lines are political boundaries, and the black box shows the central European domain. Right: MSG high-resolution cloud optical depth of the central European domain on April 1, 2018, at 15:20 UTC. From 2017 - 2019, at each available timestep, 128x128 subdomains a - h are cropped out

also provide insights into the different stages of stratocumulus to cumulus systems. Geographically, downwind trades of the North Atlantic during the dry season effectively represent these shallow convection transitions.

Given the fine-scale nature of these energetically crucial low-lying clouds, we advocate for resolving them using the ICOsahedral Nonhydrostatic - Large Eddy Simulations (ICON-LES), similar to the approach in Schulz and Stevens (2023). These simulations offer a large spatial and temporal extent, exceeding 1,000 km and 500 km in the zonal and meridional directions at the hm-scale resolution, providing a reasonable context to analyze how transition variability differs between observation and ICON-LES states.

To delve deeper into the analysis, we propose a two-step approach. We first employ cloud optical flow software (Bradski, 2000) to track cloud systems downwind of the trades. Secondly, we utilize tracked emulated images (using satellite simulators such as RTTOV), which are then ingested into our trained networks N1 and N2. Their resulting features allow us further to investigate the dynamic changes in the feature space, as shown for sugar to flower transition in Section 4.5

6.3.4 Transition probability

As a final demonstration, we aim to showcase how the trained N2 model can be utilized to generate the probability of occurrence for a specific CR to occur across nearby geographical domains in subsequent time intervals. This process can be replicated with emulated model outputs to assess the consistency between observations and numerical simulations.

We divided our central European domain into 8 subdomains (SD) (a to h of 128 x 128 pixels) (Fig. 6.3) and collected 5,60,630 samples from April to September 2017 – 2019. We ingest them in N2 and obtain their CR labels (between 1 to 7).

As a demonstration (see Fig. 6.4), we choose the SD 'a' and 'b' and ask if

a CR (1 to 7) has occurred over 'a', what is the probability of any other class occurring over b in the next 30 minutes? The same could be repeated for any pair of subdomains and choice of time intervals. We can also ask about the probability of the same class occurrence in nearby domains after some time lag.

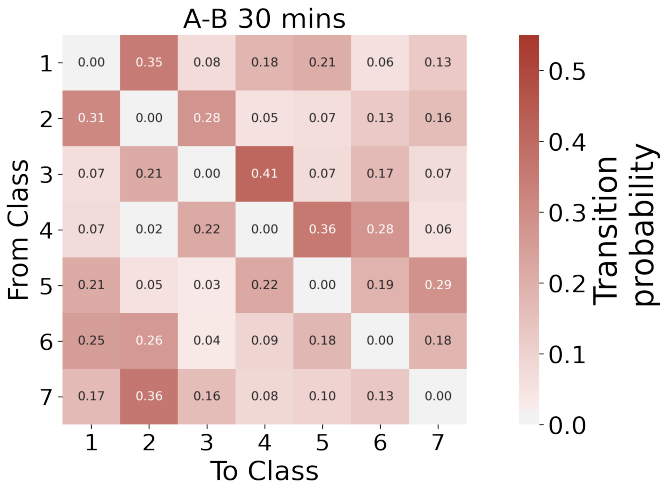


Figure 6.4: Between subdomain 'a' and 'b' as defined in Figure 6.3, the matrix represents the probability of occurrence among classes in the next 30 minutes. We used data from 2017 to 2019 for this analysis.

Among the higher probabilities, the results show that CR 3 in SD a is followed by CR4 in SD b in the next 30 minutes. CR 3 and 4 are both low-level clouds with overcast conditions. Further, CR4 is followed by CR 5, where CR 5 is a deep convective class.

The results and deep learning pathways developed in this thesis can likely foster synergies among rich satellite data, ground-based measurements, and next-generation kilometer-scale climate simulations, potentially leading to better representations of the state of the atmosphere. This interdisciplinary approach enables the extraction of essential information from large, high-dimensional datasets and leverages the internal mechanisms of AI to generate quantifiable insights for process understanding.

References

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In J. Van den Bussche & V. Vianu (Eds.), *Database theory — icdt 2001* (pp. 420–434). Berlin, Heidelberg: Springer Berlin Heidelberg.

AlSkaif, T., Dev, S., Visser, L., Hossari, M., & van Sark, W. (2020). A systematic analysis of meteorological variables for pv output power estimation. *Renewable Energy*, *153*, 12-22. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0960148120301725` doi: https://doi.org/10.1016/j.renene.2020.01.150

Althaus, P., Redder, F., Ubachukwu, E., Mork, M., Xhonneux, A., & Müller, D. (2022). Enhancing building monitoring and control for district energy systems: Technology selection and installation within the living lab energy campus. *Applied Sciences*, *12*(7). Retrieved from `https://www.mdpi.com/2076-3417/12/7/3305` doi: 10.3390/app12073305

Andrews, R., Stein, J., Hansen, C., & Riley, D. (2014, 06). Introduction to the open source pv lib for python photovoltaic system modelling package. In (p. 0170-0174). doi: 10.1109/PVSC.2014.6925501

Anvari, M., Lohmann, G., Wächter, M., Milan, P., Lorenz, E., Heinemann, D., ... Peinke, J. (2016, jun). Short term fluctuations of wind and solar power systems. *New Journal of Physics*, *18*(6), 063027. Retrieved from `https://dx.doi.org/10.1088/1367-2630/18/6/063027` doi: 10.1088/1367-2630/18/6/063027

Araujo, A., Norris, W., & Sim, J. (2019). Computing receptive fields of convolutional neural networks. *Distill*, *4*(11), e21.

Atito, S., Awais, M., Farooq, A., Feng, Z., & Kittler, J. (2021). *Mc-ssl0.0: Towards multi-concept self-supervised learning.* arXiv. Retrieved from `https://arxiv.org/abs/2111.15340` doi: 10.48550/ARXIV.2111.15340

Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., ... Norouzi, M. (2021). Big self-supervised models advance medical image classifications. *CoRR*, 3458-3468. Retrieved from `https://arxiv.org/pdf/2101.05224.pdf` doi: https://doi.org/10.1109/ICCV48922.2021.00346

Barnes, W. L., Xiong, X., Guenther, B. W., & Salomonson, V. (2003). Development, characterization, and performance of the eos modis sensors. *Earth Observing Systems VIII*, *5151*, 337–345.

Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *IEEE Potentials*, *13*(4), 27-31. doi: 10.1109/45.329294

Benas, N., Finkensieper, S., Stengel, M., van Zadelhoff, G.-J., Hanschmann, T., Hollmann, R., & Meirink, J. F. (2017, July). The MSG-SEVIRI-based cloud property data record CLAAS-2. *Earth System Science Data*, *9*(2), 415–434. Retrieved from https://doi.org/10.5194/essd-9-415-2017 doi: 10.5194/essd-9-415-2017

Berg, L. K., & Kassianov, E. I. (2008). Temporal variability of fair-weather cumulus statistics at the acrf sgp site. *Journal of Climate*, *21*(13), 3344 - 3358. Retrieved from https://journals.ametsoc.org/view/journals/clim/21/13/2007jcli2266.1.xml doi: 10.1175/2007JCLI2266.1

Bergman, J. W., & Salby, M. L. (1996). Diurnal variations of cloud cover and their relationship to climatological conditions. *Journal of Climate*, *9*(11), 2802 - 2820. Retrieved from https://journals.ametsoc.org/view/journals/clim/9/11/1520-0442_1996_009_2802_dvocca_2_0_co_2.xml doi: 10.1175/1520-0442(1996)009<2802:DVOCCA>2.0.CO;2

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023, 07 01). Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, *619*(7970), 533–538. Retrieved from https://doi.org/10.1038/s41586-023-06185-3 doi: 10.1038/s41586-023-06185-3

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S., ... John, V. O. (2011). Cosp: satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. doi: 10.1175/2011BAMS2856.1

Bomidi, M. (2020). *Hd(cp)2 short term observation data of pyranometer network, hope campaign by tropos.* World Data Center for Climate (WDCC) at DKRZ. Retrieved from https://doi.org/10.26050/WDCC/HOPE_TROP_PYRNET doi: 10.26050/WDCC/HOPE_TROP_PYRNET

Bony, S., Colman, R., Kattsov, V. M., Allan, R. P., Bretherton, C. S., Dufresne, J.-L., ... Webb, M. J. (2006). How well do we understand and evaluate climate change feedback processes? *Journal of Climate*, *19*(15), 3445 - 3482. Retrieved from https://journals.ametsoc.org/view/journals/clim/19/15/jcli3819.1.xml doi: 10.1175/JCLI3819.1

Bony, S., Dufresne, J., Le Treut, H., Morcrette, J., & Senior, C. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, *22*, 71–86. doi: 10.1007/s00382-003-0369-6

Bony, S., Dufresne, J., Treut, H. L., Morcrette, J. J., & Senior, C. A. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, *22*, 71-86. Retrieved from https://api.semanticscholar.org/CorpusID:56077074

Bony, S., & Dufresne, J.-L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*(20). doi: 10.1029/2005GL023851

Bony, S., Lau, K.-M., & Sud, Y. (1997). Sea surface temperature and large scale circulation influences on tropical greenhouse effect and cloud radiative forcing. *Journal of Climate*, *10*, 2055–2077.

Bony, S., Schulz, H., Vial, J., & Stevens, B. (2020). Sugar, gravel, fish, and flowers: Dependence of mesoscale patterns of trade-wind clouds on environmental conditions. *Geophysical Research Letters*, *47*(7), e2019GL085988. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085988` (e2019GL085988 10.1029/2019GL085988) doi: https://doi.org/10.1029/2019GL085988

Bony, S., Stevens, B., Ament, F., Bigorre, S., Chazette, P., Crewell, S., ... Wirth, M. (2017, November). EUREC4A: A Field Campaign to Elucidate the Couplings Between Clouds, Convection and Circulation. *Surveys in Geophysics*, *38*(6), 1529–1568. doi: 10.1007/s10712-017-9428-0

Bottou, L. (2012). Stochastic gradient descent tricks. In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade: Second edition* (pp. 421–436). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from `https://doi.org/10.1007/978-3-642-35289-8_25` doi: 10.1007/978-3-642-35289-8_25

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., ... Zhang, X. (2013). Clouds and aerosols. In T. Stocker et al. (Eds.), *Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Bridle, J. S. (1989). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Nato neurocomputing*. Retrieved from `https://api.semanticscholar.org/CorpusID:59636530`

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1-27. Retrieved from `https://www.tandfonline.com/doi/abs/10.1080/03610927408827101` doi: 10.1080/03610927408827101

Caron, M., Bojanowski, P., Joulin, A., & Douze, M. (2019). *Deep clustering for unsupervised learning of visual features.*

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A. (2020). Unsupervised learning of visual features by contrasting cluster assignments. *CoRR*, *abs/2006.09882*. Retrieved from https://arxiv.org/abs/2006.09882

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. *CoRR*, *abs/2104.14294*. Retrieved from https://arxiv.org/abs/2104.14294

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). *Emerging properties in self-supervised vision transformers.*

Carrera, B., Sim, M.-K., & Jung, J.-Y. (2020). Pvhybnet: a hybrid framework for predicting photovoltaic power generation using both weather forecast and observation data. *IET Renewable Power Generation*, *14*(12), 2192-2201. Retrieved from https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2018.6174 doi: https://doi.org/10.1049/iet-rpg.2018.6174

Charlock, T. P., & Ramanathan, V. (1985). The albedo field and cloud radiative forcing produced by a general circulation model with internally generated cloud optics. *Journal of Atmospheric Sciences*, *42*, 1408–1429.

Chatterjee, D., Acquistapace, C., Deneke, H., & Crewell, S. (2023). Understanding cloud systems structure and organization using a machine's self-learning approach. *Artificial Intelligence for the Earth Systems*. Retrieved from https://journals.ametsoc.org/view/journals/aies/aop/AIES-D-22-0096.1/AIES-D-22-0096.1.xml doi: https://doi.org/10.1175/AIES-D-22-0096.1

Chatterjee, D., Schnitt, S., Bigalke, P., Acquistapace, C., & Crewell, S. (2023, September). *Codes and pretrained weights for the study 'Capturing the diversity of mesoscale trade wind cumuli using complementary approaches from self- supervised deep learning'.* [Software] Zenodo. Retrieved from https://doi.org/10.5281/zenodo.8352614 doi: 10.5281/zenodo.8352614

Chen, Y., & Genio, A. D. D. (2009). Evaluation of tropical cloud regimes in observations and a general circulation model. *Climate Dyn.*, *32*, 355–369.

Cho, H.-M., Zhang, Z., Meyer, K., Lebsock, M., Platnick, S., Ackerman, A. S., ... others (2015). Frequency and causes of failed modis cloud property retrievals for liquid phase clouds over global oceans. *Journal of Geophysical Research: Atmospheres*, *120*(9), 4132–4154.

Dauhut, T., Couvreux, F., Bouniol, D., Beucher, F., Volkmer, L., Pörtge, V., ... Wirth, M. (2023). Flower trade-wind clouds are shallow mesoscale convective systems. *Quarterly Journal of the Royal Meteorological Society*, *149*(750), 325–347. doi: 10.1002/qj.4409

Denby, L. (2020a). Discovering the importance of mesoscale cloud organization through unsupervised classification. *Geophysical Research Letters*, *47*(1), e2019GL085190.

Denby, L. (2020b). Discovering the Importance of Mesoscale Cloud Organization Through Unsupervised Classification. *Geophysical Research Letters*, *47*(1), e2019GL085190. doi: 10.1029/2019GL085190

Denby, L. (2023). *Charting the realms of mesoscale cloud organisation using unsupervised learning.* doi: 10.48550/arXiv.2309.08567

Deneke, H., Barrientos-Velasco, C., Bley, S., Hünerbein, A., Lenk, S., Macke, A., ... Witthuhn, J. (2021). Increasing the spatial resolution of cloud property retrievals from meteosat seviri by use of its high-resolution visible channel: implementation and examples. *Atmospheric Measurement Techniques*, *14*(7), 5107–5126. Retrieved from `https://amt.copernicus.org/articles/14/5107/2021/` doi: 10.5194/amt-14-5107-2021

Dimet, F. L., & Talagrand, O. (1986). *Variational algorithms for analysis and assimilation of meteorological observations: Theoretic aspects.* (Vol. 38A).

Donahue, J., Krähenbühl, P., & Darrell, T. (2017). *Adversarial feature learning.*

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... Houlsby, N. (2021). *An image is worth 16x16 words: Transformers for image recognition at scale.*

Dosovitskiy, A., & Brox, T. (2016). *Inverting visual representations with convolutional networks.*

Du, X. (2023). A robust and high-dimensional clustering algorithm based on feature weight and entropy. *Entropy*, *25*(3). Retrieved from `https://www.mdpi.com/1099-4300/25/3/510` doi: 10.3390/e25030510

Durand, Y., Hallibert, P., Wilson, M., Lekouara, M., Grabarnik, S., Aminou, D., ... Verez, B. (2015, 10). The flexible combined imager onboard mtg: from design to calibration. In (p. 963903). doi: 10.1117/12.2196644

Eastman, R., & Warren, S. G. (2014). Diurnal cycles of cumulus, cumulonimbus, stratus, stratocumulus, and fog from surface observations over land and ocean. *Journal of Climate*, *27*(6), 2386 - 2404. Retrieved from `https://journals.ametsoc.org/view/journals/clim/27/6/jcli-d-13-00352.1.xml` doi: 10.1175/JCLI-D-13-00352.1

EU-Commission, for Research, D.-G., & Innovation. (2021). *European green deal – research innovation call.* Publications Office of the European Union. doi: doi/10.2777/33415

Fabel, Y., Nouri, B., Wilbert, S., Blum, N., Triebel, R., Hasenbalg, M., ... Pitz-Paal, R. (2022). Applying self-supervised learning for semantic cloud segmentation of all-sky images. *Atmospheric Measurement Techniques*, *15*(3), 797–809.

Retrieved from https://amt.copernicus.org/articles/15/797/2022/ doi: 10.5194/amt-15-797-2022

Feingold, G., Balsells, J., Glassmeier, F., Yamaguchi, T., Kazil, J., & McComiskey, A. (2017). Analysis of albedo versus cloud fraction relationships in liquid water clouds using heuristic models and large eddy simulation. *Journal of Geophysical Research: Atmospheres*, *122*(13), 7086-7102. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JD026467 doi: https://doi.org/10.1002/2017JD026467

Feng, Z., Leung, L. R., Houze Jr., R. A., Hagos, S., Hardin, J., Yang, Q., ... Fan, J. (2018). Structure and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in convection-permitting simulations over the united states. *Journal of Advances in Modeling Earth Systems*, *10*(7), 1470-1494. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001305 doi: https://doi.org/10.1029/2018MS001305

Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy*, *37*(1), 1.

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, *20*(3-4), 121–136. Retrieved 2022-08-30, from http://link.springer.com/10.1007/BF00342633 doi: 10.1007/BF00342633

Gao, J., Li, F., Wang, B., & Liang, H. (2021). Unsupervised nonlinear adaptive manifold learning for global and local information. *Tsinghua Science and Technology*, *26*(2), 163-171. doi: 10.26599/TST.2019.9010049

Ge, J., Wang, Z., Wang, C., Yang, X., Dong, Z., & Wang, M. (2021). Diurnal variations of global clouds observed from the cats spaceborne lidar and their links to large-scale meteorological factors. *Climate Dynamics*, *57*(9), 2637–2651. Retrieved from https://doi.org/10.1007/s00382-021-05829-2 doi: 10.1007/s00382-021-05829-2

Gidaris, S., Singh, P., & Komodakis, N. (2018). *Unsupervised representation learning by predicting image rotations.*

Gilpin, S., Qian, B., & Davidson, I. (2013). Efficient hierarchical clustering of large high dimensional datasets. In *Proceedings of the 22nd acm international conference on information & knowledge management* (p. 1371–1380). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2505515.2505527 doi: 10.1145/2505515.2505527

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research: Atmospheres*, *113*(D6). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD008972 doi: https://doi.org/10.1029/2007JD008972

Gordon, N., Norris, J., Weaver, C., & Klein, S. (2005). Cluster analysis of cloud regimes and characteristic dynamics of midlatitude synoptic systems in observations and a model. *Journal of Geophysical Research*, *110*. doi: 10.1029/2004JD005027

Goyal, P., Duval, Q., Reizenstein, J., Leavitt, M., Xu, M., Lefaudeux, B., ... Misra, I. (2021). *Vissl.* https://github.com/facebookresearch/vissl.

Goyal, P., Duval, Q., Seessel, I., Caron, M., Misra, I., Sagun, L., ... Bojanowski, P. (2022, February). Vision Models Are More Robust And Fair When Pretrained On Uncurated Images Without Supervision. Retrieved 2022-08-30, from http://arxiv.org/abs/2202.08360 (arXiv:2202.08360 [cs])

Greybush, S. J., Haupt, S. E., & Young, G. S. (2008). The regime dependence of optimally weighted ensemble model consensus forecasts of surface temperature. *Weather and Forecasting*, *23*(6), 1146 - 1161. Retrieved from https://journals.ametsoc.org/view/journals/wefo/23/6/2008waf2007078_1.xml doi: 10.1175/2008WAF2007078.1

Grigg, T. G., Busbridge*, D., Ramapuram, J., & Webb, R. (2021). Do self-supervised and supervised methods learn similar visual representations? In *Neurips workshop.* Retrieved from https://arxiv.org/pdf/2110.00528.pdf

Gristey, J. J., Feingold, G., Glenn, I. B., Schmidt, K. S., & Chen, H. (2020, November). On the Relationship Between Shallow Cumulus Cloud Field Properties and Surface Solar Irradiance. *Geophysical Research Letters*, *47*(22), e2020GL090152. Retrieved 2024-01-23, from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020GL090152 doi: 10.1029/2020GL090152

Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., ... Quaas, J. (2018, June). Remote sensing of droplet number concentration in warm clouds: A review of the current state of knowledge and perspectives. *Reviews of Geophysics*, *56*(2), 409–453. Retrieved from https://doi.org/10.1029/2017rg000593 doi: 10.1029/2017rg000593

Haar, T. H. V., & Suomi, V. E. (1971). Measurements of the earth's radiation budget from satellites during a five-year period. part i: Extended time and space means. *Journal of Atmospheric Sciences*, *28*(3), 305 - 314. Retrieved from https://journals.ametsoc.org/view/journals/atsc/28/3/1520-0469_1971_028_0305_moterb_2_0_co_2.xml doi: 10.1175/1520-0469(1971)028<0305:MOTERB>2.0.CO;2

Haaren, R., Morjaria, M., & Fthenakis, V. (2014, 05). Empirical assessment of short-term variability from utility-scale solar pv plants. *Progress in Photovoltaics: Research and Applications*, *22*. doi: 10.1002/pip.2302

Han, Q., Rossow, W. B., & Lacis, A. A. (1994). Near-global survey of effective droplet radii in liquid water clouds using isccp data. *Journal of Climate*, *7*(4),

465 - 497. Retrieved from https://journals.ametsoc.org/view/journals/clim/7/4/1520-0442_1994_007_0465_ngsoed_2_0_co_2.xml doi: 10.1175/1520-0442(1994)007<0465:NGSOED>2.0.CO;2

Hartmann, D. L., Ockert-Bell, M. E., & Michelsen, M. L. (1992). The effect of cloud type on earth's energy balance: Global analysis. *Journal of Climate*, *5*(11), 1281 - 1304. Retrieved from https://journals.ametsoc.org/view/journals/clim/5/11/1520-0442_1992_005_1281_teocto_2_0_co_2.xml doi: 10.1175/1520-0442(1992)005<1281:TEOCTO>2.0.CO;2

Hasenbalg, M., Kuhn, P., Wilbert, S., Nouri, B., & Kazantzidis, A. (2020). Benchmarking of six cloud segmentation algorithms for ground-based all-sky imagers. *Solar Energy*, *201*, 596–614. Retrieved from https://doi.org/10.1016/j.solener.2020.02.042 doi: 10.1016/j.solener.2020.02.042

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020, March). Momentum Contrast for Unsupervised Visual Representation Learning. Retrieved 2022-08-30, from http://arxiv.org/abs/1911.05722 (arXiv:1911.05722 [cs])

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December). Deep Residual Learning for Image Recognition. Retrieved 2022-08-30, from http://arxiv.org/abs/1512.03385 (arXiv:1512.03385 [cs])

Hendrycks, D., & Gimpel, K. (2023). *Gaussian error linear units (gelus).* Retrieved from https://arxiv.org/abs/1606.08415

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., ... Thépaut, J.-N. (2023). *Era5 hourly data on pressure levels from 1940 to present.* [dataset] Copernicus Climate Change Service (C3S) Climate Data Store (CDS). doi: 10.24381/cds.bd0915c6

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J.-N. (2020). The era5 global reanalysis [dataset]. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: https://doi.org/10.1002/qj.3803

Hoffmann, J., Bauer, P., Sandu, I., Wedi, N., Geenen, T., & Thiemert, D. (2023). Destination earth – a digital twin in support of climate services. *Climate Services*, *30*, 100394. Retrieved from https://www.sciencedirect.com/science/article/pii/S2405880723000559 doi: https://doi.org/10.1016/j.cliser.2023.100394

Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., ... Stevens, B. (2023). Icon-sapphire: simulating the components of the earth system and their interactions at kilometer and subkilometer scales. *Geoscientific Model Development*, *16*(2), 779–811. Retrieved from https://gmd.copernicus.org/articles/16/779/2023/ doi: 10.5194/gmd-16-779-2023

Hornik, K., Feinerer, I., Kober, M., & Buchta, C. (2012). Spherical k-means clustering. *Journal of Statistical Software*, *50*(10), 1–22. Retrieved from

https://www.jstatsoft.org/index.php/jss/article/view/v050i10 doi: 10.18637/jss.v050.i10

Houghton, J., Ding, Y., Griggs, D., Noguer, M., van der Linden, P., Dai, X., ... Johnson, C. (2001). *Climate change 2001: The scientific basis*. Cambridge: Cambridge University Press.

Houghton, J., Meira Filho, L., Callander, B., Harris, N., Kattenberg, A., & Maskell, K. (1996). *Climate change 1995–the science of climate change*. Cambridge: Cambridge University Press.

Howard, L. (1803). I. on the modifications of clouds, and on the principles of their production, suspension, and destruction; being the substance of an essay read before the askesian society in the session 1802–3. *The Philosophical Magazine*, *17*(65), 5–11.

Huynh, T., Kornblith, S., Walter, M. R., Maire, M., & Khademi, M. (2022). *Boosting contrastive self-supervised learning with false negative cancellation.*

Höjgård-Olsen, E., Brogniez, H., & Chepfer, H. (2020). Observed evolution of the tropical atmospheric water cycle with sea surface temperature. *Journal of Climate*, *33*(9), 3449 - 3470. Retrieved from https://journals.ametsoc.org/view/journals/clim/33/9/jcli-d-19-0468.1.xml doi: 10.1175/JCLI-D-19-0468.1

Illingworth, A. J., Hogan, R. J., O'Connor, E., Bouniol, D., Brooks, M. E., Delanoé, J., ... Wrench, C. L. (2007). Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bulletin of the American Meteorological Society*, *88*(6), 883 - 898. Retrieved from https://journals.ametsoc.org/view/journals/bams/88/6/bams-88-6-883.xml doi: 10.1175/BAMS-88-6-883

Ineichen, P., & Perez, R. (2002, 09). A new airmass independent formulation for the linke turbidity coefficient. *Solar Energy*, *73*, 151-157. doi: 10.1016/S0038-092X(02)00045-2

Iosifidis, V., & Ntoutsi, E. (2018). Dealing with bias via data augmentation in supervised learning scenarios.. Retrieved from https://api.semanticscholar.org/CorpusID:53504799

IPCC Core Writing Team, Pachauri, R., & Reisinger, A. e. (2007). *Climate change 2007: Synthesis report. contribution of working groups i, ii and iii to the fourth assessment report of the intergovernmental panel on climate change.* Geneva, Switzerland: IPCC.

Iversen, J. E. B., & Pinson, P. (2016). Resgen: Renewable energy scenario generation platform. In *Proceedings of ieee pes general meeting.*

Jakob, C., & Schumacher, C. (2008). Precipitation and latent heating characteristics of the major tropical western pacific cloud regimes. *J. Climate*, *21*, 4348–4365.

Jakob, C., & Tselioudis, G. (2003). Objective identification of cloud regimes in the tropical western pacific. *Geophysical Research Letters*, *30*(21). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GL018367 doi: https://doi.org/10.1029/2003GL018367

Jakob, C., Tselioudis, G., & Hume, T. (2005). The radiative, cloud, and thermodynamics properties of the major tropical western pacific cloud regimes. *J. Climate*, *18*, 1203–1221.

Janssens, M., Vilà-Guerau de Arellano, J., Scheffer, M., Antonissen, C., Siebesma, A. P., & Glassmeier, F. (2021, March). Cloud Patterns in the Trades Have Four Interpretable Dimensions. *Geophysical Research Letters*, *48*(5). Retrieved 2022-08-30, from https://onlinelibrary.wiley.com/doi/10.1029/2020GL091001 doi: 10.1029/2020GL091001

Jin, D., Oreopoulos, L., & Lee, D. (2016, April). Regime-based evaluation of cloudiness in CMIP5 models. *Climate Dynamics*, *48*(1-2), 89–112. Retrieved from https://doi.org/10.1007/s00382-016-3064-0 doi: 10.1007/s00382-016-3064-0

Keisler, R. (2022). *Forecasting global weather with graph neural networks.*

Kelly, A., Moyer, E., Mantziaras, D., & Case, W. (2014). Terra mission operations: Launch to the present (and beyond). *Earth Observing Systems XIX*, *9218*, 180–191.

Ketkar, N. (2017). Stochastic gradient descent. In *Deep learning with python: A hands-on introduction* (pp. 113–132). Berkeley, CA: Apress. Retrieved from https://doi.org/10.1007/978-1-4842-2766-4_8 doi: 10.1007/978-1-4842-2766-4_8

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022, jan). Transformers in vision: A survey. *ACM Computing Surveys*, *54*(10s), 1–41. doi: 10.1145/3505244

King, M., Tsay, S.-C., Platnick, S., Wang, M., & Liou, K. (1997, 12). *Cloud retrieval algorithms for modis: Optical thickness, effective particle radius, and thermodynamic phase.*

King, M. D., Platnick, S., Menzel, W. P., Ackerman, S. A., & Hubanks, P. A. (2013). Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(7), 3826-3852. doi: 10.1109/TGRS.2012.2227333

King, M. D., Platnick, S., Yang, P., Arnold, G. T., Gray, M. A., Riedi, J. C., ... Liou, K.-N. (2004). Remote sensing of liquid water and ice

cloud optical thickness and effective radius in the arctic: Application of airborne multispectral mas data. *Journal of Atmospheric and Oceanic Technology*, *21*(6), 857 - 875. Retrieved from https://journals.ametsoc.org/view/journals/atot/21/6/1520-0426_2004_021_0857_rsolwa_2_0_co_2.xml doi: 10.1175/1520-0426(2004)021<0857:RSOLWA>2.0.CO;2

Klein, S. A., & Hartmann, D. L. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, *6*(8), 1587–1606.

Kougias, I., Taylor, N., Kakoulaki, G., & Jäger-Waldau, A. (2021). The role of photovoltaics for the european green deal and the recovery plan. *Renewable and Sustainable Energy Reviews*, *144*, 111017. Retrieved from https://www.sciencedirect.com/science/article/pii/S1364032121003075 doi: https://doi.org/10.1016/j.rser.2021.111017

Krismanto, A. U., Mithulananthan, N., & Krause, O. (2018). Stability of renewable energy based microgrid in autonomous operation. *Sustainable Energy, Grids and Networks*, *13*, 134-147. Retrieved from https://www.sciencedirect.com/science/article/pii/S2352467717301856 doi: https://doi.org/10.1016/j.segan.2017.12.009

Kurihana, T., Moyer, E., Willett, R., Gilton, D., & Foster, I. (2022). Data-driven cloud clustering via a rotationally invariant autoencoder. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-25. doi: 10.1109/TGRS.2021.3098008

Kurihana, T., Moyer, E. J., & Foster, I. T. (2022). Aicca: Ai-driven cloud classification atlas. *Remote Sensing*, *14*(22). Retrieved from https://www.mdpi.com/2072-4292/14/22/5690 doi: 10.3390/rs14225690

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., ... Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416-1421. Retrieved from https://www.science.org/doi/abs/10.1126/science.adi2336 doi: 10.1126/science.adi2336

Lamer, K., & Kollias, P. (2015). Observations of fair-weather cumuli over land: Dynamical factors controlling cloud size and cover. *Geophysical Research Letters*, *42*(20), 8693-8701. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GL064534 doi: https://doi.org/10.1002/2015GL064534

Lane, D. E., Goris, K., & Somerville, R. C. J. (2002). Radiative transfer through broken clouds: Observations and model validation. *Journal of Climate*, *15*(20), 2921 - 2933. Retrieved from https://journals.ametsoc.org/view/journals/clim/15/20/1520-0442_2002_015_2921_rttbco_2.0.co_2.xml doi: 10.1175/1520-0442(2002)015<2921:RTTBCO>2.0.CO;2

Larson, K., Hartmann, D., & Klein, S. (1999). The role of clouds, water vapor, circulation, and boundary layer structure in the sensitivity of the tropical climate. *Journal of Climate*, *12*, 2359–2374.

Larsson, G., Maire, M., & Shakhnarovich, G. (2017). *Learning representations for automatic colorization.*

Lave, M., & Kleissl, J. (2010). Solar variability of four sites across the state of colorado. *Renewable Energy*, *35*(12), 2867-2873. Retrieved from https://www.sciencedirect.com/science/article/pii/S0960148110002314 doi: https://doi.org/10.1016/j.renene.2010.05.013

LeCun, Y., Bengio, Y., & Hinton, G. (2015, 05). Deep learning. *Nature*, *521*, 436-44. doi: 10.1038/nature14539

LeCun, Y., & Huang, F. J. (2005, 06–08 Jan). Loss functions for discriminative training of energy-based models. In R. G. Cowell & Z. Ghahramani (Eds.), *Proceedings of the tenth international workshop on artificial intelligence and statistics* (Vol. R5, pp. 206–213). PMLR. Retrieved from https://proceedings.mlr.press/r5/lecun05a.html (Reissued by PMLR on 30 March 2021.)

Leinonen, J., Lebsock, M. D., Oreopoulos, L., & Cho, N. (2016). Inter-regional differences in modis-derived cloud regimes. *Journal of Geophysical Research: Atmospheres*, *121*(19), 11,648-11,665. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025193 doi: https://doi.org/10.1002/2016JD025193

Lewis, J., & Derber, J. (1985). *The use of adjoint equations to solve a variational adjustment problem with advective constraints.* (Vol. 37A).

Li, Y., Thompson, D. W. J., Stephens, G. L., & Bony, S. (2014). A global survey of the instantaneous linkages between cloud vertical structure and large-scale climate. *Journal of Geophysical Research: Atmospheres*, *119*, 3770–3792. Retrieved from https://doi.org/10.1002/2013jd020669 doi: 10.1002/2013jd020669

Liou, K.-N. (1986). Influence of cirrus clouds on weather and climate processes: A global perspective. *Monthly Weather Review*, *114*(6), 1167 - 1199. Retrieved from https://journals.ametsoc.org/view/journals/mwre/114/6/1520-0493_1986_114_1167_ioccow_2_0_co_2.xml doi: 10.1175/1520-0493(1986)114<1167:IOCCOW>2.0.CO;2

Liu, H., HaoChen, J. Z., Gaidon, A., & Ma, T. (2022, May). Self-supervised Learning is More Robust to Dataset Imbalance. Retrieved 2022-08-31, from http://arxiv.org/abs/2110.05025 (arXiv:2110.05025 [cs, stat])

Lohmann, Monahan, A. H., & Heinemann, D. (2016). Local short-term variability in solar irradiance. *Atmospheric Chemistry and Physics*, *16*(10), 6365–6379. Retrieved from https://acp.copernicus.org/articles/16/6365/2016/ doi: 10.5194/acp-16-6365-2016

Lohmann, U., Lüönd, F., & Mahrt, F. (2016). *An introduction to clouds: From the microscale to climate.* Cambridge University Press.

Löhnert, U., Schween, J., Acquistapace, C., Ebell, K., Maahn, M., Barrera-Verdejo, M., ... Crewell, S. (2014, 11). Joyce: Jülich observatory for cloud evolution. *Bull. Amer. Meteor. Soc.*, *96*. doi: 10.1175/BAMS-D-14-00105.1

Ma, C.-C., Mechoso, C., Robertson, A., & Arakawa, A. (1996). Peruvian stratus clouds and the tropical pacific circulation: A coupled ocean–atmosphere gcm study. *Journal of Climate*, *9*, 1635–1645.

Mace, G. G., & Wrenn, F. J. (2013). Evaluation of the hydrometeor layers in the east and west pacific within isccp cloud-top pressure–optical depth bins using merged cloudsat and calipso data. *Journal of Climate*, *26*(23), 9429 - 9444. Retrieved from `https://journals.ametsoc.org/view/journals/clim/26/23/jcli-d-12-00207.1.xml` doi: 10.1175/JCLI-D-12-00207.1

MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations.* Proc. 5th Berkeley Symp. Math. Stat. Probab., Univ. Calif. 1965/66, 1, 281-297 (1967).

Marcos, J., De la Parra, I., García, M., & Marroyo, L. (2014). Control strategies to smooth short-term power fluctuations in large photovoltaic plants using battery storage systems. *Energies*, *7*(10), 6593–6619. Retrieved from `https://www.mdpi.com/1996-1073/7/10/6593` doi: 10.3390/en7106593

Marquez, R., & Coimbra, C. F. (2011). Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the nws database. *Solar Energy*, *85*(5), 746-756. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0038092X11000193` doi: https://doi.org/10.1016/j.solener.2011.01.007

Maskey, M., Ramachandran, R., Ramasubramanian, M., Gurung, I., Freitag, B., Kaulfus, A., ... Miller, J. (2020, 07). Deepti: Deep-learning-based tropical cyclone intensity estimation system. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *PP*, 1-1. doi: 10.1109/JSTARS.2020.3011907

McCandless, T., Haupt, S., & Young, G. (2016). A regime-dependent artificial neural network technique for short-range solar irradiance forecasting. *Renewable Energy*, *89*, 351-359. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0960148115305346` doi: https://doi.org/10.1016/j.renene.2015.12.030

Meirink, J. F., Roebeling, R. A., & Stammes, P. (2013, September). Inter-calibration of polar imager solar channels using SEVIRI. *Atmospheric Measurement Techniques*, *6*(9), 2495–2508. Retrieved from `https://doi.org/10.5194/amt-6-2495-2013` doi: 10.5194/amt-6-2495-2013

Menzel, W. P. (2005). Remote sensing applications with meteorological satellites. *The Solar Spectrum*, *3*(10).

Misra, I., Lawrence Zitnick, C., Mitchell, M., & Girshick, R. (2016, June). Seeing through the human reporting bias: Visual classifiers from noisy human-centric labels. In *Proceedings of the ieee conference on computer vision and pattern recognition (cvpr).*

Morgado, P., Li, Y., & Vasconcelos, N. (2020). *Learning representations from audio-visual spatial alignment.*

Morgado, P., Vasconcelos, N., & Misra, I. (2020). Audio-visual instance discrimination with cross-modal agreement. *CoRR*, *abs/2004.12943*. Retrieved from `https://arxiv.org/abs/2004.12943`

Morgado, P., Vasconcelos, N., & Misra, I. (2021, March). Audio-Visual Instance Discrimination with Cross-Modal Agreement. Retrieved 2022-08-30, from `http://arxiv.org/abs/2004.12943` (arXiv:2004.12943 [cs])

Muller, C. J., & Held, I. M. (2012, August). Detailed Investigation of the Self-Aggregation of Convection in Cloud-Resolving Simulations. *Journal of the Atmospheric Sciences*, *69*(8), 2551–2565. (Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences) doi: 10.1175/JAS-D-11-0257.1

Murphy, J., Sexton, D., Barnett, D., Jones, G., Webb, M., Collins, M., & Stainforth, D. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, *430*, 768–772.

Müller, R., & Pfeifroth, U. (2022, 03). Remote sensing of solar surface radiation – a reflection of concepts, applications and input data based on experience with the effective cloud albedo. *Atmospheric Measurement Techniques*, *15*, 1537-1561. doi: 10.5194/amt-15-1537-2022

Na, Y., Fu, Q., Leung, L. R., Kodama, C., & Lu, R. (2022). Mesoscale convective systems simulated by a high-resolution global nonhydrostatic model over the united states and china. *Journal of Geophysical Research: Atmospheres*, *127*(7), e2021JD035916. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JD035916` (e2021JD035916 2021JD035916) doi: https://doi.org/10.1029/2021JD035916

Nakajima, T., & King, M. D. (1990, August). Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. part i: Theory. *Journal of the Atmospheric Sciences*, *47*(15), 1878–1893. Retrieved from `https://doi.org/10.1175/1520-0469(1990)047<1878:dotota>2.0.co;2` doi: 10.1175/1520-0469(1990)047<1878:dotota>2.0.co;2

Narenpitak, P., Kazil, J., Yamaguchi, T., Quinn, P., & Feingold, G. (2021). From Sugar to Flowers: A Transition of Shallow Cumulus Organization Dur-

ing ATOMIC. *Journal of Advances in Modeling Earth Systems*, *13*(10), e2021MS002619. doi: 10.1029/2021MS002619

Narenpitak, P., Kazil, J., Yamaguchi, T., Quinn, P. K., & Feingold, G. (2023). The Sugar-To-Flower Shallow Cumulus Transition Under the Influences of Diel Cycle and Free-Tropospheric Mineral Dust. *Journal of Advances in Modeling Earth Systems*, *15*(1), e2022MS003228. doi: 10.1029/2022MS003228

NASA, LARC, SD, & ASDC. (2017, 9 19). *Ceres and geo-enhanced toa, within-atmosphere and surface fluxes, clouds and aerosols 1-hourly terra-aqua edition4a.* NASA Langley Atmospheric Science Data Center DAAC [Dataset]. Retrieved from `https://doi.org/10.5067/TERRA+AQUA/CERES/SYN1DEG-1HOUR_L3.004A`

Nesbitt, S. W., Gochis, D. J., & Lang, T. J. (2008). The diurnal cycle of clouds and precipitation along the sierra madre occidental observed during name-2004: Implications for warm season precipitation estimation in complex terrain. *Journal of Hydrometeorology*, *9*, 728–743. Retrieved from `https://doi.org/10.1175/2008jhm939.1` doi: 10.1175/2008jhm939.1

Neumann, P., Düben, P., Adamidis, P., Bauer, P., Brück, M., Kornblueh, L., ... Biercamp, J. (2019). Assessing the scales in numerical weather and climate predictions: will exascale be the rescue? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *377*(2142), 20180148. Retrieved from `https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2018.0148` doi: 10.1098/rsta.2018.0148

Nie, Y., Zamzam, A. S., & Brandt, A. (2021). Resampling and data augmentation for short-term pv output prediction based on an imbalanced sky images dataset using convolutional neural networks. *Solar Energy*, *224*, 341-354. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0038092X21004795` doi: https://doi.org/10.1016/j.solener.2021.05.095

Nordling, K., Keskinen, J.-P., Romakkaniemi, S., Kokkola, H., Räisänen, P., Lipponen, A., ... Raatikainen, T. (2024). Technical note: Emulation of a large-eddy simulator for stratocumulus clouds in a general circulation model. *Atmospheric Chemistry and Physics*, *24*(2), 869–890. Retrieved from `https://acp.copernicus.org/articles/24/869/2024/` doi: 10.5194/acp-24-869-2024

Norris, J. R., & Weaver, C. P. (2001). Improved techniques for evaluating gcm cloudiness applied to the ncar ccm3. *Journal of Climate*, *14*(12), 2540 - 2550. Retrieved from `https://journals.ametsoc.org/view/journals/clim/14/12/1520-0442_2001_014_2540_itfegc_2.0.co_2.xml` doi: 10.1175/1520-0442(2001)014<2540:ITFEGC>2.0.CO;2

Nuijens, L., & Siebesma, A. P. (2019, June). Boundary Layer Clouds and Convection over Subtropical Oceans in our Current and in a Warmer Climate.

Current Climate Change Reports, *5*(2), 80–94. doi: 10.1007/s40641-019-00126-x

Oceanic, N., & (NOAA), A. A. (2024a). *Noaa goes-16 cloud optical depth and cloud particle size distribution (codc) [dataset].* Retrieved from `https://noaa-goes16.s3.amazonaws.com/index.html#ABI-L2-CODC/andhttps://registry.opendata.aws/noaa-cdr-oceanic/` (Accessed: 2024-05-14)

Oceanic, N., & (NOAA), A. A. (2024b). *Noaa oceanic climate data records.* Retrieved from `https://registry.opendata.aws/noaa-cdr-oceanic/` (Accessed: 2024-05-14)

Ockert-Bell, M. E., & Hartmann, D. L. (1992). The effect of cloud type on earth's energy balance: Results for selected regions. *Journal of Climate*, *5*(10), 1157–1171. Retrieved 2024-04-19, from `http://www.jstor.org/stable/26197159`

Ordoudis, C., Pinson, P., Morales, J., Juan, C., & Zugno, M. (2016). *An updated version of the ieee rts 24-bus system for electricity market and power system operation studies.* Technical Report.

Oreopoulos, L., Cho, N., Lee, D., & Kato, S. (2016). Radiative effects of global modis cloud regimes. *Journal of Geophysical Research: Atmospheres*, *121*(5), 2299-2317. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JD024502` doi: https://doi.org/10.1002/2015JD024502

Oreopoulos, L., Cho, N., Lee, D., Kato, S., & Huffman, G. J. (2014, July). An examination of the nature of global MODIS cloud regimes. *Journal of Geophysical Research: Atmospheres*, *119*(13), 8362–8383. Retrieved from `https://doi.org/10.1002/2013jd021409` doi: 10.1002/2013jd021409

Orhan, A., & Lake, B. (2024, 03). Learning high-level visual representations from a child's perspective without strong inductive biases. *Nature Machine Intelligence*, *6*, 1-13. doi: 10.1038/s42256-024-00802-0

Orlanski, I. (1975). A rational subdivision of scales for atmospheric processes. *Bulletin of the American Meteorological Society*, *56*(5), 527–530. Retrieved 2024-04-21, from `http://www.jstor.org/stable/26216020`

Paletta, Q., Terrén-Serrano, G., Nie, Y., Li, B., Bieker, J., Zhang, W., ... Feng, C. (2023). Advances in solar forecasting: Computer vision with deep learning. *Advances in Applied Energy*, *11*, 100150. Retrieved from `https://www.sciencedirect.com/science/article/pii/S266679242300029X` doi: https://doi.org/10.1016/j.adapen.2023.100150

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). *Context encoders: Feature learning by inpainting.*

Pedro, H. T., & Coimbra, C. F. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, *86*(7), 2017-

2028. Retrieved from https://www.sciencedirect.com/science/article/pii/S0038092X12001429 doi: https://doi.org/10.1016/j.solener.2012.04.004

Perez, R., David, M., Hoff, T. E., Jamaly, M., Kivalov, S., Kleissl, J., ... Perez, M. (2016). Spatial and temporal variability of solar energy. *Foundations and Trends® in Renewable Energy*, *1*(1), 1-44. Retrieved from http://dx.doi.org/10.1561/2700000006 doi: 10.1561/2700000006

Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., & Vignola, F. (2002). A new operational model for satellite-derived irradiances: description and validation. *Solar Energy*, *73*(5), 307-317. Retrieved from https://www.sciencedirect.com/science/article/pii/S0038092X02001226 doi: https://doi.org/10.1016/S0038-092X(02)00122-6

Perez, R., Schlemmer, J., Hemker, K., Kivalov, S., Kankiewicz, A., & Gueymard, C. (2015). Satellite-to-irradiance modeling - a new version of the suny model. In *2015 ieee 42nd photovoltaic specialist conference (pvsc)* (p. 1-7). doi: 10.1109/PVSC.2015.7356212

Pincus, R., Batstone, C. P., Hofmann, R. J. P., Taylor, K. E., & Glecker, P. J. (2008). Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. *Journal of Geophysical Research: Atmospheres*, *113*(D14). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD009334 doi: https://doi.org/10.1029/2007JD009334

Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., ... Riedi, J. (2017, January). The MODIS cloud optical and microphysical products: Collection 6 updates and examples from terra and aqua. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(1), 502–525. Retrieved from https://doi.org/10.1109/tgrs.2016.2610522 doi: 10.1109/tgrs.2016.2610522

Poddar, S., Evans, J. P., Kay, M., Prasad, A., & Bremner, S. (2023). Assessing australia's future solar power ramps with climate projections. *Scientific Reports*, *13*(1), 11503. Retrieved from https://doi.org/10.1038/s41598-023-38566-z doi: 10.1038/s41598-023-38566-z

Prein, A. F., Liu, C., Ikeda, K., Bullock, R., Rasmussen, R. M., Holland, G. J., & Clark, M. (2020). Simulating north american mesoscale convective systems with a convection-permitting climate model. *Climate Dynamics*, *55*(1), 95–110. Retrieved from https://doi.org/10.1007/s00382-017-3993-2 doi: 10.1007/s00382-017-3993-2

PREPA — Puerto Rico Energy Authority. (2013). (www.prepa.com)

Rackow, T., Pedruzo-Bagazgoitia, X., Becker, T., Milinski, S., Sandu, I., Aguridan, R., ... Ziemen, F. (2024). Multi-year simulations at kilometre scale with the integrated forecasting system coupled to fesom2.5/nemov3.4. *EGUsphere*. doi: 10.5194/egusphere-2024-913

Ranaweera, I., Midtgård, O.-M., & Yordanov, G. (2014, 01). Short-term intermittency of solar irradiance in southern norway..

Rani, V., Nabi, S., Kumar, M., Mittal, A., & Saluja, K. (2023, 01). Self-supervised learning: A succinct review. *Archives of Computational Methods in Engineering*, *30*. doi: 10.1007/s11831-023-09884-2

Ranzato, M., Boureau, Y.-L., Chopra, S., & LeCun, Y. (2007). *A unified energy-based framework for unsupervised learning* (Vol. 2; M. Meila & X. Shen, Eds.). San Juan, Puerto Rico: PMLR. Retrieved from `https://proceedings.mlr.press/v2/ranzato07a.html`

Rasp, S., Schulz, H., Bony, S., & Stevens, B. (2020, November). Combining Crowdsourcing and Deep Learning to Explore the Mesoscale Organization of Shallow Convection. *Bulletin of the American Meteorological Society*, *101*(11), E1980–E1995. (Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society) doi: 10.1175/BAMS-D-19-0324.1

Reichler, T., & Kim, J. (2008). How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, *89*(3), 303–311. doi: 10.1175/BAMS-89-3-303

Rieck, M., Hohenegger, C., & van Heerwaarden, C. C. (2014). The influence of land surface heterogeneities on cloud size development. *Monthly Weather Review*, *142*(10), 3830 - 3846. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/142/10/mwr-d-13-00354.1.xml` doi: 10.1175/MWR-D-13-00354.1

Roebeling, R. A., Feijt, A. J., & Stammes, P. (2006). Cloud property retrievals for climate monitoring: Implications of differences between spinning enhanced visible and infrared imager (seviri) on meteosat-8 and advanced very high resolution radiometer (avhrr) on noaa-17. *Journal of Geophysical Research: Atmospheres*, *111*(D20). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006990` doi: https://doi.org/10.1029/2005JD006990

Rohden, M., Sorge, A., Timme, M., & Witthaut, D. (2012, Aug). Self-organized synchronization in decentralized power grids. *Phys. Rev. Lett.*, *109*, 064101. Retrieved from `https://link.aps.org/doi/10.1103/PhysRevLett.109.064101` doi: 10.1103/PhysRevLett.109.064101

Rosa, R. N., & Silva, A. M. (2021). Solar radiation budget in the atmosphere under broken cloudy sky. an analytical model. *Journal of Atmospheric and Solar-Terrestrial Physics*, *219*, 105643. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1364682621001024` doi: https://doi.org/10.1016/j.jastp.2021.105643

Rossow, W. B., & Schiffer, R. A. (1999). Advances in understanding clouds from isccp. *Bulletin of the American Meteorological Society*, *80*(11), 2261 -

2288. Retrieved from https://journals.ametsoc.org/view/journals/bams/80/11/1520-0477_1999_080_2261_aiucfi_2_0_co_2.xml doi: 10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2

Rother, A., Niemann, U., Hielscher, T., Völzke, H., Ittermann, T., & Spiliopoulou, M. (2021, 07). Assessing the difficulty of annotating medical data in crowdworking with help of experiments. *PLOS ONE*, *16*(7), 1-26. Retrieved from https://doi.org/10.1371/journal.pone.0254764 doi: 10.1371/journal.pone.0254764

Rousseeuw, P. J. (1987a). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65. Retrieved from https://www.sciencedirect.com/science/article/pii/0377042787901257 doi: https://doi.org/10.1016/0377-0427(87)90125-7

Rousseeuw, P. J. (1987b). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65. Retrieved from https://www.sciencedirect.com/science/article/pii/0377042787901257 doi: https://doi.org/10.1016/0377-0427(87)90125-7

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986, October). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. Retrieved from https://doi.org/10.1038/323533a0 doi: 10.1038/323533a0

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015, December). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. Retrieved 2022-08-31, from http://link.springer.com/10.1007/s11263-015-0816-y doi: 10.1007/s11263-015-0816-y

Ryan, B., & et al. (2000). Simulations of a cold front by cloud-resolving, limited-area, and large-scale models, and a model evaluation using in situ and satellite observations. *Monthly Weather Review*, *128*, 3218–3235.

Sabottke, C. F., & Spieler, B. M. (2020). The effect of image resolution on deep learning in radiography. *Radiology: Artificial Intelligence*, *2*(1), e190015. Retrieved from https://doi.org/10.1148/ryai.2019190015 (PMID: 33937810) doi: 10.1148/ryai.2019190015

Salomonson, V., Barnes, W., Xiong, J., Kempler, S., & Masuoka, E. (2002). An overview of the earth observing system modis instrument and associated data systems performance. In *Ieee international geoscience and remote sensing symposium* (Vol. 2, p. 1174-1176 vol.2). doi: 10.1109/IGARSS.2002.1025812

Sarkar, M., Zuidema, P., Albrecht, B., Ghate, V., Jensen, J., Mohrmann, J., & Wood, R. (2020). Observations pertaining to precipitation within the northeast

pacific stratocumulus-to-cumulus transition. *Monthly Weather Review*, *148*(3), 1251 - 1273. doi: 10.1175/MWR-D-19-0235.1

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., & et al. (2018). An update on the rttov fast radiative transfer model (currently at version 12). *Geoscientific Model Development*, *11*(7), 2717–2737. doi: 10.5194/gmd-11-2717-2018

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., & Ratier, A. (2002a). An introduction to meteosat second generation (msg). *Bulletin of the American Meteorological Society*, *83*(7), 977 - 992. Retrieved from `https://journals.ametsoc.org/view/journals/bams/83/7/1520-0477_2002_083_0977_aitmsg_2_3_co_2.xml` doi: 10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., & Ratier, A. (2002b). An introduction to meteosat second generation (msg). *Bulletin of the American Meteorological Society*, *83*(7), 977 - 992. Retrieved from `https://journals.ametsoc.org/view/journals/bams/83/7/1520-0477_2002_083_0977_aitmsg_2_3_co_2.xml` doi: 10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2

Schmit, T. J., Gunshor, M. M., Menzel, W. P., Gurka, J. J., Li, J., & Bachmeier, A. S. (2005). Introducing the next-generation advanced baseline imager on goes-r. *Bulletin of the American Meteorological Society*, *86*(8), 1079 - 1096. Retrieved from `https://journals.ametsoc.org/view/journals/bams/86/8/bams-86-8-1079.xml` doi: https://doi.org/10.1175/BAMS-86-8-1079

Schmit, T. J., Lindstrom, S. S., Gerth, J. J., & Gunshor, M. M. (2018). Applications of the 16 spectral bands on the advanced baseline imager (abi). *Journal of Operational Meteorology*, *6*(4), 33–46. Retrieved from `https://doi.org/10.15191/nwajom.2018.0604` doi: 10.15191/nwajom.2018.0604

Schneider, T., Kaul, C., & Pressel, K. (2019, 03). Possible climate transitions from breakup of stratocumulus decks under greenhouse warming. *Nature Geoscience*, *12*, 164-168. doi: 10.1038/s41561-019-0310-1

Schubert, E. (2023, jun). Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *ACM SIGKDD Explorations Newsletter*, *25*(1), 36–42. doi: 10.1145/3606274.3606278

Schulz, H. (2022). C^3ontext: a common consensus on convective organization during the eurec^{4}a experiment. *Earth System Science Data*, *14*(3), 1233–1256. doi: 10.5194/essd-14-1233-2022

Schulz, H., Eastman, R., & Stevens, B. (2021a). Characterization and evolution of organized shallow convection in the downstream north atlantic trades. *Journal of Geophysical Research: Atmospheres*, *126*(17), e2021JD034575.

Schulz, H., Eastman, R., & Stevens, B. (2021b). Characterization and Evolution of Organized Shallow Convection in the Downstream North Atlantic Trades. *Journal of Geophysical Research: Atmospheres*, *126*(17), e2021JD034575. doi: 10.1029/2021JD034575

Schulz, H., Eastman, R., & Stevens, B. (2021c). Characterization and Evolution of Organized Shallow Convection in the Downstream North Atlantic Trades. *Journal of Geophysical Research: Atmospheres*, *126*(17), e2021JD034575. doi: 10.1029/2021JD034575

Schulz, H., & Stevens, B. (2023). Evaluating large-domain, hecto-meter, large-eddy simulations of trade-wind clouds using eurec4a data. *Journal of Advances in Modeling Earth Systems*, *15*(10), e2023MS003648. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023MS003648` (e2023MS003648 2023MS003648) doi: https://doi.org/10.1029/2023MS003648

Sedlar, J., Riihimaki, L. D., Lantz, K., & Turner, D. D. (2021). Development of a random-forest cloud-regime classification model based on surface radiation and cloud products. *Journal of Applied Meteorology and Climatology*, *60*(4), 477 - 491. Retrieved from `https://journals.ametsoc.org/view/journals/apme/60/4/JAMC-D-20-0153.1.xml` doi: 10.1175/JAMC-D-20-0153.1

Sengupta, M., Habte, A., Kurtz, S., Dobos, A., Wilbert, S., Lorenz, E., ... Perez, R. (2015). *Best practices handbook for the collection and use of solar resource data for solar energy applications.* National Renewable Energy Laboratory (NREL).

Shang, H., Letu, H., Nakajima, T., Wang, Z., Ma, R., Wang, T., ... Shi, J. (2018, 01). Diurnal cycle and seasonal variation of cloud cover over the tibetan plateau as determined from himawari-8 new-generation geostationary satellite data. *Scientific Reports*, *8*. doi: 10.1038/s41598-018-19431-w

Simpson, J., Kummerow, C., Tao, W.-K., & Adler, R. F. (1996). On the tropical rainfall measuring mission (trmm). *Meteorology and Atmospheric Physics*, *60*(1), 19–36. Retrieved from `https://doi.org/10.1007/BF01029783` doi: 10.1007/BF01029783

Steinbach, M., Ertöz, L., & Kumar, V. (2004). The challenges of clustering high dimensional data. In L. T. Wille (Ed.), *New directions in statistical physics: Econophysics, bioinformatics, and pattern recognition* (pp. 273–309). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from `https://doi.org/10.1007/978-3-662-08968-2_16` doi: 10.1007/978-3-662-08968-2_16

Stephens. (2005). Cloud feedbacks in the climate system: A critical review. *Journal of Climate*, *18*(2), 237 - 273. Retrieved from `https://journals.ametsoc.org/view/journals/clim/18/2/jcli-3243.1.xml` doi: 10.1175/JCLI-3243.1

Stephens, G., Li, J.-L., Wild, M., Clayson, C., Loeb, N., Kato, S., ... Andrews, T. (2012, 09). An update on earth's energy balance in light of the latest global observations. *Nature Geoscience*, *5*, 691-696. doi: 10.1038/ngeo1580

Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C., ... Zuidema, P. (2020). Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds. *Quarterly Journal of the Royal Meteorological Society*, *146*(726), 141-152. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3662 doi: https://doi.org/10.1002/qj.3662

Stevens, B., Bony, S., Farrell, D., Ament, F., Blyth, A., Fairall, C., ... Zöger, M. (2021, August). EUREC^{4}A. *Earth System Science Data*, *13*(8), 4067–4119. (Publisher: Copernicus GmbH) doi: 10.5194/essd-13-4067-2021

Stevens, B., & et al. (2021). EURECA. *Earth System Science Data*, *13*(8), 4067–4119. Retrieved from https://essd.copernicus.org/articles/13/4067/2021/ doi: 10.5194/essd-13-4067-2021

Stevens, B., Farrell, D., Hirsch, L., Jansen, F., Nuijens, L., Serikov, I., ... Prospero, J. M. (2016). The barbados cloud observatory: Anchoring investigations of clouds and circulation on the edge of the itcz. *Bulletin of the American Meteorological Society*, *97*(5), 787 - 801. doi: https://doi.org/10.1175/BAMS-D-14-00247.1

Sylolypavan, A., Sleeman, D., Wu, H., & Sim, M. (2023, 02). The impact of inconsistent human annotations on ai driven clinical decision making. *NPJ digital medicine*, *6*, 26. doi: 10.1038/s41746-023-00773-3

Tan, J., Jackob, C., Rossow, W. B., & Tselioudis, G. (2015, March). Increases in tropical rainfall driven by changes in frequency of organized deep convection. , *519*. Retrieved from https://doi.org/10.1038/nature14339 doi: 10.1038/nature14339

Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, *290*(5500), 2319-2323. Retrieved from https://www.science.org/doi/abs/10.1126/science.290.5500.2319 doi: 10.1126/science.290.5500.2319

Tian, B., Shaikh, M., Azimi-Sadjadi, M., Vonder Haar, T., & Reinke, D. (1999). Errata to "a study of cloud classification with neural networks using spectral and textural features". *IEEE Transactions on Neural Networks*, *10*(3), 722-722. doi: 10.1109/TNN.1999.761731

Tomson, T. (2010). Fast dynamic processes of solar radiation. *Solar Energy*, *84*(2), 318-323. Retrieved from https://www.sciencedirect.com/science/article/pii/S0038092X09002795 doi: https://doi.org/10.1016/j.solener.2009.11.013

Trenberth, K., & Fasullo, J. (2011, 07). Tracking earth's energy: From el nio to global warming. *Surveys in Geophysics*, *33*. doi: 10.1007/s10712-011-9150-2

Tselioudis, G., Rossow, W., Zhang, Y., & Konsta, D. (2013, September). Global weather states and their properties from passive and active satellite cloud retrievals. *Journal of Climate*, *26*(19), 7734–7746. Retrieved from https://doi.org/10.1175/jcli-d-13-00024.1 doi: 10.1175/jcli-d-13-00024.1

Tselioudis, G., Rossow, W. B., Jakob, C., Remillard, J., Tropf, D., & Zhang, Y. (2021, June). EVALUATION OF CLOUDS, RADIATION, AND PRECIPITATION IN CMIP6 MODELS USING GLOBAL WEATHER STATES DERIVED FROM ISCCP-h CLOUD PROPERTY DATA. *Journal of Climate*, 1–42. Retrieved from https://doi.org/10.1175/jcli-d-21-0076.1 doi: 10.1175/jcli-d-21-0076.1

Tselioudis, G., Zhang, Y., & Rossow, W. B. (2000). Cloud and radiation variations associated with northern midlatitude low and high sea level pressure regimes. *Journal of Climate*, *13*(2), 312–327.

Tzallas, V., Hünerbein, A., Stengel, M., Meirink, J. F., Benas, N., Trentmann, J., & Macke, A. (2022, November). CRAAS: A european cloud regime dAtAset based on the CLAAS-2.1 climate data record. *Remote Sensing*, *14*(21), 5548. Retrieved from https://doi.org/10.3390/rs14215548 doi: 10.3390/rs14215548

Uriarte, F., Smith, C., VanBroekhoven, S., & Hebner, R. (2015, 11). Microgrid ramp rates and the inertial stability margin. *IEEE Transactions on Power Systems*, *30*, 1-8. doi: 10.1109/TPWRS.2014.2387700

van den Oord, A., Vinyals, O., & Kavukcuoglu, K. (2017). Neural discrete representation learning. *CoRR*, *abs/1711.00937*. Retrieved from http://arxiv.org/abs/1711.00937

van der Maaten, L., & Hinton, G. (2008a). Visualizing data using t-sne. *Journal of Machine Learning Research*, *9*(86), 2579–2605. Retrieved from http://jmlr.org/papers/v9/vandermaaten08a.html

van der Maaten, L., & Hinton, G. (2008b). Visualizing data using t-sne. *Journal of Machine Learning Research*, *9*(86), 2579–2605. Retrieved from http://jmlr.org/papers/v9/vandermaaten08a.html

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In I. Guyon et al. (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2023). *Attention is all you need.* Retrieved from https://arxiv.org/abs/1706.03762

Vial, J., Bony, S., Stevens, B., & Vogel, R. (2017). Clouds, circulation and climate sensitivity. *Surveys in geophysics*, *38*(6), 1331–1353. doi: 10.1007/s10712-017-9418-2

Vial, J., Vogel, R., & Schulz, H. (2021). On the daily cycle of mesoscale cloud organization in the winter trades. *Quarterly Journal of the Royal Meteorological Society*, *147*(738), 2850–2873. doi: 10.1002/qj.4103

Visa, A., Iivarinen, J., Valkealahti, K., & Simula, O. (1998). *Neural network based cloud classifier*. World Scientific. Retrieved from `https://www.worldscientific.com/doi/abs/10.1142/9789812816955_0035`

Vogel, R., Konow, H., Schulz, H., & Zuidema, P. (2021, November). A climatology of trade-wind cumulus cold pools and their link to mesoscale cloud organization. *Atmospheric Chemistry and Physics*, *21*(21), 16609–16630. (Publisher: Copernicus GmbH) doi: 10.5194/acp-21-16609-2021

Walther, A., & Heidinger, A. K. (2012). Implementation of the daytime cloud optical and microphysical properties algorithm (dcomp) in patmos-x [dataset]. *Journal of Applied Meteorology and Climatology*, *51*(7), 1371 - 1390. doi: 10.1175/JAMC-D-11-0108.1

Warren, S. G., Hahn, C. H., London, J., Chervin, R. M., & Jenne, R. L. (1988). *Global distribution of total cloud cover and cloud type amounts over the ocean* (Tech. Rep. Nos. NCAR/TN-317+STR). University Corporation for Atmospheric Research. doi: 10.5065/D6QC01D1

Weare, B., & Groups, A. M. (1996). Evaluation of the vertical structure of zonally averaged cloudiness and its variability in the atmospheric model intercomparison project. *Journal of Climate*, *9*, 3419–3431.

Webb, M., Senior, C., Bony, S., & Morcrette, J.-J. (2001). Combining erbe and isccp data to assess clouds in the hadley centre, ecmwf and lmd atmospheric climate models. *Climate Dynamics*, *17*, 905–922.

Wielicki, B., Barkstrom, B., Harrison, E., Lee, R., Smith, G., & Cooper, J. (1996, 05). Clouds and the earth's radiant energy system (ceres): An earth observing system experiment. *Bulletin of the American Meteorological Society*, *77*, 853-868. doi: 10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2

Williams, K. D., & Bodas-Salcedo, A. (2017). A multi-diagnostic approach to cloud evaluation. *Geoscientific Model Development*, *10*(7), 2547–2566. Retrieved from `https://gmd.copernicus.org/articles/10/2547/2017/` doi: 10.5194/gmd-10-2547-2017

Williams, K. D., & Brooks, M. E. (2008). Initial tendencies of cloud regimes in the met office unified model. *Journal of Climate*, *21*(4), 833 - 840. Retrieved from `https://journals.ametsoc.org/view/journals/clim/21/4/2007jcli1900.1.xml` doi: 10.1175/2007JCLI1900.1

Williams, K. D., Ringer, M. A., Senior, C. A., Webb, M. J., McAvaney, B. J., Andronova, N., ... Mitchell, J. F. B. (2006). Evaluation of a component of the cloud response to climate change in an intercomparison of climate models. *Climate Dynamics*, *26*(2), 145–165. Retrieved from https://doi.org/10.1007/s00382-005-0067-7 doi: 10.1007/s00382-005-0067-7

Williams, K. D., & Webb, M. J. (2008, August). A quantitative performance assessment of cloud regimes in climate models. *Climate Dynamics*, *33*(1), 141–157. Retrieved from https://doi.org/10.1007/s00382-008-0443-1 doi: 10.1007/s00382-008-0443-1

Williams, K. D., & Webb, M. J. (2009). A quantitative performance assessment of cloud regimes in climate models. *Climate Dynamics*, *33*(1), 141–157. Retrieved from https://doi.org/10.1007/s00382-008-0443-1 doi: 10.1007/s00382-008-0443-1

Wimmers, A., Velden, C., & Cossuth, J. H. (2019). Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery. *Monthly Weather Review*, *147*(6), 2261 - 2282. Retrieved from https://journals.ametsoc.org/view/journals/mwre/147/6/mwr-d-18-0391.1.xml doi: 10.1175/MWR-D-18-0391.1

Xu, W., Balaguru, K., August, A., Lalo, N., Hodas, N., DeMaria, M., & Judi, D. (2021). Deep learning experiments for tropical cyclone intensity forecasts. *Weather and Forecasting*, *36*(4), 1453 - 1470. Retrieved from https://journals.ametsoc.org/view/journals/wefo/36/4/WAF-D-20-0104.1.xml doi: 10.1175/WAF-D-20-0104.1

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. Retrieved from https://doi.org/10.1007/s13244-018-0639-9 doi: 10.1007/s13244-018-0639-9

Yung, Y. L. (2003). An introduction to atmospheric radiation. by k. n. liou. academic press. second edition, 2002. pp. xiv+583. isbn 0 12 451451 0. *Quarterly Journal of the Royal Meteorological Society*, *129*(590), 1741-1741. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/003590003102695746 doi: https://doi.org/10.1256/003590003102695746

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., ... Taylor, K. E. (2020). Causes of higher climate sensitivity in cmip6 models. *Geophysical Research Letters*, *47*, e2019GL085782. doi: 10.1029/2019GL085782

Zelinka, M. D., Randall, D. A., Webb, M. J., & Klein, S. A. (2017, October). Clearing clouds of uncertainty. *Nature Climate Change*, *7*(10), 674-678. doi: 10.1038/nclimate3402

Zhao, W., Marchand, R., & Fu, Q. (2017). The diurnal cycle of clouds and precipitation at the arm sgp site: Cloud radar observations and simulations from the multiscale modeling framework. *Journal of Geophysical Research: Atmospheres*, *122*, 7519–7536. Retrieved from `https://doi.org/10.1002/2016JD026353` doi: 10.1002/2016JD026353

Acknowledgements

Arriving in Germany amidst the pandemic's first wave, when much of life was still finding its footing, I extend my deepest gratitude to Annette (her nerves), Johannes, and dearest Jakob, who nearly matched my height over these 3.5 years. They welcomed me into their family with open arms, embracing me as one of their own. This journey owes its existence to their warmth and pillar-like support.

I extend my sincere gratitude to Prof. Dr. Susanne Crewell for her unwavering guidance throughout this thrilling and memorable journey. Her scientific passion and uplifting energy were instrumental in sustaining focus and momentum. Her leadership, vision, and commitment to the scientific community have humbled me and inspired confidence in trusting my instincts when interacting with peers. Thank you, Susanne, for your invaluable support!

I extend my gratitude to Prof. Dr. Martin Schultz for spearheading the KISTE initiative and for consistently offering insightful ideas during thesis advisory meetings.

A heartfelt thank you goes to Dr. Christoph Boehm and Nils Risse for their encouragement and uplifting support. Our discussions on machine learning are always a highlight, and I greatly value our exchanges. I'm grateful to Nils for opening his home to me during unplanned late nights. Christoph, your review of Chapter 1 was immensely helpful, and Nils, thank you for creating the cloud evolution video over Juelich!

I extend a special thank you to Dr. Claudia Acquistapace for her invaluable support in various aspects. Thank you for your assistance with the manuscripts and for reviewing Chapter 1. I hope our collaborative efforts, particularly our 2022 Christmas experiment connecting mesoscale distribution with ship-based measurements, will culminate in a published paper. Above all, I appreciate our shared fighting spirit and camaraderie.

I thank Dr. Sabrina Schnitt and Paula Bigalke for their insightful discussions on trade wind convection, which provided valuable mental exercises. Additionally, I appreciate Sabrina for introducing me to the EUREC^{4}A community.

Credit is due to Deutsche Bahn's RB 25 service, as many exciting ideas for this thesis emerged during my travels. Its diesel engine noise is deeply encoded in my memory.

I extend my gratitude to Dr. Mario Mech, Dr. Kerstin Ebell, and Mr. Rainer Lind for their unwavering positivity, fun and compassionate support throughout this journey.

A special acknowledgment goes to our admin, Dagmar Janssen, for her exceptional patience in assisting me in navigating the complexities of bureaucracy.

Finally, I want to express my heartfelt gratitude and inner solace to Ma, Bapi, and Jojo for their unconditional support and care. I feel truly fortunate to have you by my side. Of course, there are no words. A deeply noted tribute to my

grandmother (Late Gayatri Mishra) and thanks to all my three uncles, especially my "mejo" mama; your wisdom and encouragement in my primitive years are very foundational and deeply rooted in me. Additionally, I thank my undergraduate friends in India for their belief in me and for maintaining active communication despite the distance.

A special note of gratitude goes to Shreya, who entered my life during the final stages of my PhD journey. Thank you for waking me early in the morning to ensure I could complete my writing on time. Your support has been invaluable.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Dwaipayan Chatterjee

Köln, Mai 2024