



HAL
open science

Unsupervised deep learning for the analysis of remote sensing chronological series: Application to GPM

Vibolroth Sambath

► **To cite this version:**

Vibolroth Sambath. Unsupervised deep learning for the analysis of remote sensing chronological series: Application to GPM. Ocean, Atmosphere. Université Paris-Saclay, 2024. English. NNT: 2024UPASJ022 . tel-04904416

HAL Id: tel-04904416

<https://theses.hal.science/tel-04904416v1>

Submitted on 21 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unsupervised deep learning for the analysis of remote sensing chronological series : Application to GPM

*Apprentissage profond non-supervisé pour l'analyse de
séries chronologiques d'observations spatiales : Application
aux données GPM*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 129 : sciences de l'environnement d'Île-de-France
(SEIF)

Spécialité de doctorat : Géosciences

Graduate School : Géosciences, climat, environnement et planètes.

Référent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche **LATMOS (Université Paris-Saclay, CNRS)**, sous la direction de **Cécile Mallet**, Professeure, co-encadrée par **Nicolas Viltard**, Chargé de recherche, **Laurent Barthès**, Maître de Conférences et **Audrey Martini**, Maître de Conférences

Thèse soutenue à Guyancourt, le 27 septembre 2024, par

Vibolroth SAMBATH

Composition du jury

Membres du jury avec voix délibérative

Valérie Ciarletti

Professeure, UVSQ

Claire Monteleoni

Professor, University of Colorado Boulder

Christopher Kidd

Research Scientist, University of Maryland

Philippe Chambon

Ingénieur des Ponts, des Eaux et des Forêts,

Météo-France

Shruti Upadhyaya

Assistant Professor, IIT Hyderabad

Présidente

Rapportrice et examinatrice

Rapporteur et examinateur

Examinateur

Examinatrice

Titre : Apprentissage profond non-supervisé pour l'analyse de séries chronologiques d'observations spatiales : Application aux données GPM

Mots clés : précipitation, climat, observation spatiale, apprentissage non-supervisé, apprentissage profond, segmentation d'image

Résumé : Le cycle de l'eau joue un rôle essentiel pour toute vie sur Terre. L'observation et la quantification de ses divers composants sous la pression du changement climatique sont devenues de plus en plus importantes pour la société et la science. Les réponses de chaque composante du cycle de l'eau au réchauffement climatique diffèrent fortement d'une région à l'autre. Par exemple, certaines régions ont connu une augmentation des précipitations, tandis que d'autres ont connu une diminution. Il est donc nécessaire d'observer le cycle de l'eau à l'échelle globale.

La télédétection a permis d'acquérir une grande quantité de données d'observation spatiale. Concernant les précipitations, la mission Global Precipitation Measurement (GPM) (2014-présent) offre une constellation de satellites équipés de radiomètres micro-ondes, ce qui permet d'obtenir une vaste base de données de température de brillance (TB). Les TB contiennent des informations implicites sur plusieurs variables géophysiques importantes, notamment l'intensité de la pluie, la vapeur d'eau, la température de surface de la mer (SST), la neige et le contenu en eau liquide et glacé dans les nuages.

Notre présente étude utilise le radiomètre (GMI) du GPM Core Observatory, qui est le satellite de référence pour unifier les satellites de la constellation. En considérant les mesures de TB comme des images non conventionnelles, nous proposons une nouvelle méthode pour observer le cycle de l'eau à partir de ces mesures. La segmentation sémantique non-supervisée est appliquée aux données GMI pour extraire des informations sans l'aide d'une annotation ou de données auxiliaires. L'avantage d'une telle approche est qu'elle peut être appliquée à n'importe quel radiomètre de la constellation. Elle est également exempte des incertitudes liées aux données auxiliaires. Des modèles de segmentation non-supervisés de différentes complexité sont explorés. Nous avons

testé des approches qui s'appuient sur des réseaux entièrement convolutifs (FCN) pour extraire des cartes de caractéristiques à partir de l'entrée. Dans cette optique, nous avons utilisé dans un premier temps un modèle FCN relativement simple composé de quelques couches de convolution 2D. Dans un second temps, afin d'augmenter la complexité du module d'extraction de caractéristiques, nous l'avons remplacé par un réseau de type U-Net. Un modèle FCN basé sur une architecture de type auto encodeur avec un mécanisme d'attention a également été développé. Enfin, une approche par pixel de type k-moyenne a été testée.

Pour comparer les performances de ces différents modèles, nous proposons une évaluation externe où les variables géophysiques pertinentes sont utilisées pour le calcul des métriques de segmentation. D'abord, les classes sont attribuées aux variables géophysiques. Ensuite, en convertissant ces variables quantitatives (précipitation et SST) en catégories, elles peuvent être comparées à la segmentation à l'aide de différentes métriques.

Dans l'apprentissage non-supervisé, il est essentiel que chaque classe obtenue représente une variable géophysique ainsi que ses caractéristiques. Les classes de pluie identifiées montrent la variabilité intra-annuelles des précipitations régionales. La segmentation préserve également les structures de précipitations telles que les cyclones. À plus grande échelle, les classes de pluie montrent les zones de convergence intertropicale (ITCZ), tandis que les classes sur les océans peuvent représenter la saisonnalité de la SST.

Une approche non-supervisée exige une évaluation et une validation. Le travail substantiel réalisé dans cette thèse confirme que la segmentation non-supervisée de TB permet une représentation valide des variables géophysiques, ce qui est une étape critique avant de tirer des conclusions sur les changements dans le cycle de l'eau.

Title : Unsupervised deep learning for the analysis of remote sensing chronological series : Application to GPM

Keywords : precipitation, climate, remote sensing, unsupervised learning, deep learning, image segmentation

Abstract : The water cycle has a significant role in all life on Earth. Observing and quantifying its various components under the effects of climate change has become increasingly important to society and to advancing scientific knowledge. Responses of each component of the water cycle to the warming climate differ highly from one region to another. For instance, some regions have seen an increase in precipitation, while some have experienced a decrease. Consequently, there is an urgent need for global observation of the water cycle.

Remote sensing has amassed an unprecedented amount of global observation data. Regarding precipitation, the Global Precipitation Measurement (GPM) mission (2014-present) offers a constellation of satellites with microwave radiometers operating at a similar range of frequency, resulting in a large microwave brightness temperature (TB) dataset from different instruments. The TB contains implicit information on several important geophysical variables, including rain intensity, atmospheric water vapor, sea surface temperature (SST), snowfall, and cloud liquid water.

This study experiments on the TB measurement from the GPM Microwave Imager (GMI) of the Core Observatory, which is the reference satellite for unifying the observations of the constellation. We propose a novel method for observing the water cycle from TB measurements by considering them as unconventional images. Unsupervised semantic segmentation is applied to GMI data to extract information without the help of any labeling or ancillary data. Consequently, an unsupervised approach can be applied to any microwave radiometers. It is also excluded from the uncertainties involved in obtaining the ancillary data.

Unsupervised segmentation models of different levels of complexity are explored in this study. The non-pixel-wise approaches rely on fully convolutional networks (FCN) to extract feature maps from

the input. The simplest FCN-based model consists of several convolutional layers as its feature extractor. To increase the complexity, we deepen the network by using U-Net as a feature extractor. The final FCN model comprises an encoder-decoder network with an attention mechanism. Lastly, the pixel-wise approach K-means is tested.

To compare the performance of the models, this study proposes an external evaluation where relevant geophysical variables are used for the calculation of segmentation metrics. First, the classes are labeled using geophysical variables. Only classes representing precipitation and SST are used for further analysis. By converting these quantitative measurements to categorical data, they can be compared to the segmentation using traditional metrics, such as accuracy, precision, recall, and F1-score.

One of the main challenges in the unsupervised approach is the validation of the resulting segmentation. Without any supervision, it is crucial that each class obtained from the segmentation represents geophysical variables and exhibits their characteristics. Identified rain classes are shown to be able to follow intra-annual regional precipitation patterns. The segmentation also preserves precipitation structures such as cyclones. On a larger scale, the rain classes show the Inter-tropical Convergence Zones (ITCZ), while the ocean classes can represent the seasonality of SST.

Using an unsupervised deep-learning approach requires thorough evaluation and validation regarding geophysical representation. The substantial work in this thesis confirms that the unsupervised segmentation of TB results in meaningful representation of geophysical variables, which is a critical step before drawing any conclusion on the changes in the water cycle.

Acknowledgements

My favorite Khmer proverb goes, "If you want to go fast, go alone. If you want to go far, go as a team". The work that I was able to achieve is nothing short of a testament to this saying. I would like to show my gratitude to the people I have met on this journey and those whose support got me this far.

First and foremost, I would like to express my deepest appreciation to my four thesis supervisors, Cécile Mallet, Nicolas Viltard, Laurent Barthès, and Audrey Martini. I am grateful for their rigorous academic guidance. Their unwavering support and belief in my academic capability allowed me to overcome any challenges that I came across in my thesis work. I am wholeheartedly grateful to have supervisors who always make sure to be available to help and advise.

I would like to thank my thesis community members, Fatima Kabou, Pierre Lepetit, and Marjolaine Chiriaco, for monitoring the progress of my thesis during the last three years and for their constructive feedback during the thesis committee meeting. I would also like to extend my gratitude toward Aymeric Chazottes for his advice regarding machine learning algorithms.

I am extremely grateful to the reviewers who evaluated my thesis work, Claire Monteleoni and Christopher Kidd. Their feedback allows me to improve the quality of my manuscript significantly. I am also grateful to my other jury members, Valérie Ciarletti, Philippe Chambon, and Shruti Upadhyaya, whose questions led me to have a deeper reflection on my thesis.

I am also thankful for the AMX scholarship of Ecole Polytechnique. This endeavor would not have been possible without their financial support.

I would like to extend my sincere thanks to Natanaël Dubois-Quilici and Nafissa Benali, who participated in this thesis work during their Master's internship. Their hard work allowed for insightful discussions and for my thesis to progress faster.

During this journey, I had the pleasure of making many new friends whose moral support got me through the most stressful of days and who made working in LATMOS such an amazing experience. Thank you, Koyena Das, Yanjun Wang, Wei Chen, Dan Li, Qiuyu Xu, Baptiste Guigal, Maurin Zouzoua, Susana Jorquera, Filipe Toledo, and Dunya Alraddawi, for many delightful discussions. I am especially grateful to have my "CROUS Fangirls" whose company never failed to make my day a little better.

Lastly, I would like to acknowledge the support I have from my parents, Vibol Song and Sambath Hang, and my sister, Rotanak, whose unconditional belief in my academic pursuit has always uplifted my morale. I am also lucky to have the support of Vathana Thy, who has always been by my side and has never failed to cheer me up.

Table des matières

Introduction	5
1 The water cycle under climate change	9
1.1 Observed changes in the water cycle	10
1.1.1 Precipitation	12
1.1.2 Snowfall and snow cover	17
1.1.3 Droughts	17
1.1.4 Floods	18
1.1.5 Sea Surface Temperature (SST)	19
1.2 Methods of observing the evolution of the water cycle	20
1.2.1 Challenges in observing geophysical variables	20
1.2.2 Deep Learning approaches for remote sensing data	22
1.3 References	24
2 A Supervised approach for observing precipitation with remote sensing data	31
2.1 Microwave Brightness temperature (TB)	32
2.2 Deep-learning RAIN (DRAIN)	35
2.3 References	48
3 Segmenting image without labels	51
3.1 Computer Vision Toolbox	53
3.1.1 Convolutional Neural Networks (CNN)	53
3.1.2 Auto-regressive model	55
3.1.3 Transformer	55
3.2 Unsupervised image segmentation	59
3.2.1 Pixel-wise approach	59
3.2.2 Non-pixel-wise approach	62
3.2.2.1 Pseudo-labelling	63

3.2.2.2	Reconstruction of input images	66
3.2.2.3	Comparing and contrasting images within the dataset	68
3.3	Self-supervised image segmentation	69
3.3.1	Self-supervised Vision Transformer	70
3.3.2	Self-supervised CNN	72
3.4	Summary of Methods	73
3.5	Evaluation Metrics for Unsupervised Image Segmentation	75
3.5.1	Internal Evaluation	75
3.5.2	External evaluation	76
3.5.3	Spatial Correlation	78
3.6	References	80
4	Training, Comparing, and Interpreting	85
4.1	Training dataset	87
4.2	Model architectures	89
4.2.1	FCN models proposed by Kim et al. (2020)	90
4.2.2	Our proposed FCN	91
4.2.3	K-means models	93
4.3	Comparison of segmentation models without ground truth	94
4.3.1	External comparison with precipitation	95
4.3.2	External comparison with sea surface temperatures	102
4.3.3	Model choice	103
4.4	Validating the characteristics of the segmentation classes with geophysical knowledge	104
4.4.1	Precipitation patterns over two regions	104
4.4.1.1	Bay of Bengal (BOB) and the Indian Summer Monsoon (ISM)	105
4.4.1.2	Western Africa (WAF)	109
4.4.2	Precipitation over the ocean and the ITCZ precipitation band	111
4.4.3	Sea surface temperature (SST)	114
4.5	References	118
5	Analysis of GMI data from 2014 to 2024	123

5.1	Trends in brightness temperature	123
5.2	Trend in DRAIN precipitation	126
5.2.1	Rain intensity	126
5.2.2	Total rain pixels	128
5.3	Trend in the unsupervised segmentation results	130
5.3.1	Regional analysis of precipitation	130
5.3.2	Global analysis of rain classes	138
5.3.3	SST	138
5.4	References	141
6	Unsupervised domain adaptation	143
6.1	Unsupervised domain adaptation as a transfer learning approach for DRAIN	144
6.2	References	161
	Conclusion and Perspectives	163
A	Appendix : Feasibility study on unsupervised domain adaptation	169
B	Appendix : Résumé de Thèse en français	179
B.1	Approche supervisée : DRAIN	180
B.2	Approche non-supervisée	181
B.2.1	Les méthodes non-supervisées	181
B.2.2	La comparaison des modèles	183
B.2.3	Interprétation	185
B.3	Application sur une série chronologique de GPM	188
B.4	Adaptation de domaine	189
B.5	Conclusion et perspectives	191
C	Appendix : Acronyms and abbreviations	193
	List of Figures	200
	List of tables	201

Introduction

The water cycle is a crucial life-sustaining system that can vary naturally. However, increasing evidence shows that anthropogenic factors have contributed to an accelerated change over the last 50 years. According to the Sixth Assessment Report (AR6) by the Intergovernmental Panel on Climate Change (IPCC), billions of people now live in unfamiliar conditions in terms of the water cycle (Douveille et al., 2023). Changes in the water cycle include an increase in heavy precipitation, floods, and droughts, which could threaten food security and induce water scarcity. In addition, these conditions pose a threat to the global biodiversity. Observing and quantifying the changes in the water cycle is, therefore, of utmost importance.

Observing the water cycle means measuring its components. In addition to in-situ measurement, remote sensing has since created a large amount of data, especially image data. This thesis makes use of a satellite measurement called passive microwave brightness temperatures. It is one of the major indirect observations of the water cycle, which can be used to retrieve many variables, for instance, rain intensity, atmospheric water vapor, cloud liquid water, and surface wind speed. After the launch of many missions with a passive microwave aboard their satellite, such as the Tropical Rain Measuring Mission (TRMM, 1997-2015) (Kummerow et al., 1998) and the Global Precipitation Measurement Mission (GPM, 2014 to present) (Hou et al., 2014), there are multi-decades worth of satellite images containing important information about the evolution of the global water cycle.

This thesis aims to offer a novel approach to observing the water cycle and its evolution. With the ever-growing potential of using Artificial Intelligence to solve the big data problem in remote sensing, unsupervised image semantic segmentation is used to extract information about the water cycle directly from microwave brightness temperature images. The first advantage of an unsupervised approach is that it does not require any ground truth or ancillary data. Therefore, it can be applied directly to a large image dataset. This method avoids any uncertainties involved in the retrieval methods of the ancillary data and the annotations. In addition, the resulting classes from the segmentation can potentially represent at least one geophysical variable.

One of the major challenges tackled in this thesis is the validation of the classes. Without any supervision, it is essential that the segmentation classes are meaningful in their representation of geophysical variables. First, the classes are labeled using a particular variable, for example, precipitation or sea surface temperature. Then, they are compared to this variable using a reference dataset with traditional metrics used in segmentation. In order to further consolidate the segmentation results, the labeled classes are used in several case studies, such as tracking the monsoon

arrival and departure, identifying the Inter-tropical Convergence Zone (ITCZ), and regrouping isothermic sea surface temperatures.

Although the main body of work is the unsupervised approach, this thesis also includes a supervised deep-learning approach for rain intensity retrieval from TB data. In addition, my work also includes a transfer learning step, which allows machine-learning algorithms trained on one satellite to be applicable on another satellite.

In this manuscript, I described these three approaches for extracting information from TB that I explored during my thesis work. **Chapter 2** describes the supervised approach in retrieving rain intensity using TB and rain rates. **Chapter 4** details the main body of work, which is the unsupervised segmentation of TB. Finally, **Chapter 6** is about a transfer learning approach using unsupervised domain adaptation, which is necessary for the application of machine learning algorithms trained on GPM to other satellites. The content of each chapter is summarized as follows.

— **Chapter 1 The Water Cycle under Climate Change**

The observed and projected changes in the water cycle differ from one region to another. This chapter aims to report and summarize these findings with an emphasis on remote sensing and brightness temperature data. The end of the chapter presents some challenges in observing the water cycle as well as the motivation for the approach proposed in this thesis.

— **Chapter 2 A Supervised Approach in Observing Precipitation with Remote Sensing Data**

Before going into the unsupervised segmentation approach, I present a supervised deep-learning approach for retrieving rain rates from brightness temperatures. Using a precipitation radar for supervision, precipitation is retrieved from the brightness temperature images by a deep learning model (Deep Rain, DRAIN). This chapter includes a description of TB data and an article detailing the DRAIN algorithm and its evaluation.

— **Chapter 3 Segmenting Images without Labels**

Unsupervised learning of images is an emerging field in machine learning. Due to a lack of recent review articles on the topic of unsupervised semantic segmentation, Chapter 4 contributes to the categorizations of the methods in this domain and provides several example algorithms. The chapter also includes evaluation metrics for the unsupervised image segmentation.

— **Chapter 4 Training, Validating, and Interpreting**

This chapter presents the implementation of the unsupervised approach and the analysis of the result. It includes training the models, comparing the models, and interpreting the resulting segmentations. As the models are trained without any supervision, it is essential to verify that the segmenta-

tions deliver a meaningful representation of the geophysical variables. This is done by using the segmentations on simple application cases such as tracking the monsoons, identifying the ITCZ, and segmenting isothermic sea surface temperatures.

— **Chapter 5 Time Series Analysis**

This chapter presents the trend analysis of the GPM Core Observatory data from 2014 to 2024. First, the trend in the brightness temperature measurement is explored. Then, the supervised approach (DRAIN) and the unsupervised segmentation approach are applied to the brightness temperature data to analyze the trends in the resulting products.

— **Chapter 6 Unsupervised Domain Adaptation**

This chapter presents a domain adaptation approach for brightness temperature images. In this thesis, only the data from GPM Core Observatory is used. The GPM mission consists of a constellation of satellites that are essential for a better observation of the global water cycle. The satellites in the constellation may differ slightly in their configurations, resulting in differences in the brightness temperature dataset. Although unsupervised methods can directly be applied to any satellite, a transfer learning approach is necessary to harmonize the segmentation from different satellites. One possible solution is to use domain adaptation. A feasibility study on an unsupervised domain adaptation approach is explored to transform measurements from similar instruments to the domain of the GPM Core Observatory for DRAIN. For a supervised approach such as DRAIN, the transfer learning step is crucial. This chapter presents a paper on a domain adaptation technique tested for DRAIN.

References

- H. Douville, K. Raghavan, J. Renwick, R. P. Allan, P. A. Arias, M. Barlow, R. Cerezo-Mota, A. Cherchi, T. Y. Gan, J. Gergis, et al. *Water Cycle Changes*, page 1055–1210. Cambridge University Press, 2023.
- A. Y. Hou, R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi. The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 95(5) :701 – 722, 2014. doi : 10.1175/BAMS-D-13-00164.1. URL <https://journals.ametsoc.org/view/journals/bams/95/5/bams-d-13-00164.1.xml>.
- C. Kummerow, W. Barnes, T. Kozu, J. Shiue, and J. Simpson. The tropical rainfall measuring mission (trmm) sensor package. *Journal of Atmospheric and Oceanic Technology*, 15(3) :809 – 817, 1998. doi : 10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atot/15/3/1520-0426_1998_015_0809_ttrmmt_2_0_co_2.xml.

1 - The water cycle under climate change

Contents

1.1	Observed changes in the water cycle	10
1.1.1	Precipitation	12
1.1.2	Snowfall and snow cover	17
1.1.3	Droughts	17
1.1.4	Floods	18
1.1.5	Sea Surface Temperature (SST)	19
1.2	Methods of observing the evolution of the water cycle	20
1.2.1	Challenges in observing geophysical variables	20
1.2.2	Deep Learning approaches for remote sensing data	22
1.3	References	24

The water cycle is a term that covers all of the movement of water in the climate system in its liquid, gaseous, and solid form through the Earth’s surface, atmosphere, and sub-surfaces (Douville et al., 2023). Figure 1.1 shows the various components of the water cycle and its fluxes, which include water vapor, liquid water, evapotranspiration, soil moisture, vegetation, snow cover, groundwater, runoff, and the oceans (Oki et al., 1999). The primary drivers of the global water cycle are evaporation and precipitation (Douville et al., 2023). Rainfall replenishes the stock of water in rivers, soil moisture, and groundwater, which, in turn, depletes through evaporation or evapotranspiration.

The IPCC Working Group I Assessment Report 6 (AR6) reports with high confidence that the intensification of the water cycle worsens current water-related risks and vulnerability around the world (Douville et al., 2023). The overall consensus is that the increase in moisture due to the rising temperatures will result in more or more intense rainfall in wet regions. At the same time, the increase in land temperature leads to more severe droughts. Quantifying the changes happening to the water cycle is essential in mitigating extreme events such as storms, floods, and drought.

This chapter describes the recent changes in the water cycle and the methods to observe these findings. As this study is limited to microwave brightness temperature data, the scope of the water cycle elements presented here focuses mostly on

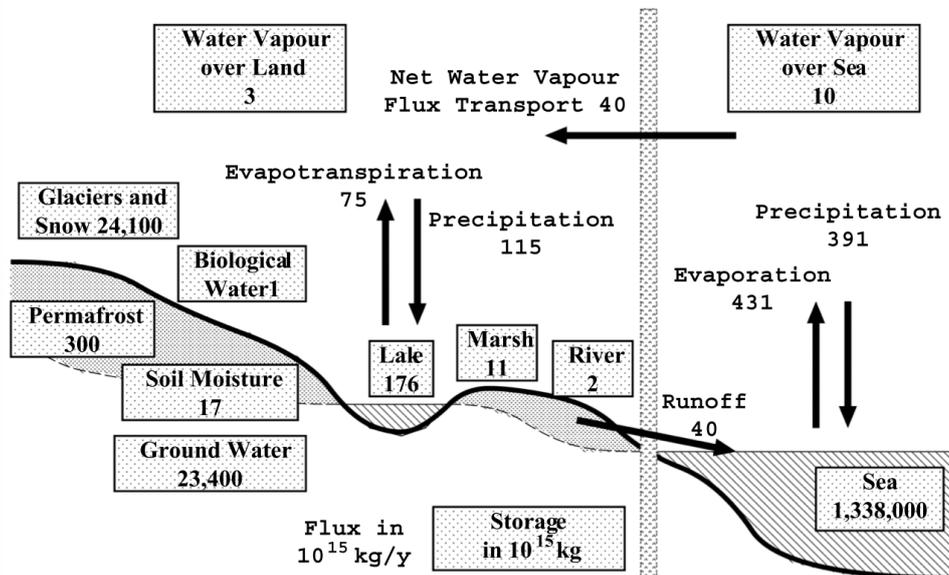


Figure 1.1 – The water cycle elements with storages and fluxes (Oki et al., 1999). The boxes represent the storage, while the arrows represent the fluxes. The values are estimated using data from 1989-1992.

the observations retrievable from the brightness temperature (TB) data, such as precipitation, snowfall, and sea surface temperature (SST).

1.1 . Observed changes in the water cycle

Although the water cycle can vary due to natural variations, there is increasing evidence that human activities are the dominating cause of the current changes to the water cycle (Allan et al., 2020). On a regional level, the aerosol has an important role in modifying the precipitation. The aerosol amount varies geographically, influencing local atmospheric circulation, which, in turn, can modify the local precipitation patterns (Caretta et al., 2022). This effect could be observed in the Indian Summer Monsoon (Singh et al., 2019). Urbanization and land use increase risks of runoff and flood as they cause a decrease in surface permeability (Caretta et al., 2022). In addition, urbanization can reduce rainfall by decreasing evaporation while increasing the risk of extreme rainfall by increasing sensible heat flux. Agriculture and irrigation can further influence local precipitation by modifying the moisture transport through (Caretta et al., 2022; Alter et al., 2015). Furthermore, they can reduce river flow and groundwater levels.

Simultaneously, river flow is dependent on snowmelt or glaciers, which are influenced by snowfall. Rising temperatures can decrease snowfall and shorten snow

seasons, which inevitably reduces the amount of water in the rivers (Caretta et al., 2022). Additionally, an increase in sea surface temperatures can lead to increasing evaporation, affecting the precipitation (Alexander et al., 2009; van der Ent and Savenije, 2013).

The key observations essential to quantifying these changes in the water cycle include water vapor, precipitation, evapotranspiration, snow, runoff, streamflow, aridity, droughts, and freshwater reservoirs (Douville et al., 2023). In addition to these variables, it is crucial to study the changes in the characteristics of large-scale phenomena such as the Inter-tropical Convergence Zones (ITCZ), the Hadley cell circulation, the Walker circulation, Monsoons, tropical cyclones, extratropical cyclones, storm tracks, atmospheric blocking, and atmospheric rivers (Douville et al., 2023).

Changes in precipitation patterns can result in further changes in other water cycle elements connected to rain. Rain has a role in transporting heat between the surface and the atmosphere (Trenberth, 2011). It can also influence the reflectivity of the Earth's surface. Thus, rain plays a crucial role in the Earth's energy balance. Furthermore, modification in rainfall patterns can also lead to other changes, such as vegetation growth (Chen et al., 2020), sea surface temperature, and ocean salinity surface (Katsaros and Buettner, 1969).

With high confidence, the AR6 reports a thermodynamic increase in moisture to the weather system, affecting the global and regional precipitation (Douville et al., 2023). As climate change leads to modification of the global and local heating and cooling patterns, it is inevitable that there will be changes in precipitation in terms of intensity, type, duration, and frequency (Pendergrass and Hartmann, 2014).

In terms of large-scale circulation, the AR6 projects multiple effects with low to medium confidence (Douville et al., 2023). The effect of the poleward shift of the Hadley cells, the regional shift of the ITCZ, and the poleward shift of the storm tracks in the Northern Hemisphere are projected with low confidence. With medium confidence are the projections about the Walker circulation weakening, the strengthening of the ITCZ core, and the poleward movement of the storm track in the Southern Hemisphere.

This section aims to further discuss recent findings regarding the changes in several water cycle components. The first subsection is about precipitation, one of the most important drivers of the water cycle. Multiple aspects of precipitation changes, such as monsoons, the ITCZ, and tropical cyclones, are examined. Next, we look at how snowfall and snow cover respond to global warming. Then, recent changes in extreme hydrological events, including droughts and floods, are briefly discussed. The last subsection describes sea surface surface temperature and its impact on the water cycle.

1.1.1 . Precipitation

Climate change is likely to affect precipitation amount and characteristics in terms of intensity, frequency, and type (Trenberth, 2011; Pendergrass and Hartmann, 2014). The characteristics of precipitation significantly impact society as they control water resources, agriculture, infrastructure, and the economy. For example, light to moderate rainfall over a long period can benefit agriculture, but heavy rain over a short time can lead to flood (Trenberth, 2011).

The warmer condition causes more evaporation. As the weather system tends to gather moisture over a large region, precipitation can intensify. However, there is only a slight increase (or a much lower increase rate) in the total amount of rainfall, which could mean a decrease in the duration and frequency of precipitation (Trenberth, 2011). In contrast to the trend in temperature, the change in precipitation is very spatially heterogeneous (Donat et al., 2013). According to the IPCC Sixth Assessment Report (AR6) from Working Group I and Working Group II, in most areas, the annual mean in precipitation has increased, while a few regions have seen a decrease (Douville et al., 2023; Caretta et al., 2022). Some of the findings reported include :

- The tropical land areas experienced the patterns of "wet convergent regimes get wetter" and "dry divergent regimes get drier." Since 1979, the convergence zone of the atmospheric circulation has received an increase in precipitation amount, while the descending branch has received a decreased amount (Douville et al., 2023).
- A decrease in rainfall amount is observed in southern Australia, South America (southern Andes), and eastern Africa. A drying effect is observed in dry summer climates such as the Mediterranean, western North America, southwestern South America, southwestern Australia, and South Africa (Caretta et al., 2022).
- An increase in rainfall amount is observed in northern Australia, southeastern South America, central and north-eastern North America.
- The intensity of heavy precipitation has increased in some regions, including the majority of North America, Europe, the Indian Subcontinent, and some parts of Asia (northern and southeastern regions), Africa (southern regions), and Australia (central, northern, and western regions) (Caretta et al., 2022).
- Heavy precipitation has decreased in several regions, including South America, Australia (eastern regions), and Africa (eastern regions) (Caretta et al., 2022).

Figure 1.2 describes visually the trends described above. For the trend between 1901 and 1984, the data used is from the Global Precipitation Climatology Center (GPCC). The GPCC is a gauge-only dataset. The errors in GPCC products could

be due to measurement errors (evaporation from gauge or strong wind near gauge) and sampling errors due to a sparse network of gauges (Schneider et al., 2008).

For the trend between 1985 and 2014, the data used is the Global Precipitation Climatology Project (GPCP). The GPCP combines rain gauge analysis from GPCC (gauge only) and satellite-based products, including passive microwave radiometers and infrared satellites (Adler et al., 2018). The quality of the GPCP Monthly analysis is affected by the differences in sensors and their estimates of precipitation. Better adjustment between sensors can lead to better estimates over the ocean, while better rain gauge analysis can improve precipitation over land.

Monsoons

Monsoons are important precipitation patterns upon which rely up to two-thirds of the world's population for water resources (Wang et al., 2020). Anthropogenic forcing resulting from greenhouse gas, aerosol, and land use could influence the global and regional monsoon (Seth et al., 2019). Understanding and quantifying the changes in monsoons is an essential step in adapting and mitigating water scarcity and natural disasters.

South, Southeast, and East Asian Monsoon

The Asian monsoon is one of the most important rainfall distributions upon which up to 60% of the world's population and many of the most vulnerable biodiversity zones depend (Buckley et al., 2014). The Asian monsoon comprises three sub-systems : the Indian Summer Monsoon (ISM), the East Asian Monsoon, and the Western North Pacific Summer Monsoon (Buckley et al., 2014). Although these three sub-systems are not explicitly correlated to one another, their interaction can modify the rainfall received over their respective regions.

Over the past 50 years, there is evidence that the Indian Summer Monsoon (ISM) over the Indian Subcontinent has seen a decrease in rainfall (Roxy et al., 2015; Samanta et al., 2020). The strength of the monsoon is strongly influenced by the temperature contrast between the Indian Subcontinent and the tropical Indian Ocean (Roxy et al., 2015; Jin and Wang, 2017). The ISM can be weakened by the increase in sea surface temperature (SST) in the tropical Indian Ocean and the western and central-eastern tropical Pacific (Singh et al., 2019). Warmer SST directly affects meridional tropospheric temperature contrast, which changes the amount and location of the rainfall. Regarding SST, El Niño-Southern Oscillation (ENSO) plays a significant role in controlling the annual precipitation in the monsoon as it contributes to the warming of the Pacific Ocean (Singh et al., 2019). More frequent El Niño in recent years could result in less precipitation over the Indian Subcontinent. Although rising SST temperatures weaken the monsoon, there is a possible revival caused by the temperature increase over the subcontinent, which reduces the contrast (Jin and Wang, 2017). However, the competition bet-

Trend in annual mean Precipitation

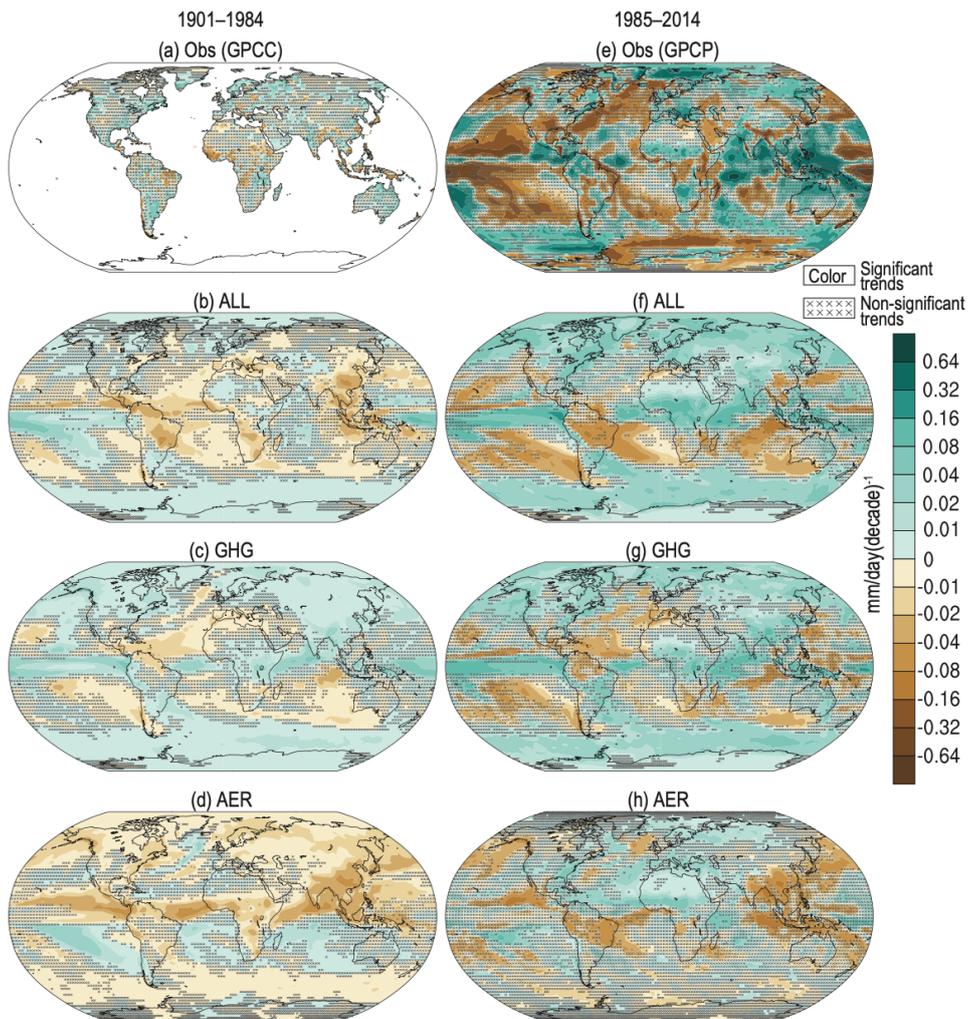


Figure 1.2 – Linear trends in annual mean precipitation per decade (mm/day per decade) based on observational dataset (a,e), CMIP6 multi-model ensemble simulations driven by : all radiative forcings (b,f), GHG-only (c,g), and aerosol only (d,h). The cross over each grid represents a non-significant trend (Douville et al., 2023)

ween the continent temperature and the SST would only end in disaster as hotter temperatures over the Indian Subcontinent lead to heat waves and flash floods, while warmer ocean brings more droughts (Roxy et al., 2015). At this stage, there are many debates on whether the ISM will weaken or strengthen in the future.

East Asian monsoon brings roughly 60% of the total rainfall received in Eastern China (Day et al., 2018). From 1951 to 2007, there is evidence of "South Flood–North Drought" over China due to the change in the frontal precipitation system (Day et al., 2018; Zhang et al., 2020). The increase in rainfall in the Southern regions, which received more moderate and heavy rain, may have been caused by greenhouse gas forcings (Zhang et al., 2020). The Northern regions received less moderate and light rain, leading to decreased total rainfall, which is linked to aerosol emissions (Zhang et al., 2020).

Southeast Asia is surrounded by the ISM, the East Asian Monsoon, and the Western North Pacific Summer Monsoon, which interact and influence the precipitation regime (Buckley et al., 2014). While some regions have seen a decrease in precipitation trend, there is an increasing risk of flood (Loo et al., 2015). Observation data from 2001 to 2010 shows that the rainy season seemed to last longer, with an earlier start date and a later end date (Suepa et al., 2016). However, there was a decrease in total rainfall in this period.

The West African Monsoon

The West African monsoon is essential to water resources and agriculture in the Sudano-Sahelian region (Sultan and Janicot, 2003). The monsoon transports moisture and precipitation from the Atlantic Ocean to West Africa, varying from interannual to interdecadal scale (Quagraine et al., 2020). The West African region experienced a period of increased rainfall between 1950 and 1970, followed by an arid period between 1970 and 1980, and then a recovery with increased precipitation in the 1990s (Douville et al., 2023). In the Sahelian region, there has been an increase in extreme rainfall since the 1980s (Taylor et al., 2017; Zhang et al., 2017). The western and central Sahelian region has seen increased rainfall amounts and extreme events, while the eastern region experienced an opposite trend (Zhang et al., 2017). Along with the increase in extreme precipitation, Western Africa is susceptible to dry spells and localized droughts (Douville et al., 2023).

North and South American Monsoon

The North American monsoon has not shown any significant trend (Seth et al., 2019). With a lot of spatial variability, observation data between 1948-2010 shows a decrease in precipitation for June to September in central and southern Mexico and an increase in northwest Mexico and the southwestern USA (Seth et al., 2019).

The South American Monsoon is projected to have some onset delay with the increase of greenhouse gas (Douville et al., 2023). The anomalous low-level moisture flux reversal over central South America determines the South American monsoon onset and withdrawal (Thome Sena and Magnusdottir, 2020). The development and strength of the monsoon are heavily influenced by the evapotranspiration over the Amazon Rain Forest (Thome Sena and Magnusdottir, 2020). Accompanied by the delays, the wet season is projected to last longer, while the Amazon becomes dryer in all seasons (Thome Sena and Magnusdottir, 2020). Although observations are very sparse, data since the 1950s suggests that the Southeast regions of South America experienced an increasing trend in precipitation (Seth et al., 2019).

Australian and Maritime Continent Monsoon

The Australian monsoon is responsible for most of the freshwater supply in Northern Australia (Heidemann et al., 2023). On an inter-annual scale, the monsoon variability is controlled by oceanic processes such as El Niño-Southern Oscillation (ENSO) (Heidemann et al., 2023). On a multi-decadal scale, it is strongly influenced by the Interdecadal Pacific Oscillation (IPO). For long-term changes, some studies suggest, with some uncertainties, a relationship between anthropogenic aerosol and an increase in precipitation over Northern Australia. Observations suggest a change in short-duration rainfall events, called burst regime, have become more frequent and longer in duration (Dey et al., 2020).

The Maritime Continent rainfall is controlled by the Asian and Australian Monsoon (Douville et al., 2023). Observation data from 1981-2014 shows a trend of increasing precipitation over many regions in the Maritime Continent. However, for regions with highly variable intra-annual precipitation, like the Maritime continent, the time period for calculating the trend can strongly influence the result.

Inter-Tropical Convergence Zone (ITCZ)

Located at the ascending branch of the Hadley cell, there is abundant and intense rainfall in the Inter-Tropical Convergence Zone (ITCZ) (Byrne and Schneider, 2016). It is an important feature of the world's precipitation, as its characteristics, including width and position, determine the tropical rain belt (Douville et al., 2023). It also has a major influence on the monsoon dynamics and the global energy budget. Based on observational and reanalysis data between 1980 and 2014, there is a trend of narrowing and intensifying precipitation in the ITCZ (Lau and Tao, 2020). The more intense precipitation in the ITCZ is accompanied by more drying areas in both hemispheres (Lau and Tao, 2020; Wodzicki and Rapp, 2016; Byrne and Schneider, 2016). It is important to examine the regional response to the strengthening of the ITCZ to understand its global effect.

Tropical cyclones

Although tropical cyclones typically have the negative connotation of extreme rainfall and flood, they are also an important source of freshwater supply and moisture transport for the coastal area in East Asia (Douville et al., 2023). Due to a lack of data coverage, it is difficult to determine how the tropical cyclones have changed due to anthropogenic forcings. Some evidence, albeit very limited, suggests an increased rain intensity and a slower translation speed. There is, however, theory and modelization that agree on the reduction of tropical cyclones' translation speed.

1.1.2 . Snowfall and snow cover

Snowfall is an essential component of the global water cycle due to its important role in the cryosphere (Tamang et al., 2020). It refills the stock of fresh water in the form of snow in winter and melts in summer to replenish the water supply (Tamang et al., 2020). A warmer climate could lead to less snowfall and earlier snowmelt, which can cause a shortage of fresh water and less streamflow.

The snowfall response to global warming differs from one region to another. As the annual mean potential snowfall areas over land in the Northern Hemisphere decreases, the Tibetan Plateau as a whole receives more snow during winter and less snowfall during summer (Douville et al., 2023).

On a large scale, passive microwave remote sensing primarily contributes to monitoring the variation in snow depth (Dai et al., 2017; Tanniru and Ramsankaran, 2023). The brightness temperatures are responsible for observing important snowpack regions such as the Himalayas and the Qinghai-Tibetan Plateau (Dai et al., 2017; Tanniru and Ramsankaran, 2023). In addition to snow depth, brightness temperature at high frequency can be used to retrieve snowfall (Edel et al., 2019).

1.1.3 . Droughts

There are several types of droughts (Caretta et al., 2022; McVicar and Jupp, 1998). Periods of persistent low precipitation lead to meteorological droughts. A meteorological drought can persist over a long period of time, causing deficits in soil moisture, streamflow, and water storage, which results in a hydrological drought that reduces the water supply. Hydrological drought can cause shortages in drinking water and substantial economic damage. The level of social and economic loss defines a socioeconomic drought. Additionally, increased atmospheric evaporative demand increases plant water stress and causes agricultural and ecological drought. Agricultural drought threatens food production, whereas ecological drought increases wildfire risk.

Agricultural drought can be monitored through soil moisture as it is a good indicator of water and energy fluxes between the land surface and the atmosphere (Liu et al., 2019). Although soil moisture can be observed by in-situ measurement and land surface modeling, low-frequency passive microwave radiometer satellites provide unprecedented advantages as they offer coverage of global soil moisture, sensitivity to soil permittivity, and the possibility of measuring in all types of weather (Liu et al., 2019; Zeng et al., 2020). Satellite soil moisture observation shows a decreasing trend for the last two decades in several regions, including central and southern America, southern and eastern South America, central Eurasia, and large areas of southern China, most of which are arid and semi-arid (Liu et al., 2019).

There is a strong link between anthropogenic factors, such as emissions of greenhouse gases and aerosol, and the trend of aridity (Douville et al., 2023). Figure 1.3, extracted from the Sixth Assessment Report of the Working Group II, shows regions that are susceptible to hazardous droughts by evaluating precipitation deficits from the Global Precipitation Climatology Centre (GPCC) for 1901–2010 (Caretta et al., 2022). The report concluded that the risks of agricultural and ecological drought trends in the Mediterranean and western North America are substantial due to an increase in evaporative demand. In addition, there is high confidence the increase in meteorological drought in southwestern Australia can be attributed to anthropogenic causes (Caretta et al., 2022). However, there is only a medium confidence that recent severe droughts occurring in southern Africa and southwestern South America are caused by human influences.

1.1.4 . Floods

The link between flood and heavy precipitation is complex (Caretta et al., 2022). While the increase in heavy rainfall causes an increase in floods in some regions, it can also result in no change in flood observation due to factors such as soil moisture conditions, land cover change, cryospheric change, and quality of river system management. In regions where rain is rare but intense when it occurs, the ground becomes hard and does not absorb water quickly, resulting in more surface runoff and severe floods (Douville et al., 2023).

Figure 1.4 shows the historical and projected flood water depths using model simulations. In the past several decades, many regions have seen vastly different trends in the frequency and magnitude of river floods (Caretta et al., 2022). The Observation data between 1985 and 2015 based on in-situ and satellite data shows that flood events have increased 4-fold in the tropics and 2.5-fold in northern mid-latitudes.

Remote sensing data contributes to indirectly observing floods by measuring the root causes of floods (Douville et al., 2023). For example, one of the main factors causing floods is extreme precipitation. In the tropics, data from an ensemble

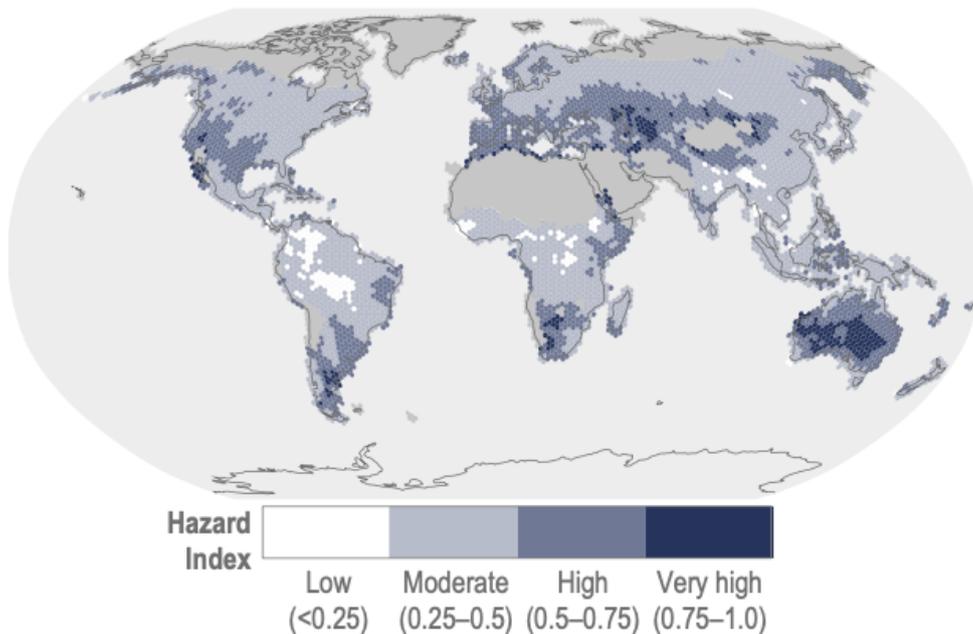


Figure 1.3 – Drought hazard between 1901 and 2020 estimated by the probability of surpassing the median of global severe precipitation deficits. The data for this estimation is the precipitation data from the GPCP data from 1901 to 2010 (Caretta et al., 2022).

of satellite products shows that extreme precipitation events are mainly caused by long-lived mesoscale convective systems (Roca and Fiolleau, 2020). Soil moisture is another important factor as it has a role in determining the location and timing of the initiation of the convective systems (Douville et al., 2023). For more direct observation, passive microwave satellites can map the flood as they can penetrate clouds, which usually cover the flood areas (De Groeve, 2010). However, monitoring through satellite observation can be limited by the satellite's revisit time and resolution.

1.1.5 . Sea Surface Temperature (SST)

The ocean is responsible for providing a large portion of the moisture necessary for continental precipitation (van der Ent and Savenije, 2013). SST variability, which affects evaporation, can significantly impact surface temperature and precipitation (Alexander et al., 2009; van der Ent and Savenije, 2013). In El Niño years, the global average of extreme temperatures is observed to be warmer (Alexander et al., 2009). However, for precipitation extremes, the response to SST is not as significant, as it is different from one region to another (Alexander et al., 2009). On a global level, there are more dry extremes during strong El Niño and more wet extremes during strong El Niña (Alexander et al., 2009).

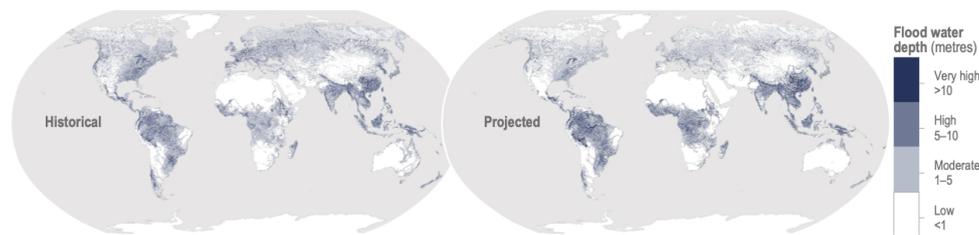


Figure 1.4 – Historical (1961-2005) and projected (2051-2070) fluvial flood depth from a combination of land surface model and inundation model in (meters). The simulation of the model is based on the reanalysis of the CMIP5 GCMs.

The rising SST is one of many important global warming indicators (Ruela et al., 2020). Model projections show that SST is increasing globally, with different intensities depending on the region (Ruela et al., 2020). In terms of observations, SST has been measured for at least 150 years through in-situ measurements such as ships and drifting buoys, and more recently through satellite observations (Alerskans et al., 2022). Satellite data offers a high-resolution and global average of SST observations (Garcia-Soto et al., 2021). Particularly, passive microwave satellites are capable of observing SST through non-precipitating clouds and small amounts of aerosol, whereas its Infrared counterpart is obstructed by aerosols and clouds (Alerskans et al., 2022).

The agreement between in-situ and remote sensing data has increased the confidence in observing the trend in SST (Garcia-Soto et al., 2021). Multiple observational datasets confirm that SST has been increasing since the 1900s, with the five warmest years recorded since 2014 (Garcia-Soto et al., 2021). The warming response is observed over most of the oceans, with the exception of the North Atlantic Ocean, which has experienced a cooling trend (Garcia-Soto et al., 2021).

1.2 . Methods of observing the evolution of the water cycle

Studying the evolution of the water cycle requires measuring its key components and understanding the changes in large-scale circulations. The major setback is the difficulty in measuring and retrieving these key components.

1.2.1 . Challenges in observing geophysical variables

Difficulties in obtaining the measurements

The first step in observing the water cycle is obtaining a high-quality dataset of its major drivers. For precipitation, the difficulty in obtaining observation is due to the fact that it has a very intermittent nature in space and time. In addition to its intermittent nature, there are differences between land and ocean in terms of

water availability and the relation between precipitation and temperature (Trenberth, 2011). Therefore, to study the daily mean accumulation or mean intensity of precipitation, it is essential to have global fine-scaled data in both times and space (Trenberth, 2011). In-situ data, such as rain gauges, are essential, but they cannot offer good enough coverage to acquire global homogeneous fine-scale datasets (Kidd et al., 2017). Ground-based instruments can also be quite lacking in some areas where the topography is complex, as well as over the ocean. They can also suffer from wind effects that make it challenging to observe light rain and snow. Additionally, there is also the issue of strict policies on data sharing in many countries (Donat et al., 2013). Remote sensing measurements (radar and space-based), on the other hand, only observe instantaneous rates but take measurements in a regular manner between the areas that the satellite overflies.

Besides precipitation, in-situ alone is also insufficient for observing other data that are less intermittent, such as SST, soil moisture, and snow cover. Satellite data is a good solution for obtaining a global dataset.

Differences among datasets

With many measurements, in-situ and remote sensing, combining these datasets for a better scale and coverage is a key process. However, it is very important that the data is homogeneous and free from inconsistency, which could show up as an outlier or extreme (Alexander et al., 2006). The mismatch of spatial and temporal scales in the dataset is a big obstacle in correctly assessing the evolution of the extremes. Furthermore, different datasets might present differences in retrieval algorithms and, therefore, differences in uncertainties and errors. In order to overcome these limitations, it is important that data sources are utilized with careful consideration (reanalyses, various satellites, radar, and remote sensing products).

For precipitation, the limitations concerning the data sets include (Alexander et al., 2006) :

- Differences between the precipitation datasets, including their representation of extremes.
- Critical gaps in the amount, quality, and consistency of the data.
- Uncertainties due to gridding methods causing problems between extreme values and scaling.
- Heterogeneous data due to, for example, when an observation station is moved.

Another approach to improving the dataset is to use reanalysis, which makes use of a wide array of historical data, remote sensing data, numerical models, and assimilation schemes (Tarek et al., 2020). Precipitation reanalysis data are limited by the lack of in-situ observations, differences between assimilated datasets, and errors in the numerical models (Tarek et al., 2020; Dorigo et al., 2021).

Combining multiple datasets is a persistent problem for many observations.

For example, in the study of drought trends, the soil moisture datasets present discrepancies due to differences in sensor design, retrieval algorithm, topology, and vegetation state (Liu et al., 2019).

Differences in definitions

The limitation to the study of temperature and precipitation extremes starts with their definitions (Alexander et al., 2006). Particularly, the inconsistency of spatial and temporal scales in the observations can impact the determination of extremes. For instance, the coordination between different meteorological services and the rigorous data quality control hinder the global gathering of homogeneous data needed for defining extremes.

1.2.2 . Deep Learning approaches for remote sensing data

Remote sensing has been used to observe many water cycle elements (Cui et al., 2018). Since the 1980s, various remote sensing missions have collected an extensive dataset of Earth observation. The large amount of data has opened up an unprecedented opportunity for the application of the emerging success of Machine Learning. Among the many types of remote sensing measurement, this thesis focuses on the Microwave Brightness Temperature (TB), which has been used to observe many water cycle elements, including precipitation, glaciers, snow, sea ice, and soil moisture. Therefore, a decade-long dataset of TB can be considered an important climate record containing insightful information about the global water cycle.

This thesis explores supervised and unsupervised approaches to retrieving information from TB. First, a supervised learning approach for retrieving rain intensity from TB data is described. Deploying this supervised approach to other satellites requires a transfer learning step due to differences in data distribution. A domain adaptation between two satellites is briefly explored in the last chapter of this thesis work.

As a complement to studying and observing each water cycle component, I propose an alternative approach where an unsupervised deep learning model extracts multiple information from satellite observations. Although there is a large dataset, these remote-sensing observations do not contain ground-truth labels. In this unsupervised approach, neither auxiliary variables nor data annotations are needed. As a result, it is also free from any uncertainties and mismatches involved in using auxiliary data. In addition, in methods where a geophysical variable is retrieved from TB, it is assumed that the relationship between TB and said variable is not modified under the condition of climate change. In the case where this hypothesis does not hold, the performance of these retrieval algorithms can be decreased. This issue does not affect the unsupervised approach where only TB is used.

The main objective of the present study is the application of unsupervised learning approaches to TB, which is a key measurement in the observation of the water cycle. Since the launch of the Tropical Rainfall Measuring Mission (TRMM) (Kummerow et al., 1998) and Global Precipitation Measurement (GPM) mission (Hou et al., 2014), there have been multiple decades of TB data collected by multiple satellites containing important information about the water cycle. Therefore, a wealth of information can be extracted from this data. The large amount of data is a great opportunity to apply a Deep Learning model.

In the proposed unsupervised approach, multiple channels of brightness temperature images are fed into an unsupervised image semantic segmentation model. This results in a labeling of each pixel of the image. Each obtained "segmentation class" represents one or a combination of geophysical variables. These classes are validated qualitatively and quantitatively before they can be used to study the evolution observed by the TB time series.

1.3 . References

- R. F. Adler, M. R. P. Sapiano, G. J. Huffman, J.-J. Wang, G. Gu, D. Bolvin, L. Chiu, U. Schneider, A. Becker, E. Nelkin, P. Xie, R. Ferraro, and D.-B. Shin. The global precipitation climatology project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, 9(4), 2018. ISSN 2073-4433. doi : 10.3390/atmos9040138. URL <https://www.mdpi.com/2073-4433/9/4/138>.
- E. Alerskans, A.-S. P. Zinck, P. Nielsen-Englyst, and J. L. Høyer. Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements. *Remote Sensing of Environment*, 281 :113220, 2022. ISSN 0034-4257. doi : <https://doi.org/10.1016/j.rse.2022.113220>. URL <https://www.sciencedirect.com/science/article/pii/S0034425722003261>.
- L. V. Alexander, X. Zhang, T. C. Peterson, J. Caesar, B. Gleason, A. M. G. Klein Tank, M. Haylock, D. Collins, B. Trewin, F. Rahimzadeh, A. Tagipour, K. Rupa Kumar, J. Revadekar, G. Griffiths, L. Vincent, D. B. Stephenson, J. Burn, E. Aguilar, M. Brunet, M. Taylor, M. New, P. Zhai, M. Rusticucci, and J. L. Vazquez-Aguirre. Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research : Atmospheres*, 111(D5), 2006. doi : <https://doi.org/10.1029/2005JD006290>.
- L. V. Alexander, P. Uotila, and N. Nicholls. Influence of sea surface temperature variability on global temperature and precipitation extremes. *Journal of Geophysical Research : Atmospheres*, 114(D18), 2009.
- R. P. Allan, M. Barlow, M. P. Byrne, A. Cherchi, H. Douville, H. J. Fowler, T. Y. Gan, A. G. Pendergrass, D. Rosenfeld, A. L. Swann, et al. Advances in understanding large-scale responses of the water cycle to climate change. *Annals of the New York Academy of Sciences*, 1472(1) :49–75, 2020.
- R. E. Alter, E.-S. Im, and E. A. Eltahir. Rainfall consistently enhanced around the Gezira Scheme in East Africa due to irrigation. *Nature Geoscience*, 8(10) : 763–767, 2015.
- B. M. Buckley, R. Fletcher, S.-Y. S. Wang, B. Zottoli, and C. Pottier. Monsoon extremes and society over the past millennium on mainland Southeast Asia. *Quaternary Science Reviews*, 95 :1–19, 2014. ISSN 0277-3791. doi : <https://doi.org/10.1016/j.quascirev.2014.04.022>. URL <https://www.sciencedirect.com/science/article/pii/S0277379114001462>.

- M. P. Byrne and T. Schneider. Energetic constraints on the width of the intertropical convergence zone. *Journal of Climate*, 29(13) :4709–4721, 2016.
- M. Caretta, A. Mukherji, M. Arfanuzzaman, R. Betts, A. Gelfan, Y. Hirabayashi, T. Lissner, J. Liu, E. L. Gunn, R. Morgan, S. Mwanga, and S. Supratid. Water. In : *Climate Change 2022 : Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. 2022. doi : 10.1017/9781009325844.006.
- Z. Chen, W. Wang, and J. Fu. Vegetation response to precipitation anomalies under different climatic and biogeographical conditions in China. *Scientific reports*, 10(1) :830, 2020.
- Y. Cui, X. Chen, J. Gao, B. Yan, G. Tang, and Y. Hong. Global water cycle and remote sensing big data : overview, challenge, and opportunities. *Big Earth Data*, 2(3) :282–297, 2018. doi : 10.1080/20964471.2018.1548052. URL <https://doi.org/10.1080/20964471.2018.1548052>.
- L. Dai, T. Che, Y. Ding, and X. Hao. Evaluation of snow cover and snow depth on the Qinghai–Tibetan Plateau derived from passive microwave remote sensing. *The Cryosphere*, 11(4) :1933–1948, 2017.
- J. A. Day, I. Fung, and W. Liu. Changing character of rainfall in eastern China, 1951–2007. *Proceedings of the National Academy of Sciences*, 115(9) :2016–2021, 2018.
- T. De Groeve. Flood monitoring and mapping using passive microwave remote sensing in Namibia. *Geomatics, Natural Hazards and Risk*, 1(1) :19–35, 2010.
- R. Dey, A. J. Gallant, and S. C. Lewis. Evidence of a continent-wide shift of episodic rainfall in Australia. *Weather and Climate Extremes*, 29 :100274, 2020. ISSN 2212-0947. doi : <https://doi.org/10.1016/j.wace.2020.100274>. URL <https://www.sciencedirect.com/science/article/pii/S2212094719302403>.
- M. G. Donat, L. V. Alexander, H. Yang, I. Durre, R. Vose, R. J. H. Dunn, K. M. Willett, E. Aguilar, M. Brunet, J. Caesar, B. Hewitson, C. Jack, A. M. G. Klein Tank, A. C. Kruger, J. Marengo, T. C. Peterson, M. Renom, C. Oria Rojas, M. Rusticucci, J. Salinger, A. S. Elayah, S. S. Sekele, A. K. Srivastava, B. Trewin, C. Villarreal, L. A. Vincent, P. Zhai, X. Zhang, and S. Kitching. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century : The HadEX2 dataset. *Journal of Geophysical Research : Atmospheres*, 118(5) :2098–2118, 2013. doi : <https://doi.org/10.1002/jgrd.50150>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50150>.

- W. Dorigo, S. Dietrich, F. Aires, L. Brocca, S. Carter, J.-F. Cretaux, D. Dunkerley, H. Enomoto, R. Forsberg, A. Güntner, et al. Closing the water cycle from observations across scales : Where do we stand? *Bulletin of the American Meteorological Society*, 102(10) :E1897–E1935, 2021.
- H. Douville, K. Raghavan, J. Renwick, R. P. Allan, P. A. Arias, M. Barlow, R. Cerezo-Mota, A. Cherchi, T. Y. Gan, J. Gergis, et al. *Water Cycle Changes*, page 1055–1210. Cambridge University Press, 2023.
- L. Edel, J.-F. Rysman, C. Claud, C. Palerme, and C. Genthon. Potential of passive microwave around 183 GHz for snowfall detection in the Arctic. *Remote Sensing*, 11(19) :2200, 2019.
- C. Garcia-Soto, L. Cheng, L. Caesar, S. Schmidtko, E. B. Jewett, A. Cheripka, I. Rigor, A. Caballero, S. Chiba, J. C. Báez, et al. An overview of ocean climate change indicators : Sea surface temperature, ocean heat content, ocean pH, dissolved oxygen concentration, arctic sea ice extent, thickness and volume, sea level and strength of the AMOC (Atlantic Meridional Overturning Circulation). *Frontiers in Marine Science*, 8 :642372, 2021.
- H. Heidemann, T. Cowan, B. J. Henley, J. Ribbe, M. Freund, and S. Power. Variability and long-term change in Australian monsoon rainfall : A review. *Wiley Interdisciplinary Reviews : Climate Change*, 14(3) :e823, 2023.
- A. Y. Hou, R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi. The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, 95(5) :701 – 722, 2014. doi : 10.1175/BAMS-D-13-00164.1. URL <https://journals.ametsoc.org/view/journals/bams/95/5/bams-d-13-00164.1.xml>.
- Q. Jin and C. Wang. A revival of Indian summer monsoon rainfall since 2002. *Nature Climate Change*, 7(8) :587–594, Aug. 2017. ISSN 1758-6798. doi : 10.1038/nclimate3348. URL <https://doi.org/10.1038/nclimate3348>.
- K. Katsaros and K. J. Buettner. Influence of rainfall on temperature and salinity of the ocean surface. *Journal of Applied Meteorology (1962-1982)*, pages 15–18, 1969.
- C. Kidd, A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum. So, how much of the earth’s surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, 98(1) :69–78, 2017.
- C. Kummerow, W. Barnes, T. Kozu, J. Shiue, and J. Simpson. The tropical rainfall measuring mission (TRMM) sensor package. *Journal of atmospheric and oceanic technology*, 15(3) :809–817, 1998.

- W. K. M. Lau and W. Tao. Precipitation–radiation–circulation feedback processes associated with structural changes of the ITCZ in a warming climate during 1980–2014 : An observational portrayal. *Journal of Climate*, 33(20) :8737 – 8749, 2020. doi : <https://doi.org/10.1175/JCLI-D-20-0068.1>. URL <https://journals.ametsoc.org/view/journals/clim/33/20/jcliD200068.xml>.
- Y. Liu, Y. Liu, and W. Wang. Inter-comparison of satellite-retrieved and Global Land Data Assimilation System-simulated soil moisture datasets for global drought analysis. *Remote Sensing of Environment*, 220 :1–18, 2019. ISSN 0034-4257. doi : <https://doi.org/10.1016/j.rse.2018.10.026>. URL <https://www.sciencedirect.com/science/article/pii/S0034425718304802>.
- Y. Y. Loo, L. Billa, and A. Singh. Effect of climate change on seasonal monsoon in Asia and its impact on the variability of monsoon rainfall in Southeast Asia. *Geoscience Frontiers*, 6(6) :817–823, 2015. ISSN 1674-9871. doi : <https://doi.org/10.1016/j.gsf.2014.02.009>. URL <https://www.sciencedirect.com/science/article/pii/S167498711400036X>. Special Issue : Geoinformation techniques in natural hazard modeling.
- T. R. McVicar and D. L. Jupp. The current and potential operational uses of remote sensing to aid decisions on drought exceptional circumstances in Australia : a review. *Agricultural Systems*, 57(3) :399–468, 1998. ISSN 0308-521X. doi : [https://doi.org/10.1016/S0308-521X\(98\)00026-2](https://doi.org/10.1016/S0308-521X(98)00026-2). URL <https://www.sciencedirect.com/science/article/pii/S0308521X98000262>. Drought Policy, Assessment and Declaration.
- T. Oki, D. Entekhabi, and T. I. Harrold. The global water cycle. *Global energy and water cycles*, 10 :27, 1999.
- A. G. Pendergrass and D. L. Hartmann. Two modes of change of the distribution of rain. *Journal of Climate*, 27(22) :8357 – 8371, 2014. doi : [10.1175/JCLI-D-14-00182.1](https://doi.org/10.1175/JCLI-D-14-00182.1). URL <https://journals.ametsoc.org/view/journals/clim/27/22/jcli-d-14-00182.1.xml>.
- K. A. Quagraine, F. Nkrumah, C. Klein, N. A. B. Klutse, and K. T. Quagraine. West African summer monsoon precipitation variability as represented by reanalysis datasets. *Climate*, 8(10) :111, 2020.
- R. Roca and T. Fiolleau. Extreme precipitation in the tropics is closely associated with long-lived convective systems. *Communications Earth & Environment*, 1(1) :18, 2020.
- M. K. Roxy, K. Ritika, P. Terray, R. Murtugudde, K. Ashok, and B. Goswami. Drying of Indian subcontinent by rapid Indian Ocean warming and a weakening land-sea thermal gradient. *Nature communications*, 6(1) :7423, 2015.

- R. Ruela, M. Sousa, M. deCastro, and J. Dias. Global and regional evolution of sea surface temperature under climate change. *Global and Planetary Change*, 190 : 103190, 2020. ISSN 0921-8181. doi : <https://doi.org/10.1016/j.gloplacha.2020.103190>. URL <https://www.sciencedirect.com/science/article/pii/S0921818120300813>.
- D. Samanta, B. Rajagopalan, K. B. Karnauskas, L. Zhang, and N. F. Goodkin. La niña's diminishing fingerprint on the central Indian summer monsoon. *Geophysical Research Letters*, 47(2) :e2019GL086237, 2020.
- U. Schneider, T. Fuchs, A. Meyer-Christoffer, and B. Rudolf. Global precipitation analysis products of the GPCP. *Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation*, 112 :3819–3837, 2008.
- A. Seth, A. Giannini, M. Rojas, S. A. Rauscher, S. Bordoni, D. Singh, and S. J. Camargo. Monsoon responses to climate changes—connecting past, present and future. *Current Climate Change Reports*, 5 :63–79, 2019.
- D. Singh, S. Ghosh, M. K. Roxy, and S. McDermid. Indian summer monsoon : Extreme events, historical changes, and role of anthropogenic forcings. *Wiley Interdisciplinary Reviews : Climate Change*, 10(2) :e571, 2019.
- T. Suepa, J. Qi, S. Lawawirojwong, and J. P. Messina. Understanding spatio-temporal variation of vegetation phenology and rainfall seasonality in the monsoon Southeast Asia. *Environmental Research*, 147 :621–629, 2016. ISSN 0013-9351. doi : <https://doi.org/10.1016/j.envres.2016.02.005>. URL <https://www.sciencedirect.com/science/article/pii/S0013935116300482>.
- B. Sultan and S. Janicot. The west african monsoon dynamics. Part II : The “preonset” and “onset” of the summer monsoon. *Journal of climate*, 16(21) : 3407–3427, 2003.
- S. K. Tamang, A. M. Ebtehaj, A. F. Prein, and A. J. Heymsfield. Linking global changes of snowfall and wet-bulb temperature. *Journal of Climate*, 33(1) : 39 – 59, 2020. doi : 10.1175/JCLI-D-19-0254.1. URL <https://journals.ametsoc.org/view/journals/clim/33/1/jcli-d-19-0254.1.xml>.
- S. Tanniru and R. Ramsankaran. Passive microwave remote sensing of snow depth : Techniques, challenges and future directions. *Remote Sensing*, 15(4), 2023. ISSN 2072-4292. doi : 10.3390/rs15041052. URL <https://www.mdpi.com/2072-4292/15/4/1052>.
- M. Tarek, F. P. Brissette, and R. Arsenault. Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, 24(5) :2527–2544, 2020.

- C. M. Taylor, D. Belušić, F. Guichard, D. J. Parker, T. Vischel, O. Bock, P. P. Harris, S. Janicot, C. Klein, and G. Panthou. Frequency of extreme Sahelian storms tripled since 1982 in satellite observations. *Nature*, 544(7651) :475–478, Apr. 2017. ISSN 1476-4687. doi : 10.1038/nature22069. URL <https://doi.org/10.1038/nature22069>.
- A. C. Thome Sena and G. Magnúsdóttir. Projected end-of-century changes in the South American Monsoon in the CESM large ensemble. *Journal of Climate*, 33(18) :7859–7874, 2020.
- K. E. Trenberth. Changes in precipitation with climate change. *Climate research*, 47(1-2) :123–138, 2011. doi : <https://doi.org/10.3354/cr00953>.
- R. J. van der Ent and H. H. Savenije. Oceanic sources of continental precipitation and the correlation with sea surface temperature. *Water Resources Research*, 49(7) :3993–4004, 2013.
- B. Wang, C. Jin, and J. Liu. Understanding future change of global monsoons projected by CMIP6 models. *Journal of Climate*, 33(15) :6471 – 6489, 2020. doi : 10.1175/JCLI-D-19-0993.1. URL <https://journals.ametsoc.org/view/journals/clim/33/15/JCLI-D-19-0993.1.xml>.
- K. Wodzicki and A. Rapp. Long-term characterization of the Pacific ITCZ using TRMM, GPCP, and ERA-Interim. *Journal of Geophysical Research : Atmospheres*, 121(7) :3153–3170, 2016.
- J. Zeng, K.-S. Chen, C. Cui, and X. Bai. A physically based soil moisture index from passive microwave brightness temperatures for soil moisture variation monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, 58(4) :2782–2795, 2020. doi : 10.1109/TGRS.2019.2955542.
- B. Zhang, B. Dong, and R. Jin. Forced decadal changes in summer precipitation characteristics over China : The roles of greenhouse gases and anthropogenic aerosols. *Journal of Meteorological Research*, 34(6) :1226–1241, 2020.
- W. Zhang, M. Brandt, F. Guichard, Q. Tian, and R. Fensholt. Using long-term daily satellite based rainfall data (1983–2015) to analyze spatio-temporal changes in the sahelian rainfall regime. *Journal of Hydrology*, 550 :427–440, 2017. ISSN 0022-1694. doi : <https://doi.org/10.1016/j.jhydrol.2017.05.033>. URL <https://www.sciencedirect.com/science/article/pii/S0022169417303220>.

2 - A Supervised approach for observing precipitation with remote sensing data

Contents

2.1	Microwave Brightness temperature (TB) . .	32
2.2	Deep-learning RAIN (DRAIN)	35
2.3	References	48

Precipitation is one of the components of the water cycle that can be retrieved from the microwave brightness temperature (TB) measures. The Tropical Rain Measuring Mission (TRMM) (1997-2015) (Kummerow et al., 1998), which aims to advance the knowledge of the distribution and variability of the precipitation in the tropics, is one of the major advancements in observing precipitation with a microwave radiometer. TRMM carried a microwave imager and a precipitation radar, which opened a new horizon for rain retrieval algorithms (Viltard et al., 2020).

The Global Precipitation Measurement (GPM) mission, the successor of TRMM, comprises a constellation of satellites, each equipped with a radiometer, with the aim of offering the global observation of precipitation. To unify the precipitation measurement in the constellation, the GPM constellation has a reference satellite called the GPM Core Observatory (GPM-CO). Aboard the GPM-CO, there is a radiometer (GPM Microwave Imager, GMI) and a precipitation radar (Dual-frequency Precipitation Radar, DPR).

This chapter presents a supervised Deep-learning Rain (DRAIN) retrieval approach for TB images. The supervised approach is possible due to the availability of more direct observations of precipitation offered by DPR to be used as targets during training. The colocated data of GMI and DPR is used for training DRAIN. The precipitation radar is not available for all satellites in the constellation, which renders the training DRAIN impossible for these satellites.

GPM Microwave Imager (GMI)

GMI is a microwave radiometer operating at multiple frequency channels ranging from 10 GHz to 183 GHz. The choice of channels is to optimize the detection of precipitation and the availability of similar frequencies on other satellites for the future possibility of transfer learning. Two channels are used as inputs to DRAIN : the 37 GHz (mostly sensitive to light to moderate precipitation over the ocean) and the 89 GHz (mostly sensitive to ice particles in precipitation events over land

and ocean) (Hou et al., 2014).

Dual-frequency Precipitation Radar (DPR)

DPR provides a more direct observation of precipitation. The sensitivity of the DPR for rain intensity is approximately 0.2 mm/h (Hou et al., 2014). Due to the difference in resolution between the DPR and the GMI, the pixels are collocated.

2.1 . Microwave Brightness temperature (TB)

A passive microwave radiometer measures the microwave radiant energy emitted from the Earth's surface and atmosphere. As a passive instrument, it is equipped with a rotating mirror that receives signals continuously. Then, the signal received is calculated as brightness temperature (TB) in Kelvin. In addition, the continuous signal is integrated over time to create discrete "pixels." A passive microwave radiometer observes the TB through multiple channels designed to be sensitive to different frequencies of microwave energy. Consequently, the observation of TB is a complex function depending on the scene that the radiometer is observing (vertical profile of the atmosphere and surface characteristics) and the channel that is used.

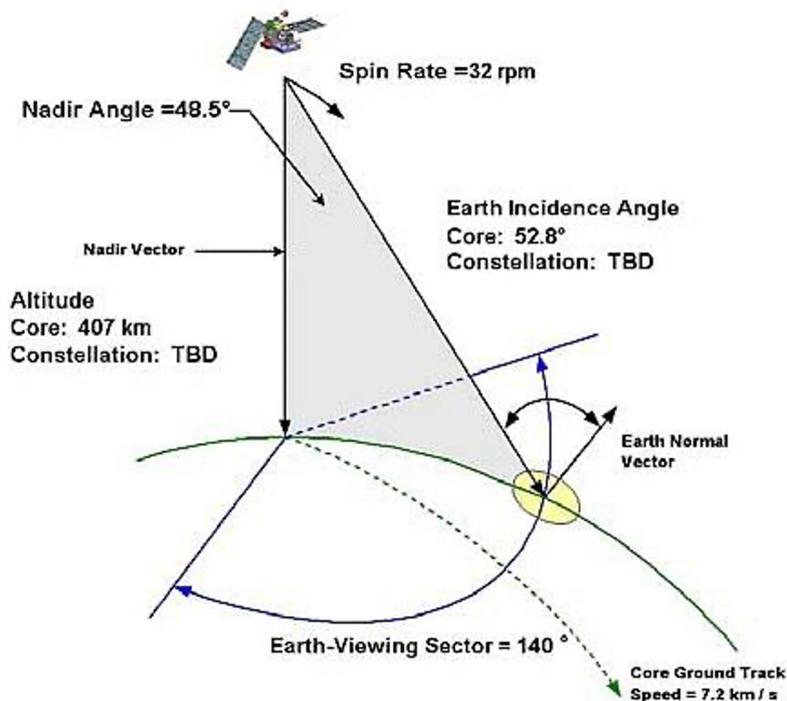
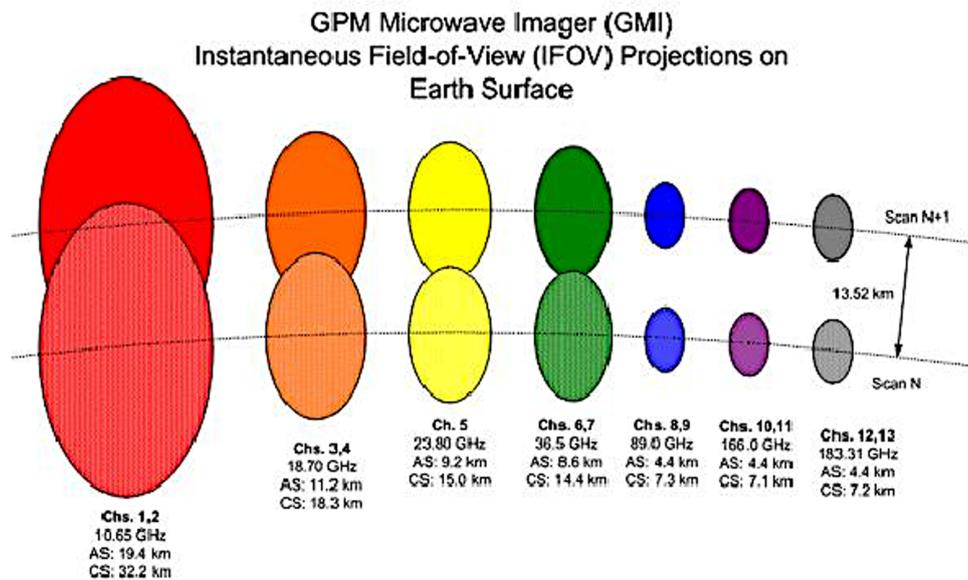


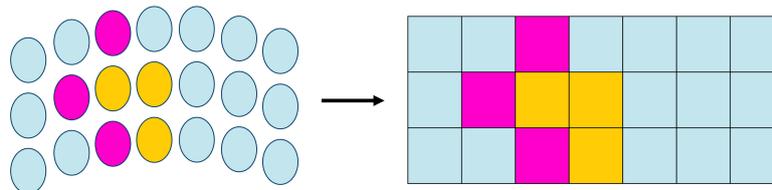
Figure 2.1 – Scanning geometry of GMI (NESDIS STAR, 2021)

Figure 2.1 shows the scanning geometry of the GMI. Scanning geometry of a satellite can determine the characteristics of the data it provides. For example,

the scanning geometry and the channel frequency determine the resulting TB pixel size. Figure 2.2 shows the pixel size for each channel frequency of the GMI. In a constellation of satellites, the differences in scanning geometry, in addition to differences in instrument configuration, can cause differences in the data distribution of the observed TB.



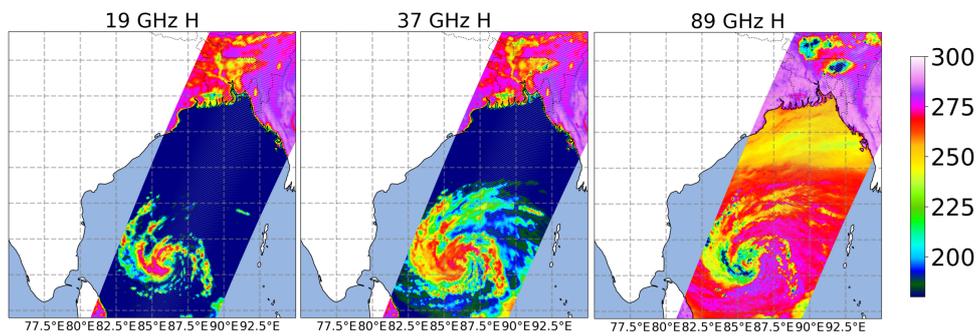
(a) The Instantaneous Field of View (IFOV) for each GMI channel along-scan direction (AS) and cross-scan direction (CS) (NESDIS STAR, 2021). The lower frequency has a bigger footprint than the higher frequency.



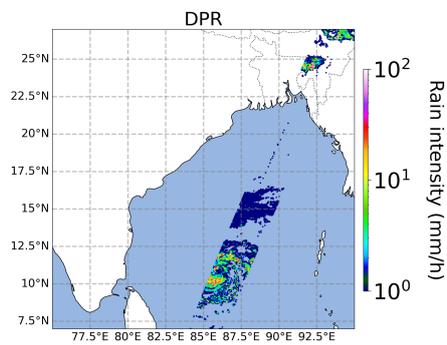
(b) The rows of pixels can be considered as a matrix and, therefore, an image in the conventional sense.

Figure 2.2 – GMI's pixel details.

Figure 2.3 shows an example of a precipitation event observed by the DPR (Figure 2.3b) and the GMI at 19 GHz, 37 GHz, and 89 GHz (Figure 2.3a). The value of TB observed depends on the observed precipitation event and the instrument frequency channel used. In this example, the precipitation zone observed using 19 and 37 appears warmer due to the emission of the water particles, while at 89 GHz, the raindrops appear colder due to the scattering of the ice particles. This precipitation event appears warmer than the ocean background in the two lower



(a) TB measurements (19GHz, 37GHz, and 89 GHz) in Kelvin of Cyclone Amphan observed by GPM-CO on 16/05/2020.



(b) DPR observation of the rain intensity of the same event (mm/h).

Figure 2.3 – GPM-CO observation of Cyclone Amphan on 16/05/2020.

frequencies due to the high emissivity of the precipitation zone. In addition, there is also a contrast between the ocean (lower emissivity) and land (higher emissivity) at 19 and 37 GHz.

The TB is an indirect observation of geophysical variables. For example, rain intensity needs to be retrieved from TB observation. A Bayesian-based pixel-wise approach for rain retrieval consists of building a large database of TB and rain intensity, subdivided using auxiliary data (cloud cover, land surface, temperature, etc.) (Kummerow et al., 2015; Viltard et al., 2006). These pixel-to-pixel approaches lack the useful information contained in the structure of precipitation events. This structural information can be taken into account in the rain retrieval algorithm using machine learning techniques such as convolution layers. It should be noted that the data used in this study is the 'Level 1C-R' product of GMI, which is the geolocated and intercalibrated TB (Hou et al., 2014). The row of pixels in this product can be considered a conventional image (Figure 2.2b), which can then be used to train deep-learning algorithms.

2.2 . Deep-learning RAIN (DRAIN)

Deep-learning Rain (DRAIN) is a deep-learning rain retrieval algorithm for the GPM-CO based on a quantile regression U-Net (Viltard et al., 2023). The inputs to DRAIN are cropped images of the channels 37 GHz and 89 GHz of GMI in both the horizontal and vertical polarization, while the targets are the images of DPR rain intensity observation. The first motivation for choosing these two channels as inputs for DRAIN is the transferability to other satellites in the constellation, as the majority of conical-scanning passive microwave radiometers in the constellation offer channel frequencies similar to those of these two. In addition, these frequencies provide good horizontal gradients and resolution.

There is a difference in resolution between GMI and DPR. Several pixels of DPR closest to a pixel of GMI are averaged to match the resolution of GMI. The DRAIN dataset includes 70,000 images for the training dataset and 33,000 images for the validation set, taken from observations between 2014 and 2018.

U-Net is a fully convolutional network (FCN) first proposed for semantic segmentation by Ronneberger et al. (2015). The convolution kernels allow it to extract structural information within the image. U-Net has three important parts : the contraction path that condenses the feature maps to a latent space, the expansion path that expands the feature maps back up, and the skip-connections that connect the contraction and expansion paths. The skip-connections are concatenation of the feature maps of the contraction path to the corresponding feature map of the expansion path. Previous study (Viltard et al., 2020) shows that these skip-connections are efficient at preserving the fine-scale structures present in the

TB images. Quantile regression loss (Koenker and Hallock, 2001) replaces the cross-entropy loss of the original U-Net paper, which allows the current DRAIN to estimate 99 quantiles of rain intensity.

In most precipitation events, DRAIN achieved similar to slightly better retrieval than the official GPM algorithm GPROF. Unlike GPROF, DRAIN does not contain different levels of errors in its estimates between land and ocean. However, as DRAIN relies on DPR as its ground truth and DPR is not very sensitive to low rain intensity, the resulting rain intensity estimation does not show very low rain intensity.

Previous improvement efforts for DRAIN (not described in the attached article) include :

- **Inclusion of 19 GHz in input**

In this case, the TB input to DRAIN is 19, 37, and 89 GHz in horizontal and vertical polarization. The brightness temperature at 19GHz is sensitive to the integrated liquid water in the atmosphere and very sensitive to the surface conditions. Results from this test show a slight improvement in rainfall estimation over the ocean and a small degradation in the estimations over the continent. This might be due to the sensitivity of the 19 GHz to surface conditions and the large number of data points over the ocean compared to land, leading to the model's better efficiency over the ocean.

- **Multi-task learning**

Multi-task learning is an approach where the deep-learning models try to predict two different outputs, aiming to improve the training of the first tasks due to implicit data augmentation, attention focusing, and eavesdropping (Ruder, 2017). In addition to precipitation intensity, U-Net also predicts types of precipitation (convective, stratiform, and other) given by DPR. These two predictions share all U-Net parameters except for the output layers. The objective function includes the loss from the prediction of rain intensity (quantile regression loss) and precipitation types (cross-entropy loss). The two losses are balanced by a coefficient that is considered a parameter of the model, which can then be learned during training (Kendall et al., 2018). The output using DPR is of a higher resolution than the TB input. In order to match the resolution, one pixel of GMI is made up of several pixels of DPR. For intensity, the pixels are averaged. However, the precipitation types are combined using the majority vote, which might reduce the quality of the data. The results of this experiment did not show any improvement in DRAIN. Future improvement efforts for this approach should include careful consideration of the suitability of the loss function, multi-task training technique, and the quality of data.

- **U-net with attention mechanism**

The model tested is the SmaAt-UNet (Trebing et al., 2021), which is a U-Net architecture that includes spatial and channel-wise attention (Woo et al., 2018) to its skip-connections. This results in reducing the total number of parameters to be updated during back-propagation. The test with SmaAt-UNet was done on a smaller dataset using only four quantiles. Results show that the SmaAt-UNet was able to achieve marginal improvement in the high rain intensity estimates. Although the improvement is very small, attention mechanism could be a potential improvement approach.

The improvement from these efforts is marginal. Therefore, the current version of DRAIN does not include any of the above modifications. The following attached article includes the implementation and evaluation of the current DRAIN algorithm.

EVALUATION OF DRAIN, A DEEP-LEARNING APPROACH...

EVALUATION OF DRAIN, A DEEP-LEARNING APPROACH TO RAIN RETRIEVAL FROM GPM PASSIVE MICROWAVE RADIOMETER.

Nicolas Viltard, Vibolroth Sambath, Pierre Lepetit*, Audrey Martini, Laurent Barthès, Cécile Mallet

LATMOS-IPSL, Université Paris-Saclay, UVSQ, CNRS, 78280, Guyancourt, France

*Météo-France, Avenue Coriolis, Toulouse

Abstract— Retrieval of rain from Passive Microwave radiometers data has been a challenge ever since the launch of the first Defense Meteorological Satellite Program in the late 1980s. Enormous progress has been made since the launch of the Tropical Rainfall Measuring Mission (TRMM) in 1997 but until recently the data were processed pixel-by-pixel or taking a few neighboring pixels into account. Deep learning has obtained remarkable improvement in the computer vision field, and offers a whole new way to tackle the rain retrieval problem. The Global Precipitation Measurement (GPM) Core satellite carries similarly to TRMM, a passive microwave radiometer and a radar that share part of their swath. The brightness temperatures measured in the 37 and 89 GHz channels are used like the RGB components of a regular image while rain rate from Dual Frequency radar provides the surface rain. A U-net is then trained on these data to develop a retrieval algorithm: Deep-learning RAIN (DRAIN). Using only the brightness temperatures from four channels as input and no other a priori information, DRAIN is offering similar or slightly better performances than GPROF, the GPM official algorithm, in most situations. These performances are assumed to be due to the fact that DRAIN works on an image basis instead of the classical pixel-by-pixel basis.

I. MOTIVATION

We have developed a rain retrieval algorithm, DRAIN, based on the use of deep-learning techniques. The architecture and principles driving DRAIN are presented in more detail in [1] but, namely, a database of co-located rain rates from the Dual frequency Precipitation Radar (DPR) and brightness temperatures from the Global Precipitation Measurement Microwave radiometer (GPM-GMI) were fed into a U-net [2]. This type of convolutional network was successfully used to de-clutter radar images [3] and thus appeared to be well adapted to detect the contours of rain events.

The main difference between more classical retrieval algorithms, either Bayesian-based [4, 5] or machine learning-based [6-12], and DRAIN arises from the fact that the latter processes the data as images while the formers work on a pixel-by-pixel basis. These image-based approaches are developing fast as shown also in [13] and [14]

In the present paper, we will not go into the details of DRAIN which have already been presented in [1], but we will focus on a more thorough validation of a more mature version of the algorithm. The two main

improvements that must be highlighted between the initial version presented in [1] and the current version are described hereafter. First, the database was increased from about 52,000 images to about 103,000 allowing us to build a training database of 70,000 images for training and 33,000 images for validation. Data from the whole years 2014 to 2018 and a few months from 2020 and 2021 are used but the whole year 2019 was kept separate for the performance assessment (test) and most results presented hereafter are computed for that year. This large database is meant to dampen the effects of seasonal and interannual variability of rain.

Second, DRAIN retrieves now a set of 99 quantiles instead of a simple averaged rain rate as in [1]. These quantiles represent the probability that the rain rate is below a certain threshold. Hereafter, when unspecified, the quantile 50 % (median) is used as the rain proxy. The loss function for quantile regression is the one proposed in [15]. Retrieving quantile for rain is interesting because as shown hereafter, it is possible to infer a confidence interval for the results. It would also be possible from the retrieved Cumulative Distribution Function (CDF) of rain intensity to deduce a Probability Density Function (PDF, not presented here).

Section II will give a short presentation of the database construction while section III gives some elements about the methodology and the associated cost function. Section IV offers a detailed validation of DRAIN rain rate against DPR and GPROF for the year 2019. Section V presents the comparison for the same year with Meteo-France five-minute 1x1 km² rain mosaic. Finally, section VI presents the conclusions and perspectives.

II. DATABASE

GMI is a conically-scanning radiometer with channels at 10.65, 18.7, 23.8, 36.6, 89.0, 166.0, 183.3+/-3 and 183+/-7 GHz. All the channels are measured in both Horizontal (H) and Vertical (V) polarization except for 23.8 and the two 183.3 GHz sounding channels (V only). In the present study, only the two 36.6 (hereafter noted 37 GHz) and 89.0 GHz channels were used. This choice was driven by the idea that most conical-scanning passive microwave radiometers for rain retrievals have channels in the 37 and 89 GHz region thus making a transposition of DRAIN to other platforms potentially easier. These two frequencies were selected because they offer a good spatial resolution with well-defined horizontal gradients. For the GMI, the pixel resolutions are respectively: 15.6x9.4 km² for the 37 GHz and 7.2x4.4 km² for 89 GHz [16-17] (the GMI product used here is 1C-R.GPM.GMI.XCAL2016-C, PPS V05A).

Since [1], some tests (not shown) were made to check if the addition of the 19 GHz channels would improve the performances of DRAIN but the results were inconclusive.

The DPR surface rain product results from the merged use of the Ku- (13.4 GHz) and Ka-band (35.5 GHz) radars. The DPR measures a three-dimensional reflectivity field with vertical resolution of 250 m and a horizontal resolution of 5 km. The resulting surface rain product offers a swath width of 245 km [18].

The DPR and GMI pixels are co-located spatially and temporally assuming that the effective one-minute lag between the two observations is negligible at the considered spatial resolution. To perform the co-location, the surface rain (precipRateESurface from 2A.GPM.DPR.V8-20180723, PPS V06A) of the DPR pixels falling within 5 km of a GMI pixel center position are averaged. On average, three to four DPR pixels fall into the 5 km-radius.

We will hereafter refer to training and validation database as the data used to adjust the weights and hyperparameters of the network and the test database as the one used for generalization and assessment of performances.

Because rain occurrence is naturally low, a scene selection is made as follow: images with at least 100 pixels with rain $> 0.1 \text{ mm.hr}^{-1}$ or at least 10 pixels $> 100 \text{ mm.hr}^{-1}$ are used to build the training/validation database. Each image is composed of 4 channels: the 37 and 89 in both H and V polarization while the surface rain is used as target.

III. METHOD

The general architecture of the U-net used for DRAIN is described in [2]. More specifically, the configuration used here is made of two convolutional layers as input layers. Then, a contraction path follows, made of four downward steps each containing a max pooling and two convolution blocks. Each convolution block contains a convolution layer, a batch normalization, and a ReLU. Next, four expansion steps made of a transpose convolution and two convolutional layers increase the size of the image back up. In addition, each upward step is concatenated with its corresponding downward step (skip-connections). Finally, the output layer is a 1x1-kernel convolutional layer. The initial input layer is a 4x128x128 subset of GMI orbit with a padding size of 1 pixel on each side of the image. As stated previously, a similar U-net was successfully used for detection and restoration of clutter echoes in weather radar raw data [3].

Weights, optimization method (Adam, [18]) and initial learning rate (10^{-4}) are set to the default values found in literature [19]. The trained U-net has about ~15 million parameters to be adjusted through the training phase.

The loss function used here is the classical one used for quantile regression (e.g. [15]) where the loss function is given by:

$$L_{q_j}(y_i, \hat{y}_i) = \begin{cases} q_j(y_i - \hat{y}_i) & \text{if } y_i - \hat{y}_i \geq 0 \\ (q_j - 1)(y_i - \hat{y}_i) & \text{if } y_i - \hat{y}_i < 0 \end{cases}$$

$$L_{q_j}(y, \hat{y}) = \frac{1}{N} \sum_i L_{q_j}(y_i, \hat{y}_i)$$

$$L(y, \hat{y}) = \sum_j L_{q_j}(y, \hat{y})$$

where \hat{y} is the prediction, y is the target, q_j is the j th quantile to be estimated and N the number of pixels. Besides the fact that retrieving quantiles gives access to more information than the mere retrieved average rain rate, the quantile regression is of particular interest here because of its demonstrated robustness to outliers. In the present study, we chose to retrieve percentiles because computation of the rain intensity probability density function is easier to compute through simple numerical derivation.

The network presented here results in a 460-epoch training on the learning and validation bases described above. We will use hereafter the median (50th quantile) as the retrieved rain rate.

Because the chosen architecture is made of a 16-level of filtering, the retrieval image sizes have to be multiple of 16. For the moment the retrieved images are thus limited to 208 pixels per 2960 scans from the original 221 pixels and up to 2963 scans (the effective number of scans may vary slightly from one granule to the next but remains larger than 2960).

IV. COMPARISON WITH DPR

Two example cases, excluded from the learning and validation dataset, are given on Fig. 2 and 3. The first case is super typhoon Nanmadol observed on September 16th, 2022, while close to its peak intensity over the Philippine Sea, South of Japan. The second case is a frontal band over France observed on August 18th, 2018. The first case is a good illustration of oceanic retrieval while the second case is a mixed continental and coastal case.

In addition to comparisons with the DPR in the common part of the swath, comparisons with the GPROF algorithm [4] are also presented (2A.GPM.GMI.GPROF2017v1, PPS V05A). GPROF is the GPM operational product, based on a per-pixel-Bayesian approach. It is considered as a reference in the GPM community and uses auxiliary data to constrain the solutions (2m temperature, TPW, and surface type etc...). Its native resolution is close to 11 km according to [22] and uses a priori databases that are build using different sources depending on the surface type [22]. Over ocean, the *Combined MS DPR* product is used. Over land, the *DPR-Ku surface precipitation* is used for most situation except snow covered ground when MRMS is used. Finally, additional adjustments, described in [22] were performed based on CloudSat measurements for drizzle and light rain to overcome the DPR sensitivity.

Note that for all the comparisons presented hereafter with DPR and GPROF and in part V also, we use the 50th quantile (median) as our best estimator for DRAIN. For both situations, DRAIN and DPR are qualitatively very close and resemble GPROF except that the latter exhibits a lot of light rain pixels (< 0.2-0.3 mm/hr). The DPR algorithm is estimated to have a minimal detection threshold near 0.2 mm.hr⁻¹ [20]. Because we perform a spatial average over two to four DPR pixels, DRAIN is expected to have a theoretical detection threshold between 0.03 and 0.1 mm.hr⁻¹ depending on the number of averaged pixels.

The most noticeable difference between GPROF and DRAIN is particularly visible on Fig. 2 (super typhoon Nanmadol) where the GPROF retrieves many more rainy pixels with intensities below 0.4 mm/hr than both DRAIN and DPR. DRAIN on this aspect is very consistent with the DPR as one would expect since DRAIN was developed solely from DPR data without any other assumptions concerning the brightness temperatures – rain intensities relationship. The reasons behind this light rain enhancement in GPROF is described in [20]. This light rain difference is more likely to be visible over ocean where, in addition to the higher frequency channels, GPROF uses 10 and 19 GHz channels which are directly linked to liquid water emission.

The second comparison is made at global level over the whole year 2019. A few granules with bad T_{BS} at either 37 or 89 GHz are excluded. A mask is used to check systematic differences between land and ocean. Unlike for GPROF, this mask is not used as an input in DRAIN but only once the retrieval is performed to assess the

performances over the two different surface types. This mask is the python “regionmask” package which is based on Natural Earth, a free vector and raster map data (naturalearthdata.com).

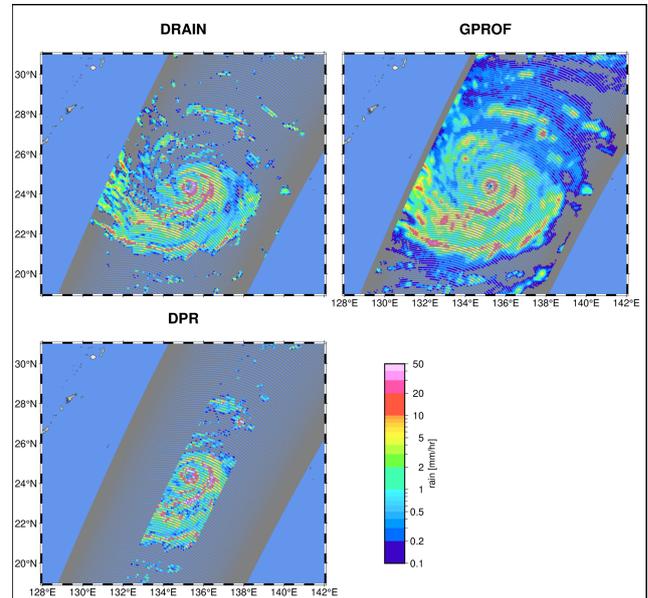


Figure 2: surface rain for super Typhoon Nanmadol (16W) on 16th September 2022 at 08:00 UTC. DPR at DRAIN resolution (top left), DRAIN (top right) and GPROF (Bottom left).

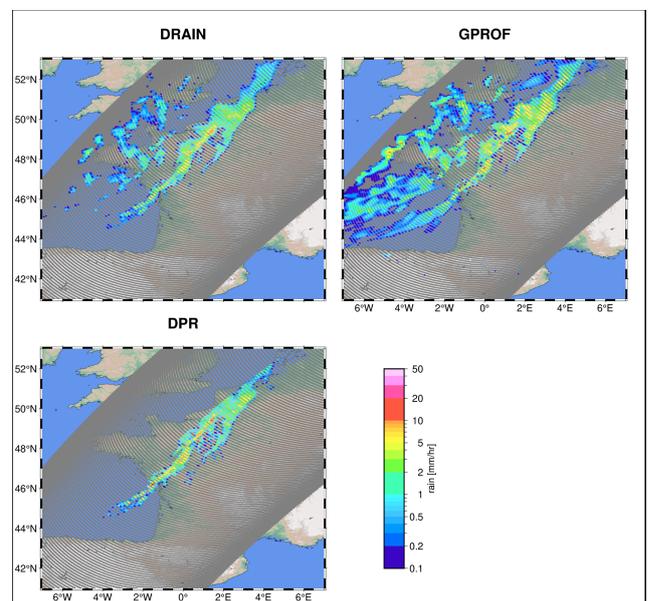


Figure 3: same as Fig. 2 but for a frontal rain band over France on 18th of August 2018.

Table 1 and 2 give respectively the contingency table for OCEAN (approximately 530 million pixels) and LAND (approximately 218 million pixels). DRAIN is well-balanced between False Alarm (FA) and Bad Detection (BD) occurrence although the False Alarm has a high

RMSE of 10.77 mm/hr (17.36 above ocean and 7.70 above land). The high probability of FA for GPROF are mostly due to the light rain treatment of the latter, as mentioned above. On the other hand, these FA have a very small RMSE of 0.30 mm/hr because they are made of very light rain rates.

OCEAN DPR	DRAIN		GPROF	
	Rain	No Rain	Rain	No Rain
Rain	6.01	1.97	5.93	2.05
No Rain	1.23	90.79	13.14	78.88

Table 1: Contingency in % table for OCEAN pixels on the whole dataset of 2019.

LAND DPR	DRAIN		GPROF	
	Rain	No Rain	Rain	No Rain
Rain	3.49	1.26	3.46	1.29
No Rain	0.91	94.34	10.44	84.81

Table 2: same as Table 1 but for LAND pixels.

	OCEAN		LAND	
	POD	FAR	POD	FAR
DRAIN	0.75	0.17	0.74	0.21
GPROF	0.74	0.69	0.72	0.75

Table 3: Probability of Detection (POD) and False Alarm Ratio (FAR) over land and ocean for DRAIN and GPROF, when comparing with DPR in the common part of the swath.

	DRAIN		GPROF	
	Bias	RMSE	Bias	RMSE
LAND	0.31	2.67	-0.24	3.39
OCEAN	0.26	2.98	0.08	3.18
TOTAL	0.27	2.92	0.01	3.22

Table 4: The numbers are computed on the respective true positive with a threshold of 10^{-4} mm/hr. Bias, in mm/hr are respectively DPR-DRAIN and DPR-GPROF.

In the case of BD for both DRAIN and GPROF, DPR average rain estimates are 0.33 mm/hr and 0.50 mm/hr respectively showing that they are consistently light rain situations.

Table 3 shows the Probability Of Detection (POD) and the False Alarm Ratio (FAR) which are often used to assess the performances of rain retrieval algorithms [21]. Perfect POD is 1 while perfect FAR is 0. Both algorithms offer similar PODs with a small advantage for DRAIN, especially over land. The low score of GPROF's FAR is logically due to the light rain over-detection with respect to DPR as mentioned above.

Table 4 shows the bias and Root Mean Square Error (RMSE) for the two algorithms, compared to DPR for their respective true positive only. A 10^{-4} mm/hr threshold is applied to make sure that no random numerical noise will contaminate the results. GPROF is always better in terms of bias but DRAIN is better in terms of RMSE. This is supported by Fig 4 that shows the corresponding scatter plots for rain intensities between 0 and 100 mm/hr. DRAIN has been trained with the DPR so, once again it is expected that its performances with respect to the latter will be optimal.

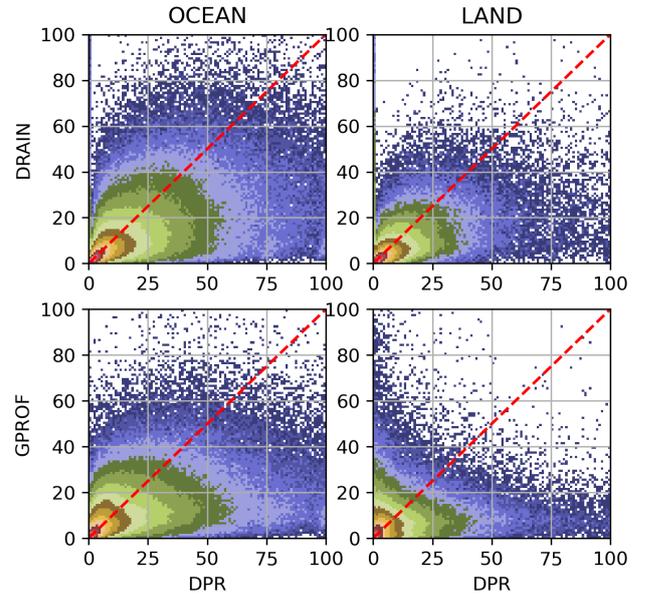


Figure 4: scatter plot of DRAIN (top row) and GPROF (bottom row) against DPR for OCEAN pixels (left hand side column) LAND pixels (right hand side column) for the whole of 2019. Colors show the density of point from red (densest) to blue (least dense) with a lognormal scale. The red dashed line on each of the graph is the $x=y$ line.

A substantial spread is however observed which is expected due to the potential parallax effects between the GMI TBs and the DPR surface rain which cannot be totally compensated. Over both land and ocean, systematic underestimation starts appearing at about 20 mm/hr and increases as the DPR rain rate increases. Over ocean, GPROF shows similar features, slightly exacerbated for DPR rain rates above 25 mm/hr. Over land, comparison is made difficult by the fact that the GPROF rain rate is made of the DPR estimates complemented by NOAA's Multi-Radar/Multi-Sensor system (MRMS) [22].

Fig. 5 shows the histograms of DRAIN, GPROF and DPR for the light rain rates, emphasizing the higher probability of rain below 0.25 mm/hr proposed by

GPROF when compared to the two other estimators. This is true for land and even more for ocean. It is to be once again mentioned that these light rain rates have been enhanced in GPROF's a priori database, based on Cloudsat statistics [21] to compensate in particular for the Ku-DPR detection threshold (~ 12 dBZ).

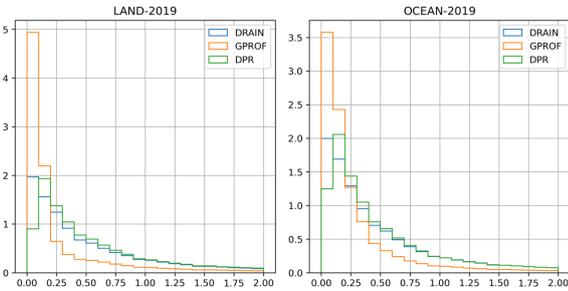


Figure 5: pdf of rain intensity for DPR, GPROF and DRAIN for the year 2019, focusing on the light rain between 0 and 2 mm.hr⁻¹ with 0.1 mm.hr⁻¹ bins.

Rain interval mm/hr	50%	90%
0 to 0.1	73.96	97.03
0.1 to 1.	56.36	93.93
1 to 10	46.49	87.01
10 and above	33.65	75.47
All	53.96	91.78

Table 5: retrieved confidence intervals as a function of rain intensities.

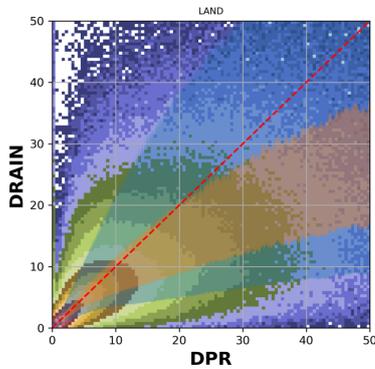


Figure 6: 2-D histogram of the retrieved rain vs. initial (DPR) rain, color indicates density of points with similar scale as Fig. 4. Blue-shaded area is the 90 % confidence interval and brow-shaded area is the 50 % confidence interval.

Since percentiles are retrieved by the network, among other applications, it is possible to define confidence intervals for the retrieved rain intensity. The rain rate, RR_j , given for the j^{th} percentile means that the a-priori probability that the DPR rain rate is between 0 and RR_j is $j\%$. In Table 5 the 50% and 90% confidence interval are verified against the DPR surface rain. For the true positive, pixels for which the DPR rain rates falls in between percentiles 25 and 75% and 5 and 95% are counted. It can be seen that overall, the confidence

interval is indeed reliable but does actually depends on the rain interval under consideration. Up to about 10 mm/hr, the confidence interval is robust but above this, the systematic underestimation of the retrieved rain rate degrades the results.

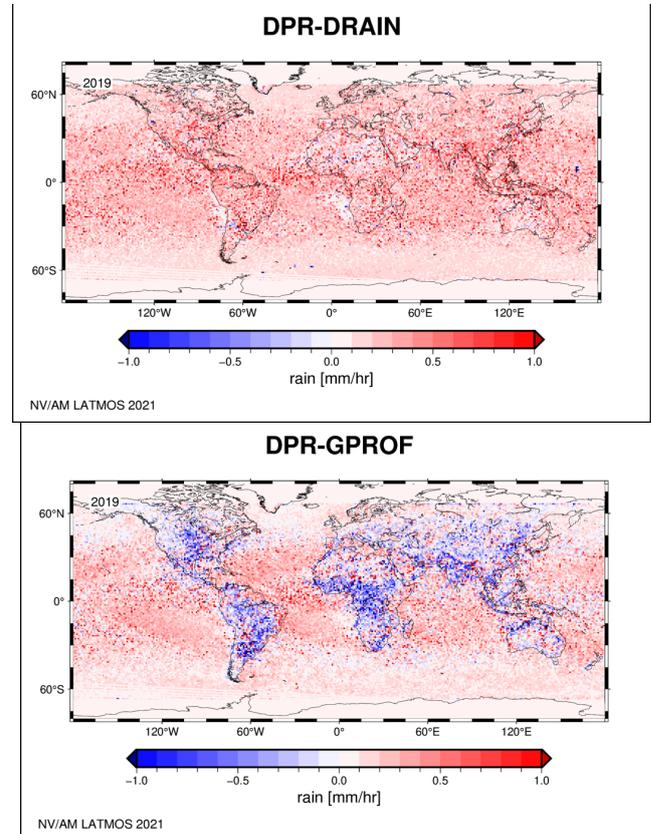


Figure7: 1°x1° averaged difference between top: DPR - DRAIN and bottom DPR - GPROF. The differences are performed at pixel level and then averaged over all the orbits of 2019.

Figure 6 is a zoom for rain rates between 0 and 50 mm/hr of DRAIN vs DPR. The scatter plot is not differentiating land and ocean. On top of the scatter plot is the 90 % confidence (shade of blue) and the 50 % confidence (shade of brown). These confidences are computed using the 5th and 95th retrieved percentiles and the 25th and 75th retrieved percentiles respectively. It can be seen that, as stated from Table 5, up to 12-15 mm/hr, the confidence interval is reliable. It then slowly degrades and after 20 mm/hr, it is just an indicator but cannot be considered as accurate. One can also see that as for GPROF and most retrieval methods, DRAIN cannot overcome the rain underestimation starting in the 20 mm/hr region even if it is mitigated to some extent.

Figure 7 shows these differences between GPROF, DRAIN and DPR on a 1°x1° global map. The three estimators are first averaged on 1-degree squares, keeping only the pixels above 10⁻³ mm/hr.

DRAIN offers a very consistent bias overall. No difference is noticeable between land and ocean or artifact over coastal area which is always very difficult to handle for such retrieval methods. However, the bias appears to be mostly positive (underestimation) while a more balanced distribution of positive and negative values was expected. This might be due to the choice of the median quantile as the rain estimator, which is somewhat arbitrary.

A slight latitude dependence of the error can be observed as the error seems smaller above 50° N and below 50° S over ocean and slightly larger in the ITCZ. Errors appear to be also larger above mountainous areas like the Tibetan plateau, the Rocky Mountains and somewhat the Andes.

GPROF's error range is very similar to DRAIN's but with a marked difference between land and ocean which was already mentioned. A slight dependence to the latitude can also be observed but less marked than DRAIN.

V. RESULTS ON MÉTÉO-FRANCE MOSAIC

An assessment of the performances was also conducted using Météo-France five-minute mosaic product which is a good reference for mid-latitude QPE. The used product is described in [23] and comes as a 1536x1536 pixels grid of 1 km resolution every 5 minutes over the whole of year 2019. Co-location in time was performed by matching the closest mosaic in time with the mid-time of GMI overpass (the GMI overpass lasts about 3 minutes). The accumulation over 5 minutes is converted in mm.hr⁻¹ by simple multiplication by a factor 12. A quality flag is associated with the Météo-France product ranging from 0 (unreliable) to 100 % (very reliable). After visual comparison on a series of cases, it appeared that a threshold of 80 % reliability should be applied in order to eliminate spurious rain estimates particularly in the mountainous areas.

The mosaic data are then co-located and averaged at the DRAIN resolution and pixels position. Between January 1st 2019 and December 31st 2019, 1565 overpasses are kept, with the condition that more than 50 DRAIN pixels fall into the mosaic domain: 8° West to 12° East and 39° North to 54° North.

First, a pixel-by-pixel performance is evaluated for the three rain estimators: DRAIN, DPR and GPROF. Contingency table and F1-scores are computed and presented Table 6, 7 and 8 respectively. The total number of pixels taken into account differs for each

estimator because of the swath difference. The three estimators show performances that are close with a few differences. GPROF shows a better POD than both DRAIN and DPR but its FAR is degraded by the excess of light rain detected. On the other hand, DPR and DRAIN miss some of the rain which degrades their respective POD but their precision remains high.

Ref.\DRAIN	Rain	No Rain	POD
Rain	4.85 %	9.81 %	0.33
No Rain	0.36 %	84.98 %	FAR
Precision	0.93		0.07
F1-score	0.49		

Table 6: Contingency table and F1-score for DRAIN with Météo-France mosaic as a reference. The total number of co-located pixels is 6645997.

Ref.\DPR	Rain	No Rain	POD
Rain	5.59 %	8.53 %	0.40
No Rain	0.44 %	84.44 %	FAR
Precision	0.93		0.07
F1-score	0.56		

Table 7: same as Table 4 but for DPR. The total number of co-located pixels is 1434964.

Ref.\GPROF	Rain	No Rain	POD
Rain	7.38 %	7.26 %	0.50
No Rain	4.49 %	80.88 %	FAR
Precision	0.62		0.38
F1-score	0.56		

Table 8: same as Table 4 but for GPROF. The total number of co-located pixels is 5611805.

The DRAIN, DPR and GPROF data are also averaged on 0.2°x0.2° grid to minimize the possible impact of localization errors between the ground and satellite-based estimates. Figure 8 shows the scatter plot for each of the estimators against the Météo-France mosaic. Only the grid-boxes where the mosaic ≥ 0 mm.hr⁻¹ are accounted for. For each scatter plot, the associated linear regression (blue-dotted line) is shown. For DRAIN the corresponding $R^2=0.29$ while it is 0.10 for DPR and 0.20 for GPROF. Below 1 mm/hr, all estimators are mostly centered on the x=y line and then the points spread out with a general underestimation. DPR appears more spread than both DRAIN and GPROF but the data sample is about four times smaller and the narrow swath might induce more important edges artifacts. GPROF appears to show a slight underestimation (~0.1 mm/hr) around 0.6-0.7 mm/hr, in the densest part of the scatter

plot. Globally however, GPROF and DRAIN appear to offer similar performances.

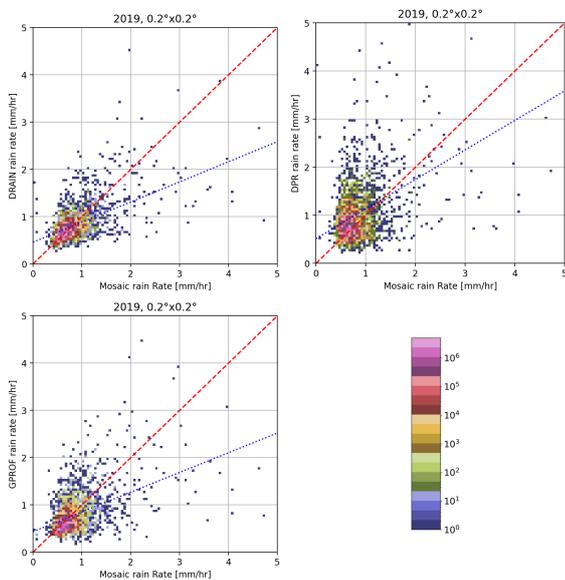


Figure 8: Scatter plot of DRAIN (top left), DPR (top right) and GPROF (bottom left) vs Météo-France mosaic over all the overpasses of 2019. In each graph, red-dashed line is the $x=y$ line, blue-dotted line is the linear regression and the color are proportional to the density of points according to the colorbar.

Figure 9 shows the maps of the biases Mosaic-DRAIN and Mosaic-GPROF, in $\text{mm}\cdot\text{hr}^{-1}$, for 2019. Except for a few grid-boxes, the difference remains between -2 and +2 $\text{mm}\cdot\text{hr}^{-1}$. The general pattern and amplitude of the difference are similar between the two estimators.

In the mountainous area of the Alps and somewhat on the Mediterranean shores, the errors are almost identical on a surprising number of boxes which might show a problem with the Météo-France estimate. It is noticeable though, that GPROF has a very clear artifact of overestimating the land part of the coastal regions. This is likely due to the difference between the land and the ocean version of GPROF. Most of DRAIN error structure appears to be much more random and better balanced, yet there is continuity from one box to the next, showing that the errors are not pure noise. This spatial continuity of the errors is also true for GPROF.

Figure 10 shows the dependance with time of the Mean Average Error for the three estimators. Environmental conditions change depending on the season and this is eventually even more pronounced over land with possible snow cover on the ground. All three estimators follow more or less the same patterns with an increase in the errors during winter. The DPR should be

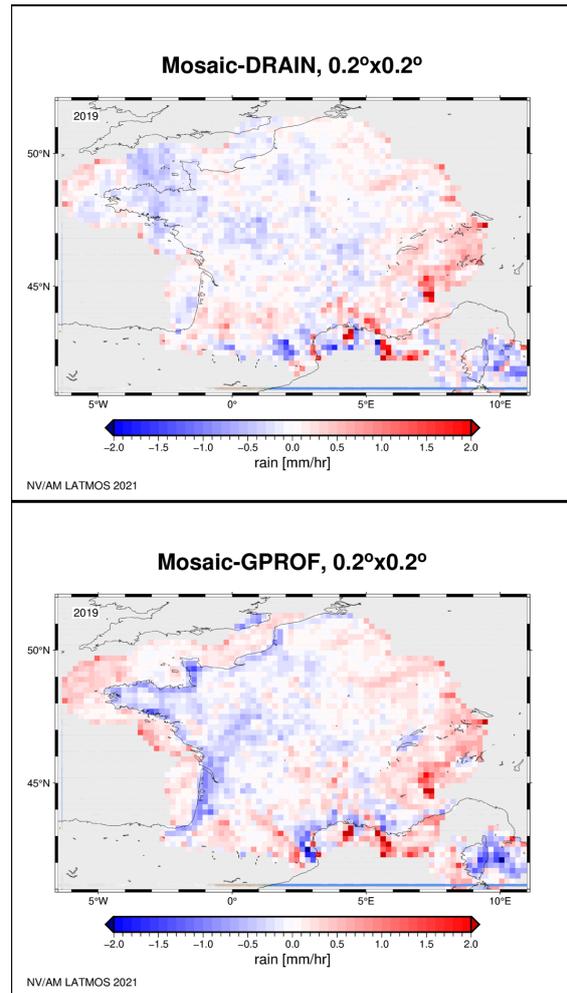


Figure 9: Bias between Météo-France mosaic and DRAIN on top and GPROF on bottom. The difference is computed over $0.2^\circ \times 0.2^\circ$ boxes over the whole year 2019.

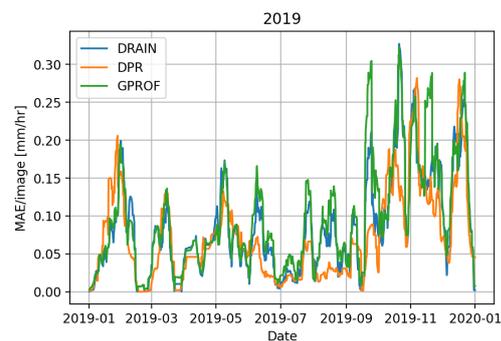


Figure 10: Mean Average Error between the DRAIN, GPROF and DPR as a function of time over 2019.

considered with some care because its coverage is substantially lower than both DRAIN and GPROF inducing some possible representativeness artifacts.

VI. CONCLUSIONS

From a set of 103,000 images of co-located data between GMI brightness temperatures and DPR surface rain, a U-net was trained to retrieve the latter over the whole swath of the radiometer. To minimize the impact of surface emissivity, work with the highest spatial resolution possible and at the same time remaining easily transposable to other instruments, only 37 and 89 GHz horizontal and vertical polarization brightness temperatures are used as an input. The strength of U-nets is their ability to process brightness temperature scenes as an image, as opposed to most existing algorithms that proceed pixel by pixel.

Evaluation of the developed algorithm is performed two ways. First, a comparison with the DPR surface rain itself on a set of images that was not in the training database (whole of 2019) is presented. DRAIN shows a good agreement in terms of structure and intensities which ensures the good quality of the generalization. When compared on the same dataset, DRAIN is generally on par with GPROF or slightly better. This is most noticeable over land where GPROF does not use only the DPR rain rate in its database.

Second, a comparison is performed with Météo-France 1 km²-resolution mosaic over the same period as a fully independent dataset. The trends are similar. DRAIN performances are close to GPROF if not slightly better. The most noticeable difference is observed in the coastal regions where GPROF tends to overestimate the rain intensities when compared to the mosaic.

These four channels used here are present on most passive microwave radiometers of the GPM constellation which will facilitate transposition of the developed U-net to other available sensors of the constellation as proposed in [24].

ACKNOWLEDGMENTS

This research work is supported by TOSCA of Centre National d'Etudes Spatiales (CNES) and INSU PNTS program. Our thanks to the reviewers for their helpful comments.

REFERENCES

- [1]. Viltard N., Lepetit P., Mallet C., Barthès L., Martini A. 2020: Retrieving Rain Rates from Space Borne Microwave Sensors Using U-Nets. *Climate Informatics 2020. 10th International Conference*, Sep 2020, Oxford, United Kingdom - [insu-02894942](https://doi.org/10.1109/TGRS.2021.3052582)
- [2]. Ronneberger O., P. Fischer and T. Brox, 2015: U-net: Convolutional networks for biomedical image segmentation. *Int. Conf. on Med. Image Comput. and Computer-Assisted Interv.* Springer, pp. 234–241.
- [3]. Lepetit P., L. Barthès, C. Mallet, C. Ly, N. Viltard, Y. Lemaître, 2022: Using Deep Learning for Restoration of Precipitation Echoes in Radar Data, *IEEE Transac. Geosci. And R. Sensing*, **60**, pp.5100914. ([10.1109/TGRS.2021.3052582](https://doi.org/10.1109/TGRS.2021.3052582)) - [insu-03151106](https://doi.org/10.1109/TGRS.2021.3052582)
- [4]. C. D. Kummerow, D.L. Randel, M. Kulie, N. Wang, R. Ferraro, S. Joseph Munchak, and V. Petkovic, 2015: The Evolution of the Goddard Profiling Algorithm to a Fully Parametric Scheme. *J. Atmos. Oceanic Technol.*, **32**, 2265–2280
- [5]. N. Viltard, C. Burlaud and C. D. Kummerow, 2006: Rain retrieval from TMI brightness temperature measurements using a TRMM PR–based database. *J. Appl. Meteor. Climatol.*, **45**, 455–466
- [6]. E. Moreau, Mallet C., Thiria S., Mabboux X., Badran F., Klapisz C., 2002: Atmospheric Liquid Water Retrieval Using a Gated Expert Neural Network, *J. of Atmos. and Ocean. Tech.*, **19**, 457–466
- [7]. Y. Hong, K. Hsu, S. Sorooshian and X. Gao, 2004: Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *J. of Appl. Meteor.*, **43**, 1834–1853
- [8]. C. Surussavadee and D. H. Staelin, 2008: Global Millimeter-Wave Precipitation Retrievals Trained With a Cloud-Resolving Numerical Weather Prediction Model, Part I: Retrieval Design. *IEEE Transactions on Geosci. and R. Sens.*, **46**, 99–108
- [9]. P. Sanò G. Panegrossi, D. Casella, A.C. Marra, L.P. D'Adderio, J.F. Rysman and S. Dietrich, 2018: The Passive Microwave Neural Network Precipitation Retrieval (PNPR) Algorithm for the CONICAL Scanning Global Microwave Imager (GMI) Radiometer. *Remote Sens.* **10**, 1122
- [10]. C. Mahesh, S. Prakash, V. Sathiyamoorthy, R.M. Gairola, 2011: Artificial neural network-based microwave precipitation estimation using scattering index and polarization corrected temperature. *Atmos. Res.*, **102**, 358–364
- [11]. Y. Tao, X. Gao, A. Ihler, K. Hsu and S. Sorooshian, 2016: Deep neural networks for precipitation estimation from remotely sensed information. *2016 IEEE Congress on Evol. Comp. (CEC)*, Vancouver, BC, 1349–1355
- [12]. Y. Tao, X. Gao, A. Ihler, S. Sorooshian and K. Hsu, 2017: Precipitation identification with bispectral satellite information using deep learning approaches. *J. of Hydrometeorol.*, **18**, 1271–1283
- [13]. Sanò, P., Casella, D., Camplani, A., D'Adderio, L. P., & Panegrossi, G. (2022): A Machine Learning Snowfall Retrieval Algorithm for ATMS. *Remote Sensing*, **14**(6), 1467, <https://doi.org/10.3390/rs1406146>
- [14]. Pfreundschuh, S., Brown, P. J., Kummerow, C. D., Eriksson, P., and Norrestad, T., 2022: GPROF-NN: a neural-network-based implementation of the Goddard Profiling Algorithm, *Atmos. Meas. Tech.*, **15**, 5033–5060, <https://doi.org/10.5194/amt-15-5033-2022>
- [15]. Koenker, R., and K. F. Hallock, 2001: Quantile Regression. *Journal of Economic Perspectives*, **15** (4): 143–156. DOI: [10.1257/jep.15.4.143](https://doi.org/10.1257/jep.15.4.143)
- [16]. E. A. Smith, G. Asrar, Y. Furuhashi, A. Ginati, C. Kummerow, V. Levizzani, A. Mugnai, K. Nakamura, R. Adler, V. Casse, M. Cleave, M. Debois, and J. Durning,

- 2007: International Global Precipitation Measurement (GPM) Program and Mission: An Overview. *Springer - Measuring Precipitation from Space - EURAINSAT and the future*. Vol. **28**, pp. 611-653
- [17].Petty, G. W. and Bennartz, R., 2017: Field-of-view characteristics and resolution matching for the Global Precipitation Measurement (GPM) Microwave Imager (GMI), *Atmos. Meas. Tech.*, **10**, 745–758, <https://doi.org/10.5194/amt-10-745-2017>
- [18].P. D. Kingma and J. Ba, 2014: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*
- [19].X. Glorot and Y. Bengio, 2010: Understanding the difficulty of training deep feedforward neural networks. *Proc. 13th Int. Conf. on Artificial Intelligence and Statistics*, pp 249-256
- [20].Skofronick-Jackson G., D. Kirschbaum, W. Petersen, G. Huffman, C. Kidd, E. Stocker, and R. Kakar, 2018: The Global Precipitation Measurement (GPM) mission's scientific achievements and societal contributions: Reviewing four years of advanced rain and snow observations. *Quart. J. Roy. Meteor. Soc.*, **144**, 27–48, <https://doi.org/10.1002/qj.3313>
- [21].Tan J., W. A. Petersen, G. Kirchengast, D. C. Goodrich, and D. B. Wolff, 2018: Evaluation of Global Precipitation Measurement Rain fall Estimates against Three Dense Gauge Networks. *J. of Hydromet.*, **19**, 517-532 DOI: 10.1175/JHM-D-17-0174.1
- [22].Passive Microwave Algorithm Team Facility, GPROF2017 Version 1 and Version 2, ATBD, 2018, https://gpm.nasa.gov/sites/default/files/2020-05/ATBD_GPM_V5B_April15_2018.pdf
- [23].Figueras I Ventura J. and P. Tabary, 2013: The New French Operational Polarimetric Radar Rainfall Rate Product. *J. Appl. Meteor. And Climato.*, **52**, 1817-1835
- [24].Sambath V., N. Viltard, L. Barthès, A. Martini, C. Mallet, 2022: Unsupervised domain adaptation for Global Precipitation Measurement satellite constellation using Cycle Generative Adversarial Nets. *Environmental Data Science*, **1**, pp.e24.

2.3 . References

Reference of the attached article :

N. Viltard, V. Sambath, P. Lepetit, A. Martini, L. Barthès, and C. Mallet. Evaluation of DRAIN, a deep-learning approach to rain retrieval from GPM passive microwave radiometer. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

Additional references for this chapter :

- A. Y. Hou, R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi. The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, 95(5) :701 – 722, 2014. doi : 10.1175/BAMS-D-13-00164.1. URL <https://journals.ametsoc.org/view/journals/bams/95/5/bams-d-13-00164.1.xml>.
- A. Kendall, Y. Gal, and R. Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491, 2018.
- R. Koenker and K. F. Hallock. Quantile regression. *Journal of economic perspectives*, 15(4) :143–156, 2001.
- C. Kummerow, W. Barnes, T. Kozu, J. Shiue, and J. Simpson. The tropical rainfall measuring mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, 15(3) :809 – 817, 1998. doi : 10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atot/15/3/1520-0426_1998_015_0809_ttrmmt_2_0_co_2.xml.
- C. D. Kummerow, D. L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. Joseph Munchak, and V. Petkovic. The evolution of the goddard profiling algorithm to a fully parametric scheme. *Journal of atmospheric and oceanic technology*, 32 (12) :2265–2280, 2015.
- NESDIS STAR. MiRS sensors : GPM GMI overview. <https://www.star.nesdis.noaa.gov/mirs/gpmgmi.php>, 2021. Accessed : 14-10-2024.
- O. Ronneberger, P. Fischer, and T. Brox. U-net : Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015 : 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

- S. Ruder. An overview of multi-task learning in deep neural networks. *arXiv e-prints*, pages arXiv-1706, 2017.
- K. Trebing, T. Stańczyk, and S. Mehrkanoon. SmaAt-UNet : Precipitation now-casting using a small attention-UNet architecture. *Pattern Recognition Letters*, 145 :178–186, 2021.
- N. Viltard, C. Burlaud, and C. D. Kummerow. Rain retrieval from TMI brightness temperature measurements using a TRMM PR-based database. *Journal of applied meteorology and climatology*, 45(3) :455–466, 2006.
- N. Viltard, P. Lepeit, C. Mallet, L. Barthès, and A. Martini. Retrieving rain rates from space borne microwave sensors using u-nets. In *Proceedings of the 10th International Conference on Climate Informatics*, pages 30–36, 2020.
- N. Viltard, V. Sambath, P. Lepeit, A. Martini, L. Barthès, and C. Mallet. Evaluation of DRAIN, a deep-learning approach to rain retrieval from gpm passive microwave radiometer. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon. Cbam : Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

3 - Segmenting image without labels

Contents

3.1 Computer Vision Toolbox	53
3.1.1 Convolutional Neural Networks (CNN)	53
3.1.2 Auto-regressive model	55
3.1.3 Transformer	55
3.2 Unsupervised image segmentation	59
3.2.1 Pixel-wise approach	59
3.2.2 Non-pixel-wise approach	62
3.3 Self-supervised image segmentation	69
3.3.1 Self-supervised Vision Transformer	70
3.3.2 Self-supervised CNN	72
3.4 Summary of Methods	73
3.5 Evaluation Metrics for Unsupervised Image Segmentation	75
3.5.1 Internal Evaluation	75
3.5.2 External evaluation	76
3.5.3 Spatial Correlation	78
3.6 References	80

In Computer Vision, there are three types of image segmentation : semantic, instance, and panoptic image segmentation (Figure 3.1) (Jung et al., 2022). A semantic segmentation classifies each pixel in an image into a class. It does not distinguish between two different instances. For example, it classifies pixels of two different cats into the same class. An instance segmentation only labels the foreground of the image. In addition, it differentiates between two different instances, i.e., cat 1 and cat 2. Panoptic segmentation combines semantic and instance segmentations. It offers a pair of labels : a semantic label and an instance label for all pixels in the image.

In this study, semantic segmentation best serves our objective. We aim to extract information from the brightness temperature images by classifying the pixels in the images into a number of classes. These classes should represent geophysical variables observable in brightness temperature.

Supervised semantic segmentation could be considered as a process of classifying all pixels of an image based on a known ontology (Hamilton et al., 2022). In

supervised semantic image segmentation, the objective function is calculated using the ground truth labels available for each pixel in the image. Although it can isolate the object very efficiently, it requires a large amount of precise ground truth labeling that is often lacking due to the lack of human annotators or the complexity in the image where the precise boundary is unknown (medicine, biology, and physics) (Hamilton et al., 2022).

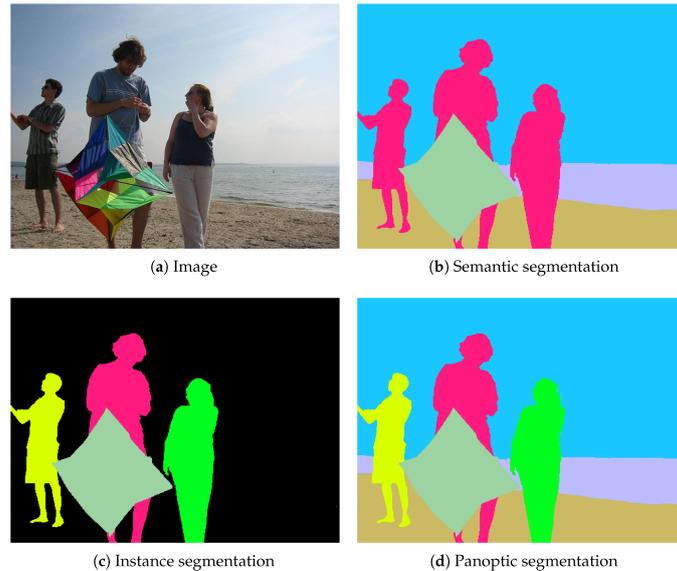


Figure 3.1 – Different types of image segmentation (Jung et al., 2022). An example of semantic (b), instance (c), and panoptic segmentation (d).

The focus of this thesis is segmenting unconventional images without any ground truth labels. We turn to two types of training approaches in deep learning : unsupervised learning and self-supervised learning.

Unsupervised image segmentation

Unsupervised image segmentation has historically been considered a clustering problem (Xia and Kulis, 2017). The pixel clustering is usually based on color, brightness, or texture. With the introduction of Machine Learning, more innovative approaches to creating an objective function without the supervision of ground truth labels are emerging.

Self-supervised image segmentation

Self-supervised image segmentation tries to learn representation from the data via a pretext task (Chen et al., 2022). Contrary to unsupervised approaches, where the objective function is clearly defined to train the model to segment the image, the loss function of the self-supervised models is a pretext task that does not provide a segmentation. A downstream task is needed to get the model to classify

each pixel. Sometimes, a small set of labeled data is required for the downstream task.

This chapter aims to provide an overview of unsupervised and self-supervised approaches in image segmentation. Before presenting these various training approaches, we present a few tools and architectures commonly used for dealing with image data in Computer Vision (Section 3.1). Next, we delve into the two learning approaches : Unsupervised Learning (Section 3.2) and Self-supervised Learning (Section 3.3). Each section presents a non-exhaustive list of models categorized by its training framework. Section 3.4, the Summary of Methods, provides some commentary on the listed methods regarding the application to the brightness temperature images. Finally, Section 3.5 provides details on the evaluation metrics for assessing model performances for an unsupervised approach.

3.1 . Computer Vision Toolbox

The particularity of image data is that it contains spatial information. It is important to consider the relationship between adjacent pixels. In this section, we describe three techniques that allow the model to analyze the structure of the image data : Convolutional Neural Networks (CNN), Auto-regressive models, and Transformers.

3.1.1 . Convolutional Neural Networks (CNN)

Ever since the introduction of CNN for written digit recognition in [Le Cun et al. \(1997\)](#), it has played a major role in the performance of computer vision models. CNN is advantageous in computer vision for its capacity for feature extraction of images and the reduction of the number of parameters when compared to a multilayer perceptron (MLP).

In [Li et al. \(2021\)](#), the authors provide a detailed description of CNN and its application. The most essential component of the CNN is the convolution kernel, whose size is a hyperparameter of the model. This kernel contains shared weights that glide through the image with a certain length of stride. A bigger stride results in a lower density of convolving. In order to deal with the size of the input and avoid losing information around the border of the image, padding can be added around the image. The padding could be, for example, extra rows of zeros or a row reflecting the values in the image. The convolutional kernel glides through the image from left to right, creating a feature map. To avoid redundancy and reduce the number of weights needed to deal with the feature map, a pooling, either calculating the average or choosing the maximum value, is applied to the feature map. A summary of this process is shown in Figure 3.2. It should be noted that the number of layers in a kernel needs to be equal to the number of layers in the

input image. This will then result in a single-layer feature map (Figure 3.3).

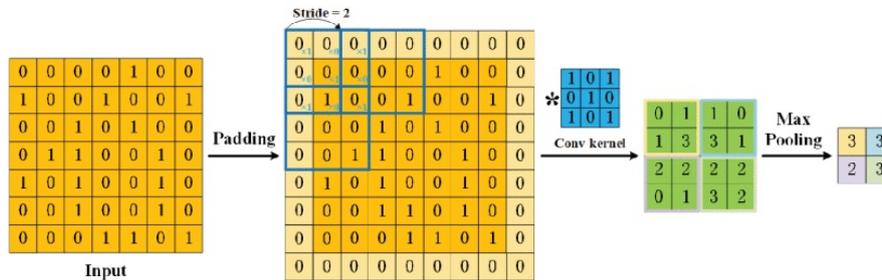


Figure 3.2 – An example of a 2-D convolution kernel Li et al. (2021). The hyper-parameters of a convolutional kernel include padding of the input, stride of the kernel, size of the kernel, and Max pooling size. The parameters of the model are the kernel values.

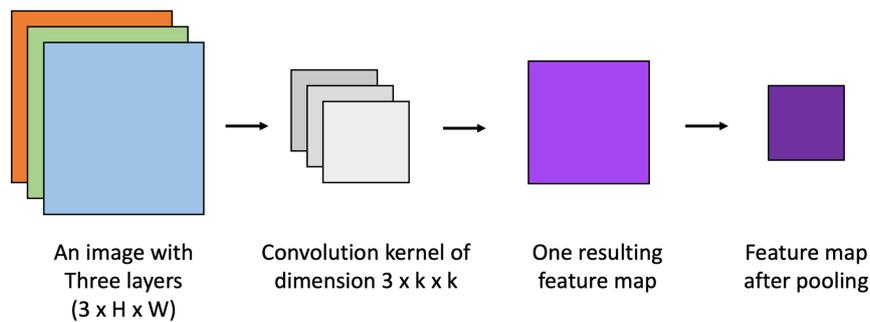


Figure 3.3 – The application of 2D convolutional layers on an image with three layers.

The structure of CNN offers weight sharing, local connection, and dimensionality reduction that are highly advantageous in solving image-level problems (classification) and pixel-level problems (object detection and image segmentation) (Li et al., 2021). Some downsides of CNN include computational cost and the need for a large amount of diverse data (Li et al., 2021). An efficient CNN for a complex problem often requires a deep and wide network, which could be challenging to deploy in real-time devices. Sometimes, a model compression technique is needed. CNN also requires many examples of a particular item to be able to efficiently recognize it with rotation and size changes. Another critique of a common technique in CNN is the pooling layer, which, although it offers many advantages, ignores the relationship between the local part of the image and the whole image and can sometimes lead to loss of spatial information during training (Li et al., 2021).

3.1.2 . Auto-regressive model

An auto-regressive model for image segmentation is based on the concept that a pixel x_i is influenced by the previous pixels that come before it (Menick and Kalchbrenner, 2018). In other words, it treats an image as a sequence of pixels,

$$p(x) = \prod_{i=1}^{n^2} p(x_i | x_1, x_2, \dots, x_{i-1}) \quad (3.1)$$

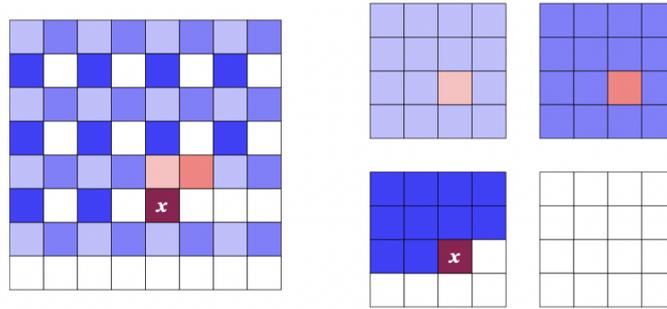


Figure 3.4 – Subscale Pixel for an image (Menick and Kalchbrenner, 2018). The input image is subsampled to allow better information sharing among pixels.

The auto-regressive model is often used to generate a pixel in an image using previously known pixels. It has the advantage of forcing its loss function to consider the entire data distribution as it uses the negative log-likelihood function. However, it is not sufficient for producing high-resolution images. In addition, in an image, a pixel is related to all the other pixels surrounding it, not only the pixels behind it. Pixel Subscale Network, a model for generating images proposed by Menick and Kalchbrenner (2018), makes use of all the pixels by dividing the image into sub-images (Figure 3.4). It increases the use of spatial information while reducing the computation cost.

The auto-regressive approach is often used with the help of a convolution kernel. For example, PixelCNN used masked convolution for pixel generation to not take into account the previous pixels (Van Den Oord et al., 2016).

3.1.3 . Transformer

The Transformer mechanism was first introduced in the paper "Attention is all you need" with impressive application on machine translation (Vaswani et al., 2017) and later became the foundation of many Natural Language Processing (NLP) models. A Transformer model is a sequence-to-sequence type of deep learning model that relies on attention mechanisms without needing any recurrent structure to consider the previous elements in a sequence. The attention mechanism for sequential models is first introduced by Bahdanau et al. (2014). The attention mechanism allows the model to learn how much attention to give to the sequence that comes before.

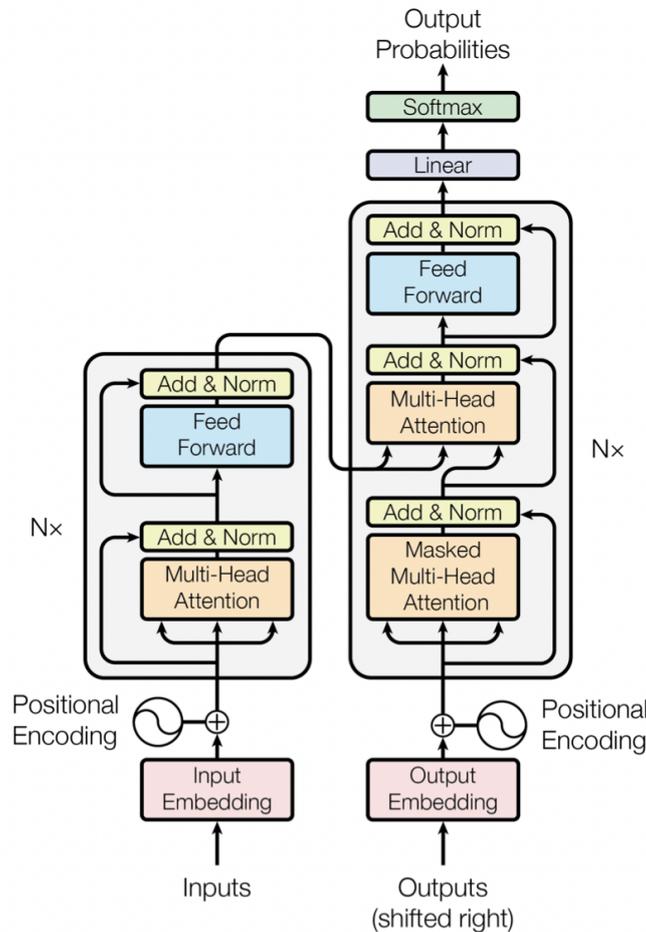
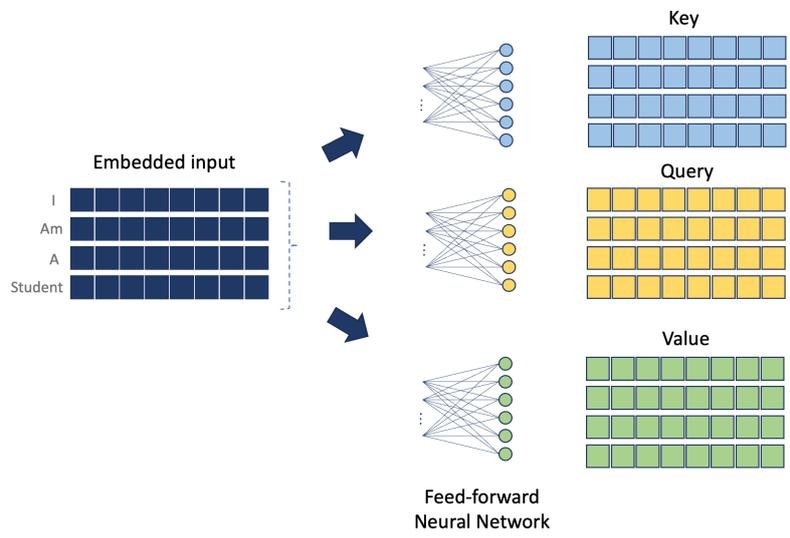


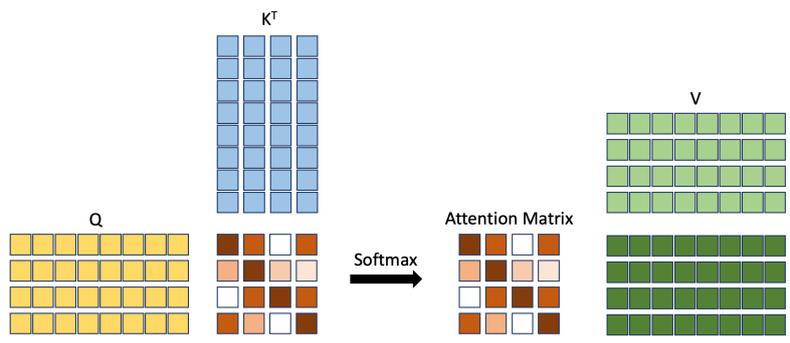
Figure 3.5 – Transformer architecture as presented in (Vaswani et al., 2017)

Figure 3.5 shows the architecture of the Transformer proposed by Vaswani et al. (2017). In NLP, the input is a phrase or a subset of a phrase and, therefore, needs to be embedded (this process is also called tokenization). As the Transformer does not contain any convolution or recurrent network, the positional encoding, such as cosine positional encoding, allows it to take into account the position or order of tokens. The input to the Transformer is the sum of the word embedding and the positional embedding. As a machine translation algorithm, the architecture has an encoder and a decoder. There is a cross-attention where the output of the encoder is used as an input to the multi-head attention blocks of the decoder.

The key element in the Transformer introduced by (Vaswani et al., 2017) is the self-attention module. Self-attention is a type of attention mechanism that relies only on the input. The first step is to calculate the Key (K), Query (Q), and Value (V) matrices from the embedded input. In the application of NLP, each token is a vector of values. The K , Q , and V are obtained by feeding the embedded input

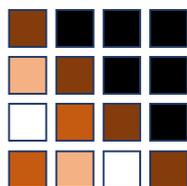


(a) Calculating the key, Query, and value from input



(b) Calculating the key, Query, and value from input

Masked Attention



(c) Masked attention

Figure 3.6 – The calculation of attention in attention mechanism.

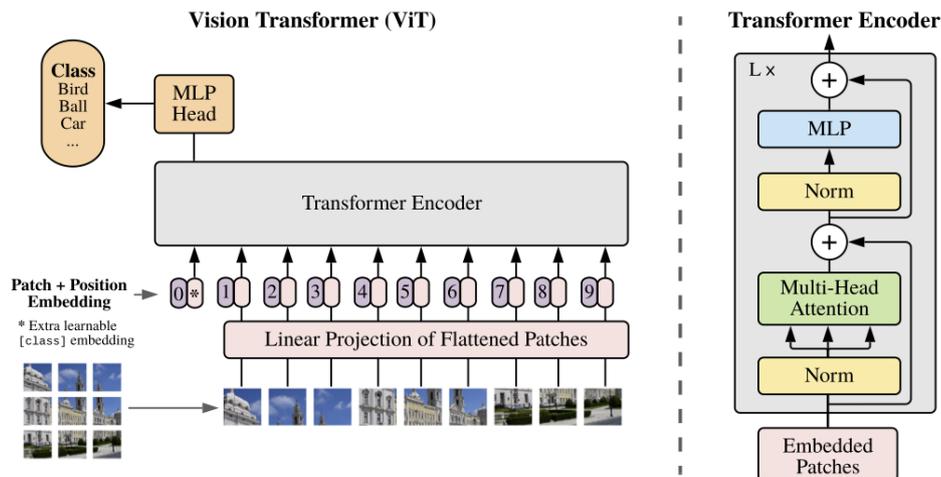


Figure 3.7 – Architecture of the Vision Transformer (ViT) (Dosovitskiy et al., 2020)

row by row into a different feed-forward neural network for each matrix (Figure 3.6a). Next, the softmax of the multiplication of Q and K^T gives the attention matrix. The softmax allows the resulting matrix to have each row summing to one and with values between 0 and 1 (Figure 3.6b). The intuition behind the attention matrix is that the parameters in the feed-forward neural network can be trained, and consequently, the model will be able to pay attention to relevant parts of the sequence. Then, the value V is multiplied by the attention matrix, which acts as the coefficient for each row. Similarly, the multi-head attention block is just the attention block divided into multiple parts that can then focus its attention on different things. The number of heads is a hyper-parameter of the model. The masked attention (figure 3.6c) replaces the upper triangle of the attention matrix with 0 to not consider the subsequent words in the sequence.

The Vision Transformer (ViT) is a modified Transformer for applying to image data (Dosovitskiy et al. (2020)). As an imitation of NLP data, the input image is transformed into fixed-size patches, which go through a linear projection to be flattened and then given a positional embedding. The embedded flattened patches are then fed into the Transformer Encoder. In the paper that proposed Vision Transformer (Dosovitskiy et al. (2020)), the goal is image recognition. Therefore, the output of the Transformer Encoder is fed into a Multilayer Perceptron (MLP) that classifies the image. The performance of ViT depends on pre-training on a large dataset. However, supervised training on a large dataset can be very costly.

It should be noted that besides the three approaches presented above, using graphs is another method, albeit not very popular, for treating an image while considering the surrounding pixels. For example, a Markov Random Field (MRF) is

a graph-based approach for image segmentation (Kato et al., 2012). For example, in an image, each pixel is a node. The MRF segments images based on energy function that favors smoothness. Consequently, it has a tendency to minimize the difference between two adjacent pixels.

3.2 . Unsupervised image segmentation

This section is divided into two categories : the pixel-wise and the non-pixel-wise approaches. The pixel-wise approaches refer to algorithms where the clustering relies on a form of distance-based metrics (spectral distance) between pixels without taking into consideration the structural information of the neighboring pixels. The algorithms that will be detailed here are K-Means, DBSCAN, OPTIC, and SLIC. On the other hand, non-pixel-wise models take into account the structure of the image, for example, as in convolution neural networks. The non-pixel-wise approaches are regrouped by the construction of their objective function, which can serve to train models without labeled data.

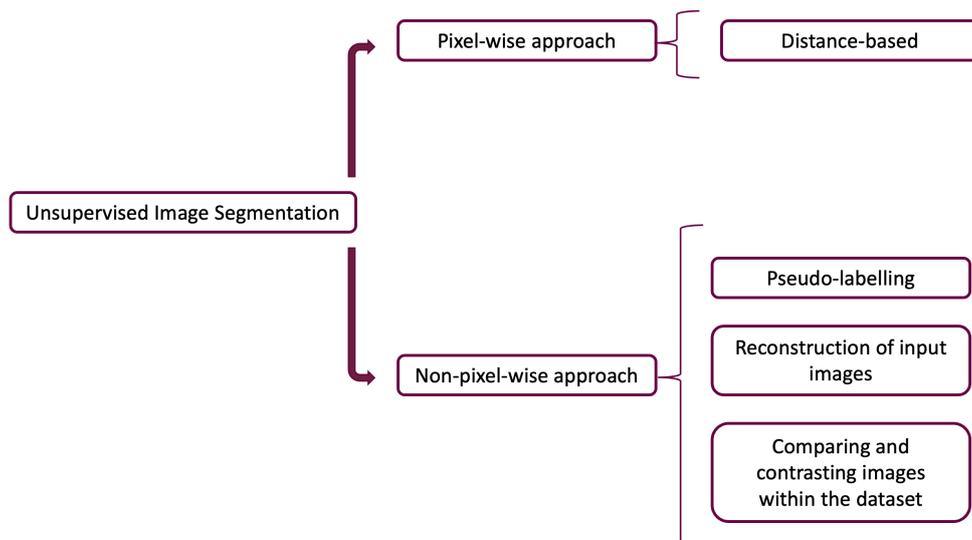


Figure 3.8 – The categories of unsupervised image segmentation approaches presented in this chapter

3.2.1 . Pixel-wise approach

The common thread in the algorithms here is that they classify the pixels together based on the distance, either between points or to the centroid of the cluster. The brightness temperature images are treated by pixels, with each pixel containing the chosen channels of brightness temperatures. We present two pixel-wise algorithms that are applicable to very large samples : K-means and DBSCAN.

The third approach, SLIC, is not very suitable for a large dataset of brightness temperature images. It is included due to its usefulness in aiding the non-pixel-wise approach (Section 3.2.2).

K-means

K-means clustering is a method for grouping unlabeled data points into a predetermined number of clusters using the distance between points. To separate the pixels, it relies on minimizing the inertia or the intra-cluster sum of squares of points from the centroids (Hastie et al., 2001). The K-means algorithm begins by choosing the centroids based on the predefined number of centroids and the specified approach. One basic initialization is to randomly select the centroids from the data. A more innovative approach, greedy k-means++ (Arthur et al., 2007), chooses the centroid to be very distant from each other. After the centroids are initialized, the points are assigned to the closest centroid. After this step, a new centroid is calculated based on the points in the clusters. These two steps are repeated until a criterion is reached; for example, there is no significant change between the old and the new centroid.

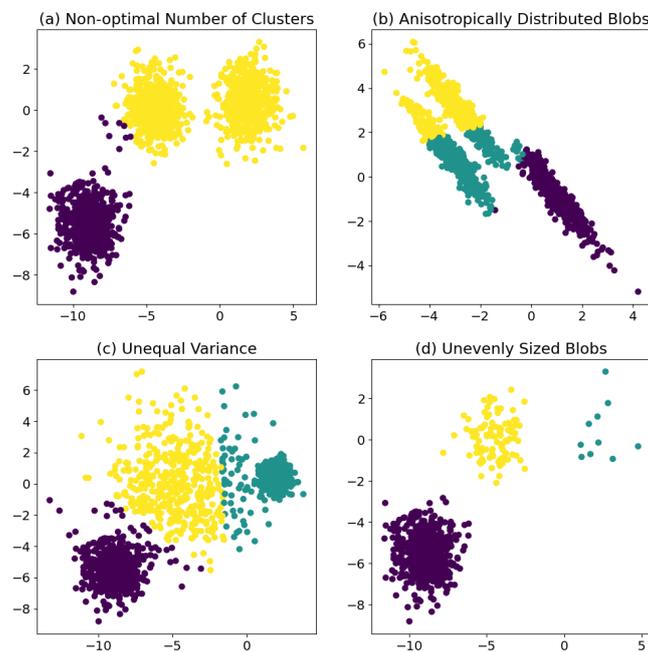


Figure 3.9 – Examples of possible issues in K-means clustering. Figure obtained using the Scikit-Learn Python Machine Learning library (Pedregosa et al., 2011)

However, as K-means minimizes inertia, it only works well on convex and isotropic data points (Sciikit-Learn User Guide, [Pedregosa et al. \(2011\)](#)). In addition, as a non-normalized metric, the inertia encounters problems in a higher dimension. Therefore, a dimension reduction algorithm is recommended. Figure 3.9 shows some problematic cases of the K-means algorithm. The K-means will not give a good partition if the number of clusters is insufficient. Therefore, it is necessary to understand the data well and perform silhouette and elbow tests to determine the number of clusters. For uneven sizes (Figure 3.9 (d)), the possible workaround is multiple initializations of K-means to avoid local minima. However, unequal variance and anisotropically distributed points are limitations of K-means.

When K-means is used for image segmentation, it classifies each pixel into a cluster based on its values. It does not take into account the structure within the image, which might be essential in the study of the brightness temperature images. For a large dataset, the mini-batch variation of K-means is more suitable for reducing computation time.

Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN considers clusters as high-density areas separated by low-density areas ([Ester et al., 1996](#); [Schubert et al., 2017](#)). As a result, unlike K-means, which only works well with convex shapes, it can deal with various shapes (Sciikit-Learn User Guide, [Pedregosa et al. \(2011\)](#)). DBSCAN works by finding sample cores, which are defined as a sample in the dataset that has a predefined number of core sample neighbors within a predefined distance. This results in core samples being in densely populated areas in the data. A cluster is then built by taking one core sample, finding its neighbors, then its neighbors' neighbors, and so on. A cluster also has non-core samples around its fringes. This approach is deterministic, giving the same clusters if the same order of data is given. However, this algorithm can be slow for very large datasets, which is the case in most image segmentation tasks, as it requires large memory.

Simple linear iterative clustering (SLIC) on super-pixels

Super-pixel can be defined as a group of pixels with similar characteristics, generally a color-based determination. Super-pixel is useful for segmentation because it carries more information than a single pixel. In addition, it can be effective in terms of computational power ([Achanta et al., 2010](#)). On the other hand, it can be challenging to determine the number of pixels and how to regroup them into clusters. If this process is not done correctly, it is impossible to have sharp edges in segmentation. Approaches involving super-pixels in image segmentation can be a method to regroup pixels based on their properties or an intermediate step for another unsupervised algorithm.

SLIC Super-pixel was introduced by (Achanta et al., 2010). Each super-pixel is approximately the same size. The super-pixels are regrouped together based on their similarity and proximity. The SLIC algorithm starts by initializing the cluster by taking pixels at a regular grid step. Then, iteratively, the pixels are associated with the nearest cluster center, and the new cluster centers are recalculated. These two steps are repeated until a certain preset threshold is met. This algorithm differs from K-means as it avoids the calculation of redundant distances. It is one among many speed-up schemes for the K-means algorithms.

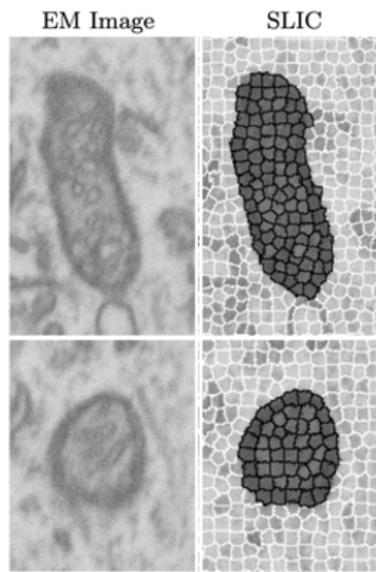


Figure 3.10 – SLIC algorithm on medical image

3.2.2 . Non-pixel-wise approach

In this section, non-pixel-wise approaches are discussed. These methods rely on the technique previously described in Section 3.1 : CNN, Auto-regressive models, and Transformers. These types of models are preferable as they can take into calculation the gradient of the brightness temperatures of the neighboring pixels as well as the structural information.

The main difference between the deep learning models described here is the training framework. Therefore, we will categorize the models by their training method, including creating pseudo-labels, reconstructing input images, and comparing/contrasting input images. In each category, the objective is to create the loss function to optimize the parameters. To better illustrate the framework, these training methods will then be discussed with a few examples.

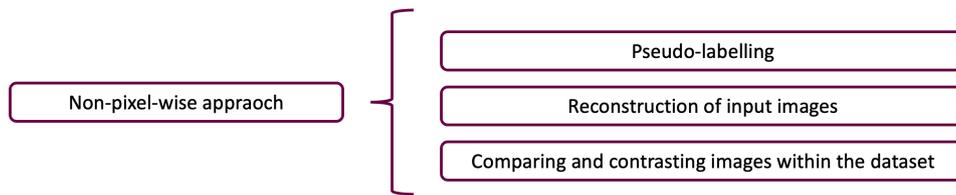


Figure 3.11 – Categorization of the non-pixel-wise approach.

3.2.2.1 Pseudo-labelling

In unsupervised learning, where there are no ground truth labels, models that rely on pseudo-labels create temporary labels that act as supervision in their training. These models use a simple approach (K-Means and SLIC) to create labels that serve a very similar role to the target labels in supervised learning. These models often contain feature extraction blocks.

The following examples are architectures that make use of pseudo-labeling. In the first example, the creation of the pseudo-labels is outside of the model and directly relies on the inputs (Separated pseudo-labeling), while in the two following examples, the creation of the pseudo-labels is integrated into the model and relies on the features extracted by the model (Integrated pseudo-labeling).

Separated pseudo-labeling

In this example, [Barthakur and Sarma \(2019\)](#) combines Seg-Net and K-means clustering. Seg-Net ([Badrinarayanan et al., 2015](#)) is an encoder-decoder network very similar to U-Net ([Ronneberger et al., 2015](#)). Instead of a concatenation of the feature maps like in U-Net, Seg-Net only uses the max-pooling indices from the contraction path and, therefore, requires less memory usage.

K-means is applied to the color space that is converted from the image. The regrouped clusters are then labeled. Next, these labels are used to train a Seg-Net model. The final result is the output of the Seg-Net containing a label for each pixel. Figure 3.12 shows the architecture of this model.

However, this model takes only one image as input. Consequently, it cannot be generalized to multiple images with different contents in the image.

Integrated pseudo-labelling

[Kanezaki \(2018\)](#) proposed a method for unsupervised image segmentation by combining a convolutional approach and super-pixels regrouping of pixels. The pseudo-labeling is a refinement of the output of the model using SLIC super-pixels. It is developed based on three criteria : the constraint on feature similarity, the constraint on spatial continuity, and the constraint on the number of unique

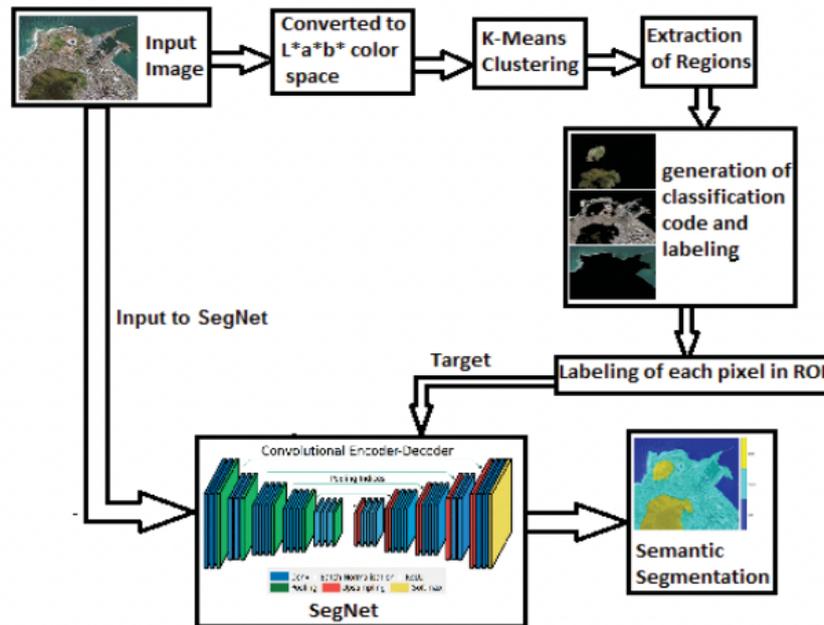


Figure 3.12 – Architecture with CNN and K-Means for unsupervised image segmentation (Barthakur and Sarma, 2019).

cluster labels.

The unsupervised training of the model by back-propagation works in two steps : the forward pass to get the predictions of the clusters, followed by a backward process of training the parameter. More precisely, the model gets the feature maps from the input using convolution layers. Then, these so-called response maps are normalized. Next, cluster labels can be obtained by an argmax function. The next step is a super-pixel refinement step in order to satisfy the spatial continuity criterion. Using the SLIC method (Achanta et al., 2010), the pixels are regrouped into super-pixels, and the most frequent label among the containing pixels is assigned. A softmax loss is calculated between the obtained cluster labels and the normalized response map. All the parameters of the network are updated by this loss. Figure 3.13 shows the architecture of this network.

The disadvantage of this method is that it passes over the images one by one. As a result, there is a problem with generalization as we cannot get the same class labeled as the same thing over different images.

In a second example of the integrated pseudo-labeling, (Kim et al., 2020) proposed a novel CNN method for unsupervised image segmentation, which is an update on their last architecture (Kanezaki, 2018). It is very similar to the previous method but without the reliance on super-pixel. The outputs of the feature

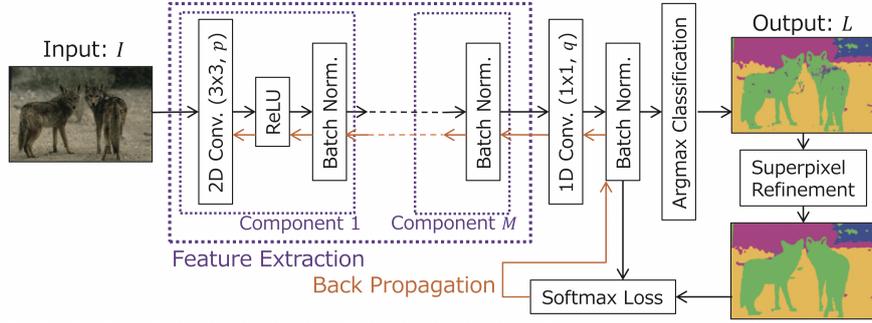


Figure 3.13 – The architecture for the unsupervised image segmentation by back-propagation using SLIC (Kanezaki, 2018).

extraction blocs pass through another batch normalization to become the normalized response map. The model uses this normalized response map pass through an argmax to create the pseudo-target to train in an unsupervised approach. The architecture of the model is shown in Figure 3.14.

The loss function is calculated using the response map. It consists of two terms, one for feature similarity and one for spatial continuity (Equation 3.2).

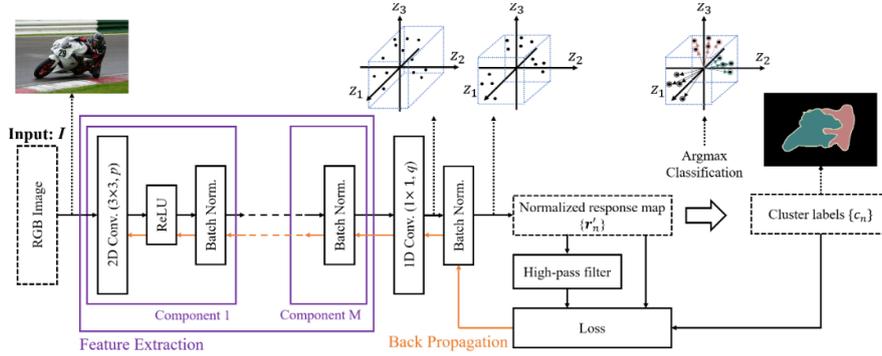


Figure 3.14 – The architecture proposed by Kim et al. (2020) as an improvement to Figure 3.13

$$L = L_{sim}(\{r'_n, c_n\}) + \mu L_{con}(\{r'_n\}) \quad (3.2)$$

The feature similarity loss compares the response map r'_n to the cluster labels c_n obtained by applying the argmax function on the response map. In comparison to the supervised segmentation approach, the cluster labels c_n are used as a pseudo-target. The loss, as a result, is a cross-entropy loss between c_n and r'_n .

The spatial continuity loss imposes similarity between adjacent pixels. It is the sum of the pixel difference in the vertical and horizontal direction in the response map r'_n (Equation 3.3).

$$L_{con}(\{r'_n\}) = \sum_{\xi=1}^{W-1} \sum_{\eta=1}^{H-1} \|r'_{\xi+1,\eta} - r'_{\xi,\eta}\| + \|r'_{\xi,\eta+1} - r'_{\xi,\eta}\| \quad (3.3)$$

3.2.2.2 Reconstruction of input images

The reconstruction of input images is another method for creating the loss function and adding extra information for the unsupervised segmentation. Usually, in this approach, the segmentation happens in an intermediate step before the reconstruction of the input image.

Two examples are presented here. In the first example, the objective function is the weighted sum of the segmentation loss and the construction loss. In the second example, the objective function of Generative Adversarial Nets (GAN) (Goodfellow et al., 2014) is used.

Weighted sum loss

W-Net (Xia and Kulis, 2017) makes use of the very successful supervised semantic segmentation U-Net (Ronneberger et al., 2015) by extending it from a U-shape to a W-shape. It contains two fully convolutional networks, with the first one (U_{Enc}) encoding the input image into a k-way segmentation and the second one (U_{Dec}) reversely reconstructing the image (Figure 3.15). The first loss is the soft normalized cut loss for the segmented images used only for updating U_{Enc} . The second loss is the reconstruction loss comparing the generated image and the original image used for updating the whole W-Net.

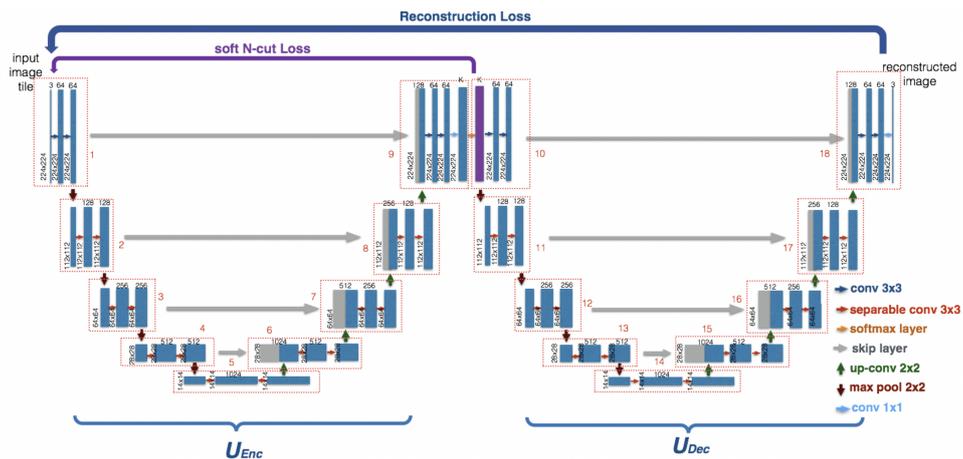


Figure 3.15 – W-Net architecture (Xia and Kulis, 2017).

Normalized cut loss is the application of graph theory in image segmentation. Each pixel and its value become the nodes in a graph. The soft normalized cut loss

(Shi and Malik, 2000) is the sum of weights of the ratio between cut (two disjoint subsets of a graph) and association (the total connections from nodes in a cut to all the nodes in the graph). The weights, on the other hand, signify the distance in locations and color values between nodes. In other words, the N-Cut loss could also be viewed as the ratio between the weights of the nodes in a cut to the nodes outside of the cut and the weights between these nodes to all nodes in the graphs, including those in the same cut.

Adversarial loss

GAN is very useful in image reconstruction as it is one of the most common methods for image-to-image generation. GAN is a generative model via an adversarial process first proposed by Goodfellow et al. (2014). It consists of two components. The generator, G , creates fake images resembling the input data and aims to learn the distribution of the input data. The discriminator, D , is trained to classify if an image is generated by G or if it is from the training data. The generator G aims to fool the discriminator D . As a result, the two functions have an adversarial relationship. Once convergence is achieved, the training samples and the generated samples are indistinguishable from each other. This adversarial process became the basis of many unsupervised image segmentation methods.

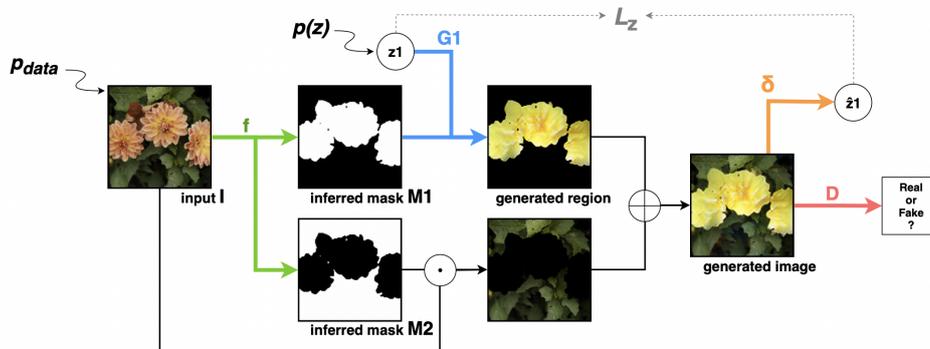


Figure 3.16 – ReDO architecture (Chen et al., 2019)

ReDO (Chen et al., 2019) is a GAN-based approach to object segmentation (Figure 3.16). It consists of three main steps : the segmentation step, the generation step, and the discrimination step. The concept of ReDO is based on the assumption that there is information in the segmented image for image generation. The generative process itself has three steps : the composition step for the segmentation, the generating process for each region, and the assembling step, where the two generated regions are combined. Then, the discriminator tries to correctly label the real and fake images. The final product is only the segmenter that provides labeling for the pixels of the object of interest and the background.

3.2.2.3 Comparing and contrasting images within the dataset

Comparing or contrasting images can allow the model to find the relevant information for the semantic segmentation of an image. The comparison can be made to an augmented version of the input image or a different image in the dataset.

Comparing the augmented inputs

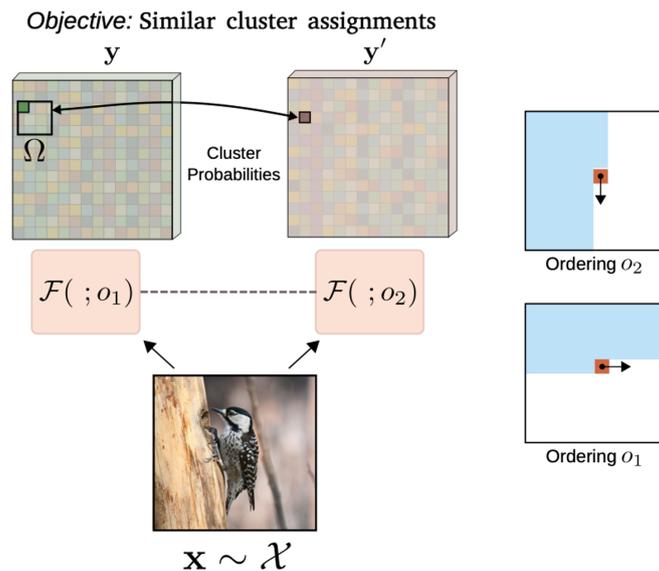


Figure 3.17 – Autoregressive Unsupervised Image Segmentation (Ouali et al., 2020). Two ordering of the image o_1 and o_2 goes through the encoder-decoder \mathcal{F} . The objective is to have similar cluster assignments regardless of ordering.

Built on Mutual Information, Ouali et al. (2020) proposed to compare two different views of the inputs obtained from different ordering of the pixels. In an Autoregressive Clustering approach, the two views are fed into an encoder-decoder network (\mathcal{F}) that produces probability distribution of cluster assignment for all the pixels (Figure 3.17). The pixels are then re-ordered to their original position. The goal is to have the same distribution regardless of the views. There are many possible views or ordering of the pixels, as it can also be a zig-zag pattern.

Comparing input with a different image in the database

STEGO (Self-supervised Transformer with Energy-based Graph Optimization) is trained using an image similar to the input and an image completely different (Hamilton et al., 2022). The architecture of STEGO (Figure 3.18) contains first a backbone for extracting the global image features used for finding and comparing similar and different images (Hamilton et al., 2022). This backbone does not require to be re-trained or fine-tuned, therefore allowing STEGO to be very efficient in

training. The feature extraction in the backbone is done by global average pooling (GAP). GAP was first proposed to replace the fully connected layer in CNN for classification (Lin et al., 2013). Instead of flattening the feature map and feeding it into a fully connected MLP, each feature map is averaged. The advantage of GAP is that it allows a more native representation of the convolution structure by enforcing correspondences between categories and feature maps. In addition, it also reduces over-fitting as there is no parameter to train. Based on the cosine similarity, a look-up table of each image's K-Nearest Neighbors is constructed from the backbone's feature space. The output of the backbone is also used for the input of the Segmentation Head, which is an MLP (Figure 3.18).

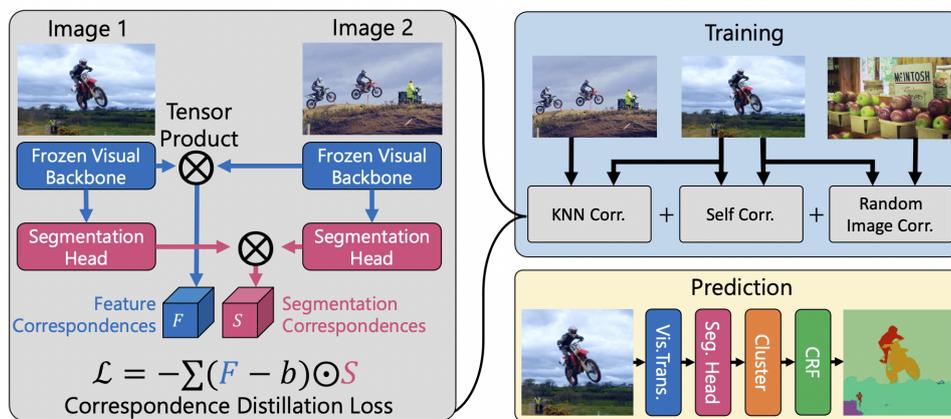


Figure 3.18 – STEGO architecture and prediction scheme (Hamilton et al., 2022)

In the prediction phase, the outputs of the Segmentor head, which are segmentation features, go through a clustering step (cosine distance-based minibatch K-Means) in order to obtain cluster labels. Next, there is a refinement step using Conditional Random Field (Krähenbühl and Koltun, 2011).

3.3 . Self-supervised image segmentation

Self-supervised learning learns one part of the input from another part of the input, a process called pretext learning. It uses the supervisory signals from the input data. One way to do this is by training a predictive model that tries to predict the missing part of the data from the available ones, for example, predicting the next part of a video given the previous sequence. This can be done by using an energy-based model. For instance, if the sequence x and y are compatible, they are trained to have low energy, and if they are not, the energy is high. This is done through a two-part process (LeCun and Misra, 2021) :

- (1) When two compatible sequences x and y are shown to the model, it is

trained to produce a low energy,

(2) Impose on the model for a particular x , an incompatible value y produce higher energy than a more compatible value y

The difference between self-supervised learning and unsupervised learning is that they have different objectives. Self-supervised learning aims to learn the data feature, whereas unsupervised learning has a more direct and specific goal. Another interesting approach is the weakly supervised semantic segmentation, which does not rely on pixel-level labels as a fully supervised learning approach. It can use weakly labeled information such as bounding boxes, scribbles, and picture-level labels (Wang et al., 2020).

Using self-supervised learning means defining pretext tasks and downstream tasks. In computer vision, a pretext task means learning the features within the images to use in another task. These pretext tasks in treating images include in-painting and colorization. Self-supervision in pretext tasks can also act as a pretraining step that could be used to transfer knowledge to downstream tasks such as object detection, image segmentation, image classification, etc.

This section explores some models in self-supervised learning for semantic image segmentation. We will discuss two main types of architectures in self-supervised learning : Vision Transform (ViT) and Convolutional Neural Networks (CNN). One prominent approach that links both ViT and CNN models is to use an attention mechanism, especially self-attention in the pretext task. In each category of architecture, we also describe several training frameworks based on contrastive learning, which has gained a lot of success in self-supervised learning of image representation.

3.3.1 . Self-supervised Vision Transformer

Despite its success in NLP, the Transformer encountered many challenges in computer vision. ViTs, in particular, are competitive in performance with CNN. Many publications have proposed supervised learning approaches for image classification using ViTs. However, ViT applications are still very limited compared to CNN models, as they demand a large database for training and are very costly in computation. The success of Transformer in NLP might be due to the self-supervised pre-training step. Caron et al. (2021) proposed to self-train DINO (self-distillation with **no** labels) without using any labels.

DINO (Caron et al., 2021) made use of the ViT and Knowledge Distillation (Hinton et al., 2015) to implement self-supervision. Self-training DINO is done by using two identical ViTs, with different parameters, called the student and the teacher network (Figure 3.19). The student and teacher networks are given two different crops of the input image. The loss is the similarity of the two outputs of the teacher and student network. During training, the teacher is only updated

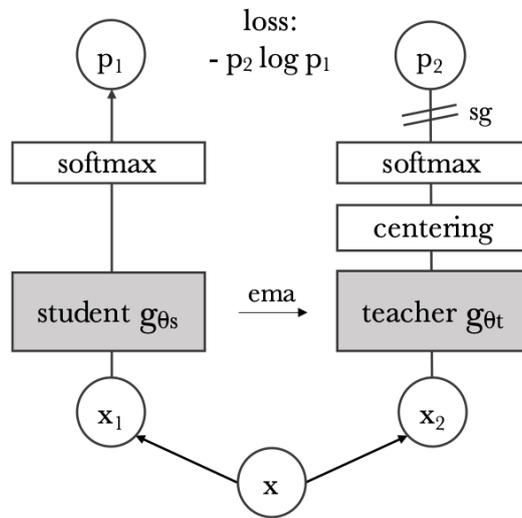


Figure 3.19 – DINO framework (Caron et al., 2021). The "sg" (stop gradient) in the teacher network indicates that the weights are not updated by the gradient descent. Instead, the teacher's weights are updated through exponential moving average (ema) from the student network.

using the exponential moving average (EMA) of the student network.

In self-training methods, where the model tries to compare and contrast the input image, the data is augmented for the local and global crops. The issue with data augmentation is that most of the conventional approaches used on RGB data cannot be applied to brightness temperatures. These conventional data augmentation techniques involving the color space include changing the color value, grayscaling, and blurring, which can cause the physical meaning in TB to change. One simple solution is to test the omission of these types of data augmentations and use only some of the geometric augmentations that do not change the containing structures, such as random flips, random crops, and random rotations. Other solutions include transforming TB data in a way that preserves its gradient and values, for example, randomly rotating the cropped images.

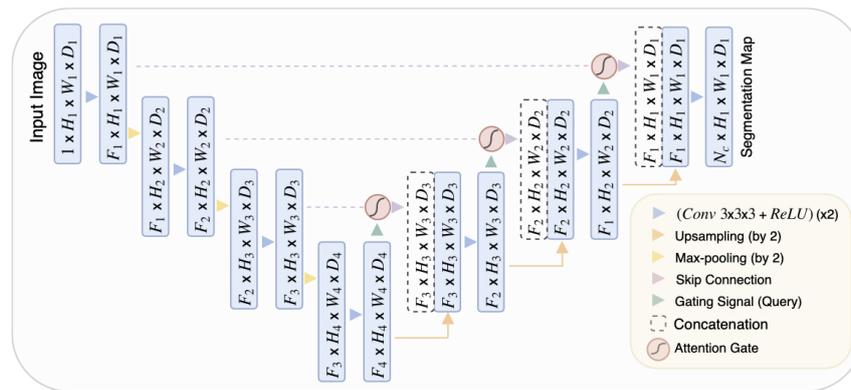
The obtained features from the self-supervised training of DINO contain explicit information about the input data that can then be used for a downstream task, such as image classification and segmentation. These emerging features appear more clearly than supervised ViTs and convolutional neural networks (Caron et al., 2021). In DINO, one of the downstream tasks is image classification. Using k-Nearest-Neighbors (k-NN), which requires some labeled data, the features are classified into different categories.

In our case, to do semantic segmentation without labeled data, we need methods that analyze and make use of the information in the attention output. For

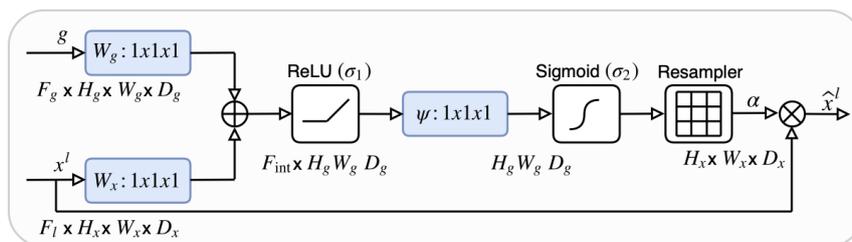
example, Grad-CAM analyzes the attention map (Selvaraju et al., 2017). Other approaches include using K-means (Ziegler and Asano, 2022).

3.3.2 . Self-supervised CNN

The attention mechanism is a fundamental part of the Transformers. However, it can be adapted and applied to CNN. For instance, Attention U-Net is built upon the architecture of U-Net (Oktay et al., 2018). In the skip connection part, the current layer and the previous block layer are fed into an attention gate before concatenating to the upsampled layer. The attention gate is shown in figure 3.20b. The two inputs, x^l and g go through a 1×1 convolution that equalizes the number of channels without changing the size. Their sum then passes through a ReLU activation, another 1×1 convolution layer, and a Sigmoid activation to flatten the result to between 0 and 1.



(a) Architecture of U-Net with attention mechanism



(b) Attention gate

Figure 3.20 – U-Net with attention mechanism (Oktay et al., 2018).

Self-supervised learning on images aims to learn the visual representations in an image. Techniques in training a model for this task without human supervision are mainly based on generative and discriminative approaches (Chen et al., 2020). These generative approaches can be computationally expensive as they need to generate many pixels in the input. In discriminative approaches, contrastive learning in the latent space has gained a lot of popularity. Chen et al. (2020) proposed a

training framework, SimCLR (Simple Framework for Contrastive Learning of Visual Representations), where the input goes through two different stochastic data augmentation, then two different encoders. The outputs of the encoder are projected (using MLP) to a latent space where the contrastive loss is then estimated. Similarly to the contrastive learning applied to ViT, data augmentation has an important role in the performance of the model (Chen et al., 2020).

Also focused on the representation in the latent space, Bootstrap Your Own Latent (BYOL) is a self-supervised learning framework with a smaller drop in performance than SimCLR when only random crop is used as the data augmentation approach (Grill et al., 2020). BYOL, similar to SimCLR, also creates two different representations and projections of the augmented input by two encoders and projectors with the same architectures and different weights. One side is called the online networks, where the weights are updated for each batch. The other side is the target networks that are updated by the moving average of the online networks.

3.4 . Summary of Methods

Some of the unsupervised segmentation approaches presented in the previous section (Section 3.2) are tested. Here, we summarize the models into three categories in Table 3.1 : (1) if a model is selected for further study (Selected), (2) if it is not suitable (Not suitable), or (3) if it has potential but was not tested (Potential).

The chosen models are presented in detail in terms of training and validation in Chapter 4.3. The selected models moving forward are K-means and the model proposed by Kim et al. (2020) called "Unsupervised learning of image segmentation based on differentiable feature clustering". They offer promising and coherent segmentation.

Table 3.1 – Summary of unsupervised approaches presented

Pixel-wise		K-means	Selected
		DBSCAN	Not suitable
		SLIC	Not suitable
Non-pixel-wise	Pseudo-labeling	Barthakur and Sarma (2019)	Not suitable
		Kanezaki (2018)	Not suitable
		Kim et al. (2020)	Selected
	Reconstruction	Xia and Kulis (2017)	Not suitable
		Chen et al. (2019)	Not suitable
	Compare and contrast	(Ouali et al., 2020)	Potential
(Hamilton et al., 2022)		Potential	

Among the "Not suitable" models, the model by Kanezaki (2018), the first iteration of the model by Kim et al. (2020), was not retained due to its inferior results to the updated model and its issue with generalization. Similarly, (Bartha-

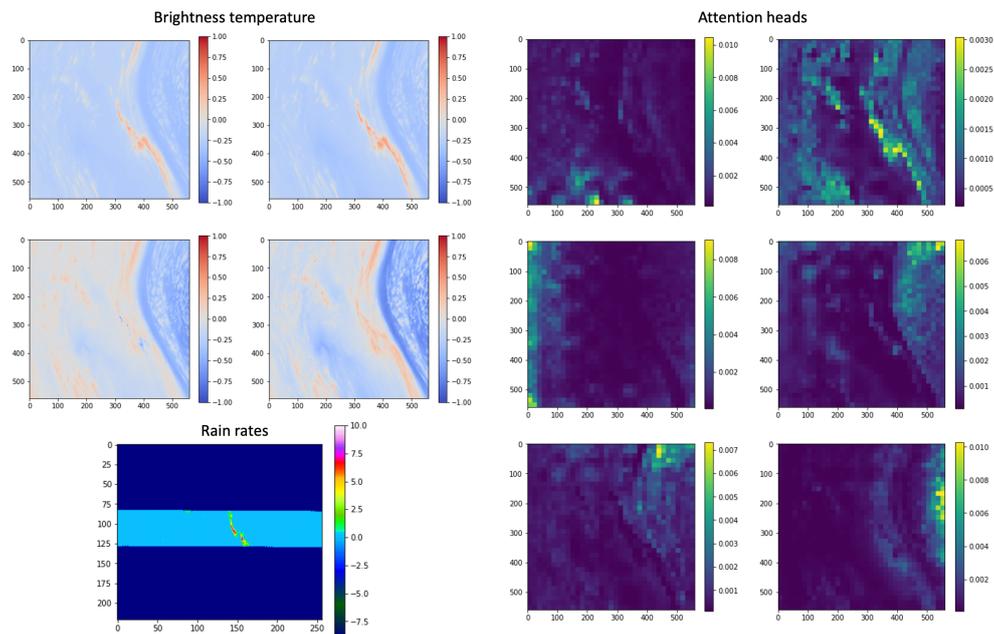


Figure 3.21 – Preliminary result of DINO. On the left are cropped brightness temperatures from GMI observation used as inputs, and the DPR used just for comparison. On the right are the attention outputs from the multi-attention heads of DINO.

kur and Sarma, 2019) also has generalization problem. On the other hand, W-Net (Xia and Kulis, 2017) was tested, and the resulting segmentation did not show any coherent structure. Another model in the "Reconstruction" category, the ReDO model (Chen et al., 2019), was not tested but deemed unsuitable due to the fact that it relies on the shape of the segmented objects in the image to reconstruct the input. Structures found in brightness temperature are highly irregular and, therefore, could be undesirable in this technique. In the pixel-wise approach, DBSCAN was tested and was too slow compared to its counterpart, Mini-batch K-means. SLIC is also not applicable as it is better suited for the segmentation of one image at a time.

The rest of the models were not tested. Nonetheless, their architectures and training frameworks might hold some potential for the application of brightness temperature. For example, the "Compare and contrast" approaches are both marked as "Potential" because they are quite different from the previous approaches and might offer a new perspective.

The self-supervised models are not as extensively explored as the unsupervised approaches in this study. However, preliminary results show that they have the potential to offer geophysically meaningful segmentation. Figure shows the preli-

minary results from the DINO algorithm. The images on the left in Figure 3.4 show the brightness temperatures used as inputs to DINO and the rain rates used for comparison. The images on the right are the attention head (self-attention block, Section 3.1.3) output of the DINO model (Section 3.3.1). The data augmentation is omitted in this test. We could observe some promising similar structures present in the input brightness temperatures and the attention head outputs. An important notion of the self-supervised approaches is that they require a downstream task that makes use of these attention outputs to create a segmentation.

Attention mechanisms normally used in self-supervised learning can also serve as a backbone to an unsupervised approach. For example, U-Net architecture used in unsupervised segmentation can be replaced by Attention U-Net (Okta et al., 2018).

3.5 . Evaluation Metrics for Unsupervised Image Segmentation

In the case of unsupervised image segmentation, where there is no ground truth readily available, the evaluation of the results is a very challenging task. The inputs to the model consist of only the brightness temperature images. The resulting segmentation regroups the brightness temperature in a way that is unknown to us. Consequently, the evaluation of the segmentation is done in two steps : the internal and external evaluations. In internal evaluation, the same variables used for training are used in order to find the interclass and intraclass dispersions between each class. External evaluation, on the other hand, compares the segmentation to other relevant variables, such as precipitation, sea surface temperatures, etc. To compare to external variables, the segmentation classes are labeled with these geophysical variables.

These two approaches described above evaluate the segmentation pixel-by-pixel. Approaches that use a spatial correlation may provide more information regarding the structure of the obtained segmentation and an external variable.

3.5.1 . Internal Evaluation

The internal evaluation uses only the variable used for the segmentation, which is the brightness temperature in our case. It is mainly comprised of the distance between the intra-class and inter-class classes. However, this approach might be more favorable to pixel-to-pixel approaches such as K-means. It also depends on the number of segmentation classes. Two measures are presented here : Calinski-Harabasz (Calinski and Harabasz, 1974) and Davies-Bouldin index (Davies and Bouldin, 1979).

Calinski-Harabasz

Calinski-Harabasz index is the ratio between the sum of inter-cluster dispersion and intra-cluster dispersion (Calinski and Harabasz, 1974; Pedregosa et al., 2011). Therefore, a higher Calinski-Harabasz score means a better-defined cluster. For a dataset E of size n_E and center c_E , clustered into k clusters, the index is defined as,

$$s = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_E - k}{k - 1} \quad (3.4)$$

where $tr(B_k)$ is the trace of the between-group dispersion with

$$B_k = \sum_{q=1}^k n_q (c_q - c_E)(c_q - c_E)^T \quad (3.5)$$

and $tr(W_k)$ the trace of within-cluster dispersion with

$$W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q)(x - c_q)^T \quad (3.6)$$

Here, for a cluster q , C_q is the set of points, and c_q is the center.

Davies-Bouldin index

The Davies-Bouldin index measures the similarity between clusters. The similarity for cluster i and j , R_{ij} , is defined by the distance between clusters and the size of the clusters (Davies and Bouldin, 1979; Pedregosa et al., 2011). The Davies-Bouldin index is then the average of the similarity between each cluster and its most similar cluster. It is defined as,

$$DB = \frac{1}{k} \sum_i \max_{i \neq j} R_{ij} \quad (3.7)$$

where $R_{ij} = \frac{s_i + s_j}{d_{ij}}$. R_{ij} , therefore, is a trade-off between the sum of cluster diameters s_i and s_j (average distance between each point of the cluster to its centroid) and the distance d_{ij} (distance between cluster centroids c_i and c_j). The lower score indicates a better-defined cluster, with zero as the lowest possible score.

3.5.2 . External evaluation

External evaluation involves using a new variable that was not in the training database. This comparison should be made to relevant data that can be retrieved from the brightness temperatures, such as precipitation and sea surface temperature. Due to the nature of these data, this approach requires a conversion step of a quantitative variable to a categorical one. For instance, when precipitation is

compared to the resulting segmentation, the rain intensity needs to be converted to a categorical variable, for example, Rain vs. Non-rain.

After the conversion of the quantitative variables into categorical variables, the metrics for comparison are the same as those used in a supervised segmentation approach. These metrics include the Rand index, the Jaccard index (or IoU), the Dice coefficient (or F1-score for images), accuracy, and precision.

Rand index

The Rand Index (Hubert and Arabie, 1985) allows a comparison between two segmented images with different labeling. It counts the number of pairs of pixels being in the same class for both images as well as the pairs that are in different classes for both images. Given two set S and S' with labels l_i and l'_i , Rand index is given by,

$$R(S, S') = \frac{1}{\binom{N}{2}} \sum_{i,j,i \neq j} [I(l_i = l_j \wedge l'_i = l'_j) + I(l_i \neq l_j \wedge l'_i \neq l'_j)] \quad (3.8)$$

where I is an identity function. However, the Rand Index can be sensitive to images with slightly different segmentation borders due to the comparison between pairs of pixels. In addition, in the case of comparison with precipitation, rain structures are small and rare in the images. These small structures can be overlooked in the score. It is also very dependent on the number of clusters. For example, a random segmentation with a high number of clusters can have a high Rand index (Wagner and Wagner, 2007).

The Adjust Rand Index is the Rand Index adjusted for chance (Hubert and Arabie, 1985). The Rand Index is adjusted using the expected similarity provided by a random model. Adjusted Rand Index is advantageous when the number of predicted clusters is much higher than the number of ground truth clusters or vice versa (Sciikit-Learn User Guide, Pedregosa et al. (2011)).

IoU and DICE

Intersection over Union (IoU), or the Jaccard Index, is the ratio between the intersection of the pixels of a class over the union of this class in two segmentations of the same image. The Dice or Sørensen-Dice coefficient for comparing the predicted and ground-truth segmentations is the ratio between twice the intersection of the predicted and ground-truth pixels divided by the total number of pixels. For binary variables, the Dice score is also known as the F1-score in image segmentation. Although it is used in more scientific publications, Dice penalizes under and over-segmentation less compared to IoU (Müller et al., 2022).

Accuracy, precision, recall, F1-score

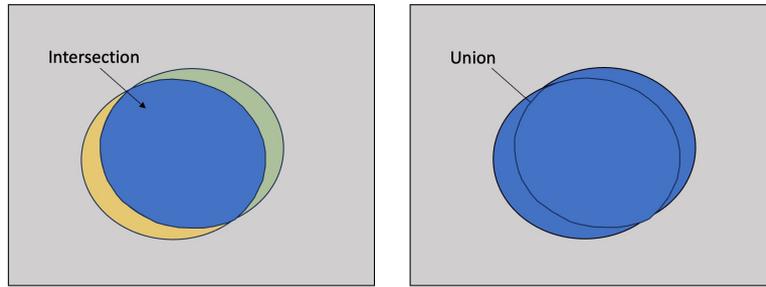


Figure 3.22 – Visual representation of intersection and Union.

Accuracy, precision, recall, and F1-score are the traditional metrics used in the evaluation of image segmentation. Accuracy is simply the ratio between the correctly predicted pixels ($TP + TN$) over the total number of pixels. Accuracy is a good indicator of performance only when the classes are balanced. In the case of imbalanced classes, i.e., a low number of rain pixels, a relatively high accuracy might be misleading as the model might miss most of the minority class (Rain pixels). Precision is the ratio between the TP and all the predicted positive pixels, whereas Recall is the ratio between the True Positives and all the positive pixels. For an imbalanced dataset, a segmentation might achieve a good precision score and low recall score if it only captures a few positive pixels but correctly predicts most of those pixels. On the other hand, we can have a low precision but high recall if the segmentation captures a large percentage of TP but also has a lot of FP . F1-score is a better metric for an imbalanced dataset as it tries to balance between Precision and Recall.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

With TP for True Positive, FP for False Positive, and FN for False Negative.

3.5.3 . Spatial Correlation

The purpose of using spatial correlation is to compare the occurrence of each class obtained from the segmentation model to various geophysical variables. It is thus also an external evaluation.

V-measure differs slightly from the previously presented methods as it does not compare pixels to the corresponding pixels. Instead, it looks at pixels in each partition. V-measure is an entropy-based external evaluation that aims to compare two clusterings using homogeneity and completeness (Rosenberg and Hirschberg, 2007). An advantage of the V-measure is that it is independent of the labeling of the two clusterings, i.e., a permutation of the cluster label value will not change the score. It is also symmetric. For the same domain, we define a set of classes C and a set of clusters K to be compared. Homogeneity is satisfied when all points from a cluster belong to the same class. On the contrary, completeness is satisfied when all points of the same class belong to the same cluster. These two values change in opposite directions : increasing completeness leads to decreasing homogeneity and vice versa.

Nowosad and Stepinski (2018) proposed a modified V-measure to apply to a spatial context. Instead of Information Theory, the authors reformulate the method to an analysis of variance. As a result, the modified V-measure provides more information about the comparison, including indices of global association and local association indicators. For example, it can compare the segmentation from the unsupervised to a segmentation based on another variable, for example, precipitation.

Conclusion

This chapter offers an overview of architectures and training frameworks as well as some evaluation metrics for unsupervised image segmentation. The main categories are pixel-wise and non-pixel-wise approaches. Only distance-based pixel-wise approaches are presented. For non-pixel approaches, the categorization is based on its training framework : pseudo-labeling, reconstruction of inputs, and comparing/contrasting inputs. In addition, we include some pros and cons of the application of presented models on brightness temperature data. In the next chapter, the selected algorithms (details in Section 3.4) will be implemented.

3.6 . References

- R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels. *Technical report, EPFL*, June 2010.
- D. Arthur, S. Vassilvitskii, et al. k-means++ : The advantages of careful seeding. In *Soda*, volume 7, pages 1027–1035, 2007.
- V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet : A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015. URL <http://arxiv.org/abs/1511.00561>.
- D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *arXiv*, 2014.
- M. Barthakur and K. K. Sarma. Semantic segmentation using k-means clustering and deep learning in satellite image. In *2019 2nd International Conference on Innovations in Electronics, Signal Processing and Communication (IESPC)*, pages 192–196, 2019. doi : 10.1109/IESPC.2019.8902391.
- T. Calinski and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1) :1–27, 1974.
- M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021.
- M. Chen, T. Artières, and L. Denoyer. Unsupervised object segmentation by re-drawing. *Advances in neural information processing systems*, 32, 2019.
- Q. Chen, L. Yang, J.-H. Lai, and X. Xie. Self-supervised image-specific prototype exploration for weakly supervised semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4288–4298, 2022.
- T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.
- D. L. Davies and D. W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2) :224–227, 1979. doi : DOIifavailable.

- A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words : Transformers for image recognition at scale. *arXiv preprint arXiv :2010.11929*, 2020.
- M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.
- I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks, 2014.
- J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33 :21271–21284, 2020.
- M. Hamilton, Z. Zhang, B. Hariharan, N. Snavely, and W. T. Freeman. Unsupervised semantic segmentation by distilling feature correspondences. *arXiv preprint arXiv :2203.08414*, 2022.
- T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., 2001.
- G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv :1503.02531*, 2015.
- L. Hubert and P. Arabie. Comparing partitions. *Journal of classification*, 2 : 193–218, 1985.
- S. Jung, H. Heo, S. Park, S.-U. Jung, and K. Lee. Benchmarking deep learning models for instance segmentation. *Applied Sciences*, 12(17), 2022. ISSN 2076-3417. doi : 10.3390/app12178856. URL <https://www.mdpi.com/2076-3417/12/17/8856>.
- A. Kanezaki. Unsupervised image segmentation by backpropagation. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1543–1547. IEEE, 2018.
- Z. Kato, J. Zerubia, et al. Markov random fields in image segmentation. *Foundations and Trends® in Signal Processing*, 5(1–2) :1–155, 2012.
- W. Kim, A. Kanezaki, and M. Tanaka. Unsupervised learning of image segmentation based on differentiable feature clustering. *IEEE Transactions on Image Processing*, 29 :8055–8068, 2020.

- P. Krähenbühl and V. Koltun. Efficient inference in fully connected CRFS with gaussian edge potentials. *Advances in neural information processing systems*, 24, 2011.
- Y. Le Cun, L. Bottou, and Y. Bengio. Reading checks with multilayer graph transformer networks. In *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 151–154. IEEE, 1997.
- Y. LeCun and I. Misra. Self-supervised learning : The dark matter of intelligence, 2021. URL <https://ai.facebook.com/blog/self-supervised-learning-the-dark-matter-of-intelligence/>.
- Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou. A survey of convolutional neural networks : analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 2021.
- M. Lin, Q. Chen, and S. Yan. Network in network. *arXiv preprint arXiv :1312.4400*, 2013.
- J. Menick and N. Kalchbrenner. Generating high fidelity images with subscale pixel networks and multidimensional upscaling, 2018.
- D. Müller, I. Soto-Rey, and F. Kramer. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1) :210, 2022.
- J. Nowosad and T. F. Stepinski. Spatial association between regionalizations using the information-theoretical V-measure. *International Journal of Geographical Information Science*, 32(12) :2386–2401, 2018.
- O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al. Attention U-Net : Learning where to look for the pancreas. *arXiv preprint arXiv :1804.03999*, 2018.
- Y. Ouali, C. Hudelot, and M. Tami. Autoregressive unsupervised image segmentation. In *Computer Vision—ECCV 2020 : 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*, pages 142–158. Springer, 2020.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.
- O. Ronneberger, P. Fischer, and T. Brox. U-net : Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted*

intervention–MICCAI 2015 : 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, pages 234–241. Springer, 2015.

- A. Rosenberg and J. Hirschberg. V-measure : A conditional entropy-based external cluster evaluation measure. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, pages 410–420, 2007.
- E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu. DBSCAN revisited, revisited : why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 42(3) :1–21, 2017.
- R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam : Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8) :888–905, 2000.
- A. Van Den Oord, N. Kalchbrenner, and K. Kavukcuoglu. Pixel recurrent neural networks. In *International conference on machine learning*, pages 1747–1756. PMLR, 2016.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- S. Wagner and D. Wagner. *Comparing clusterings : an overview*. Universität Karlsruhe, Fakultät für Informatik Karlsruhe, 2007.
- Y. Wang, J. Zhang, M. Kan, S. Shan, and X. Chen. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12275–12284, 2020.
- X. Xia and B. Kulis. W-Net : A Deep Model for Fully Unsupervised Image Segmentation. *arXiv preprint arXiv :1711.08506*, Nov. 2017. URL <https://arxiv.org/abs/1711.08506v1>.
- A. Ziegler and Y. M. Asano. Self-supervised learning of object parts for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14502–14511, 2022.

4 - Training, Comparing, and Interpreting

Contents

4.1	Training dataset	87
4.2	Model architectures	89
4.2.1	FCN models proposed by Kim et al. (2020)	90
4.2.2	Our proposed FCN	91
4.2.3	K-means models	93
4.3	Comparison of segmentation models without ground truth	94
4.3.1	External comparison with precipitation	95
4.3.2	External comparison with sea surface temperatures	102
4.3.3	Model choice	103
4.4	Validating the characteristics of the segmentation classes with geophysical knowledge	104
4.4.1	Precipitation patterns over two regions	104
4.4.2	Precipitation over the ocean and the ITCZ precipitation band	111
4.4.3	Sea surface temperature (SST)	114
4.5	References	118

As described in Chapter 1, it has become increasingly important to study the evolution of the water cycle. However, detecting and quantifying changes in precipitation variability, and more generally in the evolution of the water cycle, remains highly challenging and uncertain. Ground-based measurements such as rain gauges are lacking in terms of coverage ([Kidd et al., 2017](#)), while datasets derived from reanalyses have significant limitations ([Dorigo et al., 2021](#); [Tarek et al., 2020](#)). In order to analyze the variability of the Earth’s climate based on spatial observations, this chapter proposes an analysis method using unsupervised segmentation of remote sensing images. The first step is to verify that the obtained segmentations are linked to geophysical variables. As spatial data have been available for several decades, the study of the temporal evolution of the various segments will eventually enable us to quantify the climate evolution observed by satellite.

This chapter focuses on the unsupervised segmentation of passive radiometric remote sensing observation of microwave brightness temperatures. The microwave spectrum provides diverse transmission conditions. Lower frequencies, near 10 GHz,

are very sensitive to liquid water emission and nearly transparent to the atmosphere, even with clouds and moderate rainfall, making them ideal for surface observations. In contrast, higher frequencies, near 90 GHz, are highly sensitive to iced precipitation through scattering and clouds but have reduced sensitivity to surface features due to significant atmospheric attenuation. These distinct characteristics make radiometric remote sensing observation well-suited for water cycle applications. For instance, microwave brightness temperatures are used for retrieving precipitation (Skofronick-Jackson et al., 2017), land surface temperature, sea surface temperature, wind direction, and snow cover.

After decades of observations, remote sensing big data has become an emerging field with numerous applications (Cui et al., 2018). Missions such as the Tropical Rain Measurement Mission (TRMM) (Kummerow et al., 1998) and the Global Precipitation Measurement (GPM) mission (Hou et al., 2014) have contributed to the accumulation of microwave brightness temperature images, which offer a valuable source of information about the Earth's water cycle over the last three decades. To extract this information, elements of the water cycle are retrieved from the radiometric observations with state-of-the-art algorithms (Viltard et al., 2023) (Kummerow et al., 2001; Pfreundschuh et al., 2022). Then, the resulting retrieved data is analyzed. If a Bayesian approach is used, the retrieval algorithm requires many auxiliary variables. The novelty of the proposed approach is that using an unsupervised model to analyze various elements of the water cycle does not require any help from any other variables besides the brightness temperature images. These images are inputs to unsupervised semantic segmentation models, and the resulting segmentation classes are studied.

These deep-learning models are trained and evaluated on conventional RGB images. In these images, objects have consistent structures or forms, such as cars, trees, pedestrians, animals, etc. This chapter aims to show the importance and the challenges in evaluating unsupervised semantic segmentation on unconventional images. Objects such as clouds, precipitation systems, and snow-covered surfaces do not have a predefined or consistent form in brightness temperature images.

In addition, in terms of evaluating the performance of the models, the challenge lies in labeling the obtained segmentation classes to a geophysical variable. This step requires comparing each class to multiple datasets to determine its meaning. Therefore, the quality of the segmentation is controlled through two steps : the comparison of segmentation from different model implementations and the validation of the segmentation. The first step aims to choose the best segmentation model by comparing its ability to find geophysical variables within an image compared to a reference dataset and the level of detail it provides in the segmentation. The second step consists of validating the segmentation capacity to find patterns and variability in geophysical variables. Here, precipitation and sea surface temperature are chosen as external variables.

This chapter aims to show that an unsupervised approach can offer a geophysically significant segmentation of brightness temperatures and tackle the challenges involved in validating and interpreting the segmentation. The structure of this chapter is as follows. Section 4.1 describes the data used in this study. Section 4.2 presents the three approaches in unsupervised segmentation tested in this study : a fully convolutional neural network (FCN) model proposed by Kim et al. (2020) (previously discussed in Chapter 3), an FCN-based model with attention mechanism, and a K-mean model. The hyper-parameters tested in the implementation of each approach are also presented. Next, Section 4.3 shows how the models are compared in the absence of ground truth labeling. After the model comparison, only one model is chosen for the next step. Finally, the characteristics and behavior of the segmentations obtained from the chosen model are analyzed in Section 4.4.

4.1 . Training dataset

The Global Precipitation Measurement (GPM) mission contains a constellation of satellites whose ambitions include unifying and improving the remote sensing measurement of precipitation from space, observing the global water cycle variability, and improving weather forecast (Hou et al., 2014). The cooperation between different agencies allows the GPM constellation to have both conical-scanning and cross-track-scanning microwave radiometers with channels between 6 and 183 GHz (Hou et al., 2014). The constellation contributes to better coverage by lessening the revisit time of precipitation observation. The GPM data has served many societal applications, including hurricane track prediction, flood monitoring, agricultural drought monitoring, and even disease tracking (Kirschbaum et al., 2017). This is done by, for example, using GPM data for assimilation and contributing to global precipitation datasets such as IMERG (Integrated Multi-Satellite Retrievals for GPM, Pradhan et al. (2022)).

The amount of brightness temperatures images observed using similar frequencies from the whole constellation offers an unprecedented opportunity for analysis by Computer Vision. The present study experiments with data from the GPM Core Observatory as a first step in processing brightness temperature images using an unsupervised approach. The GPM Core Observatory (GPM-CO), launched in February 2014, is a reference of the constellation and serves as a standard for unifying the satellites within the constellations. It carries a Microwave Imager (GMI) and a Dual-frequency Precipitation Radar (DPR). Since its launch, the GPM-CO has amassed up to 10 years of satellite observations. The availability of the more direct observation of precipitation by DPR offers the target data for a deep-learning rain retrieval (previous work, Viltard et al. (2023)).

GMI is a conically scanning radiometer that measures the brightness temperatures at a wide range of frequencies in both vertical and horizontal polarization (Hou

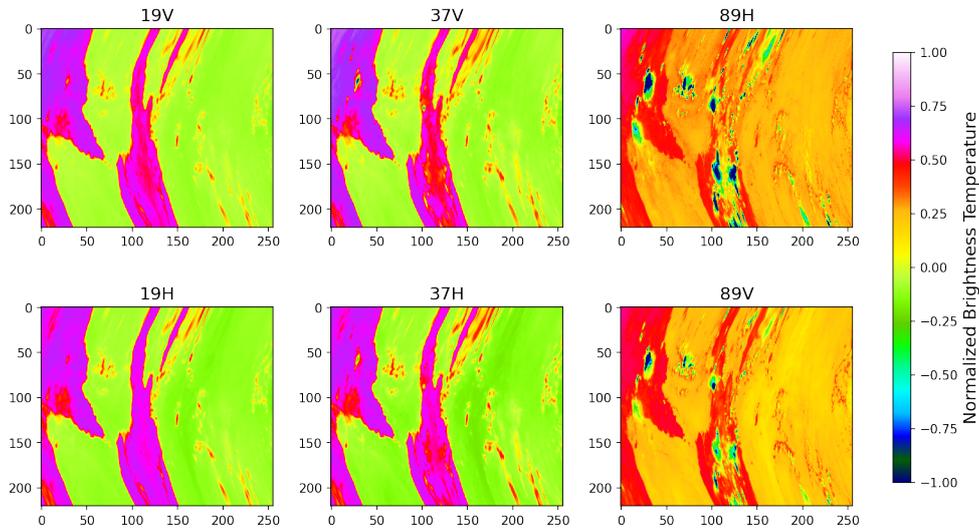


Figure 4.1 – Training example from cropped satellite observation on 13/01/2017.

et al., 2014). Its channels include 10.65, 18.7, 23.8, 36.6, 89.0, 166.0, 183.3+/-3, and 183+/-7 GHz. Its swath is approximately 885 km. The GPM-CO is on a non-sun-synchronous orbit with a 65° inclination, which gives its radiometer, GMI, a coverage of latitudes between 68°N and 68°S. This configuration, which allows it to complete approximately 16 revolutions around Earth per day, was chosen because it enables the satellite to observe precipitation between the Arctic Circle and the Antarctic Circle at different times of the day.

In the present study, we compare two choices of GMI channels as inputs to the model. The first choice is the 36.6 GHz (hereafter referred to as 37 GHz) and the 89 GHz in horizontal and vertical polarizations. The resolution of these channels is 8.6 km by 15.0 km and 4.4 km × 7.3 km, respectively. This choice is motivated by the resolution of the observation, the availability of similar channels across the satellites in the GPM constellation, and the previous study by Viltard et al. (2023) that has shown that these channels are very efficient for rain retrieval. The second choice is the inclusion of the 18.7 GHz (hereby 19 GHz, resolution 11.2 km × 18.3 km) to the previous two channels. This lower frequency channel is sensitive to surface emissivity and exhibits different sensitivities depending on the surface type : for continental surfaces, it is sensitive to variables such as soil moisture and vegetation, while for oceanic surfaces, it is primarily influenced by sea surface temperature (Carver et al., 1985).

The training database consists of 30,000 cropped satellite images of 221 pixels by 256 pixels, measured during 2017 and 2018. The measurements taken during 2019 are reserved as the test dataset for evaluating the performance. The brightness temperature images are normalized using their mean and standard deviation. Figure

4.1 shows an example of the cropped orbit used as the input data.

The results of the unsupervised segmentation are evaluated using two external variables : precipitation and sea surface temperatures. The details of these data are described in their respective comparison sections.

4.2 . Model architectures

The architectures of unsupervised segmentation models tested in this study include approaches based on fully convolutional networks (FCN) and mini-batch K-means (Scikit-Learn implementation, [Pedregosa et al. \(2011\)](#)). The fully convolutional network (FCN) models (labeled M_i and U_i), proposed by [Kim et al. \(2020\)](#), consist of a simple architecture based primarily on a feature extraction block. Next, we proposed to test another FCN model (labeled L_i), which takes inspiration from [Kim et al. \(2020\)](#) and includes an attention mechanism. These models use convolution layers, allowing them to consider the spatial structures in an image. The last algorithm tested was K-means (labeled K_i), which is a pixel-to-pixel approach.

The choice of the first FCN [Kim et al. \(2020\)](#) is due to preliminary testing and its simplicity. The second FCN is more complex as it has a deeper network and includes an attention mechanism. The last choice, K-means, is to compare the two FCN architectures to a simple clustering method. The objective is to compare models of different complexities and evaluate their performance.

The application of the unsupervised approach on TB images has two important axes : the training of the models and the evaluation of the resulting segmentation. The implementation of each algorithm can add up to many different models by changing their respective hyper-parameters. The architectures and hyper-parameters are described in this section. However, choosing the best hyper-parameters is a complex task as the output labeling alone does not provide any indication of the geophysical variable that it represents. In addition, the training loss is not indicative of a good model. In the model by [Kim et al. \(2020\)](#), the loss function continues to decrease as the training progresses until multiple classes collapse into a single class. Early stopping is activated when the number of unique clusters reaches a predetermined number of classes. If the number of classes exceeds the minimum, the training continues until completion. The minimum number of classes is set empirically to be 15 due to qualitative evaluation using numerous training. Below 15 classes, the segmentation is not detailed enough to be useful.

In this study, we propose to train multiple models with different hyperparameters. Then, the choice of the most suitable model depends on comparison to external variables, described in Section 4.3.

4.2.1 . FCN models proposed by [Kim et al. \(2020\)](#)

The unsupervised segmentation model proposed by Kim et al. (2020) contains multiple convolution blocks serving as feature extractors. Each convolution block consists of a convolution kernel, a ReLU, and a batch normalization. The model relies on the creation of pseudo targets c_n to train. These targets are obtained by first applying a 1D convolution kernel and batch normalization on the output of the feature extractor, resulting in a normalized response map r'_n . Next, a softmax function converts the r'_n into cluster labels, the so-called pseudo targets c_n .

Table 4.1 – Hyper-parameters for the simple FCN model proposed by Kim et al. (2020). The total number of models tested is 12. These models are labeled M_i .

Input	Number of output class	Spatial continuity	Feature similarity
37 GHz, 89 GHz	32	1	1
37 GHz, 89 GHz	64		2
			5

Table 4.2 – Hyper-parameters for the FCN model proposed by Kim et al. (2020) with U-Net as the feature extractor. The total number of models tested is 20. These models are labeled U_i .

Input	Number of output class	Spatial continuity	Feature similarity
37 GHz, 89 GHz	32	1	1
37 GHz, 89 GHz	64	3	2
		5	4

The objective function of this model includes two losses : the feature similarity loss and the spatial continuity loss. The feature similarity loss is obtained by comparing the normalized response map r'_n and the pseudo labels c_n using cross-entropy. It aims to induce similarity within the same cluster. The second loss, spatial continuity, imposes similarity between adjacent pixels. Using the response map, it is the sum of the differences between pixels in vertical and horizontal directions. A coefficient balances the two losses. Further information about this model can be found in Kim et al. (2020).

Better performance was achieved by replacing the feature extraction block with the U-Net architecture (Ronneberger et al., 2015), which has proven to be very efficient on rain retrieval from microwave brightness temperature (Viltard et al., 2023). U-Net contains a contracting path where convolutional layers reduce the size of feature maps and increase the number of feature channels. Then, its expanding path upsamples the feature maps and decreases the number of feature channels.

Between these two paths, there are skip connections where corresponding feature maps are cropped from the contracting paths and concatenated to the expanding path. After the feature extraction block, the rest of the architecture remains the same as the original paper (Kim et al., 2020).

Multiple hyper-parameters are tested for this architecture : the number of output classes, the architecture of the feature extractor, the input channels of the brightness temperature, and the coefficient that balances the two losses. The list of hyper-parameters tested is presented in Table 4.1. For the model with U-Net as a feature extractor, the hyper-parameters are shown in Table 4.2. As the training time per model is short, the hyper-parameters in both tables are chosen for further analysis due to their promising results in a few case studies.

In line with the findings from previous deep-learning models applied to brightness temperatures (Viltard et al., 2023; Sambath et al., 2022), data augmentation (random crop and random flip) is added to the training to keep the model from picking up any particular patterns within the satellite orbits.

4.2.2 . Our proposed FCN

Our proposed FCN model is based on an autoencoder architecture. In our case, it involves reconstructing the brightness temperature maps presented as inputs at the output (Figure 4.2). Let $\{x_{j,k} \in \mathbb{R}^t\}^{j,k \in \mathbb{N}^{H,W}}$ be a set of t -dimensional brightness temperature vectors of maps pixels, where H and W denotes the height and width of an input map and $\{c_{j,k} \in [1 \dots q]\}^{j,k \in \mathbb{N}^{H,W}}$ the cluster labels to assign to each pixel. In this study, we set empirically the number of classes to $q = 40$ (using qualitative evaluation over several trainings). The developed model consists of 3 parts : encoder, decoder, and attention class module. The feature extraction performed by the encoder is composed of 2D convolution layers associated with ReLU activation functions and batch normalizations (Table 4.3). It outputs m feature maps (with $m = 32, m = 64, \text{ or } 128$) of the same sizes as the input ones. The decoder is slightly more complex than a standard decoder. It is composed of t sub-decoders of identical architecture ($t = 4$ if the input channels are 37 GHz and 89 GHz; $t = 6$ if the input channels are 19 GHz, 37 GHz, and 89 GHz), each dedicated to a brightness temperature channel. Each of them outputs q maps of the same size as the input maps. The last module, called the attention class module, generates for each class and each pixel a "probability" that the latter belongs to this class. Finally, each decoder is connected to the operator $\sum \Pi$ in order to reconstruct the corresponding input map. For each pixel, the operator $\sum \Pi$ sums the q values weighted by its probability of belonging to the corresponding class provided by the class attention module :

$$output_{j,k} = \sum_{i=1}^q A_{i,j,k} \otimes Y_{i,j,k} \quad (4.1)$$

where \otimes denotes element-wise multiplication with A and $Y \in \mathbb{R}^{q,H,W}$. The last step consists of stacking the t outputs in the same way as the input maps. The segmentation map is finally derived from the class attention module as follows :

$$c_{j,k} = \operatorname{argmax}_{i \in [1..q]} A_{i,j,k} \quad (4.2)$$

Table 4.3 – Our proposed architecture. Conv2d(K) means 2D convolution with a KxK kernel size.

Encoder	Decoder	Attention class	DoubleConv
DoubleConv(K1,K2)	DoubleConv(3,3)	DoubleConv(3,3)	Conv2d(K1)
DoubleConv(3,3)	Conv2d(3,3)	DoubleConv(3,3)	BatchNorm
Conv2d(3,3)	BatchNorm	Softmax	Relu
BatchNorm			Conv2d(K2)
			BatchNorm
			ReLU

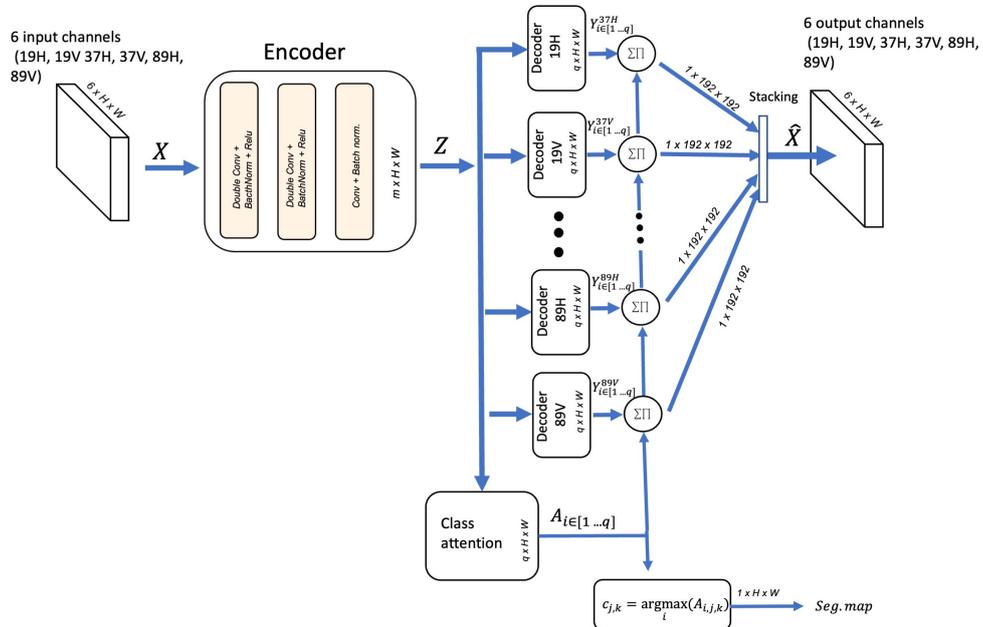


Figure 4.2 – The proposed architecture. The choice of input channels $t = 6$, i.e 19H, 19V, 37H, 37V, 89H, 89V, is shown as an example.

The loss function is divided into 2 parts. The first part consists of a constraint on maps quality reconstitution defined by a standard mean square error loss L_{MSE} between input and output maps, while the second part L_C is to favor each pixel one class among the q classes.

$$L = L_{MSE} - \alpha L_C$$

With α positive or null constant

L_C loss aims to maximize the variance in the attention output.

The training was done on the same dataset as the previous model using an Adam optimizer and a fixed learning rate (5×10^{-4}).

Table 4.4 – List of hyper-parameters tested for our proposed model. The total number of models tested is 162. These models are labeled L_i .

Input	α	m	K1	K2
37 GHz, 89 GHz	0.1	32	3	3
19 GHz, 37 GHz, 89 GHz	0.01	64	5	5
	0	128	7	7

4.2.3 . K-means models

Table 4.5 – Kmeans hyper-parameters tested. The total number of models is 18. These models are labeled K_i .

Input	Output class	Minibatch size
37 GHz, 89 GHz	10	16
19GHz, 37 GHz, 89 GHz	13	32
	15	64

Due to the large number of pixels, the mini-batch K-means is more suitable as it works significantly faster. Mini-batch K-means offers a faster computing speed as it uses randomly sampled subsets of input data, called mini-batches, to optimize the same objective function as K-means (Sculley, 2010). The algorithm works in two steps. First, a pre-determined number of samples are drawn from the training dataset and assigned to the closest centroid. Next, the centroids are updated using the streaming average of the current sample and all the previous samples. The initialization used is "K-Means++" (Arthur et al., 2007). Although Mini-batch K-means is faster than K-means, it comes with a slight reduction in performance.

K-means is a pixel-to-pixel approach. Therefore, its inputs contain the chosen dimensions (4 or 6 brightness temperature channels) in the spectral dimension and

a single dimension in the spatial dimension. The input is normalized the same way as for the convolutional models described above. The hyper-parameters tested for mini-batch K-means include the number of TB channels (4 and 6), the mini-batch size (32, 16, and 64), and the number of clusters (10, 13, and 15) (Tabel 4.5). There are 18 total models trained.

4.3 . Comparison of segmentation models without ground truth

To evaluate the performance of the models, the segmentations can be compared using two approaches : an internal and an external comparison. In internal comparison, the only variable used for comparison is the model's inputs, which are the brightness temperature in our case. Metrics for internal comparison mainly rely on intra-class and inter-class dispersions, such as Calinski-Harabasz (Calinski and Harabasz, 1974) and Davies-Bouldin index (Davies and Bouldin, 1979). However, metrics relying on pixel-to-pixel distances in the feature space tend to favor pixel-to-pixel models (K-Means) as they calculate the distances between classes pixel-to-pixel using the input data. The internal scores can be influenced by the difference in the number of classes. In addition, they do not offer any indication of the model's performance on the important geophysical variables. Therefore, we choose the most suitable model based on the external evaluation.

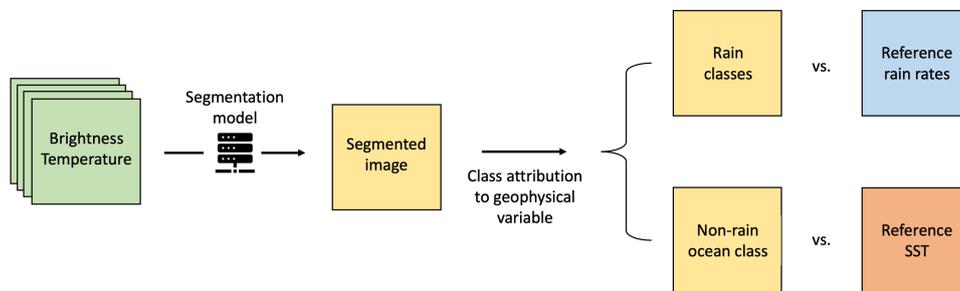


Figure 4.3 – The process of external comparison. The first step is to label the classes obtained from the segmentation. Those that represent precipitation and SST are compared to reference data using traditional segmentation metrics such as accuracy, recall, and F1-score. The classes that represent other geophysical variables are not used for model evaluation in this study.

In external comparison, variables that the model has not seen during training are used to evaluate the segmentation. These variables should be relevant ; for instance, geophysical quantities are generally retrieved from the brightness temperature. The present study compares the clusters to precipitation (over land and ocean) and sea surface temperature (Figure 4.3). However, as the labeling of the outputs does not directly provide any indication of their geophysical significance, the first step is to

label the class to a geophysical variable (Figure 4.3). Then, they can be compared using traditional segmentation metrics such as accuracy and F1-score.

In the absence of ground truth, labeling the classes in the segmentation to the geophysical variable that it represents can be a very challenging task. The class labels in the output of the segmentation models can change with each run, meaning that the labels are not fixed and can be assigned differently each time. For example, a precipitation structure in one model can have a completely different label number than another model. Even in multiple initializations, the label number for the same geophysical variable can be completely different. Therefore, comparing all the models presented above cannot be done manually. A prerequisite to model comparison is automatically labeling the class labels to a physical variable.

This section aims to describe the process of efficiently comparing the segmentation between many different models to choose the model with classes that represent geophysical variables well. The evaluation requires an identification of the characteristics of the segmented class and choosing appropriate metrics to compare it to the corresponding geophysical variable. A good segmentation model must fulfill two criteria : the accuracy of the segmented objects compared to the geophysical variables and the level of detail contained in the segmentation. The selected model will then be used in the application in Section 4.4.

4.3.1 . External comparison with precipitation

The different implementations are compared using qualitative and quantitative approaches. The quantitative approach serves mainly as a pre-selection step, as there are too many models to evaluate qualitatively first. The goal is to keep only similar-performance models that are then judged on several different case studies of vastly different natures.

The reference precipitation used for the external comparison is the retrieved rain rates from GMI using a deep learning rain retrieval algorithm (DRAIN) (Viltard et al., 2020, 2023). In DRAIN, 37 and 89 GHz channels of brightness temperatures are used as the sole input to a quantile regression U-Net. The model is trained in a supervised approach using the Precipitation Radar (DPR) aboard the same satellite as the target rain rates. Extensive evaluation of DRAIN can be found in Viltard et al. (2023).

The first step of comparison is class labeling. Due to the number of models to be compared, each class's geophysical property must be identified automatically. Using a separate set of data, the pixels of the segmentation are compared to the pixels of the retrieved rain rates of the same observation. Next, the number of rain pixels per class is counted. If the percentage of rain pixels shared by a class is superior to a pre-defined threshold (15 %), that class is designated as a rain class (1). Classes below the threshold become non-rain classes (0). This step results in

a binary rain/non-rain segmentation of the image.

After identifying the rain classes, a score for each model is calculated by comparing the now binary segmented images and the precipitation images (converted to binary data using a threshold of 0.1 mm/h for rain) using the Intersection over Union (IoU), accuracy, recall, and F1-score, which are metrics commonly used in segmentation evaluation. Accuracy is the ratio between the correct predicted pixels over all pixels. Recall is the ratio of the True Positives over all actual positive pixels. Finally, the F1-score is a preferred metric for the case of imbalanced data. For example, in our case, there is an overwhelming amount of non-rain pixels and very few rain pixels. F1-score balances the precision score (the ratio of True Positive over all the predicted positive pixels) and the recall score. For all the four metrics presented, higher scores represent better segmentation agreement.

The external comparison is first used as a pre-selection of models to choose two best-performing models for each approach before the in-depth comparison (the scores of this step are not presented here). As there are many models per approach, this is a comparison step for comparing the same algorithm but with different hyper-parameters. These top two models per each approach (models M_i , U_i , L_i , and K_i) are chosen based on their average F1-score per 450 orbits. All models struggle to distinguish the precipitation and the presence of snow and ice on the surface. Due to this issue, the scores are limited to latitudes between 40°S to 40°N. In this step, the hyper-parameters of each approach are investigated separately to study their influence on the segmentation. The most notable finding in this step is that, across all approaches, models with 19, 37, and 89 GHz channels as inputs perform consistently better than their counterparts where the 19 GHz is not included.

Table 4.6 – Top two models for the algorithm proposed by Kim et al. (2020). M1 and M2 are the models with simple feature extractors. U1 and U2 have U-Net as their feature extractor.

Model Name	Input	Output	Spatial Continuity	Feature Similarity
M1	19 GHz, 37 GHz, 89 GHz	32	1	1
M2	19 GHz, 37 GHz, 89 GHz	64	1	5
U1	19 GHz, 37 GHz, 89 GHz	64	1	1
U2	19 GHz, 37 GHz, 89 GHz	32	3	1

The inter-comparison between different approaches uses another set of 450 orbits chosen randomly from the validation set (within 40°S to 40°N latitudes).

Table 4.7 - The hyper-parameters of the top two models of our proposed architecture.

Model name	a	m	K1	K2
L1	0.1	32	3	5
L2	0	128	5	5

Table 4.8 - The hyper-parameters for the top two K-Mean models.

Model name	Input	Output	Minibatch size
K1	19 GHz, 37 GHz, 89 GHz	13	64
K2	19 GHz, 37 GHz, 89 GHz	15	16

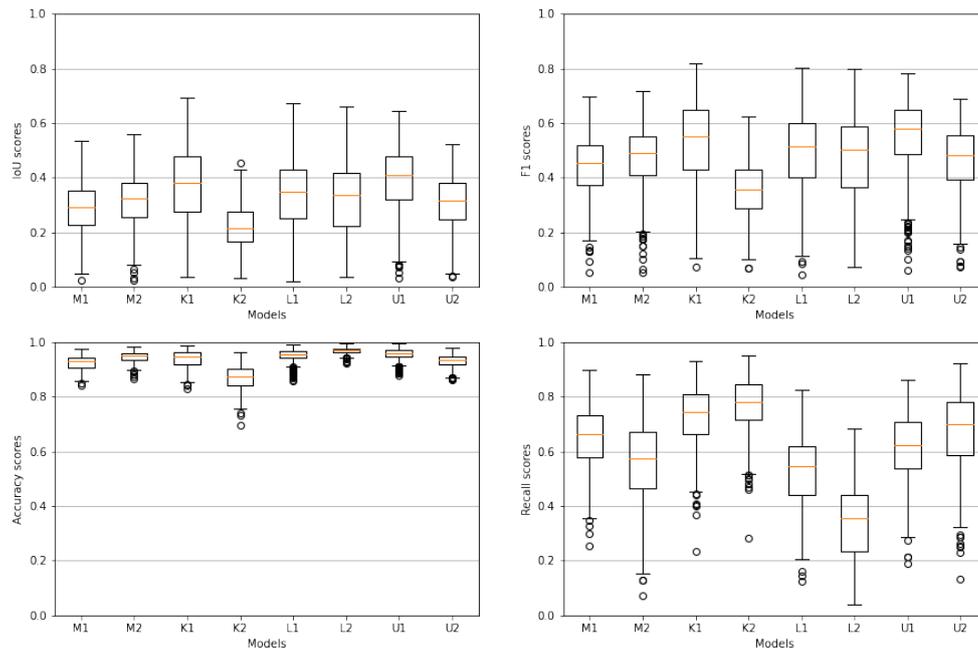


Figure 4.4 - Comparison of two of the best models from each approach. IoU, F1-score, Accuracy, and Recall score are calculated on 450 orbits observed in 2019. M1 and M2 are models from Kim et al. (2020) with simple convolution layers as feature extractors. K1 and K2 are K-Means models. L1 and L2 are our proposed models. U1 and U2 are models from Kim et al. (2020) with U-Nets as feature extractors.

Table 4.9 – Average and standard deviation of F1-scores for the 450 orbits shown in Figure 4.4.

Model name	Average F1-scores	Standard deviation of scores
U1	0.55	0.13
K1	0.52	0.15
L1	0.49	0.13
M2	0.47	0.11
L2	0.47	0.14
U2	0.46	0.11
M1	0.43	0.10
K2	0.35	0.10

Figure 4.4 shows the scores for the top two models of each type. Here, M1 and M2 are the best models from Kim et al. (2020) approach (Table 4.6), U1 and U2 are the models with Kim et al. (2020) approach using U-Nets as feature extractor (Table 4.6), L1 and L2 are the models proposed above (Table 4.7), and K1 and K2 are K-means models (Table 4.8).

The boxplots can offer more information as they show how much the score varies from orbit to orbit. A good model should have a higher mean score and smaller variance between orbits. All models seem to be able to achieve similar performance for all metrics considered. However, the models with U-Net as feature extractors (U1 and U2) have relatively higher median and smaller variance. Table 4.9 shows the average and standard deviation of the F1-scores for the 450 orbits.

The selected best models of each type are then compared again on several case studies (Figures 4.5, 4.6, and 4.7). The qualitative approach consists of comparing the segmented image to the reference precipitation images. The segmented images are evaluated based on their structure compared to the precipitation images, their level of detail, and the number of classes. Ideally, the segmented images should have classes that represent different types of precipitation (for example, convective or stratiform rain) and their intensity. It should also contain classes representing the continent and the ocean with a certain level of detail. For the first case study, figure 4.5, which shows Typhoon Nanmadol (16/09/2022), all the best models selected above are able to segment the precipitation structure. However, only models U1, U2, M1, and L2 show more than three levels of rain intensity. In all three case studies, K-means segmentation is very pixelated. This might hinder object analysis in the long run as the isolated pixels cannot be easily regrouped together to form an object. Similarly, in case 2 (Figure 4.6), all models can capture the precipitation zone, with K-means models offering the coarsest segmentation. Finally, case 3 shows a frontal system over France (Figure 4.7). The models seem to struggle more to segment the precipitation structure. Most models show the precipitation

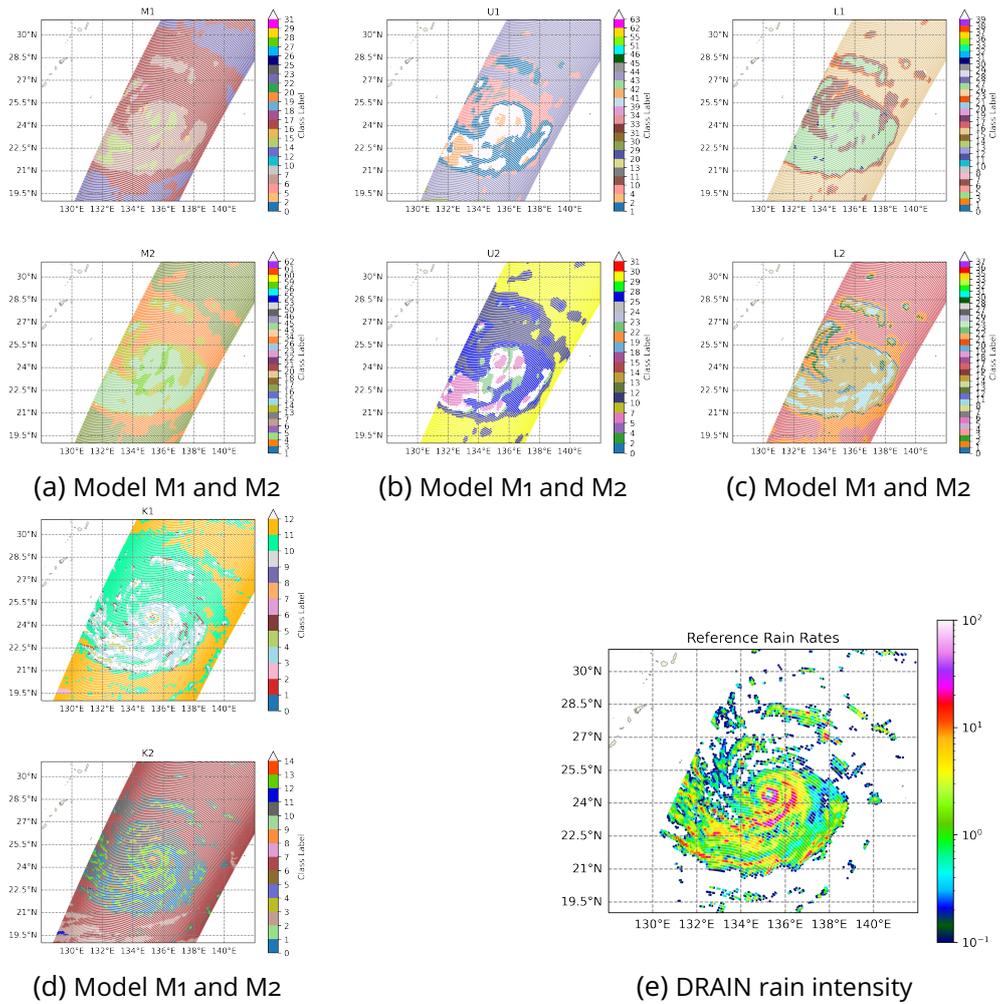


Figure 4.5 – Comparison between segmentation and reference rain rate for Typhoon Nanmadol observed by GMI on 16th September 2022.

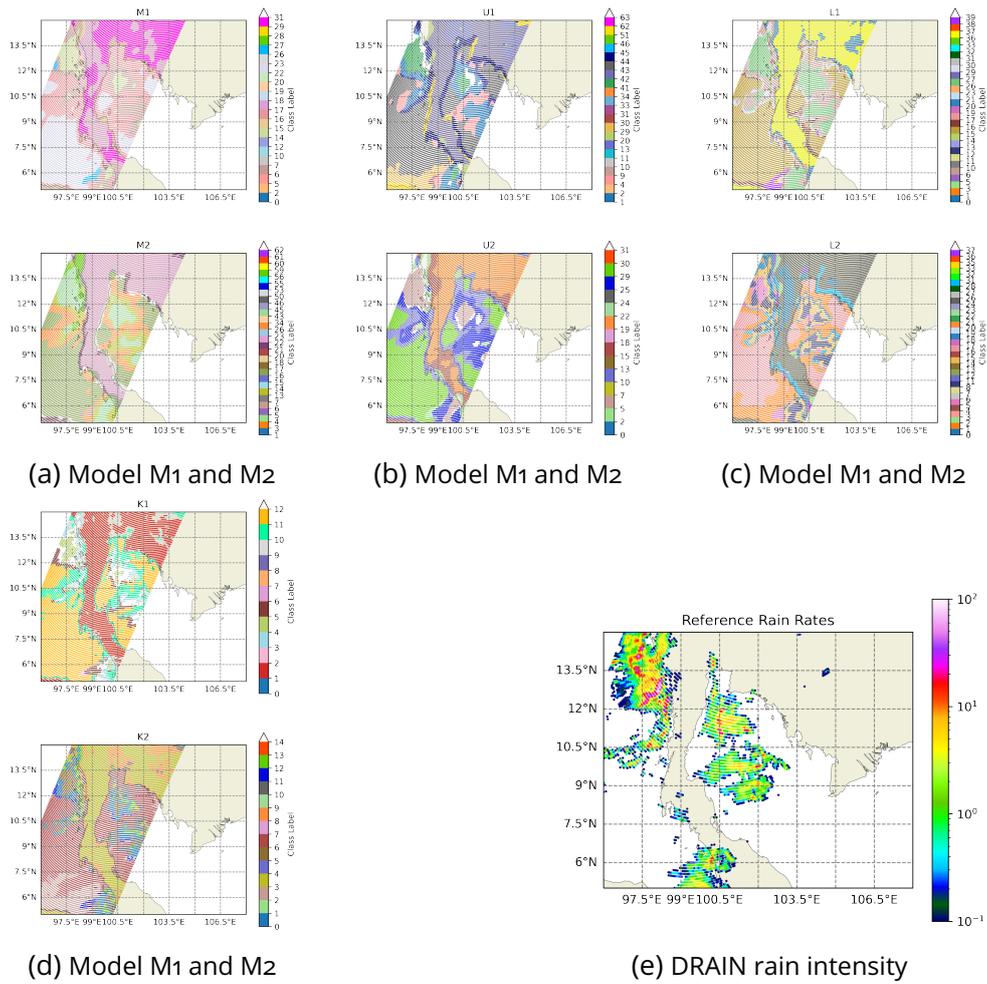


Figure 4.6 – Comparison between segmentation and reference rain rate for a rain structure observed by GMI on 29th May 2017.

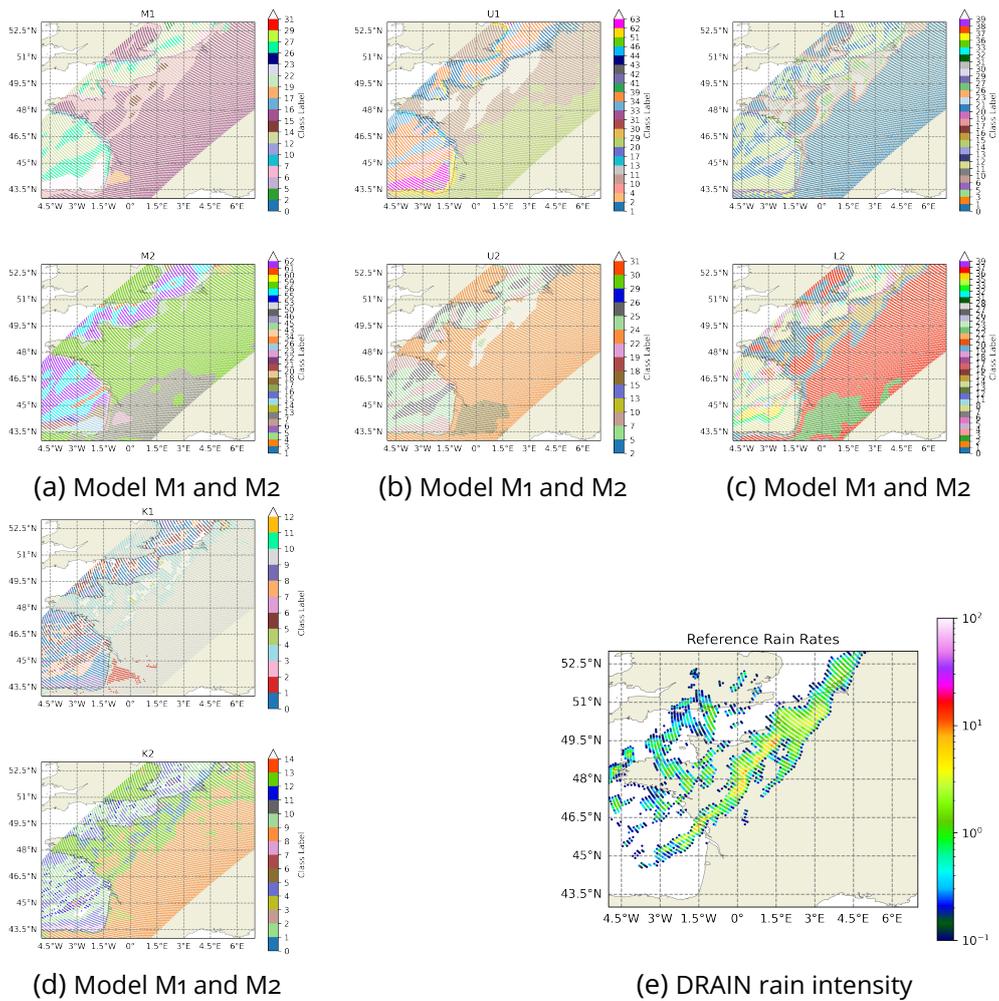


Figure 4.7 – Comparison between segmentation and reference rain rate for a frontal system observed by GMI on 18th August 2018.

zone except for model M2. Models M1, U2, L1, L2, and K2 offer the most similar structure to the reference image.

When compared to the precipitation, the preselected best models have very similar performance. The main difference is that K-means segmentation is very noisy as it does not take into account the structure of the brightness temperatures. The overall top model is model U1, as it is among the top scorers across all metrics considered, and it has a smaller variance in its scores for the 450 orbits. In terms of precipitation structure, this model contains a sufficient amount of detail in its segmentation and follows the structure of the reference data well.

4.3.2 . External comparison with sea surface temperatures

The sea surface temperature (SST) data to be compared to the segmentation is the product of GMI brightness temperature by the Group for High-Resolution Sea Surface Temperature (GHRSSST) (Remote Sensing Systems, 2017). This product is the temperature of ocean water's top layer (approximately -1 mm), ranging from -3 to 34.5 degrees Celsius. The data is available for each orbit of the GPM Core Observatory from its launch up to the present day with a resolution of 0.25 by 0.25 degrees. Data unavailability in this dataset can be caused by high wind speed, rain, cloud cover, or pixels near sea ice and land.

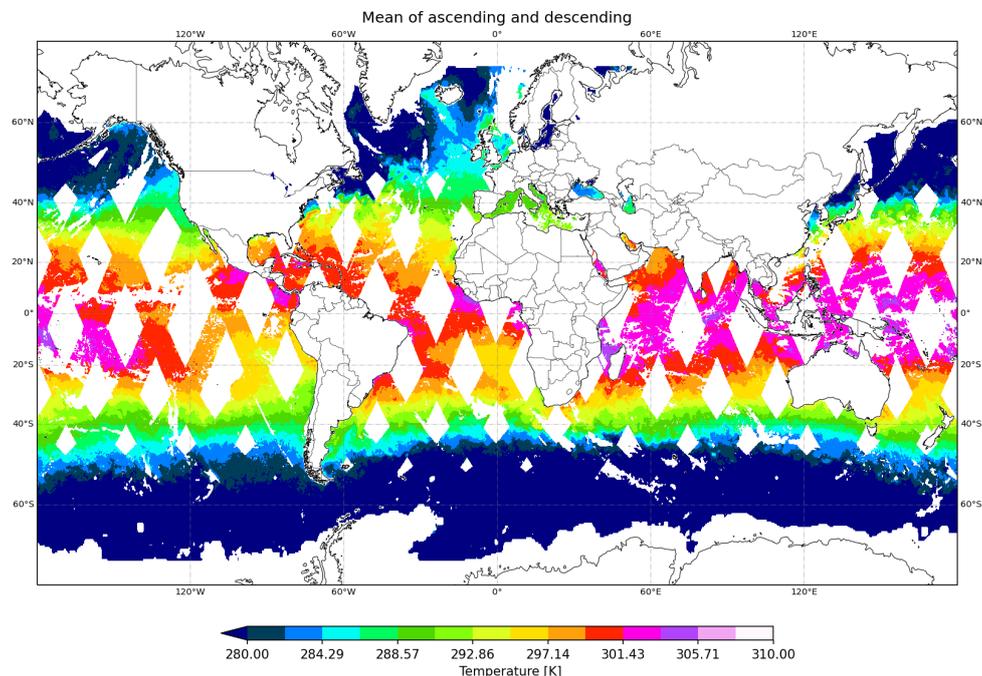


Figure 4.8 – An example of SST obtained from GMI brightness temperatures orbits by the GHRSSST.

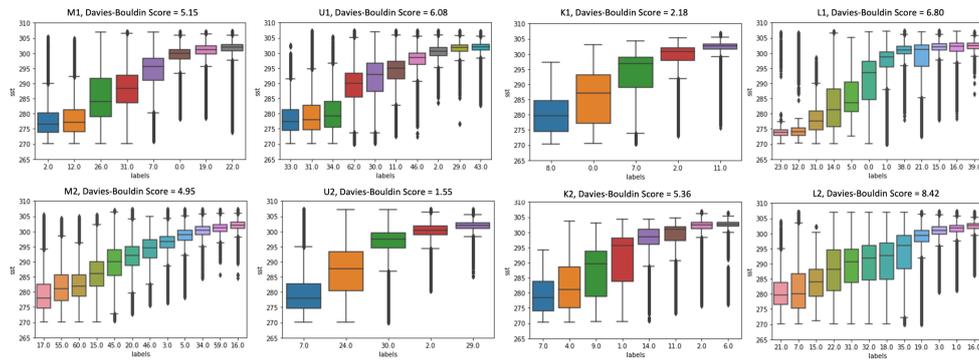


Figure 4.9 – The sea surface temperatures (K) of the non-rain ocean classes for the top models described in Section 4.3.1. The result of each model is obtained from the same 100 orbits.

Similarly to the previous comparison, the evaluation of the segmentation results using SST as the reference needs to be automated. Ocean classes are simply defined as non-rain classes that only occur over the ocean. Next, the corresponding SST pixels for each class are regrouped. Figure 4.9 shows the SST values in each class observed over 150 orbits from 2019 for the chosen models from the previous section.

We evaluate the segmentation performance on SST based on whether the SST in each cluster is densely populated and well-separated. Calinski-Harabasz index is the ratio of the sum of between-cluster dispersion and within-cluster dispersion (Calinski and Harabasz, 1974). Therefore, a higher Calinski-Harabasz score means a better-defined cluster. Another metric is the Davies-Bouldin index, which compares the compactness and separation of clusters by measuring the similarity between clusters by comparing the distance between clusters with the size of the clusters themselves (Davies and Bouldin, 1979). The lower score indicates a better-defined cluster, with zero as the lowest possible score.

The models with more output classes tend to have multiple classes representing the same SST range, which results in lower Davies-Bouldin scores. Similar clusters can be regrouped together during post-processing to deal with this issue. If the mean SST of two clusters is similar enough, they will be combined into one cluster. This results in an improvement of scores. Consequently, all best-performing models achieve similar scores in terms of SST clustering as models with more classes (lower scores) can have their SST classes regrouped to achieve a better score. A model with more classes also has the advantage of offering a better representation of different SST ranges.

4.3.3 . Model choice

The quantitative and qualitative comparisons with the reference precipitation and SST show that all the top-performing models of different approaches perform

similarly. They are all capable of segmenting geophysical structures in the brightness temperature images. In addition to the classes representing precipitation and SST that are evaluated here, there are other classes that show similarity to other geophysical variables, such as soil humidity and snow-covered surfaces. The model choice lies in the minor detail of the performance. With more priority placed on precipitation, the chosen model for the next step is the best model for precipitation under the condition that it achieves similar performance on SST among its peers. The model that fulfills these criteria is model U1. It should be noted that, regarding calculation time, the convolutional models run approximately 10 times faster than the K-means models. To segment one orbit, the FCN models take approximately 0.05 seconds, while mini-batch K-Means require 0.5 seconds.

4.4 . Validating the characteristics of the segmentation classes with geophysical knowledge

The previous section selects the top-performing model by identifying rain and SST classes and comparing them to reference data using traditional segmentation metrics. In this section, the classes are examined in terms of their characteristics and behavior. It is important to verify that we can observe precipitation and SST patterns regionally and globally using only these classes. Using the chosen model, three application case studies are investigated. The first application is on a regional scale (Section 4.4.1). Two regions from the IPCC reference regions [Iturbide et al. \(2020\)](#) are observed using the GMI data from 2019. Next, on a global scale, GMI data from 2019 is used to study the seasonal rain structure over the ocean and the position of the Inter-tropical Convergence Zone (Section 4.4.2). Finally, we look at the seasonal gradient of the SST using the non-rain ocean classes (Section 4.4.3).

4.4.1 . Precipitation patterns over two regions

The previously identified rainy classes are studied in two different regions. To facilitate the choice of regions, the application relies on the updated IPCC reference regions [Iturbide et al. \(2020\)](#). The IPCC reference regions are homogenous gridboxes with consistent regional climate features. The study is applied to two reference regions, the Bay of Bengal (BOB) and Western Africa (WAF), as shown in figure 4.10. These two regions have vastly different features. The BOB grid box contains mainly the ocean and is driven by the Indian Monsoon. On the other hand, WAF mostly consists of land with a large coastal area.

For each region, the segmentation of a cyclone observed near or in the region observed by the GMI is compared to the retrieved rain intensity by DRAIN. Then, the rain pixels are counted using both the unsupervised segmentation and the supervised retrieval approach, DRAIN. The aim is to analyze the pattern of monsoon

precipitation in the two regions.

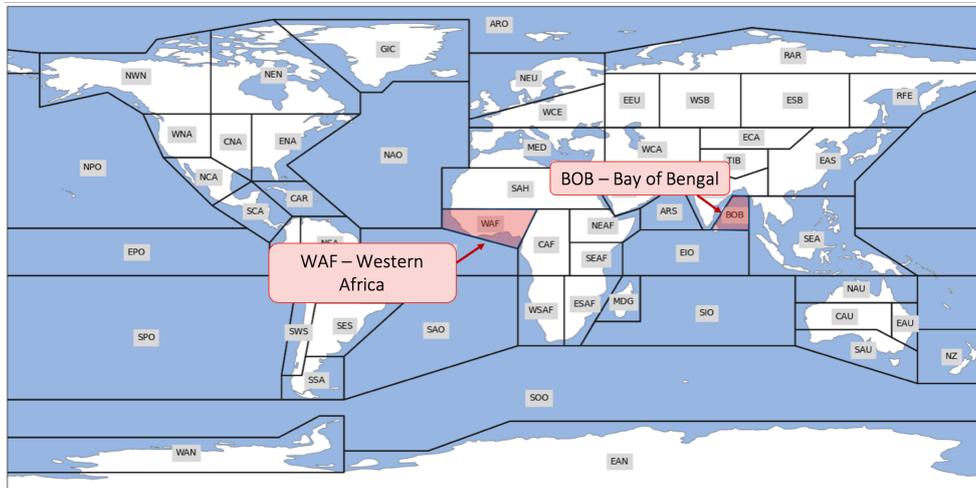


Figure 4.10 – IPCC reference region *Iurbide et al. (2020)*. The two chosen regions, BOB and WAF, are highlighted.

4.4.1.1 Bay of Bengal (BOB) and the Indian Summer Monsoon (ISM)

Tropical cyclones are among the extreme weather phenomena that occur over tropical or sub-tropical waters. It can last at least 6 hours and has wind speed up to 34 kt (*Bhardwaj and Singh, 2020*). On average, the Bay of Bengal (BOB) region saw 152 tropical cyclones between 1972 and 2017 (*Bhardwaj and Singh, 2020*). During this period, an average of 3.3 tropical cyclones formed yearly without any significant trend. However, there is an intensification of approximately 40% of all tropical cyclones observed. The number of tropical cyclones peaked twice, first in May and then higher in November. The highest number of tropical cyclones form during post-monsoon, while the lowest number occurs during winter.

Figure 4.11 shows an example of segmentation done on Cyclone Amphan on May 16th, 2020, as it was observed in its early stage by the GPM Core Observatory (*Kelly, 2020*). The structure of the cyclone is composed of four classes (Figure 4.11a). Compared to the precipitation retrieved by DRAIN (Figure 4.11b), the rain classes represent different intensity ranges. For example, class 41 shows the highest intensity, while class 4 shows the lowest intensity in the cyclone. The segmentation is representative of the structure of the Cyclone. The next step is to verify if it can also capture the precipitation patterns in this region.

In terms of precipitation over the BOB region, it is heavily influenced by the Indian Monsoon. The summer Monsoon over India, occurring from June to September, is responsible for providing at least 70% of rainfall (*Wang et al., 2009*).

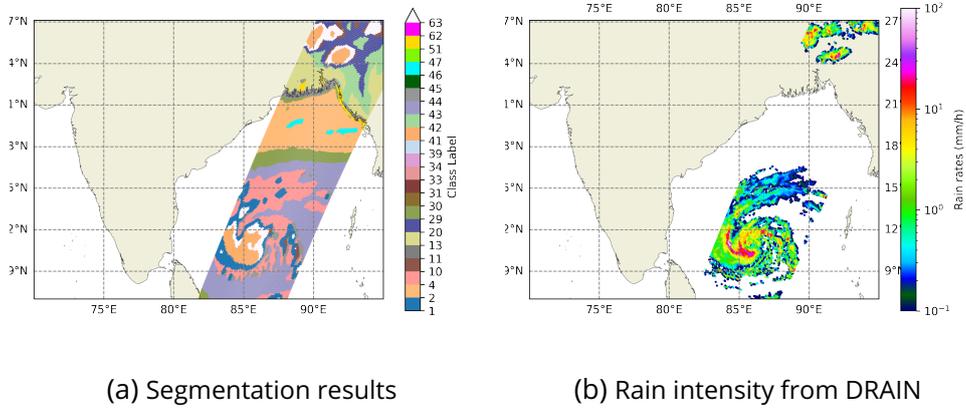


Figure 4.11 – Cyclone Amphan observed by the GPM Core Observatory on May 16th, 2020. (a) shows the segmentation results. (b) shows the rain intensity retrieved by DRAIN.

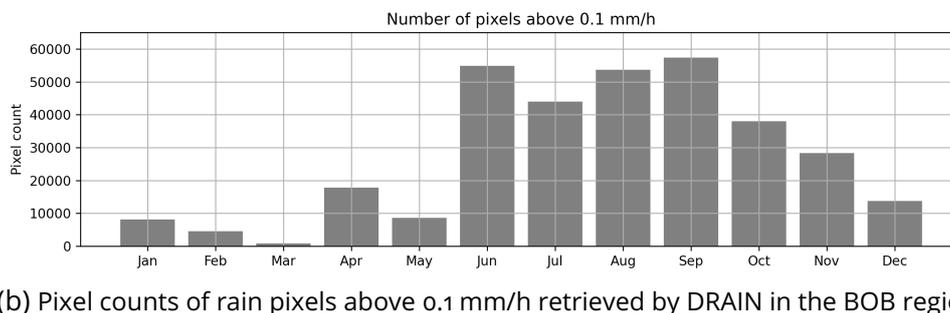
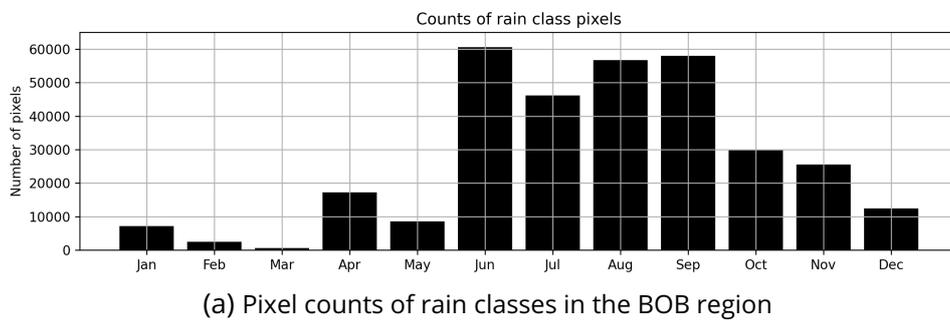


Figure 4.12 – Count of rain pixels per month inside the BOB region in 2019. (a) The count of pixels of the rain class was obtained from the segmentation. (b) The pixel of rain above 0.1 mm/h retrieved by DRAIN.

Predicting its onset and retreat dates significantly impacts agriculture and food production. The beginning of the Indian Summer Monsoon (ISM) is marked by the precipitation over the Bay of Bengal region, the southern Indian peninsula. The precipitation then moves inland from South to Northern India. On the other hand, the end of ISM is marked by the retreating rainfall from North to South India. The authors in [Bombardi et al. \(2020\)](#) reviewed the detection methods of Monsoon onset dates globally. They categorized the techniques into two groups : based on a local scale and based on a regional-to-large scale. The local-scale approach is characterized by the data from a grid point data or single rain gauge. For ISM, it includes, for example, the surface moist static energy exceeding a certain threshold. For a regional scale method, the Indian Meteorological Department defined the onset dates of the Monsoon as the amount and persistence of precipitations over Kerela (a state on India's tropical Malabar Coast), the depth of the westerlies, zonal wind speed, and outgoing longwave radiation (or OLR). ([Joseph et al., 2006](#)) defined ISM using zonal wind average and OLR over Kerela.

In Section 4.3, we have identified classes that contain rain. It is necessary to verify that these identified rain classes can demonstrate the intra-annual precipitation pattern (Figure B.5). Figure B.5b shows the number of rain pixels per month (above 0.1 mm/h) retrieved by DRAIN in the BOB region for the 2019 GMI observations. The number of pixel counts per month in the two cases is nearly identical. We can observe the preonset (April-May) and the monsoon season (June-July-August-September) ([Rai et al., 2015](#)) from both the segmentation approach (Figure B.5a) and the retrieval approach (Figure B.5b).

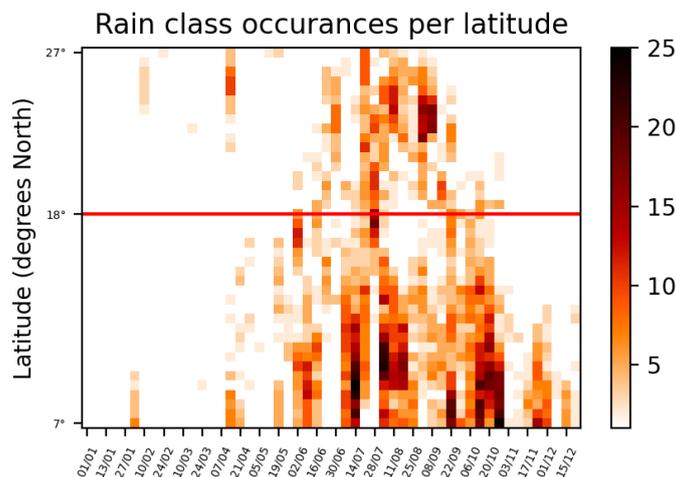


Figure 4.13 – A Hovmöller diagram of the advancing and retreating Monsoon precipitation over India in 2019. Each date on the x-axis represents the start of the week. The count is the sum of rain class occurrences from the first day of the week to the seventh day.

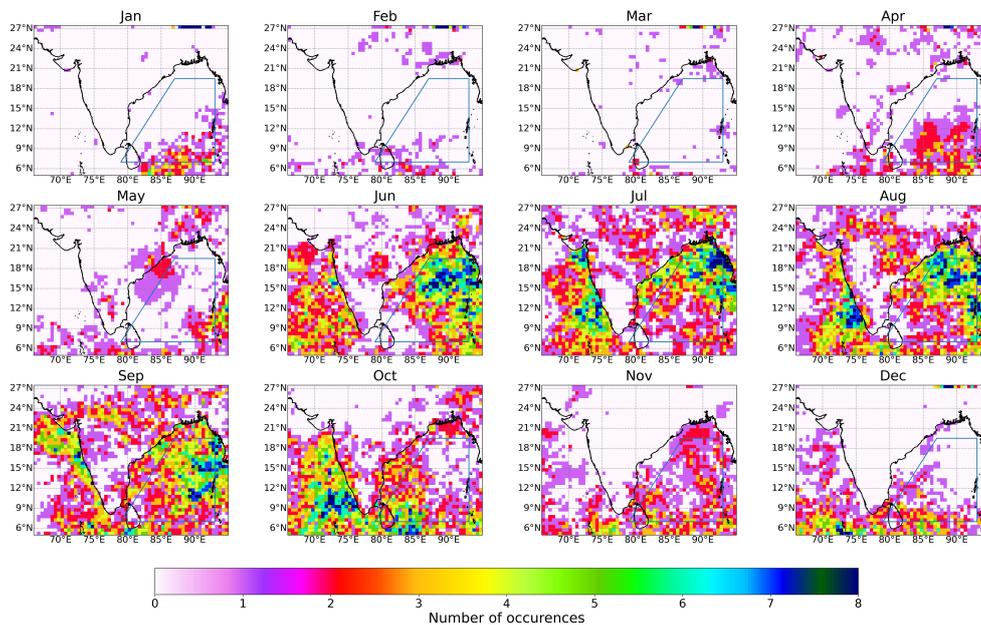


Figure 4.14 – Using data from 2019, the number of rain class occurrences for each $0.5^\circ \times 0.5^\circ$ gridpoint is counted for each month. The area shown is between 5°N - 27°N and 6°E to 100°E . The blue line indicates the BOB region.

According to Skofronick-Jackson et al. (2017), GPM constellation data is capable of tracking the movement of the ISM from the Bay of Bengal to Northern India, thanks to the abundance of rainfall. The authors believe that, with multiple years of GPM precipitation records, one can identify the interannual variation of the ISM resulting from large-scale oceanic or atmospheric patterns or climate change.

Using rain pixel counts per latitude, Figure 4.13 shows a Hovmöller diagram of the onset and withdrawal dates of the ISM in 2019. The number of rain class occurrences per pixel is counted over a box bounded by latitudes 7°N to 27°N and longitudes 73°E to 80°E . However, as only one satellite is used to create the time-latitude graph, the time interval for summing the pixel counts is one week. Reports from the Indian Department of Meteorology stated that the 2019 ISM started on May 19th for the Southern BOB and on June 8th for Kerala (Balachandran et al., 2019). The retreat began on October 15th-16th, with full withdrawal on the 16th of October (Balachandran et al., 2019). Using the time-latitude plot, we observe some preonset rain around April and some rain around the week of May 19th for the very Southern region. The full onset of Monsoon began in the first week of June. The red line in Figure 4.13 represents the latitude 18°N . No rain above this latitude after the week of October 20th, which coincides with the retreat date of the 2019 ISM. The 2019 ISM is particular due to the strong connection between the ISM and the El Niño Southern Oscillation (ENSO) over the Pacific, (Gadgil et al., 2019). The authors remarked on a weaker monsoon in June and above-average

rainfall in July.

Figure 4.14 is a supplementary result of the time-latitude plot. It shows, on a map, the occurrence of rain classes per gridpoint of 0.5 by 0.5 degrees. It shows less rainfall in June than in July, August, and September. There was no more rain over India after October. It should be noted that the result would be clearer if more satellites were added to the analysis.

The results of the segmentation in BOB show that the precipitation classes represent both the structure of the precipitation event in an image and the precipitation pattern. Although the data is only from one satellite, the segmentation shows the approximate dates of the arrival and departure of the monsoon.

4.4.1.2 Western Africa (WAF)

The rainfall in Western Africa is controlled by the West African Monsoon, which transports moisture inland from the Atlantic Ocean (Quagraine et al., 2020). Most rainfall occurs between June and September, while the dry season is from January to March. The WAM can vary from interannual to interdecadal (Quagraine et al., 2020). In addition, agriculture in the Sudano-Sahelian region heavily depends on the arrival of the Monsoon (Sultan and Janicot, 2003).

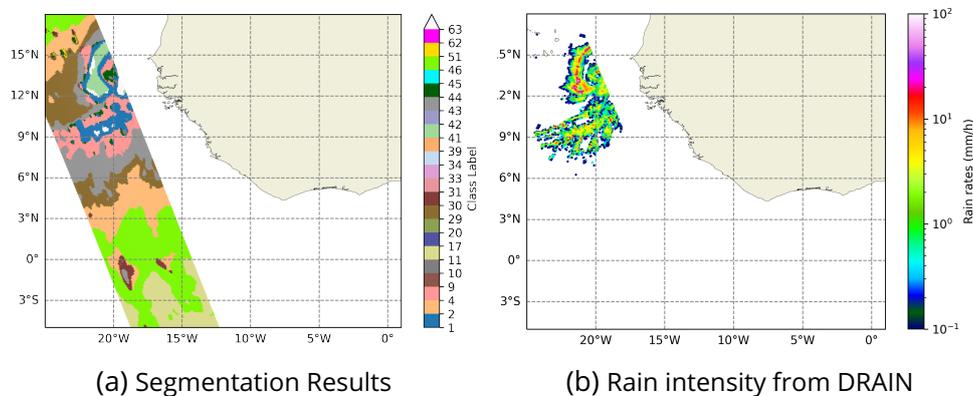


Figure 4.15 – Tropical storm Fred observed by the GPM Core Observatory on August 30th, 2015. (a) shows the segmentation results, (b) shows the rain intensity obtained by DRAIN on the brightness temperature data.

Similarly to the previous section, first, a single case of satellite observation is examined. Figure 4.15 shows the tropical storm Fred on August 30th, 2015, before it developed into a hurricane near the coast of WAF (Kekesi et al., 2023). Compared to the precipitation obtained by DRAIN (Figure 4.15b), The structures of the tropical storm can be found in the segmentation (class labels 1, 4, 41, 63). Class 41 again represents high-rain intensity, while class 4 represents low-intensity rain. The next step is to again verify the precipitation pattern.

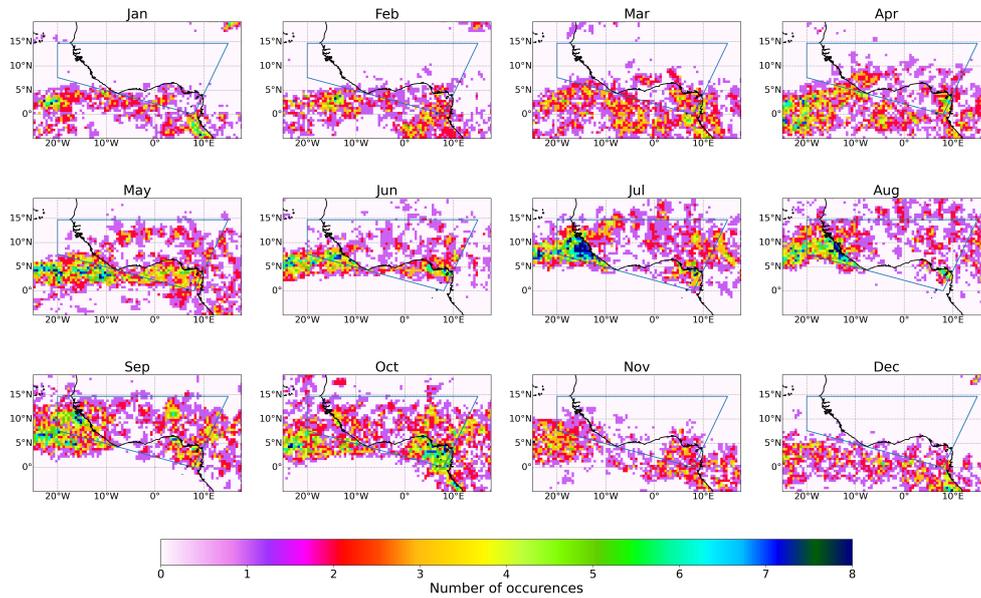
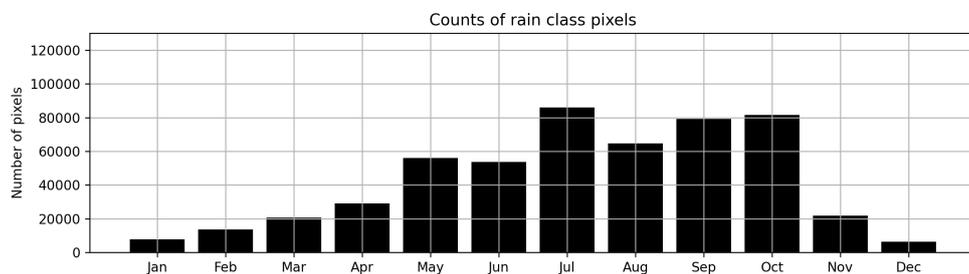
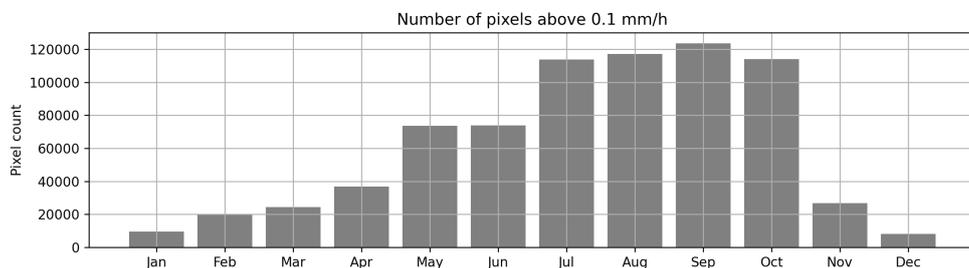


Figure 4.16 – Using data from 2019, the number of rain class occurrences for each $0.5^\circ \times 0.5^\circ$ gridpoint is counted for each month. The area shown is between 5°S to 20°N and 15°W to 15°E . The blue line indicates the WAF region.



(a) Pixel counts of rain classes in the WAF region



(b) Pixel counts of rain pixels above 0.1 mm/h retrieved by DRAIN in the WAF region

Figure 4.17 – Rain pixel counts per month in 2019 inside the WAF region. Top Figure (a) shows the counts obtained by the segmentation. Bottom Figure (b) shows the count from rain pixels above 0.1 mm/h retrieved by DRAIN.

There are multiple ways to define the preonset and the onset of the WAM. The preonset of WAM can be characterized by rainfall arriving in the Sudano-Sahelian zone in mid-May (Sultan and Janicot, 2003). The onset of WAM is signaled by the abrupt change in the position of the ITCZ, from its position along the 5°N latitude to around 10°N in mid to end of June (Sultan and Janicot, 2003). From Figure 4.16, which shows the occurrences of rain classes per gridpoint on a map, we can observe the rain over the Sudano-Sahelian region in May as well as the shift of the ITCZ from 5°N to 10°N in May and June. Furthermore, the figure also shows the position of the ITCZ at its highest at around 10°N in July and August (Sultan and Janicot, 2003).

The rain pixels obtained through the segmentation and DRAIN for the pixels inside the WAF region are compared in Figure 4.17. The rain count shows the same pattern as the count from DRAIN rain retrieval. However, between July and October, there are fewer counts using the segmentation as the model seems to struggle more to find rain pixels on land than on the ocean.

The tropical storm case study and the monthly precipitation count result show that the precipitation class can again represent rain structure and pattern in WAF (mainly continent area), which is a very different region from BOB (mainly ocean area).

4.4.2 . Precipitation over the ocean and the ITCZ precipitation band

Figure 4.18 shows the total occurrences of rainfall pixels per each 0.5x0.5 gridpoint in 2019 obtained from the precipitation classes. As the satellite passes over the tropics less often than in higher latitudes in the Northern and Southern hemispheres, the number of occurrences is divided by the total annual satellite passages per latitude. The rain class pixel counts correctly indicate wet areas and dry areas in the ocean. With the exception of the Indian Ocean, the dry areas are located in the subtropics on the eastern basins (Yuan and Miller, 2002). These dry regions have an almost triangular shape, with bigger dry regions along the eastern coast of the basin. In the Indian Ocean, the dry region is found in the western part of the basin. Most wet regions are found in the Intertropical Convergence Zone (ITCZ) and the Northwestern Pacific and Atlantic Oceans storm tracks (Yuan and Miller, 2002). One of the most important characteristics of rainfall over the ocean is the ITCZ. It is located at the ascending branch of the Hadley cell and is subjected to abundant and intense rainfall (Byrne and Schneider, 2016b). It stretches from the eastern equatorial Indian Ocean through the equatorial Pacific Ocean to the eastern equatorial Atlantic Ocean (Yuan and Miller, 2002). In the Indian Ocean, the wet regions of the ITCZ extend North, while in the Pacific Ocean, the areas extend to the South Pacific Convergence Zone (SPCZ) (Yuan and Miller, 2002). It should be noted that the higher occurrence of rain classes in the high latitudes

is due to snow on the surface being misclassified as rain.

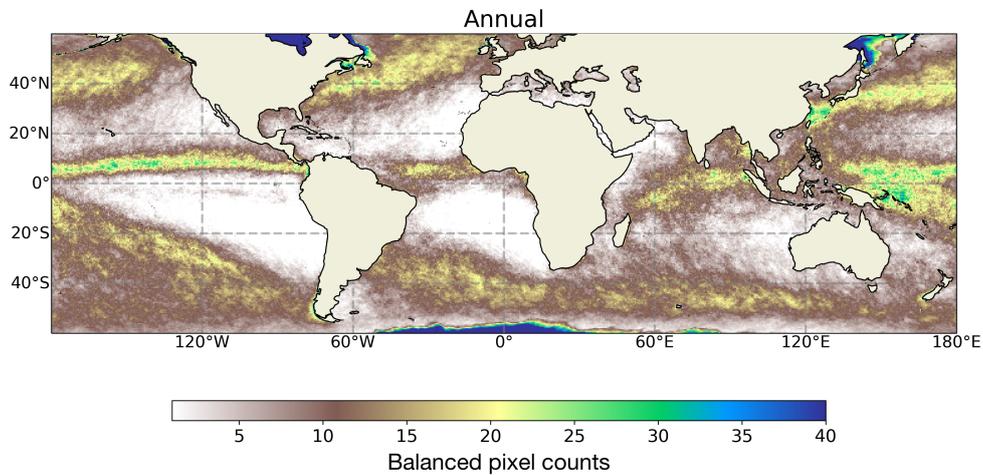
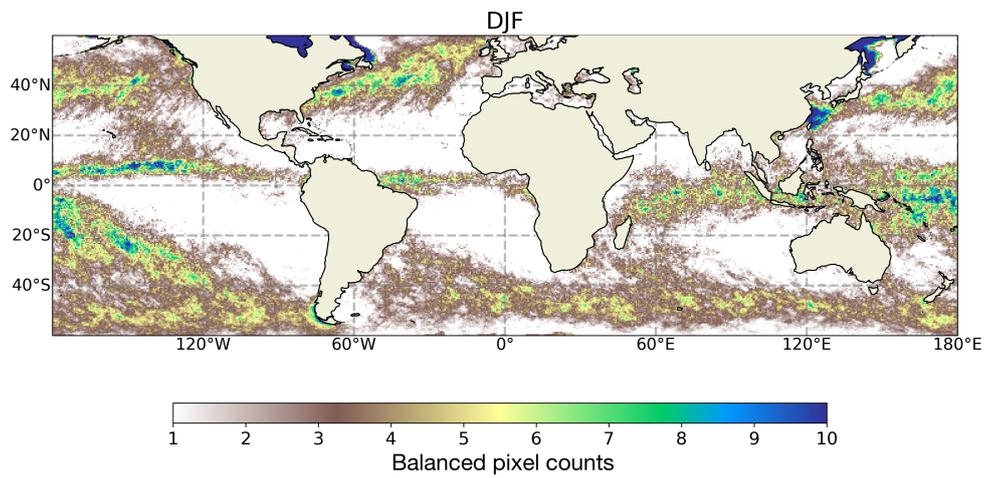


Figure 4.18 – Annual Precipitation over the Ocean in 2019 using pixel counts of the same rain classes as the previous sections. As there are more satellite passages near the poles than in the tropics, the pixel counts are adjusted by the total number of pixels per latitude.

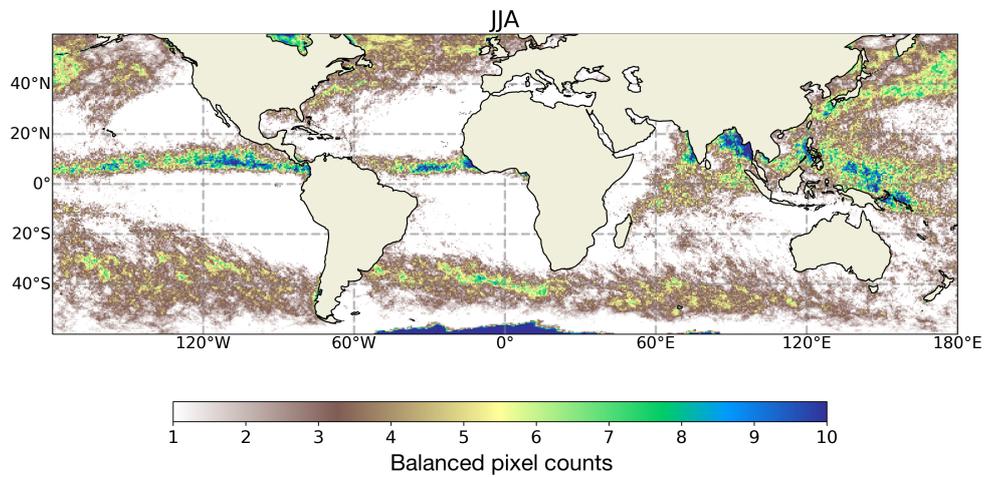
An understanding of the structure of the ITCZ is critical in answering many scientific questions, such as its implications in the monsoon dynamics, its evolution in seasonal cycles, its interannual variability, and its role in the global energy budget. Based on observational and reanalysis data between 1980 and 2014, (Lau and Tao, 2020) found that there is a multidecadal trend of a narrowing and intensifying ITCZ precipitation. The precipitation intensity in the climatological ITCZ has increased, accompanied by drying areas in both hemispheres (Lau and Tao, 2020) Wodzicki and Rapp (2016). Using data from the Coupled Model Intercomparison Project Phase 5 (CMIP5), the authors found evidence of the ITCZ width narrowing as the climate warms (Byrne and Schneider, 2016a). The width of the ITCZ getting smaller means a substantial impact on the regional water cycle, especially around the edge of the ITCZ.

The annual mean latitude of the ITCZ has been the subject of interest in numerous studies, while the width (meridional extent) has gained relatively smaller interest (Byrne and Schneider, 2016b). There are several definitions of the ITCZ width. (Dias and Pauluis, 2011) defined the ITCZ as a cross-equatorial meridional distance between latitudes with brightness temperature being equal to a certain threshold. Due to the sensitivity to brightness temperatures, another definition of the ITCZ width is proposed in (Byrne and Schneider, 2016b) using the pressure velocity.

Figure B.6 shows the occurrences of rain class over the ocean in DJF and JJA.



(a) ITCZ from segmented objects over the ocean in DJF



(b) ITCZ from segmented objects over the ocean in JJA

Figure 4.19 – Rainfall on the ocean using the precipitation classes. Figure (a) is the data from DJF (December 2018 and January-February 2019). Figure (b) shows data from JJA (June- July- August 2019).

From the count of precipitation class occurrences, we can observe a seasonal shift between DJF and JJA. In JJA, the ITCZ precipitation band is thicker than in DJF. In the Pacific Ocean, the precipitation occurs closer to the coast of Mexico. In addition, the SPCZ sees more rainfall in DJF than in JJA. In the Atlantic Ocean, the precipitation band shifts North from DJF to JJA. In the Indian Ocean, there is more rainfall in JJA. The wet area also moves up North in JJA.

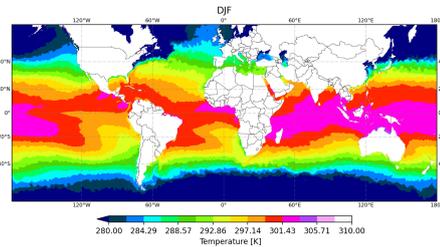
The pixel count of rain classes over the ocean shows a coherent annual precipitation structure over the ocean. These classes can also capture the seasonality of precipitation in DJF and JJA. The global analysis is limited to only the ocean due to the models' issue in distinguishing between precipitation and snow-covered surfaces at higher latitudes.

4.4.3 . Sea surface temperature (SST)

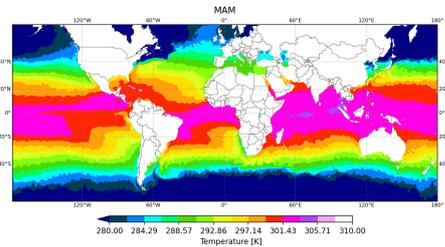
Sea surface temperature is a key variable in determining the interaction between the atmosphere and the ocean (O'carroll et al., 2019). It is essential in quantifying large ocean gyres and atmospheric circulation cells, which have major societal impacts by influencing climate, weather systems, local phenomena (sea breeze and convective clouds), and severe storms. Obtaining a daily map of SST leads to more accurate maritime forecasting, while a decadal record of SST allows for an understanding of the role of the ocean in climate variability. Efforts in mapping the SST include retrieval algorithms for infrared and microwave brightness temperatures (Chen et al., 2022).

The sea surface temperature (SST) data to be compared to the segmentation is the ESA SST CCI and C3S reprocessed sea surface temperature analyses provided by the E.U. Copernicus Marine Service Information (DOI : <https://doi.org/10.48670/moi-00169>). The seasonal means (DJF, MAM, JJA, and SON) are calculated from the daily global analyzed SST of the dataset. The seasonal average of SST in Kelvin is shown in Figure 4.20 (a)-(d).

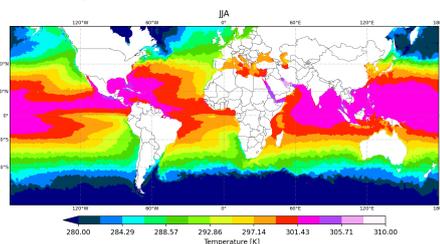
Figure 4.20 (e)-(h) shows how the ocean classes represent the SST. The label for each gridpoint is obtained by the majority vote based on the number of occurrences of each class. This representation shows a north-to-south gradient with isothermal zones varying between seasons. The temperatures decrease further away from the equator, which reflects the heating received from the Sun. Across all seasons, in the Atlantic Ocean, we observe higher temperatures in the ocean basin of the eastern border of the United States in both Figures 4.20 (a)-(d) and 4.20 (e)-(h). Similarly, the temperature is higher in the North Western of the Pacific Ocean. The western border of most ocean basins is warmer due to the effects of the ocean gyres transporting warmer water from the tropics to higher latitudes (Merchant et al., 2019). The eastern basins, on the other hand, receive the return flow from the Polar regions to the equator (Merchant et al., 2019).



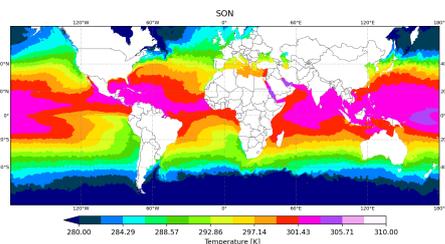
(a) SST averaged over December 2018 to February 2019



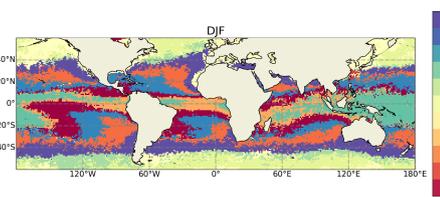
(b) SST averaged over March to May 2019



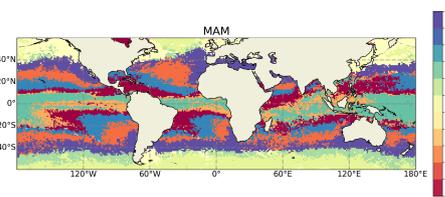
(c) SST averaged over June to July 2019



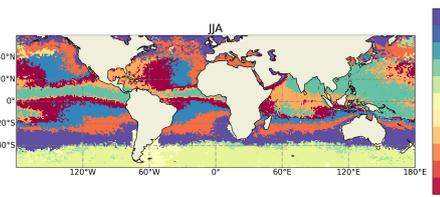
(d) SST averaged over September to November 2019



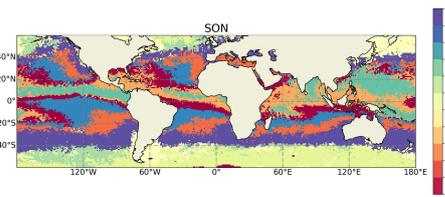
(e) Segmentation majority votes for December 2018 to February 2019



(f) Segmentation majority votes for March to May 2019



(g) Segmentation majority votes for June to July 2019



(h) Segmentation majority votes for September to November 2019

Figure 4.20 – Seasonal SST comparison. (a)-(d) Seasonal average SST using data from the ESA SST CCI and C3S reprocessed sea surface temperature analyses. (e)-(h) SST represented by the ocean classes from the segmentation. The label for each pixel is obtained by the majority vote from the occurrences of each class.

To validate the ocean classes, we compare them to some seasonal characteristics of the SST. For example, the SST in the Gulf of Mexico is uniformly warm in JJA, while in other seasons, the SST is less warm with a slightly tilted North-to-South gradient (Allard et al., 2016). The ocean classes in Figure 4.20 (e)-(h) show a seasonal pattern over the Gulf of Mexico. In DJF and MAM, the area is represented by lower SST classes. The gradient in SST can be observed in MAM and SON. In JJA, the representation in the region is uniform, as suggested by (Allard et al., 2016). However, for DJF, the gradient is unclear, which might be due to an under-representation of the cold SST ranges. The seasonal cycle of SST can also be observed in the Oman Sea and the Bay of Bengal. The higher temperature classes move up to the coast in JJA and SON, while DJF and MAM see lower temperature classes.

Historically, El Niño is described as warmer SST than usual around the coast of Peru in DJF (Cai et al., 2020). As 2019 is an El Niño year, we observed higher temperatures along the coast of Peru in Figure 4.20 (a). In Figure 4.20 (e), the third highest temperature class (Class 0) extends to the coast of Peru in DJF.

These results show that the classes corresponding to SST are able to capture its global structure and seasonality. Each class regroups SST at similar levels. Therefore, depending on the objective, more SST classes allow for the representation of more ranges of SST.

Conclusion

This study sets out to develop a tool to extract information from brightness temperature (TB) images without any ancillary data or labeling. The chosen approach is an unsupervised semantic segmentation of TB images. In this study, we tested FCN models (M_i , U_i , and L_i) that take into account the spatial information of an image, as well as a pixel-wise model (K_i) that only considers the TB values.

The first challenge to solve is the comparison of models' performance and the selection of the top model. In this study, we propose to evaluate the segmentation using two external variables : precipitation and SST. First, the classes corresponding to precipitation and SST are identified. Then, they are compared to the reference precipitation and SST data using traditional segmentation metrics. All models tested achieved very similar scores. The top-performing model is chosen based on its overall performance on 450 orbits and the level of detail in segmentation that it can offer. The next step is to verify that the classes can capture the patterns and seasonality of these geophysical variables.

The results on precipitation in BOB and WAF, as well as precipitation over the ocean, show that the precipitation classes are able to show intra-annual patterns of

precipitation. They also provide a good structure for precipitation events. The SST classes show the seasonality of the ocean's temperature. These results confirm the validity of the segmentation classes.

As these classes have proven to be meaningful, the next step is to apply the unsupervised model to a longer dataset and analyze the evolution of these classes. It is interesting to further explore what kind of information is contained in these classes and study how they evolved in the 10 years of GMI data. In addition to GMI, an unsupervised approach is key to analyzing data from other microwave remote-sensing instruments. It is an approach that does not require any ancillary data besides the brightness temperature images while offering observation of several geophysical variables.

In this study, rain classes are defined simply as classes that correspond to rain pixels when compared to the reference precipitation data. However, these classes might hold more information. For example, they could represent rain of different intensities and types (convective or stratiform). In addition, the segmentation class might represent more than one geophysical variable. They could, for example, be a mixture of two geophysical variables. Furthermore, segmentation offers other classes besides precipitation and SST, which have not yet been explored in this study. For example, there are classes that closely resemble soil moisture and permafrost. In future studies, the properties of all classes need to be analyzed.

Another thing to address in future studies is the improvement of the model, as it does not perform well for precipitation at higher latitudes. Without any supervision, the signal in the brightness temperature is not enough for the model to distinguish between heavy to moderate rain and snow on the ground. This limits the study for precipitation to the latitudes between 40°S to 40° N.

4.5 . References

- J. Allard, J. V. Clarke III, and B. D. Keim. Spatial and temporal patterns of in situ sea surface temperatures within the Gulf of Mexico from 1901-2010. *American Journal of Climate Change*, 5(3) :314–343, 2016.
- D. Arthur, S. Vassilvitskii, et al. k-means++ : The advantages of careful seeding. In *Soda*, volume 7, pages 1027–1035, 2007.
- S. Balachandran, B. Geetha, K. Ramesh, R. Deepa, Y. Mourya, and K. Rakhil. Southern Peninsular India : Southwest Monsoon - Report, 2019. URL https://mausam.imd.gov.in/chennai/mcdata/sw_monsoon_2019.pdf. Accessed on May 17th, 2014.
- P. Bhardwaj and O. Singh. Climatological characteristics of Bay of Bengal tropical cyclones : 1972–2017. *Theoretical and Applied Climatology*, 139 :615–629, 2020.
- R. J. Bombardi, V. Moron, and J. S. Goodnight. Detection, variability, and predictability of monsoon onset and withdrawal dates : A review. *International Journal of Climatology*, 40(2) :641–667, 2020.
- M. P. Byrne and T. Schneider. Narrowing of the ITCZ in a warming climate : Physical mechanisms. *Geophysical Research Letters*, 43(21) :11–350, 2016a.
- M. P. Byrne and T. Schneider. Energetic constraints on the width of the intertropical convergence zone. *Journal of Climate*, 29(13) :4709–4721, 2016b.
- W. Cai, M. J. McPhaden, A. M. Grimm, R. R. Rodrigues, A. S. Taschetto, R. D. Garreaud, B. Dewitte, G. Poveda, Y.-G. Ham, A. Santoso, et al. Climate impacts of the El Niño–southern oscillation on south america. *Nature Reviews Earth & Environment*, 1(4) :215–231, 2020.
- T. Calinski and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1) :1–27, 1974.
- K. Carver, C. Elachi, and F. Ulaby. Microwave remote sensing from space. *Proceedings of the IEEE*, 73(6) :970–996, 1985. doi : 10.1109/PROC.1985.13230.
- Z. Chen, R. Jin, Q. Li, G. Zhao, C. Xiao, Z. Lei, and Y. Huang. Joint inversion algorithm of sea surface temperature from microwave and infrared brightness temperature. *IEEE Transactions on Geoscience and Remote Sensing*, 60 :1–13, 2022.

- Y. Cui, X. Chen, J. Gao, B. Yan, G. Tang, and Y. Hong. Global water cycle and remote sensing big data : overview, challenge, and opportunities. *Big Earth Data*, 2(3) :282–297, 2018. doi : 10.1080/20964471.2018.1548052. URL <https://doi.org/10.1080/20964471.2018.1548052>.
- D. L. Davies and D. W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2) :224–227, 1979. doi : DOIifavailable.
- J. Dias and O. Pauluis. Modulations of the phase speed of convectively coupled kelin waves by the ITCZ. *Journal of the atmospheric sciences*, 68(7) :1446–1459, 2011.
- W. Dorigo, S. Dietrich, F. Aires, L. Brocca, S. Carter, J.-F. Cretaux, D. Dunkerley, H. Enomoto, R. Forsberg, A. Güntner, et al. Closing the water cycle from observations across scales : Where do we stand? *Bulletin of the American Meteorological Society*, 102(10) :E1897–E1935, 2021.
- S. Gadgil, P. Francis, and P. Vinayachandran. Summer monsoon of 2019. *Current Science*, 117(5) :783–793, 2019.
- A. Y. Hou, R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi. The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, 95(5) :701 – 722, 2014. doi : 10.1175/BAMS-D-13-00164.1. URL <https://journals.ametsoc.org/view/journals/bams/95/5/bams-d-13-00164.1.xml>.
- M. Iturbide, J. M. Gutiérrez, L. M. Alves, J. Bedia, E. Cimadevilla, A. S. Cofiño, R. Cerezo-Mota, A. Di Luca, S. H. Faria, I. Gorodetskaya, et al. An update of IPCC climate reference regions for subcontinental analysis of climate model data : definition and aggregated datasets. *Earth System Science Data Discussions*, 2020 :1–16, 2020.
- P. Joseph, K. Sooraj, and C. Rajan. The summer monsoon onset process over South Asia and an objective method for the date of monsoon onset over Kerala. *International Journal of Climatology : A Journal of the Royal Meteorological Society*, 26(13) :1871–1893, 2006.
- A. Kekesi, G. Skofronick Jackson, G. Huffman, D. Kirschbaum, R. Gran, R. Fitzgibbons, and R. Gutro. NASA’s Scientific Visualization Studio, 2023. URL <https://svs.gsfc.nasa.gov/4354>. Accessed on May 17th, 2014.
- O. Kelly. NASA’s Scientific Visualization Studio, 2020. URL <https://gpm.nasa.gov/applications/weather/gpm-sees-cyclone-amphan-approach-india>. Accessed on May 17th, 2014.

- C. Kidd, A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum. So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, 98(1) :69–78, 2017.
- W. Kim, A. Kanezaki, and M. Tanaka. Unsupervised learning of image segmentation based on differentiable feature clustering. *IEEE Transactions on Image Processing*, 29 :8055–8068, 2020.
- D. B. Kirschbaum, G. J. Huffman, R. F. Adler, S. Braun, K. Garrett, E. Jones, A. McNally, G. Skofronick-Jackson, E. Stocker, H. Wu, and B. F. Zaitchik. NASA's remotely sensed precipitation : A reservoir for applications users. *Bulletin of the American Meteorological Society*, 98(6) :1169 – 1184, 2017. doi : 10.1175/BAMS-D-15-00296.1. URL <https://journals.ametsoc.org/view/journals/bams/98/6/bams-d-15-00296.1.xml>.
- C. Kummerow, W. Barnes, T. Kozu, J. Shiue, and J. Simpson. The tropical rainfall measuring mission (TRMM) sensor package. *Journal of atmospheric and oceanic technology*, 15(3) :809–817, 1998.
- C. Kummerow, Y. Hong, W. Olson, S. Yang, R. Adler, J. McCollum, R. Ferraro, G. Petty, D.-B. Shin, and T. Wilheit. The evolution of the goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. *Journal of Applied Meteorology and Climatology*, 40(11) :1801–1820, 2001.
- W. K. M. Lau and W. Tao. Precipitation–radiation–circulation feedback processes associated with structural changes of the ITCZ in a warming climate during 1980–2014 : An observational portrayal. *Journal of Climate*, 33(20) :8737 – 8749, 2020. doi : <https://doi.org/10.1175/JCLI-D-20-0068.1>. URL <https://journals.ametsoc.org/view/journals/clim/33/20/jcliD200068.xml>.
- C. J. Merchant, P. J. Minnett, H. Beggs, G. K. Corlett, C. Gentemann, A. R. Harris, J. Hoyer, and E. Maturi. Global sea surface temperature. In *Taking the Temperature of the Earth*, pages 5–55. Elsevier, 2019.
- A. G. O'carroll, E. M. Armstrong, H. M. Beggs, M. Bouali, K. S. Casey, G. K. Corlett, P. Dash, C. J. Donlon, C. L. Gentemann, J. L. Høyer, et al. Observational needs of sea surface temperature. *Frontiers in Marine Science*, 6 :420, 2019.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.

- S. Pfreundschuh, P. J. Brown, C. D. Kummerow, P. Eriksson, and T. Norrestad. GPROF-NN : a neural-network-based implementation of the goddard profiling algorithm. *Atmospheric Measurement Techniques*, 15(17) :5033–5060, 2022. doi : 10.5194/amt-15-5033-2022. URL <https://amt.copernicus.org/articles/15/5033/2022/>.
- R. K. Pradhan, Y. Markonis, M. R. Vargas Godoy, A. Villalba-Pradas, K. M. Andreadis, E. I. Nikolopoulos, S. M. Papalexiou, A. Rahim, F. J. Tapiador, and M. Hanel. Review of GPM IMERG performance : A global perspective. *Remote Sensing of Environment*, 268 :112754, 2022. ISSN 0034-4257. doi : <https://doi.org/10.1016/j.rse.2021.112754>. URL <https://www.sciencedirect.com/science/article/pii/S0034425721004740>.
- K. A. Quagraine, F. Nkrumah, C. Klein, N. A. B. Klutse, and K. T. Quagraine. West African summer monsoon precipitation variability as represented by reanalysis datasets. *Climate*, 8(10) :111, 2020.
- A. Rai, S. K. Saha, S. Pokhrel, K. Sujith, and S. Halder. Influence of preonset land atmospheric conditions on the Indian summer monsoon rainfall variability. *Journal of Geophysical Research : Atmospheres*, 120(10) :4551–4563, 2015.
- Remote Sensing Systems. GHRSSST Level 3U Global Subskin Sea Surface Temperature from GMI onboard GPM satellite, 2017. URL <https://podaac.jpl.nasa.gov/dataset/GMI-REMSS-L3U-v8.2a>.
- O. Ronneberger, P. Fischer, and T. Brox. U-net : Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015 : 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.
- V. Sambath, N. Viltard, L. Barthès, A. Martini, and C. Mallet. Unsupervised domain adaptation for global precipitation measurement satellite constellation using Cycle Generative Adversarial Nets. *Environmental Data Science*, 1 :e24, 2022. doi : 10.1017/eds.2022.16.
- D. Sculley. Web-scale k-means clustering. In *Proceedings of the 19th international conference on World wide web*, pages 1177–1178, 2010.
- G. Skofronick-Jackson, W. A. Petersen, W. Berg, C. Kidd, E. F. Stocker, D. B. Kirschbaum, R. Kakar, S. A. Braun, G. J. Huffman, T. Iguchi, P. E. Kirstetter, C. Kummerow, R. Meneghini, R. Oki, W. S. Olson, Y. N. Takayabu, K. Furukawa, and T. Wilheit. The global precipitation measurement (GPM) mission for science and society. *Bulletin of the American Meteorological Society*, 98(8) :1679 – 1695, 2017. doi : <https://doi.org/10>.

- 1175/BAMS-D-15-00306.1. URL <https://journals.ametsoc.org/view/journals/bams/98/8/bams-d-15-00306.1.xml>.
- B. Sultan and S. Janicot. The West African monsoon dynamics. Part II : The “preonset” and “onset” of the summer monsoon. *Journal of climate*, 16(21) : 3407–3427, 2003.
- M. Tarek, F. P. Brissette, and R. Arsenault. Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, 24(5) :2527–2544, 2020.
- N. Viltard, P. Lepetit, C. Mallet, L. Barthès, and A. Martini. Retrieving rain rates from space borne microwave sensors using u-nets. In *Proceedings of the 10th International Conference on Climate Informatics*, pages 30–36, 2020.
- N. Viltard, V. Sambath, P. Lepetit, A. Martini, L. Barthès, and C. Mallet. Evaluation of DRAIN, a deep-learning approach to rain retrieval from gpm passive microwave radiometer. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- B. Wang, Q. Ding, and P. V. Joseph. Objective definition of the Indian Summer Monsoon onset. *Journal of Climate*, 22(12) :3303 – 3316, 2009. doi : <https://doi.org/10.1175/2008JCLI2675.1>. URL <https://journals.ametsoc.org/view/journals/clim/22/12/2008jcli2675.1.xml>.
- K. Wodzicki and A. Rapp. Long-term characterization of the Pacific ITCZ using TRMM, GPCP, and ERA-Interim. *Journal of Geophysical Research : Atmospheres*, 121(7) :3153–3170, 2016.
- J. Yuan and R. L. Miller. Seasonal variation in precipitation patterns to the global ocean : An analysis of the GPCP version 2 data set. *Global biogeochemical cycles*, 16(4) :50–1, 2002.

5 - Analysis of GMI data from 2014 to 2024

Contents

5.1	Trends in brightness temperature	123
5.2	Trend in DRAIN precipitation	126
5.2.1	Rain intensity	126
5.2.2	Total rain pixels	128
5.3	Trend in the unsupervised segmentation results	130
5.3.1	Regional analysis of precipitation	130
5.3.2	Global analysis of rain classes	138
5.3.3	SST	138
5.4	References	141

In the previous chapters, we present the supervised approach DRAIN for retrieving rain intensity from TB images and the unsupervised approach (Chapter 2), which relies on an unsupervised image segmentation model to extract multiple geophysical variables from the TB images (Chapter 4.3). In this chapter, these two approaches are used to analyze the GMI data between 2014 and 2024. The results from DRAIN are used to study the precipitation intensity and occurrence in the dataset. For the unsupervised segmentation, the evolution of the identified rain and SST classes is examined. In addition, we include a trend analysis using the brightness temperature measurements directly.

The Mann-Kendall test adjusted for seasonality (Hirsch et al., 1982) is applied to the trend in the monthly data using the Python package proposed by Hussain and Mahmud (2019). The slope and intercept of the trend are obtained using Seasonal Theil-Sen’s Slope Estimator (Hipel and McLeod, 1994) and Conover method (Conover, 1999), respectively. For annual data, the original Mann-Kendall test is used (Mann, 1945; Kendall, 1948). The Mann-Kendall tests have been used in several studies to analyze the trend in precipitation and hydrological cycles (Mondal et al., 2012; Kampata et al., 2008; Xu et al., 2010).

5.1 . Trends in brightness temperature

The TB at 89 GHz is sensitive to ice particle scattering in both land and ocean precipitations (Hou et al., 2014). The coldest 89GHz in horizontal polarization (T_{BH89}) for only rain pixels is an indicator of the convective nature of the precipi-

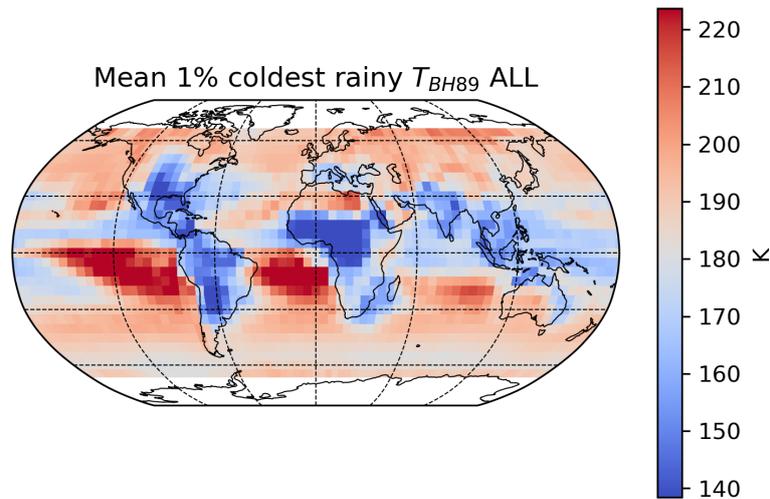


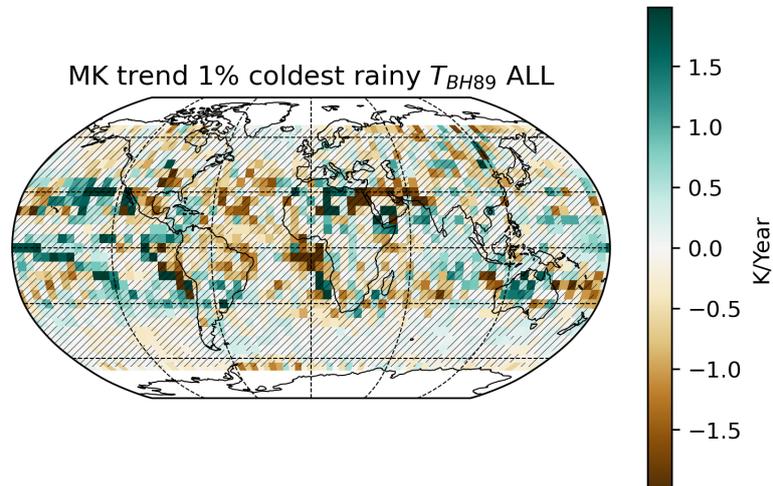
Figure 5.1 – Mean of 1% of the coldest 89 GHz (T_{BH89}) in horizontal polarization for $5^\circ \times 5^\circ$ grids using rain pixels between 2014 and 2024.

tation as it is almost proportional to the amount of ice produced in the convective parts of clouds.

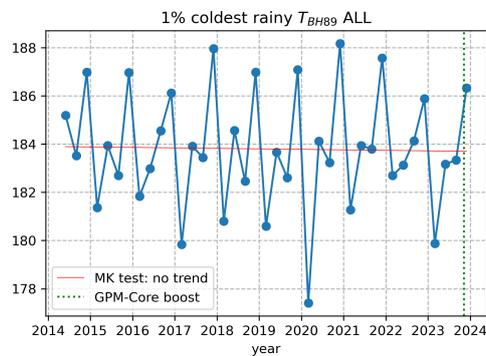
Using only rain pixels from 2014-2024, Figure 5.1 shows the average of the 1% coldest 89 GHz in horizontal polarization (T_{BH89}) for $5^\circ \times 5^\circ$ grids. The average coldest T_{BH89} corresponds to regions where convective precipitations are most likely to occur, which includes most of the tropics and subtropics regions (Liu et al., 2013). These regions include South America (particularly over the Amazon forest), southern parts of North America (including the Great Plains), central Africa and the southern Sahaelian region, the regions influenced by the Indian Summer Monsoon (ISM), the ITCZ, and the South Pacific Convergence Zone (SPCZ).

Figure 5.2a shows the map of trend per $5^\circ \times 5^\circ$ grid box for the seasonal average of the 1% coldest T_{BH89} of rain pixels between 2014 and 2024. However, there are only a few significant pixels based on the Mann-Kendall test. In some regions, we could observe a significant decreasing trend, such as in the Amazon forest, the south of North America, and the Gulf of Guinea. Around the ITCZ, one can observe some warming trends of the T_{BH89} , with very few significative grid boxes. This might suggest less convective activities around the edges of the ITCZ. The summary of water cycle changes in Chapter 1 also points to more drying areas on both sides of the hemispheres separated by the ITCZ.

Figure 5.2b shows the trend of the global seasonal average of the 1% coldest T_{BH89} of rain pixels between 2014 and 2024. There is a non-significant decreasing trend. On a global scale, there seems to be no significant increase in ice particle amounts during rainstorms.



(a) The trend in the seasonal average (DJF, MAM, JJA, SON) coldest T_{BH89} for $5^\circ \times 5^\circ$ grid boxes using only rain pixels from 2014-2024. The solid-colored grid boxes represent a significant Mann-Kendall trend.



(b) Trend in the global seasonal average (DJF, MAM, JJA, SON) of the coldest T_{BH89} for only rain pixels in the same period. The Mann-Kendall test found no significant trend.

Figure 5.2 – Trend of the seasonal average of the 1% coldest T_{BH89} of rain pixels between 2014 and 2024. Figure 5.2a shows the trend for each $5^\circ \times 5^\circ$ grid box. Figure 5.2b shows the global trend.

It should be noted that, due to saturation, the observed T_{BH89} cannot descend below a certain threshold even if there are more ice particles present in the clouds. Consequently, a decrease in the average 1% coldest T_{BH89} of the TB distribution could be due to an increase in the number of cold T_{BH89} pixels observed and not necessarily due to the decrease in the temperatures themselves.

5.2 . Trend in DRAIN precipitation

The DRAIN algorithm, presented in Chapter 2, is applied to the GMI data from 2014 to 2024 to retrieve rain intensity. Some preliminary results, including an analysis of the trend of rain intensity and the number of rain pixels, are presented here.

5.2.1 . Rain intensity

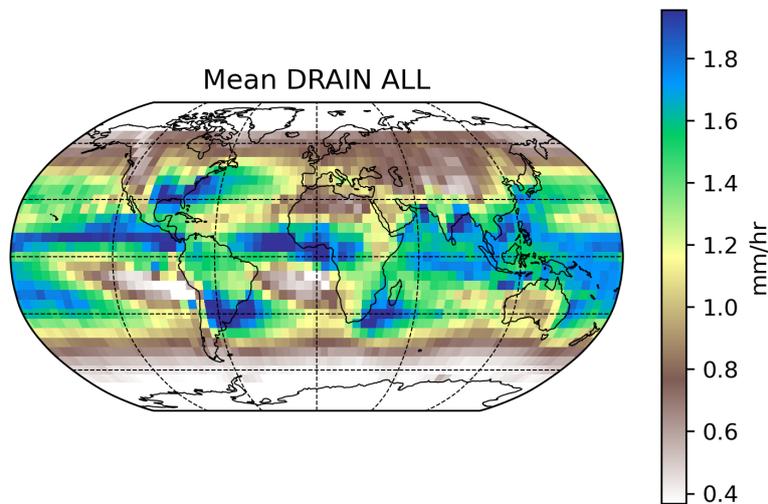
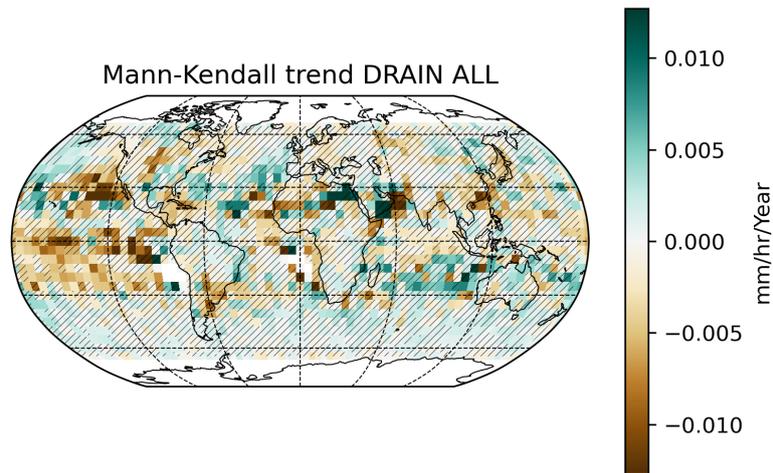


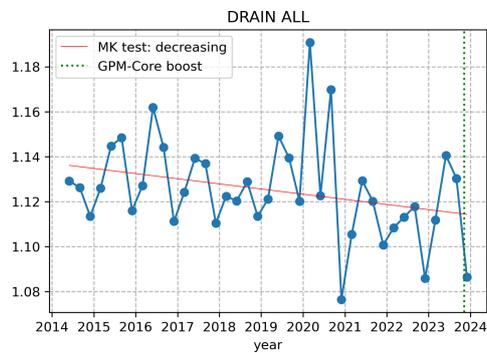
Figure 5.3 – Mean DRAIN rain intensity in a $5^\circ \times 5^\circ$ grid for GMI data from 2014 to 2024.

Figure 5.3 shows the average rain intensity per grid box of $5^\circ \times 5^\circ$ for 2014-2024. Next, we analyzed the trend in each grid box using the Mann-Kendall test. The grid boxes with stripes have no significant trend (Figure 5.4a). There are more grid boxes with downward trends than upward trends. The grid boxes with a negative slope are located mostly in the East Pacific Ocean, while those with positive slopes are in the East Atlantic Ocean, the Arabian Sea, and the Indian Ocean. However, very few of these grid boxes contain significant trends.

On a global scale, the seasonal average rain intensity is decreasing with a significant trend (Figure 5.4b). Nonetheless, as this analysis is limited to only one



(a) The trend in rain intensity for 5° x 5° grid boxes using DRAIN data from 2014-2024. The solid-colored grid boxes represent a significant Mann-Kendall trend.



(b) Trend in the global seasonal average (DJF, MAM, JJA, SON) rain intensity for the same period

Figure 5.4 – Trend in rain intensity using DRAIN data from 2014-2024. (a) shows the trend for each grid box, while (b) shows the global average rain intensity trend.

satellite, the cause of this decrease is unclear. It could be due to a general decrease in rain intensity for all precipitation events. It is also possible that the occurrences of low rain events increase, which leads to a decrease in global average intensity.

5.2.2 . Total rain pixels

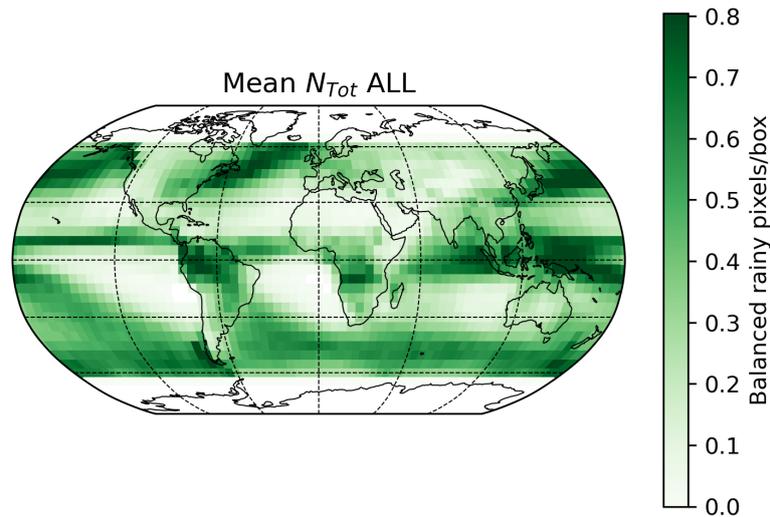


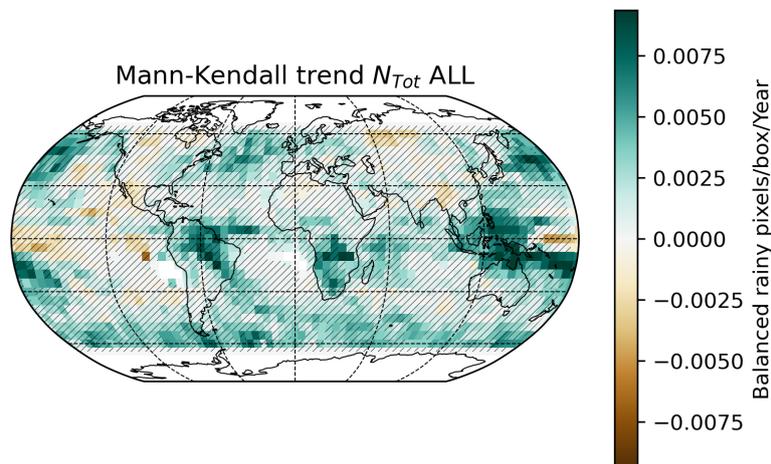
Figure 5.5 – Average occurrences of rain pixels seen by DRAIN from 2014-2024 for grid boxes of $5^\circ \times 5^\circ$. The mean number of rain pixels in each grid is adjusted for the difference in satellite passage over each latitude.

Observing the total accumulation of rain requires determining the number of occurrences, the intensity, and the duration of precipitation events. With only GMI, neither the total occurrences nor the duration is available for analysis. As a result, the number of rain pixels is used as a proxy to observe the trend in precipitation.

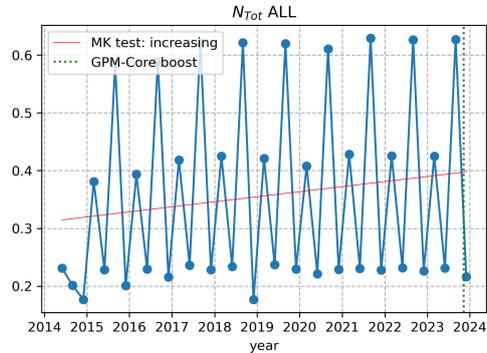
The rain pixels are pixels with rain intensity above 0.1 mm/h. Figure 5.5 shows the annual average number of rain pixels per $5^\circ \times 5^\circ$ grid box for 2014-2024. Due to the number of satellite passages per latitude in the high latitudes being more than in the tropics (more land area in the tropics than in the high latitudes), the average number of rain pixels is adjusted by a coefficient based on the total number of pixels observed by the satellites per latitude.

The average number of rain pixels per grid box is representative of the global wet and dry regions. Over the ocean, the wet regions are normally found in the ITCZ, the SPCZ, and the storm tracks in the Northwestern Pacific and Atlantic Oceans (Yuan and Miller, 2002).

Figure 5.6a shows the trend in average rainy pixels per year for the period 2014-2024 in $5^\circ \times 5^\circ$ grid boxes. Similarly to Figure 5.4a, the striped boxes are non-significant. Across the globe, most of the grid boxes show a positive trend.



(a) Trend for the annual average rain pixel counts for 2014-2024. The solid-colored grid points are statistically significant based on the Mann-Kendall test.



(b) Trend for the global average rain pixels counts for 2014-2024. The average is done per 3-month period (DJF, MAM, JJA, and SON).

Figure 5.6 – Trend in rain pixel counts using DRAIN. (a) shows the trend per $5^\circ \times 5^\circ$ grid boxes. (b) shows the trend in the global seasonal average annually.

The seasonal average of global rain pixels has a significant increasing trend (Figure 5.6b). There are many possibilities for the cause of the increase in total rain pixel counts. For example, the precipitation events could be becoming larger, last longer, or occur more frequently.

5.3 . Trend in the unsupervised segmentation results

Identified rain classes and SST classes from the chosen optimal model presented in Chapter 4.3 are examined for the GMI data between 2015-2023. The trend analysis is done for regional precipitation and global SST.

5.3.1 . Regional analysis of precipitation

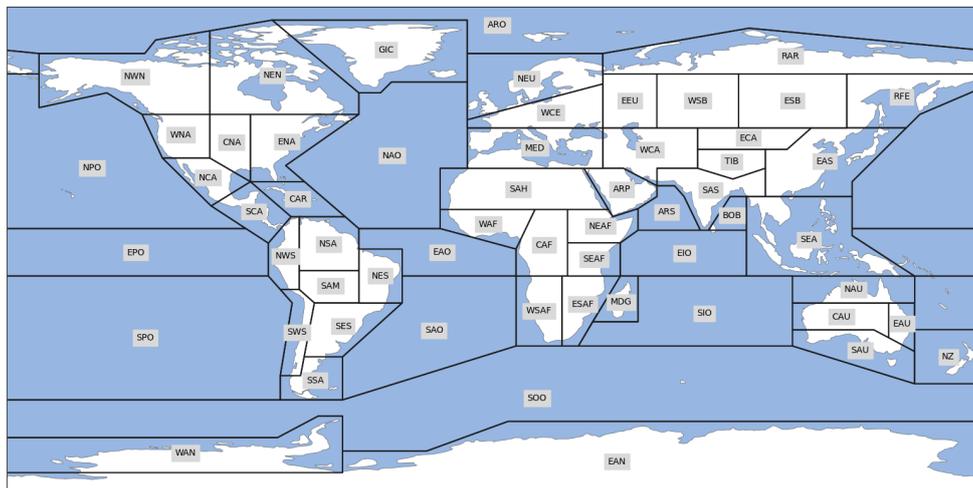


Figure 5.7 - The IPCC reference region (Iturbide et al., 2020).

Based on the summary of changes in the water cycle presented in Chapter 1, changes in precipitation differ from one region to another. Assessing the global change of precipitation could lead to averaging processes that cancel each other out.

Due to the regionalized response of the water cycle to climate change, we choose to analyze several regions for this preliminary result. Using the IPCC reference regions (Iturbide et al., 2020), also presented in Chapter 4.3, the study focuses on three regions : Western Africa (WAF), Southeast Asia (SEA), and the Bay of Bengal (BOB) (Figure 5.7). According to the trend in annual mean precipitation from 1985-2014 (Figure 1.2, Douville et al. (2023)) for these three regions include,

- Western Africa (WAF)

WAF is one of the regions that have experienced an increase in mean annual precipitation between 1985 and 2014. It should be noted that in Figure 1.2, not all trends in the gridpoints are statistically significant. For WAF, the significant pixels are located in the center of the region.

— **Bay of Bengal (BOB)**

The ocean region BOB also saw an increase in annual precipitation in 1985-2023. The trend in BOB is higher than in WAF, with a higher ratio of significant grid points.

— **Southeast Asia (SEA)**

Most of the SEA region experienced an increasing trend between 1985 and 2015. However, some parts of the region experienced a decreasing trend.

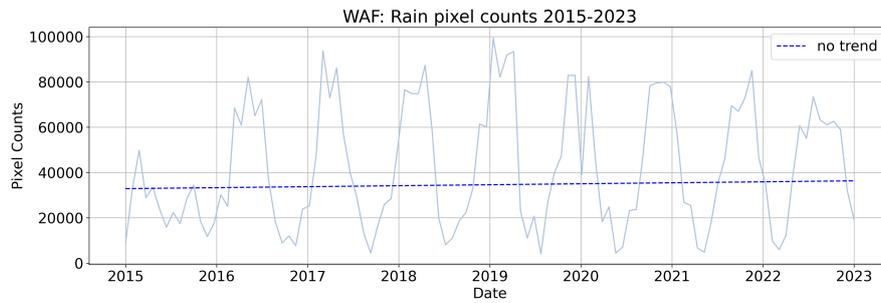
Rain pixel count

We extend the application of rain pixel count presented in Chapter 4.3 to the GMI data from 2015-2023. Figure 5.8 shows the monthly rain pixel count of all rain classes combined for WAF (Figure 5.8a), SEA (Figure 5.8b), and BOB (Figure 5.8c). For the regions WAF and BOB, there is a clear seasonality to precipitation as these regions are influenced by a monsoon regime. On the other hand, the region SEA is influenced by several regimes of Monsoons, including the Indian Summer Monsoon (ISM), the East Asian Monsoon, and the Australian Monsoon. It also includes a region where it rains almost all year round.

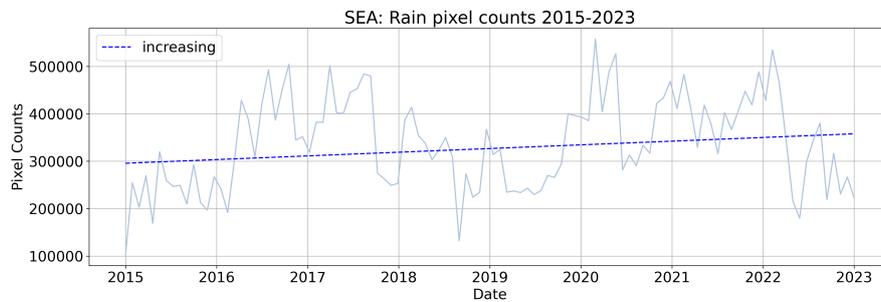
There is a positive slope in the trend for the monthly precipitation in all three regions. However, only SEA (Figure 5.8b) and BOB (Figure 5.8c) show a significant increasing trend according to the Mann-Kendall test.

Figure 5.9 shows the total pixel count annually for 2015-2023. We could observe some annual variability. The slopes of the trends in all three regions are increasing. However, there is no significant trend using the Mann-Kendall test. The time series, consisting of only nine data points, might be too short to determine any trend. The increasing trend observed in the monthly count for BOB and SEA might be due to the availability of more data points. In addition, there might be a compensational effect of increasing and decreasing rainfall occurrence in monsoon months and non-monsoon months.

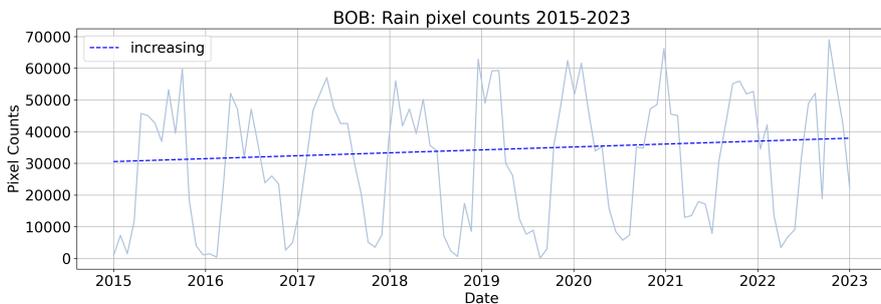
Next, the pixels are counted separately for each rain class. Compared with DRAIN rain retrieval, the four precipitation classes contain different rain intensities (Figure 5.10). Although the range differs slightly, the order of rain intensities in the rain classes is the same in WAF, which comprises mainly land, and in BOB, which comprises mainly ocean. The two classes with high rain intensity (class 41 and 63) contain relatively higher rain intensity in BOB than in WAF. The lower



(a) WAF



(b) SEA



(c) BOB

Figure 5.8 – Monthly rain pixel counts from 2015 to 2023. The pixels of the rain classes are counted for the IPCC regions : (a) WAF, (b) SEA, and (c) BOB. The dashed lines show the trend in the data. 'No trend' represents a non-significant trend, while 'increasing' represents a significant increasing trend.

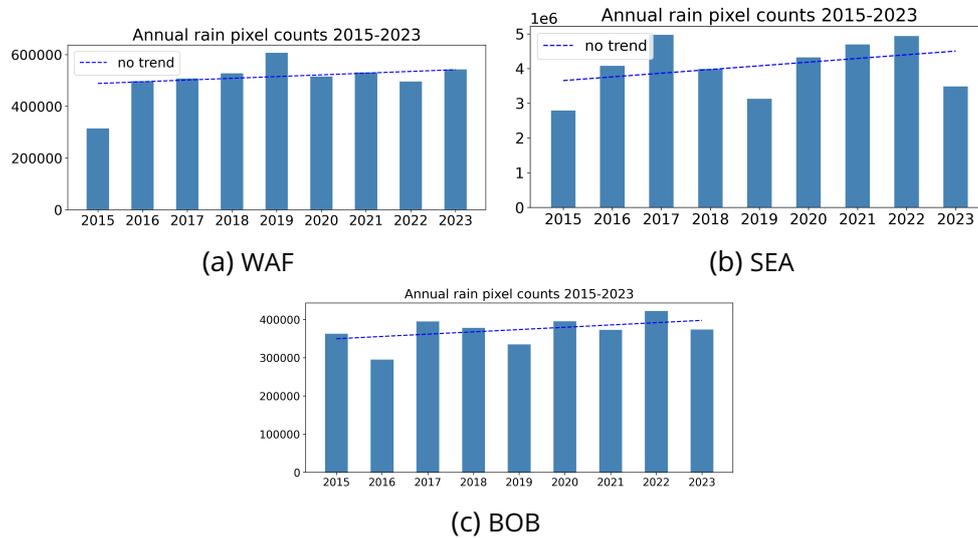


Figure 5.9 – Annual rain pixel counts from 2015 to 2023. The pixels of the rain classes are counted for the IPCC regions : (a) WAF, (b) SEA, and (c) BOB. The Mann-Kendall test shows no significant trends.

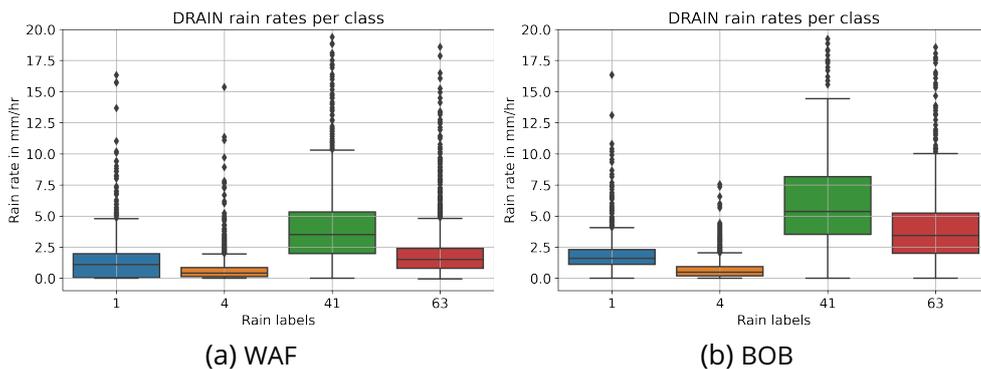
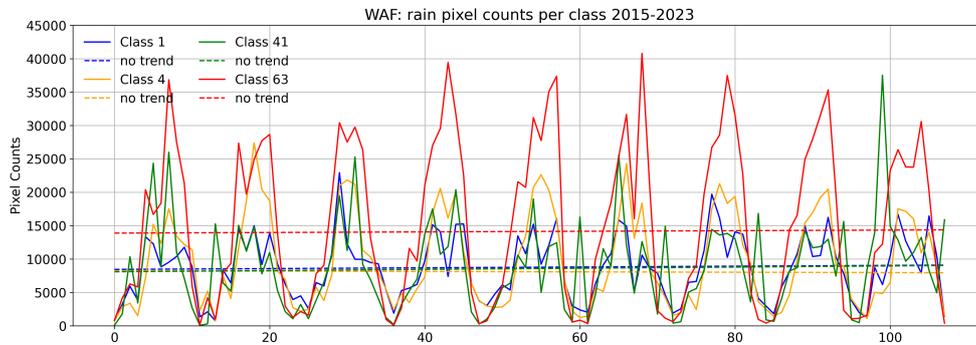
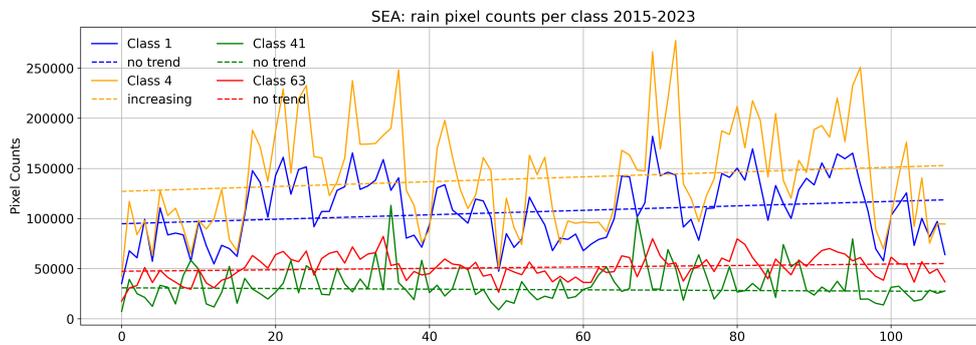


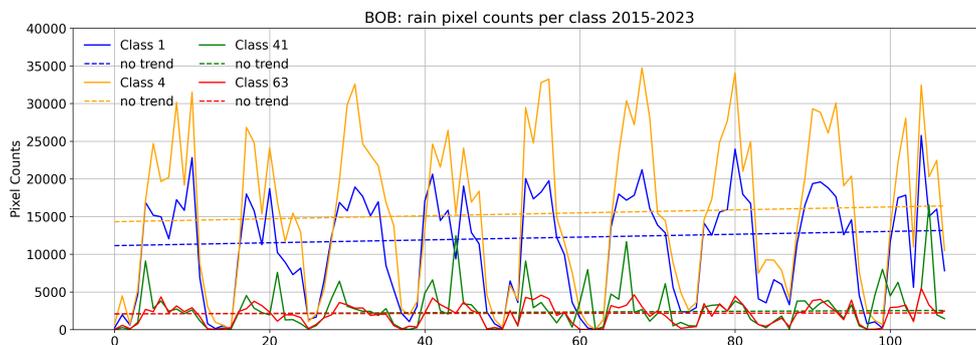
Figure 5.10 – Rain intensity per class by comparison to DRAIN data. The order of the range of rain intensity in each class from highest to lowest is class 4, 1, 63, and 41 for both regions.



(a) WAF



(b) SEA



(c) BOB

Figure 5.11 – Monthly pixel counts for each rain class from 2015 to 2023 : (a) WAF, (b) SEA, and (c) BOB. The order of rain intensity from lowest to highest is Class 4, 1, 63, 41. The Mann-Kendall test shows a significant increasing trend in SEA for Class 4.

rain intensity classes (classes 1 and 4) are also lower in WAF.

Figure 5.11 shows the pixel count for each rain class in the three regions. The SEA and BOB regions are more dominated by the lower intensity class (Class 1 and 4, approximately 0.1 to 1 mm/h), while WAF is dominated by the medium intensity class (Class 63, approximately 1 to 2.5 mm/h).

In WAF, the rain classes do not have any trend. In BOB and SEA, the medium and low-intensity classes (classes 1 and 4) show an upward trend. However, only the low-intensity rain (class 4) shows a significant increasing trend. The significant trend in total monthly rain pixels in BOB and SEA might be contributed to by the low rain intensity classes (classes 1 and 4).

The trends in the results of the unsupervised approach tend to go in the same direction as the results from TB and DRAIN, especially the increase in low rain intensity.

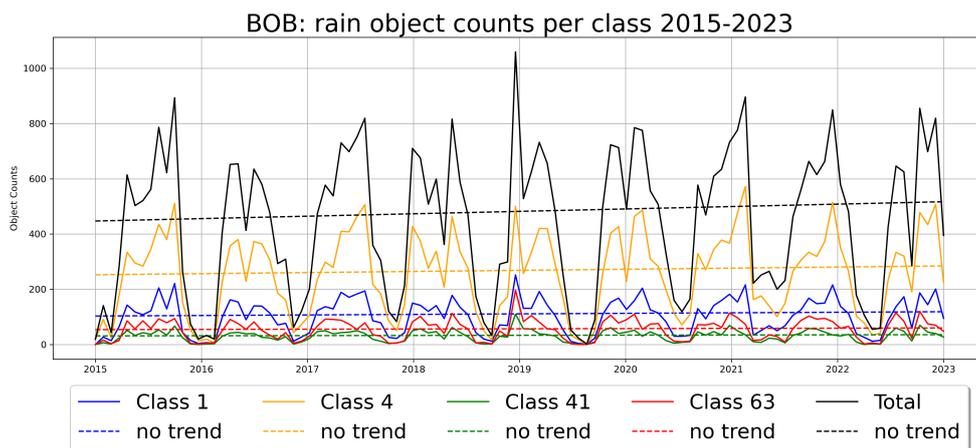
Rain object analysis

In addition to analyzing the pixel count of the rain classes, it is possible to consider occurrences of precipitation events as separate objects. This approach allows us to study the size of precipitation classes and their occurrence as separate events instead of accumulating all pixels together.

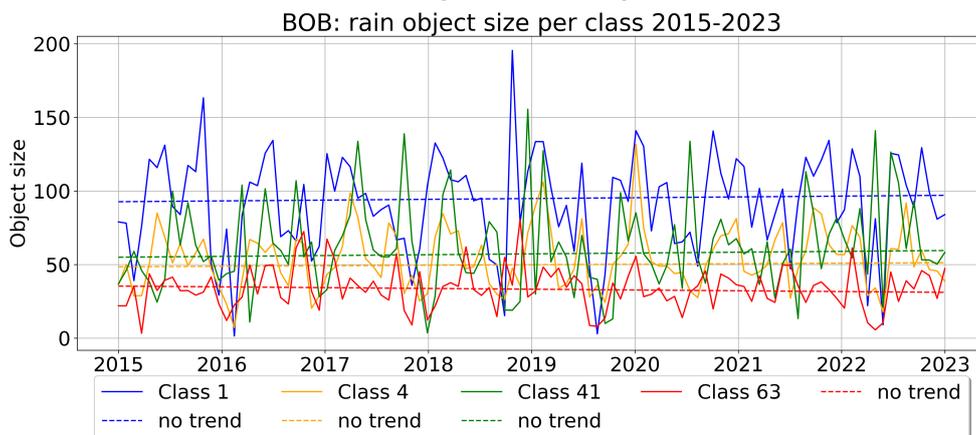
The isolation of rain objects for each rain class is done by drawing a contour around the object using the Python package OpenCV (Bradski, 2000). In order to optimize the contour detection, the rain classes are treated one by one, and the segmentation is converted to binary (1 for the selected class and 0 for other classes). The contour is defined as a continuous curve with points of the same values. However, an issue with this method is that if an object contains another object of a different class inside, the contour considers all the pixels inside the contour of such object as its size. In the results presented here, the issue with object size has not yet been resolved and will be addressed in the next iteration.

For BOB (Figure 5.12), there is a positive slope, albeit non-significant, in the trend for the total object counts. There is no clear trend in the average size of the objects. For WAF (Figure 5.13), there is no observed trend in the number of rain objects and the average rain object size.

Although the number of rain objects and the average rain object size do not show any significant trend, the total rain pixel counts per month in BOB present a significant increasing trend. There is likely a small increase in the number of objects and sizes that results in a significant increase in the total pixel counts.

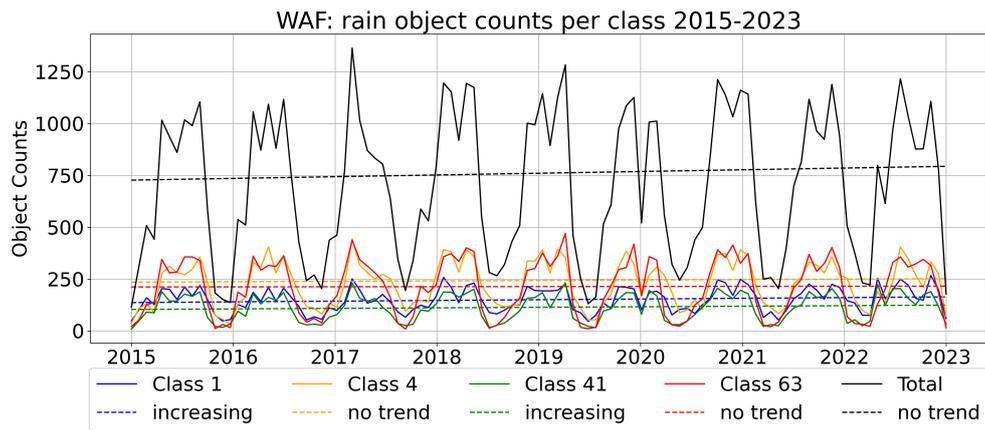


(a) Monthly number of objects

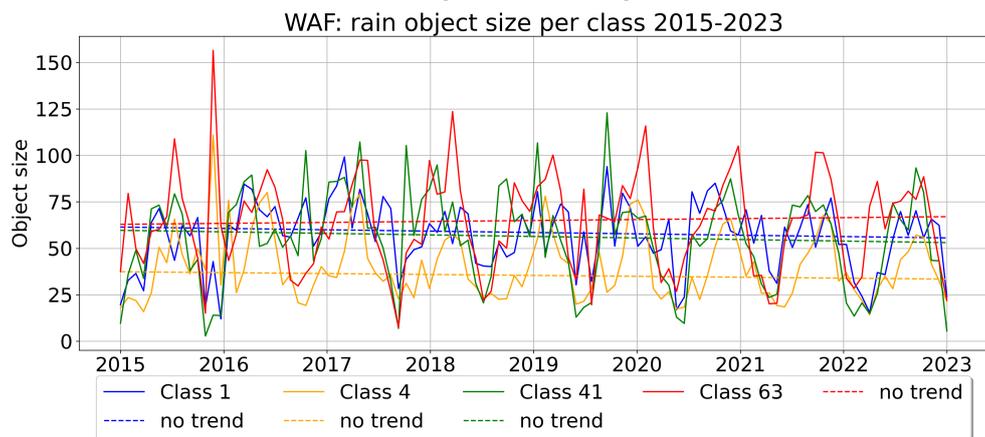


(b) Average object size per month for each rain class (To be updated)

Figure 5.12 – Monthly rain object counts per class (a) and monthly average object size for BOB (b). There is no significant trend based on the Mann-Kendall test.



(a) Monthly number of objects



(b) Average object size per month for each rain class (To be updated)

Figure 5.13 – Monthly rain object counts per class (a) and monthly average object size for BOB (b). There is no significant trend based on the Mann-Kendall test.

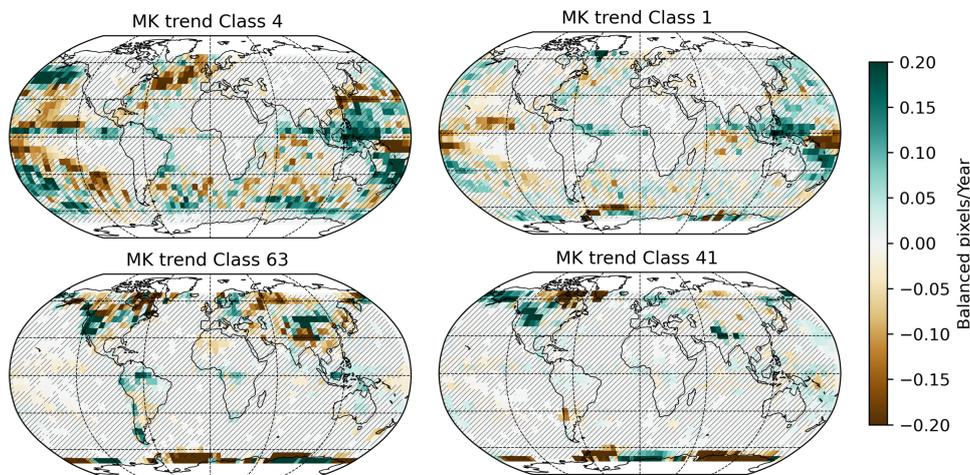


Figure 5.14 – Mann-Kendall per each grid box of 5° by 5° for each rain class. In order of increasing rain intensity representation are Class 4, Class 1, Class 63, and Class 41. The solid-colored grid points are statistically significant based on the Mann-Kendall test.

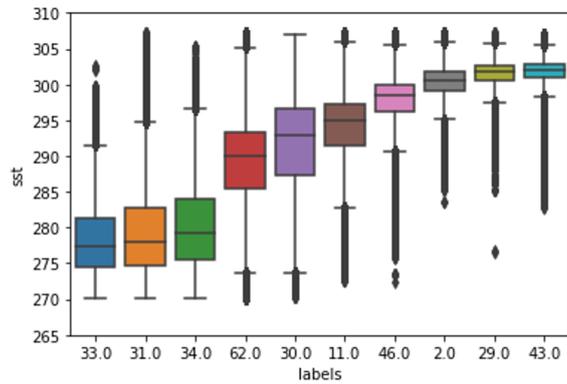
5.3.2 . Global analysis of rain classes

Similarly to the global trend analysis using DRAIN rain pixels, the Mann-Kendall trend per 5°x5° grid box is also done for each rain class of the unsupervised segmentation. The difference between the global trend here and that of DRAIN is the number of decreasing trends in the low rain intensity class (Class 4) found by the segmentation. It is possible that the segmentation found more decreasing-trend grid boxes than DRAIN because DRAIN is not very sensitive to very low rain intensity. The general structure of trends is otherwise similar for the two analysis approaches; for example, the drying trend around the ITCZ and the increasing trend in the Maritime continent. However, due to the confusion between rain and ice on the surface of the unsupervised model, some trends should not be solely interpreted as precipitation trends. This includes mainly regions that are normally covered by ice, such as the Himalayas mountain range, the permafrost, and the Arctic sea.

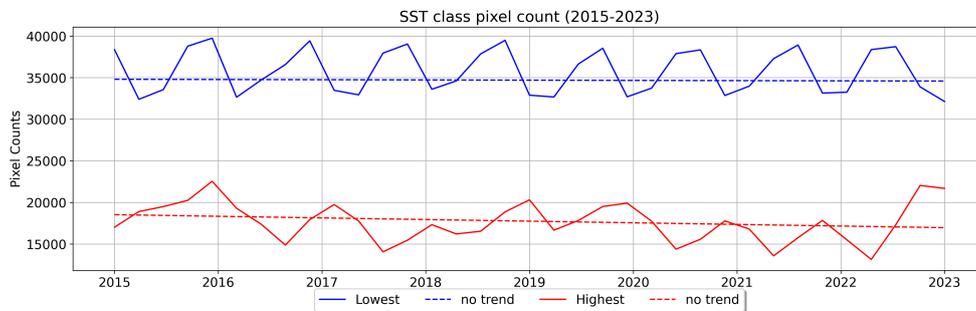
5.3.3 . SST

Similar to the previous section, the pixels of SST classes are counted for the GMI data from 2015 to 2023. The pixel counts are averaged for DJF, MAM, JJA, and SON each year. Figure 5.15a shows the ranges of SST in each SST class. Figure 5.15b only shows the highest SST (Class 43) and the lowest SST (Class 33). There is a slight downward non-significant trend for the highest SST.

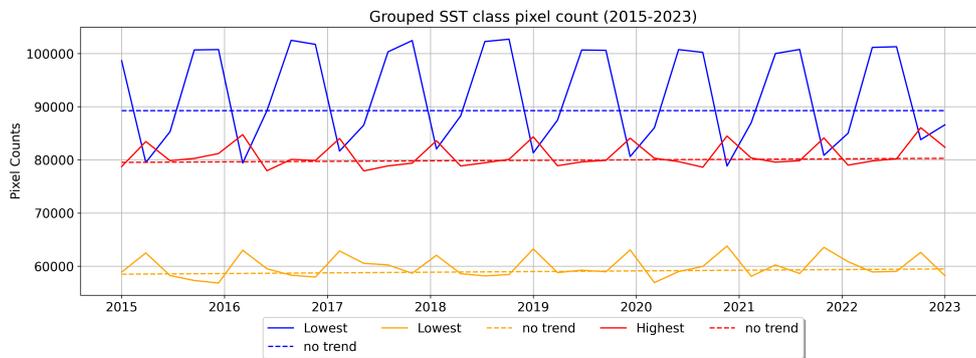
Figure 5.15c regrouped SST of similar ranges into three categories : Low SST



(a) The SST ranges in each class using comparison with GHRSSST data (Remote Sensing Systems, 2017).



(b) SST class pixel counts 2015-2023.



(c) Regrouped SST per season : Low SST (Class 22, 31, and 34), medium SST (Class 62, 30, 11, and 46), high SST (Class 2, 29, and 43).

Figure 5.15 – Seasonal mean of SST class pixel count for GMI data from 2015-2023.. Figure (a) shows the average pixel count for DJF, MAM, JJA, and SON for each class. Figure (b) shows the ranges in SST per class when compared to GHRSSST data (Remote Sensing Systems, 2017). Figure (c) shows the sum of regrouped SST classes based on their ranges of temperatures. The Mann-Kendall test does not show any significant trend.

(Class 22, 31, and 34), medium SST (Class 62, 30, 11, and 46), high SST (Class 2, 29, and 43). Instead of pixel counts, we present the proportion of each category per season to avoid the problem of missing pixels due to SST pixels being replaced by precipitation pixels in certain months. No trend is observed for any of the categories.

Conclusion

In this chapter, the trend in the coldest T_{BH89} for the rain pixels between 2014 and 2024 is analyzed due to its interpretability. We could observe some cooling in the coldest T_{BH89} . However, most regional and global trends are non-significant based on the Mann-Kendall test. The analysis of the trend is limited without an intermediate step to extract the information available in the TB dataset.

Preliminary results on DRAIN retrieval for 2014–2024 show an increase in rain pixels and a decrease in rain intensity on the global scale. Using the unsupervised segmentation classes for precipitation, the regional analysis of the GMI data from 2015–2023 shows trends that are coherent with the results from DRAIN as well as the trends from 1985 to 2014 presented in the IPCC AR6 (Douville et al., 2023). All three regions, WAF, SEA, and BOB show an upward trend, although only a few cases are statistically significant.

It should be noted that these results are obtained from a single satellite. GMI could only observe instantaneous precipitation. For wet regions, monthly counts of precipitation events could be representative of the rainfall received. For example, the results in Chapter 4.3 show that monthly pixel counts can capture monsoon seasonality in BOB and WAF. However, it could not offer a complete picture of the observed trends. For instance, the cause of the increase in the number of rain pixels or decrease in intensity is not known. The increase in the number of rain pixels could be hypothesized to be due to either an increase in precipitation events, an increase in size, or an increase in the life cycle of precipitation events. Using multiple satellites from the GPM constellation is an essential step in analyzing the trends. An unsupervised domain adaptation model to adapt between the TB observation of two satellites is presented in Chapter 6.

For SST, no particular trend is observed. The changes in the temperature might be too small, while the segmentation model does not offer enough nuances to capture the change.

5.4 . References

- G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
- W. J. Conover. *Practical nonparametric statistics*, volume 350. john wiley & sons, 1999.
- H. Douville, K. Raghavan, J. Renwick, R. P. Allan, P. A. Arias, M. Barlow, R. Cerezo-Mota, A. Cherchi, T. Y. Gan, J. Gergis, et al. *Water Cycle Changes*, page 1055–1210. Cambridge University Press, 2023.
- K. W. Hipel and A. I. McLeod. *Time series modelling of water resources and environmental systems*. Elsevier, 1994.
- R. M. Hirsch, J. R. Slack, and R. A. Smith. Techniques of trend analysis for monthly water quality data. *Water Resources Research*, 18(1) :107–121, 1982. doi : <https://doi.org/10.1029/WR018i001p00107>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/WR018i001p00107>.
- A. Y. Hou, R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi. The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, 95(5) :701 – 722, 2014. doi : 10.1175/BAMS-D-13-00164.1. URL <https://journals.ametsoc.org/view/journals/bams/95/5/bams-d-13-00164.1.xml>.
- M. Hussain and I. Mahmud. pyMannKendall : a python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software*, 4(39) : 1556, 7 2019. ISSN 2475-9066. doi : 10.21105/joss.01556. URL <http://dx.doi.org/10.21105/joss.01556>.
- M. Iturbide, J. M. Gutiérrez, L. M. Alves, J. Bedia, R. Cerezo-Mota, E. Gimenez, A. S. Cofiño, A. Di Luca, S. H. Faria, I. V. Gorodetskaya, et al. An update of IPCC climate reference regions for subcontinental analysis of climate model data : definition and aggregated datasets. *Earth System Science Data*, 12(4) : 2959–2970, 2020.
- J. Kampata, B. Parida, and D. Moalafhi. Trend analysis of rainfall in the head-streams of the Zambezi River Basin in Zambia. *Physics and Chemistry of the Earth, Parts A/B/C*, 33(8) :621–625, 2008. ISSN 1474-7065. doi : <https://doi.org/10.1016/j.pce.2008.06.012>. URL <https://www.sciencedirect.com/science/article/pii/S1474706508001332>. Integrated Water Resources Management - From Concept to Practice.

- M. G. Kendall. Rank correlation methods. 1948.
- P. Liu, C. Li, Y. Wang, and Y. Fu. Climatic characteristics of convective and stratiform precipitation over the Tropical and Subtropical areas as derived from TRMM PR. *Science China Earth Sciences*, 56 :375–385, 2013.
- H. B. Mann. Nonparametric tests against trend. *Econometrica : Journal of the econometric society*, pages 245–259, 1945.
- A. Mondal, S. Kundu, A. Mukhopadhyay, et al. Rainfall trend analysis by Mann-Kendall test : A case study of north-eastern part of Cuttack district, Orissa. *International Journal of Geology, Earth and Environmental Sciences*, 2(1) :70–78, 2012.
- Remote Sensing Systems. GHRSSST Level 3U Global Subskin Sea Surface Temperature from GMI onboard GPM satellite, 2017. URL <https://podaac.jpl.nasa.gov/dataset/GMI-REMSS-L3U-v8.2a>.
- K. Xu, J. D. Milliman, and H. Xu. Temporal trend of precipitation and runoff in major Chinese Rivers since 1951. *Global and Planetary Change*, 73(3-4) : 219–232, 2010.
- J. Yuan and R. L. Miller. Seasonal variation in precipitation patterns to the global ocean : An analysis of the GPCP version 2 data set. *Global biogeochemical cycles*, 16(4) :50–1, 2002.

6 - Unsupervised domain adaptation

Contents

6.1	Unsupervised domain adaptation as a transfer learning approach for DRAIN	144
6.2	References	161

The Global Precipitation Measurement (GPM) mission aims to provide a high-quality global observation of precipitation (Hou et al., 2014). To achieve this purpose, the GPM mission relies on the collaboration between multi-national space agencies to create a constellation of satellites that offer unprecedented global coverage and revisit time. Most of these satellites share a similar configuration for their instruments : they carry either conical-scanning or cross-track-scanning microwave radiometers with frequencies ranging between 6 and 183 GHz. However, not all of these satellites carry aboard the precipitation radar as the GPM Core Observatory. The difference in the data distribution in the constellation degrades the performance of DRAIN. In the absence of the precipitation radar, it is impossible to retrain DRAIN in the same way as on GMI. In addition, using overpasses between GMI and other satellites offers a very limited rainfall dataset due to differences in trajectory and the intermittency of rain. Therefore, a transfer learning step is essential.

The unsupervised segmentation model can be trained to any of the microwave radiometers in the constellation and be analyzed separately. However, to take advantage of the increased number of observations, it is important to harmonize the classes obtained from different satellites. This could be done after the segmentation by comparing the properties of the classes and regrouping similar ones. Another possible approach is to do a domain adaptation approach to adapt the brightness temperature images from other satellites to the domain of the GMI before applying the segmentation model. This can ensure that the properties of the resulting segmentations are the same.

Unsupervised domain adaptation is advantageous in our case because it requires a large dataset for each domain but does not require the images in each dataset to be of the same observations. Consequently, it can be easily applied to the constellation without requiring overpass data.

This Chapter presents a study on a transfer learning approach using unsupervised domain adaptation for DRAIN. The transferability of DRAIN is a good

indicator of the transferability of the unsupervised segmenter.

6.1 . Unsupervised domain adaptation as a transfer learning approach for DRAIN

The Deep Learning algorithm for rain retrieval from the brightness temperature (DRAIN) described in Chapter 2 is trained on the GPM Core Observatory. Its input database is from GMI. Consequently, the slight difference in radiometer configuration in the GPM constellation does not allow a direct application of DRAIN on new satellites. In order to test the transferability of DRAIN to a new satellite, a transfer learning step is explored for SSMI/S, a microwave imager aboard the DMSP-F18 Defense Meteorological Satellite Program.

The SSMI/S is chosen due to its similarity to GMI, including similar frequency channels, particularly those for training DRAIN. However, there are also differences in data distribution between these two satellites. First, SSMI/S orbits at higher altitudes than GMI, which causes its pixel size to be bigger. Due to the complex nature of precipitation processes, this difference could lead to the field of view of the radiometers not observing the same uniform rain field (Mallet et al., 2023; Harris and Foufoula-Georgiou, 2001). This introduces a non-linear relationship between the two data domains. In addition, as the SSMI/S orbits at higher altitudes, it also has a larger swath, which can cause differences due to the curvature of the Earth. Another difference is due to their orbitography. SSMI/S is a Sun-synchronous satellite and has global coverage, while GMI is non-sun-synchronous and does not cover latitudes higher than 70 degrees. This leads to differences in the datasets due to different meteorological situations observed.

There are many possible approaches to transfer a deep-learning model trained on one domain to another. Some approaches require a small dataset annotated ground truth. For example, fine-tuning is a process where some or all of the hyper-parameters of the previously trained model are frozen during training on a new dataset. This method is not suitable in this case since the re-training dataset is insufficient. The re-training database is created from the overpasses between the two satellites. However, these overpasses are not evenly distributed as the majority happen at latitudes higher than 60 degrees. As rain is very intermittent in time and space, the resulting overpass dataset does not contain a sufficient amount of rain observations. Moreover, most of these rain observations are of lower rain intensity happening at higher latitudes. Consequently, the dataset is not representative enough of rain events to be used as a retraining dataset. The most advantageous approach for transfer learning between DRAIN and SSMI/S is one that is not limited to training on overpasses. One way to solve this issue is to use the unsupervised domain adaptation, where only the data from each radiometer is required.

The attached article is about the application of Cycle Generative Adversarial Nets (CycleGAN) as an unsupervised domain adaptation model that adapts the brightness temperature images of the SSMI/S satellite to the domain of the GMI. The 37 GHz and 91 GHz (horizontal and vertical polarization) of the SSMI/S are adapted to the 37 GHz and 89 GHz (horizontal and vertical polarization) of the GMI, respectively. CycleGAN is a generative approach that does not require corresponding images between the two domains. It contains two GANs, each adapting from one domain to another. Most of the hyper-parameters and architecture choices are from the previous feasibility study presented in **Appendix A**. With two GANS working together, CycleGAN is very unstable and difficult to converge. The best generators are saved during each training epoch. These best generators are then tested on the rain retrieval downstream task.

In order to evaluate the rain retrieval performance of DRAIN after the domain adaptation is applied, two tactics are used : a qualitative comparison over several case studies and a quantitative analysis using overpasses. This overpass comparison includes overpasses with GMI and its precipitation radar DPR. Next, a regional evaluation of France is done using the Mosaic Rain Product provided by Météo-France (Figueras i Ventura and Tabary, 2013). The qualitative evaluation shows an improvement in precipitation structure when the data is adapted, while the quantitative evaluation shows a decrease in rain intensity errors.

The significant increase in the performance of DRAIN on SSMI/S when the data is adapted proves that deep learning models are viable approaches for quantitative precipitation retrieval in the constellation of satellites. Nonetheless, a generative approach like CycleGAN is very difficult to train. Moreover, it can only adapt between two satellites at a time. A constellation will require multiple CycleGANs, which could have a varying performance. Future applications should focus on a more stable approach that could work on multiple domains at once.

The following attached article contains the implementation and evaluation of CycleGAN for domain adaptation between GMI and SSMI/S.

Unsupervised Domain Adaptation to mitigate out-of-distribution problem of spatial radiometer images: Application to Quantitative Precipitation Estimation

Vibolroth Sambath, Natanaël Dubois-Quilici, Nicolas Viltard, Audrey Martini, Cécile Mallet

Abstract—A major issue limiting the successful deployment of deep learning algorithms in geophysical applications is their inability to generalize to new contexts. Regarding the quantitative precipitation estimation (QPE) from the Global Precipitation Mission (GPM) satellite constellation, the GPM Microwave Imager (GMI) contains enough co-located brightness temperatures and rain rates data to train a deep learning inverse model to retrieve precipitation intensity. However, the difference in instrumental configurations makes it impossible to directly apply this inverse operator to another space-borne radiometric imager. A domain adaptation is thus necessary to solve the domain shift problem encountered when applying the model trained on one satellite to another satellite. The present paper tests a method to map the SSMI/S data to the GMI data. In the absence of sufficient paired images between the two satellites, we applied a Cycle consistent Generative Adversarial Network (CycleGAN), which allows for an Unsupervised Domain Adaptation approach. Evaluating the quality of adapted images is a complex problem. This paper employs two tactics: a brief evaluation of adapted radiometric images and a qualitative/quantitative evaluation of rain retrieval. Over several case studies, the results show that the domain adaptation step produces adapted SSMI/S images that retain the majority of the rain structure. Next, the rain detection score and intensity bias are then compared using 847 overpasses. The same analysis is carried out over mainland France by comparing the results with rainfall products supplied by Météo-France. In both comparisons, the adapted images allow the inverse operator to provide a better score in rain detection and intensity.

Index Terms—Precipitation estimation, Unsupervised domain adaptation, Passive Microwave Radiometer, Satellite.

I. INTRODUCTION

A homogeneous and fine-scaled global precipitation map is essential to a wide range of scientific research and societal applications [1]. For instance, global precipitation estimates lead to a better understanding of the water budget on a regional and global scale, to monitoring the changes in the water cycle due to climate change, and to following floods, storms, and dry

spells. However, constructing such a map is very challenging due to the intrinsically intermittent nature of precipitation and the scarcity of precipitation observations. Rain intensity varies drastically over time and space. Ground-based instruments alone can only cover small regions and are insufficient for complex terrain, over the oceans, or in developing countries. Therefore, the estimation of rainfall fields from space using remote sensing satellites is a key issue both for the global Quantitative Precipitation Estimates (QPE) and for its evolution in the context of climate change. The Global Precipitation Measurement (GPM) Mission is an international collaboration that aims to provide global precipitation products from a heterogeneous constellation of passive microwave radiometers. The consistency of the constellation is reached through the support of the Core Observatory (GPM-Core), launched in 2014, whose orbit was chosen so that it would cross the path of all the other satellites. The Core Observatory carries a conical-scanning passive microwave radiometer (GPM Microwave Imager, GMI) and a Dual-frequency Precipitation Radar (DPR). In contrast, the other satellites in the constellation carry mainly a passive microwave radiometer for rain retrieval [1].

Recently, a supervised deep learning model of the U-Net type for rain retrieval has achieved excellent performance in the field [2]. The images of brightness temperatures from four GMI channels are used as the sole input, and the precipitation deduced from the collocated radar observations (DPR) are used as the targets for a quantile regression. The resulting algorithm is called Deep-RAIN (DRAIN), and its complete evaluation can be found in [3]. The DRAIN model, which has learned to generalize well on GMI data, performs poorly on another radiometer due to the differences in instrument characteristics (channel frequency, angle of view, pixel resolution). Using co-located data between GPM-Core and the other satellites of GPM is, in theory, possible to build a supervised learning approach to deploy a version of DRAIN for each constellation member. However, these overpasses are not evenly distributed in latitude, with a much higher probability of occurrence near 60° North or South. An alternate solution is to set up a transfer learning step to convert brightness temperature images of any satellite of the constellation into a GMI-like image on the closest channel basis.

Other rain retrieval algorithms for the GPM constellation face the same challenge due to the differences in characteristics between the various passive microwave radiometer instru-

Manuscript submitted 03/08/2023

Vibolroth Sambath (corresponding author) is with the Laboratoire Atmosphères, Milieux et Observations Spatiales (LATMOS), Université de Versailles-Saint Quentin-en-Yvelines Paris-Saclay, 78280, Guyancourt, France. (email: vibolroth.sambath@latmos.ipsl.fr)

Natanaël Dubois-Quilici is with LATMOS. (natanael.dubois-quilici@latmos.ipsl.fr)

Nicolas Viltard is with LATMOS. (nicolas.viltard@latmos.ipsl.fr)

Audrey Martini is with LATMOS. (audrey.martini@latmos.ipsl.fr)

Cécile Mallet is with LATMOS. (cecile.mallet@latmos.ipsl.fr)

ments. The most prominent among these algorithms is the operational NASA passive microwave precipitation retrieval algorithm, GPROF [4], which is a Bayesian-based retrieval algorithm. In a recent effort to use neural networks for rain retrieval, a new version, GPROF-NN, has been proposed. The authors rely on simulated brightness temperatures for the other sensors in the constellation [5]. The neural networks for each instrument use the same dataset as the original version of GPROF to train. In GPROF-NN, there are two approaches to estimating rain intensity: a one-dimensional model (in spectral dimension) where a single pixel of brightness temperature is considered and a three-dimensional model (the spectral dimension plus the two spatial dimensions in the horizontal plane) where two-dimensional brightness temperatures (with multiple channels of brightness temperature images) are considered. In both cases, GPROF-NN uses ancillary data coming from global numerical models' (re-)analyses, such as total column water vapor, surface type, and surface temperature. For the one-dimensional model, a new training database of brightness temperatures is simulated from the observations by taking into account the different viewing geometries and resolutions. On the other hand, for the three-dimensional model, a convolutional neural network model is trained to simulate a GMI swath from another satellite swath. Then, this simulated image is mapped into the GMI viewing geometries. The authors noted that the simulation does not guarantee that the distribution of brightness temperatures for the database matches the observations. Therefore, they are corrected with surface types and total column water vapor. It also should be noted that both DRAIN and the 3D version of GPROF-NN are convolutional models.

Transfer learning is an emerging field in machine learning that deals with the scarcity of data. It relies on the transfer of knowledge of a model obtained from a larger dataset to a slightly different but smaller dataset. Transfer learning can take many different approaches, such as fine-tuning the trained model, combining a domain adaptation step and the predictive model within one algorithm, or adapting only the data [6]. Fine-tuning is a supervised approach where a small labeled dataset is needed in order to retrain some or all hyper-parameters of the model. The dataset size depends on the complexity of the task and the number of hyper-parameters to retrain. Models guided by the targeted task, for instance, [7], try to adapt the new data to the known domain under the constraint of the target task's error. However, while this approach sounds advantageous, it still requires a set of labeled data. Domain adaptation aims to solve the domain shift problem encountered and is extensively studied in many applications, such as computer vision, speech recognition, and medical imaging. In [8], the authors presented a review of Deep Learning domain adaptation from theoretical and practical points of view and mentioned some applications in the geoscience field. These applications concerning discriminative tasks (classification, segmentation) deal with satellite images in the visible and IR domain. [9] provides an overview of unsupervised domain adaptation on remote sensing data. However, to our knowledge, no application on microwave radiometric images, whose multi-frequency and multi-scale

structures are very specific, and no evaluation of the ability of unsupervised domain adaptation to fill the gap of distribution for regression task application has been made. The challenge for domain adaptation in regression tasks is that pixel values in the adapted images need to be as precise as possible in order to obtain a good performance of the downstream regression task.

Image-to-image translation refers to a class of problems that aims to learn the correspondence between a source domain image and a target domain image using a training set of image pairs [10]. In the application considered in this paper, co-located images between two satellites are scarce and do not provide a representative dataset; paired training images are thus not available. Our goal is to learn a mapping from source to target domain without paired images. Consequently, we turned to unsupervised domain adaptation, which does not assume the availability of labeled target data [11].

The choice of architecture here is based on the previous feasibility study [12], where CycleGAN was used to transform between the 89 GHz horizontal and 89 GHz vertical polarisation of the GMI. CycleGAN [13] is an architecture developed for computer vision problems and trained on conventional RGB images. The previous feasibility study analyzed the applications and limitations of CycleGAN on brightness temperature images and rain retrieval in a simple case.

This article explores the domain adaptation between two different satellites and an in-depth validation of rain retrieval. The second satellite, the DMSP-F18 Defense Meteorological Satellite Program, carries a microwave imager SSMI/S that measures brightness temperatures at frequencies similar to the GMI. An in-depth analysis of a domain adaptation model between these two radiometers, which enables the brightness temperature measurements of all GPM constellation satellites to be converted into GMI-like observations, is a crucial step toward the deployment of a single precipitation restitution algorithm for the whole constellation of satellites. The main objective of this study is to extend the use of DRAIN to all satellites in the GPM constellation. As a result, the choice of channels for DRAIN and for the domain adaptation here depends on the common or similar channels aboard all satellites in the constellations, i.e., the 19, 37, and 89 GHz. In the DRAIN paper [3], it was shown that the 36.6 and 89 GHz channels in vertical and horizontal polarizations are sufficient for rain retrieval. Therefore, in this paper, we choose to adapt these two similar channels of the SSMI/S.

The organization of this paper is as follows. Section 2 describes the characteristics, similarities, and differences between the two instruments, GMI and SSMI/S. It also includes the details of the training and validation database. Section 3 describes the unsupervised domain adaptation model, CycleGAN, and its implementation details. The results of the domain adaptation are shown in sections 4 and 5. First, section 4 offers a visual evaluation of the adapted brightness temperatures. Next, section 5 presents an in-depth evaluation of the rain retrieval performance on the adapted data. Section 6 discusses the advantages and limitations of using CycleGAN as a domain adaptation algorithm for brightness temperatures. Finally, section 7 provides the conclusion and perspective.

TABLE I
TECHNICAL INFORMATION ON THE GMI CHANNELS OF INTEREST [14]

Central Frequency	Bandwidth	Polarisation	NEDT	IFOV	Pixel	Incidence angle
36.64 GHz	1000 MHz	V	0.65 K	8.6 × 14 Km	6.0 × 13.4 Km	52.8°
36.64 GHz	1000 MHz	H	0.65 K	8.6 × 14 Km	6.0 × 13.4 Km	52.8°
89.0 GHz	6000 MHz	V	0.57 K	4.4 × 7.2 Km	3.0 × 13.4 Km	52.8°
89.0 GHz	6000 MHz	H	0.57 K	4.4 × 7.2 Km	3.0 × 13.4 Km	52.8°
Conical: 53° zenith angle; useful swath: 850 km - Scan rate: 32 scan/min = 13.4 km/scan						
Near-global coverage in 2 days; high latitudes (> 70°) not covered.						

TABLE II
TECHNICAL INFORMATION ON THE SSMI/S CHANNELS OF INTEREST [15]

DMSP-F18 Defense Meteorological Satellite Program - SSMI/S						
Central Frequency	Bandwidth	Polarisation	NEDT	IFOV	Pixel	Incidence angle
37.0 GHz	1580 MHz	H	0.24 K	27.5 × 44.2 Km	25.0 × 12.5 Km	53.1°
37.0 GHz	1580 MHz	V	0.24 K	27.5 × 44.2 Km	25.0 × 12.5 Km	53.1°
91.655 GHz	2829 MHz	H	0.19 K	13.1 × 14.4 Km	12.5 × 12.5 Km	53.1°
91.655 GHz	2829 MHz	V	0.19 K	13.1 × 14.4 Km	12.5 × 12.5 Km	53.1°
Conical: 53.1° zenith angle, swath 1700 km – Scan rate: 31.9 scan/min = 12.5 km/scan						
Sun-synchronous orbit, Altitude 850 km, Global coverage once/day.						

II. DATA

A. Brightness temperature images

The application of CycleGAN on brightness temperature scenes is not straightforward due to the differences in nature from the conventional images used by the computer vision community in previous applications. In passive microwave remote sensing, a radiometer records the radiant energy arising from the earth's surface and atmosphere. The signal is detected by an electro-optical scanner using a rotating mirror and a set of detectors. The power of the received signal is expressed as a temperature in Kelvin but is called "brightness temperature" (TB) to be distinguished from the physical temperature. The TB in Earth observation results from three main effects: emission and scattering from the atmosphere after reflection by the earth's surface, emission by the earth's surface, and emission and scattering from the atmosphere. The resulting TB is a frequency-dependent complex function of the vertical profiles of the atmosphere and the surface characteristics.

Around 90 GHz, TB responds mainly to scattering from atmospheric ice in the upper part of the clouds; the TB decreases somewhat proportionally in the presence of atmospheric ice over both ocean and land. Near 37 GHz, TB responds on the one hand to emission from atmospheric liquid water (rain or cloud below freezing level) and, when the ice content aloft increases, to scattering by the latter. Depending on precipitation characteristics, the TB of a precipitating cloud will appear either warmer (emission-dominated) or colder (scattering-dominated) than the environment.

In addition, passive microwave radiometers observe the atmosphere with an Earth incidence angle near 50°; the observed atmospheric volume is thus a tilted column. The signal measured by the radiometer appears in a pixel corresponding to the beam intercepting the precipitation column. The higher up in the atmosphere for a given channel, the more horizontal shift should be expected. This angle effect, called parallax shift, is well documented in several publications [16] [17] [18]

[19]. Consequently, a rainy system with a particular 3D spatial structure observed with the same sensor but from different directions may give different measured TB and a frequency-dependent localization of the rainy pixels.

B. Differences between the GMI and SSMI/S

The Instantaneous Field-Of-View (IFOV) is determined by the size of the main reflector, the altitude of the satellite, and the considered channel. However, a fixed receiver integration time and sampling period are used for each radiometric channel. The resulting samples are averaged in the along-scan direction to improve the Noise Equivalent Difference Temperature (NEDT), which characterizes the radiometric sensitivity. The relative motion of the IFOV thus reduces the effective spatial resolution because of the integration time.

The GPM-Core is a non-Sun-synchronous satellite orbiting at an altitude of 407 km and at 65° inclination at the equator [20]. Its main instrument, the GMI, is a conical-scanning microwave radiometer with channels at 10.65, 18.7, 23.8, 36.6, 89.0, 166.0, 183.3+/-3 and 183+/-7 GHz. Among these channels, only the 36.5 GHz and 89 GHz in their horizontal (H) and vertical (V) polarizations are chosen for the domain adaptation. This choice is motivated, first by the performance of these four channels in the training of the deep learning rain retrieval algorithm DRAIN [3] and second, by the number of conically scanning radiometer instruments in the GPM constellation measuring at similar frequencies.

The Defense Meteorological Satellite Program-F18 (DMSP-F18), operated by the National Oceanic and Atmospheric Administration (NOAA), has a set of channel frequencies very similar to GPM-Core. It is, however, a Sun-synchronous satellite at 98.7° inclination at the Equator, carrying on-board an instrument called Special Sensor Microwave Imager/Sounder (SSMIS) with channels at 19.35, 22.235, 37, 91.67, 150, 183+/-1, 183+/-3, 183+/-6.6 GHz plus some additional channels in the 50-60 GHz for temperature sounding (www.eoportal.org).

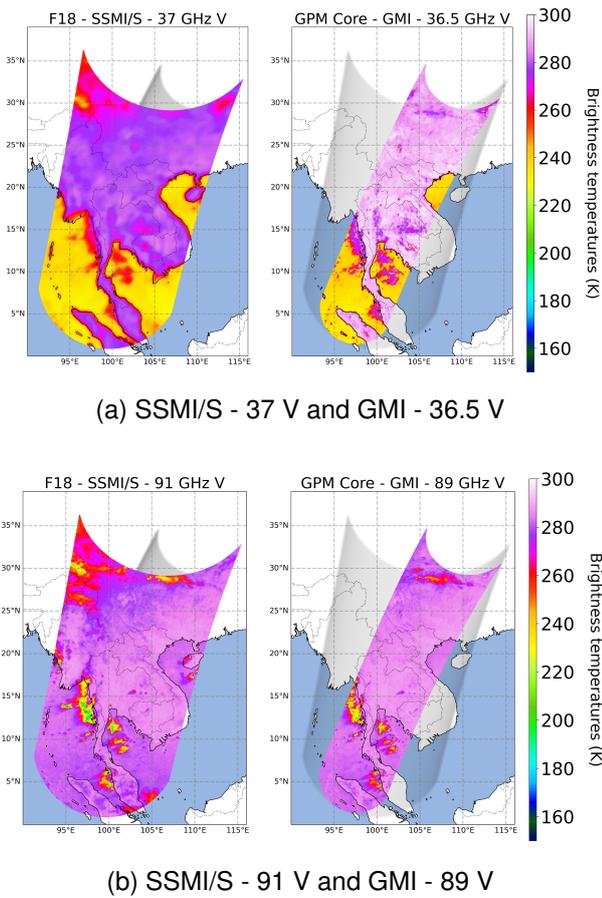


Fig. 1. Overpass example between GMI and SSMI/S that occurred on 29th May, 2017 over South-East Asia. The first row (a) shows the 37 V channel of SSMI/S and the 36.5 V of GMI. The second row (b) shows the 91 V of SSMI/S and the 89 V of GMI. The geometry at the Earth’s surface is taken into account (see text for explanation). The grey area on each image corresponds to the swath of the other satellite.

Tables I and II give a summary of the characteristics of the channels relevant to the present study for the two instruments. It can be seen that SSMI/S and GMI are very close in terms of central frequency channels (36.6 GHz instead of 37 GHz and 89 instead of 91.6 GHz) with similar bandwidths. Though there is a slight degradation of the radiometric sensitivity of GMI compared to SSMI/S (0.6 K instead of 0.2 K), DMSP-F18 being about twice higher in altitude than GPM-Core, SSMI/S pixel sizes are thus about twice as large as GMI ones.

This difference in pixel size, combined with the highly intermittent and varying nature of the underlying precipitating processes, has important consequences for the observed TB values. The beam-filling effect is a well-known effect [21], [22] introduced when the field of view of the radiometer is not filled with a uniform rain field. Due to the complex relationships between the atmospheric parameters, the ground parameters, and the observed TB, the difference in spatial resolution necessarily introduces non-linear relationships between TB observed at different resolutions, making it difficult to estimate the relationships between the two scales.

Another side effect of this higher altitude is the much broader swath of SSMI/S when compared to GMI. This

induces an adverse effect on the learning database. Because the training is performed on geometrically uncorrected images, the deformation due to Earth sphericity is not exactly the same for the two instruments. Figure 1 shows an overpass case between the two satellites over Southeast Asia while coastal convection is developing. The time difference between the two overpasses is less than 5 minutes. It illustrates the different swath sizes and slight differences in TB for the SSMI/S and GMI channels. One should note that these types of highly favorable overpasses in the Tropics are relatively rare due to the respective orbit inclination of the two satellites. The TB in the 36-37 GHz (figure 1.a) are in the 280-300 K range over land and 240 K over the ocean. The rain systems appear as warm spots at about 260-270 K along the coast over the ocean. This increase in TB is due to emissions by the liquid water. In the 89-91 GHz channels (figure 1.b), the contrast between the land and the ocean background is not as clear as for the 36-37 GHz channels. The overall TB is near 280 K, and the rain systems appear as colder spots going down to 210-220 K, which is due to ice scattering in the cloud. In figure 2, when this overpass is flattened and plotted in a regular grid instead of on a map, the deformation effect due to the instrumental differences as described above is particularly visible when looking at the land masses (yellow-green at 36-37 GHz, darker purple at 89-91 GHz).

Another critical difference in the context of domain adaptation is the difference in orbitography since SSMI/S aboard the DMSP-F18 has global coverage, whereas the GMI aboard the GPM-Core does not overfly high latitudes. One consequence is that the areas overflown are different; even considering long observation periods, the statistical distributions of the observed meteorological situations are different. Thus, the observed TB distributions will reflect the difference in the precipitation distribution. A second consequence is that although the two satellites fly over the same area, their trajectories are different, and the same scene is not observed from the same azimuthal view angle. Thus, the three-dimensional variability of the precipitation systems will be reflected in different ways in the observed two-dimensional TB fields.

C. Training and validation datasets

Due to the relatively rare occurrence of rain in an image, the set of GMI images goes through a pre-selection step to only keep those images with non-zero rain rates from the DPR [2]. The selection criteria are either images with at least 100 pixels containing more than 10 mm/h rain or images with at least 10 pixels containing more than 100 mm/h rain rates. The pre-selection ensures that the transformations could work on rain pixels even though they are quite rare in terms of proportions. On the other hand, there is no rain pre-selection of rain images for F18 as it does not carry a precipitation radar instrument. In addition, as the F18 overfly higher latitudes than the GPM core observatory, F18 measurements from polar regions (latitude greater than 70° N or lower than 70° S) are excluded from the dataset to maintain the similarity of the types of meteorological situations observed. After the pre-selection, the training data set consists of 12,000 images for each domain, making 24,000

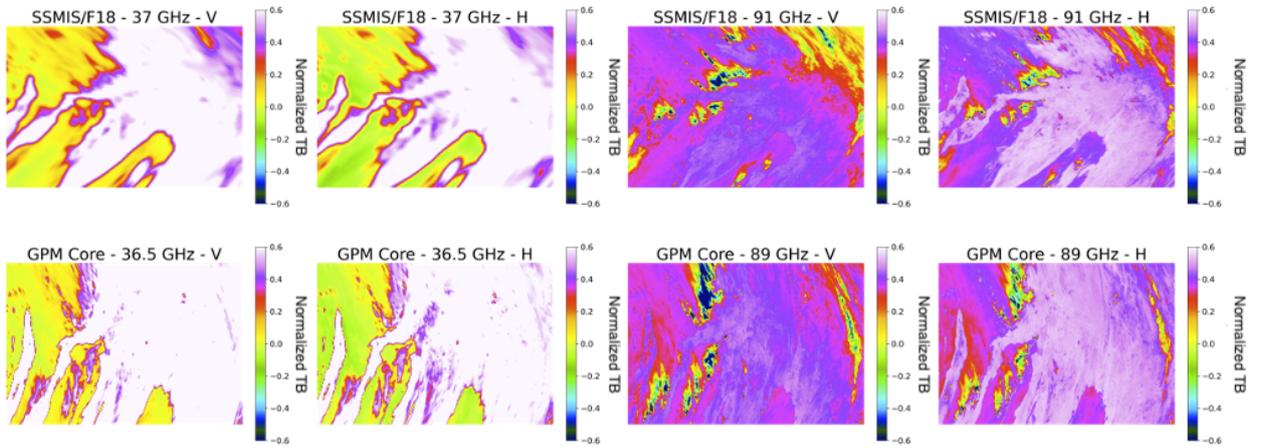


Fig. 2. The same overpass case as in figure 1, but for the four channels of interest for each sensor respectively: top row SSMI/S, bottom row GMI. The brightness temperatures are normalized, and the latitude/longitude geometry is disregarded.

TABLE III
SUMMARY OF TRAINING AND VALIDATION DATASET

	GPM-GMI		F18-SSMI/S	
	Period of observation	Number of images	Period of observation	Number of images
Training dataset	2017	12 000	2017	12 000
Validation dataset	2017	4 000	2017	4 000

images in total of different observations measured during 2017 for both satellites. The validation set comprises 4,000 images for each domain taken from the same period. Table III summarizes the training and validation datasets. None of the images are a corresponding pair. These images contain 221 x 256 pixels and 180 x 256, respectively, for the GMI and SSMI/S images. Next, the training and validation sets undergo the pre-processing step: data normalization, random crop (to 128 x 128 pixels), and random rotation. The random crop (cropping randomly within the image) and random rotation (choosing an angle at random to rotate the image) are added as a data augmentation method to increase the difficulty of the task for the CycleGAN model.

III. METHOD

The deep learning rain retrieval algorithm DRAIN is trained on the GMI TB as input and DPR rain rates as targets [3]. In the DRAIN paper, it was shown that only four channels of TB, 36.6 GHz and 89 GHz in vertical and horizontal polarization, were sufficient (figure 3a). The architecture in DRAIN is based on the U-Net architecture, which is a fully convolutional neural network first introduced for medical imagery semantic segmentation [23]. The cross-entropy loss for image segmentation is changed to quantile regression loss to estimate the rain intensity. The complete evaluation of DRAIN can be found in [3].

The DRAIN model was trained in a supervised approach, which is optimized on the TB from GMI. Consequently, it performs best on the unseen TB drawn from the same distribution as its training set. However, due to the differences in SSMI/S and GMI described in section II, the TB from the two satellites

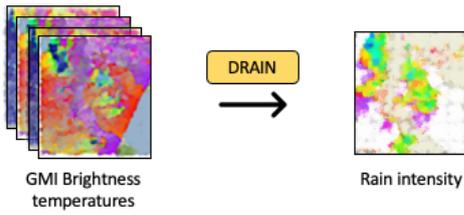
is similar but not enough to apply DRAIN directly. We chose an unsupervised domain adaptation approach to solve this problem for two main reasons. Firstly, the dataset of SSMI/S is unlabeled, which is to say, without the reference rain rates. As a result, neither fine-tuning nor retraining DRAIN on SSMI/S is possible. Secondly, the number of colocated overpasses between the two satellites is limited, which eliminates the supervised domain adaptation approach. We thus proposed to test a Cycle-consistent Generative Adversarial Network (CycleGAN) approach [13] for the unsupervised domain adaptation between GMI and SSMI/S (figure 3b).

Once the CycleGAN is trained, only the generator that transforms SSMI/S to GMI images is used. The adapted images are then used as inputs to DRAIN, which was previously trained on GMI data (figure 3c). The resulting precipitation estimates will then be compared to the colocated DRAIN rain intensity, the DPR rain intensity, as well as the Mosaic product from Météo-France [24].

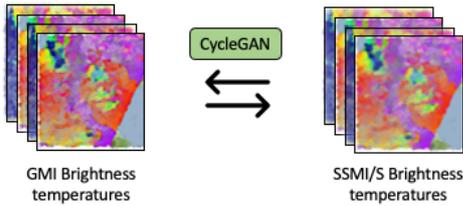
A. CycleGAN architecture

CycleGAN is an unsupervised domain adaptation method based on the Generative Adversarial Networks (GANs) [13]. It contains two GANs working together, one for each domain. Given two domains X and Y , the first generator G transforms images from its source domain X to its target domain Y . The second generator F works inversely by transforming data from its source domain Y to its target domain X . To complete the GAN structure, two discriminators, D_X and D_Y , try to classify real and generated images of their respective domains.

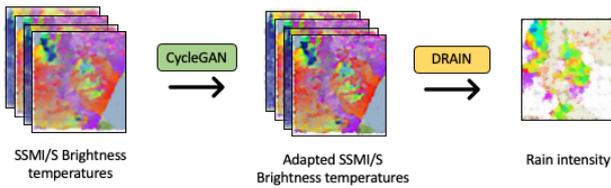
To train the two GAN generators in an unsupervised approach, the weights of the models are updated using the sum



(a) The rain retrieval algorithm DRAIN is trained with GMI brightness temperature as inputs and DPR rain rates as targets.



(b) The CycleGAN has two generators that can transform back and forth between the GMI and SSMI/S brightness temperatures.



(c) The SSMI/S brightness temperatures need to be adapted to the domain of the GMI before applying the DRAIN model for rain retrieval.

Fig. 3. The summary of the proposed approach. (a) shows the DRAIN rain retrieval on GMI. (b) shows the training of CycleGAN. (c) shows the domain adaptation and precipitation estimation on SSMI/S.

of identity loss, cycle consistency loss, and adversarial loss. The identity loss is the difference between an image and a transformation to its domain, for example, $|x - F(x)|$ for x in X . Based on the seminal paper [13], the identity loss helps to regularize the generator to be near an identity mapping; it preserves the color of the input paintings in computer vision applications. In the same way, the feasibility study [12] shows that the identity loss allows the model to conserve the correct intensity of the TB. On the other hand, cycle consistency is the difference between the original image and its complete transformation; that is, $|x - F(G(x))|$ for x in X . Considering that a complete transformation of an image should be close to the original image, the training of the generators does not require paired images. The adversarial loss is based on the performance of the discriminators and operates on the same principle as in a conventional GAN.

B. Implementation details

Since the U-Nets architecture has proven suitable for the rain/TB problems [2], two such networks are used as the two generators (F and G) in the CycleGAN. The skip-connections in U-Net are very efficient in preserving the fine-scaled structure within the brightness temperature images, as previously stated in [2]. As described in [13], the discriminators are based on the PatchGAN discriminator [25], whose architecture is a series of convolution layers.

The results of our experiments show that the convergence of the CycleGAN is not trivial. The adapting domain performed with unpaired data does not rely on any task-specific, and the problem is thus under-constrained. Given the specificity of our images, unlike computer vision images, visually determining whether the result is realistic is impossible. In our case, the resulting TB fields must retain the essential properties of the precipitation characteristics that cannot be assessed by simple visual analysis. During training, the optimization of the hyperparameters relies on the learning curves to obtain the behavior shown in figure 4. An empirical approach based on the following principles is used: keeping similar loss for both domains, keeping discriminator loss slightly lower than generator loss, and stabilizing the training curve.

In order to obtain similar losses for both domains, different learning rates are necessary. The optimization of the generator depends in parts on its ability to deceive the discriminator. Therefore, the quality of the generated images depends on the performance of the discriminator. Each time the generator deceives the discriminator, the discriminator needs to find another criterion to differentiate between real and generated images. The rivalry principle is then essential. Based on the results of our experiments, the discriminator should always be the best of the two models. Still, the generator must sometimes be able to catch up to stimulate the discriminator. In numerous runs of the model, the discriminator converges quite easily and quickly. To solve this problem, the generator's and the discriminator's learning rates are set differently due to the two-time-scale update rule (TTUR) [26]. It has been shown that GANs converge better and with fewer epochs using this method. The learning rate of the generator is twice that of the discriminator due to the complexity of its task and to increase the rivalry. In addition, the task for the generator and discriminator pair for one domain (generating a distribution corresponding to a higher spatial resolution image from one of a lower resolution image, i.e., generating GMI domain from SSMI domain) is more complicated than the opposite (from GMI to SSMI domain). For this reason, the learning rate for the SSMI to GMI pair is twice as high as for the GMI to SSMI pair. Therefore, the learning rate for the generator and discriminator pair for GMI to SSMI/S is $2e-4$ and $1e-4$, respectively, while the pair for SSMI to GMI is $4e-4$ and $2e-4$, respectively. This results in a similar loss for both domains, as shown in the training loss in figure 4.

The choice of cost functions, normalization, and weighting coefficients is fundamental to achieving a balanced reduction in the three terms of loss. Many different configurations were tested before an acceptable one was found.

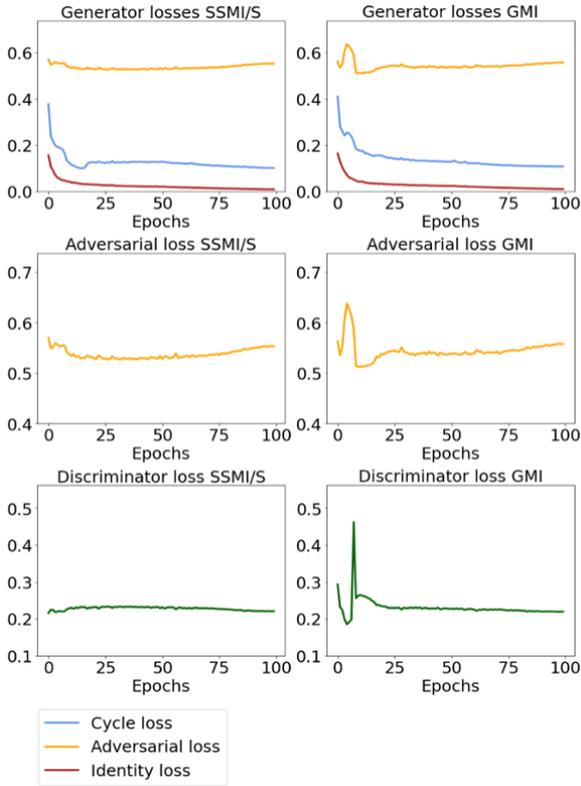


Fig. 4. Training loss for the generators and discriminators for the two domains (left column for the GMI and right column for the SSMI/S). The first row shows the three losses of the generator, while the last row shows the loss of the discriminator. The middle row shows the adversarial loss of the generators alone for better comparison to the discriminator loss.

To train CycleGAN, the L1 loss is used for the cycle and identity losses, while the Mean Squared Error (MSE) loss is used for the discriminator and the adversarial loss. For the discriminator, due to the value of the labels (0 for generated images and 1 for real images), these errors are, therefore, between 0 and 1, as shown in figure 4. With MSE, squaring the error will then lead to a decrease in the value used for backpropagation. To increase the impact of the adversarial loss in the generator, the MSE loss has been replaced by the L1 loss. Results from multiple training in this study using this approach show better-generated satellite images. On the other hand, replacing the MSE loss with the L1 loss for the discriminator did not result in any improvement, as the discriminator encountered more difficulty in converging. As there are three losses for training the generators, they are balanced by three coefficients. In both the feasibility study [12] and the proposed model here, the same coefficients as proposed by [13] were used (10-5-1 as the coefficients for the cycle-identity-adversarial losses, respectively).

Therefore, our full loss function for the generators is $L_{generator} = 10L_{cycle} + 5L_{identity} + L_{adversarial}$, where:

$$L_{cycle} = E_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] + E_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1]$$

$$L_{identity} = E_{x \sim p_{data}(x)} [\|F(x) - x\|_1] + E_{y \sim p_{data}(y)} [\|G(y) - y\|_1]$$

$$L_{adversarial} = E_{x \sim p_{data}(x)} [\|D_Y(G(x)) - 1\|_1] + E_{y \sim p_{data}(y)} [\|D_X(F(y)) - 1\|_1]$$

In order to stabilize the training of CycleGAN, spectral normalization introduced by [27] was used to replace the batch normalizations. In [27], spectral normalization is only proposed for the discriminator. However, according to [28], spectral normalization can also be applied to the generator. In the present study, spectral normalization was thus used to replace the normalization methods for the two U-Nets (generators) and the discriminators. The effect of the spectral normalization results in a relatively smoother loss curve, as shown in figure 4. This normalization results in the best score in the validation by rain intensity shown in section V. Spectral normalization is a method for controlling the variations of the model function under a constant by normalizing the weights to satisfy the 1-Lipschitz constraint. This greatly reduces the large variations and thus makes the training more stable, limits the impact of local minima, and, as a result, allows the model to converge faster. It has also been shown that this method avoids the vanishing gradient problem.

After defining the implementations using the empirical approach described above that makes the training loss smoother and retains a balanced loss between the two domains, CycleGAN is still very sensitive to initialization. Several models have been trained with different initial conditions. As shown in figure 4, the discriminator and adversarial loss for the two domains exhibit extremely slow convergence. In each training, the current model was saved along with the best generator model. The best generator here is defined as having the lowest cycle and identity losses on the validation set. As the problem is under-constrained, the downstream rain retrieval task results differ for each training. Therefore, the best generator in each training (corresponding to different initial conditions) was tested on the rain retrieval task to evaluate the whole performance of the process as defined in figure 1.c. The model presented here was trained over 100 epochs. However, the best generator on the validation set appeared at around epochs 20. It is also the generator that gives the best score in the rain retrieval task. More discussions on that aspect are presented in section VI.

IV. EVALUATION OF ADAPTED BRIGHTNESS TEMPERATURE IMAGES

The main difficulty in the present application of CycleGAN on TB fields is the assessment of the quality of the adaptation since the images that are used are somewhat abstract. Unlike in the original paper [13], one can not easily recognize an object inside the image and immediately and subjectively measure how well it was transformed. Concerning the spatial properties, CycleGAN is a model that cannot make any drastic changes to the structure of the image [13]. It contains the Cycle Loss and the Identity Loss, which are pixel-to-pixel differences between the images. Cycle loss is the pixel-to-pixel differences between an image and its complete cycle adaptation; for example, an SSMI/S adapted to GMI then adapted back to SSMI/S. It

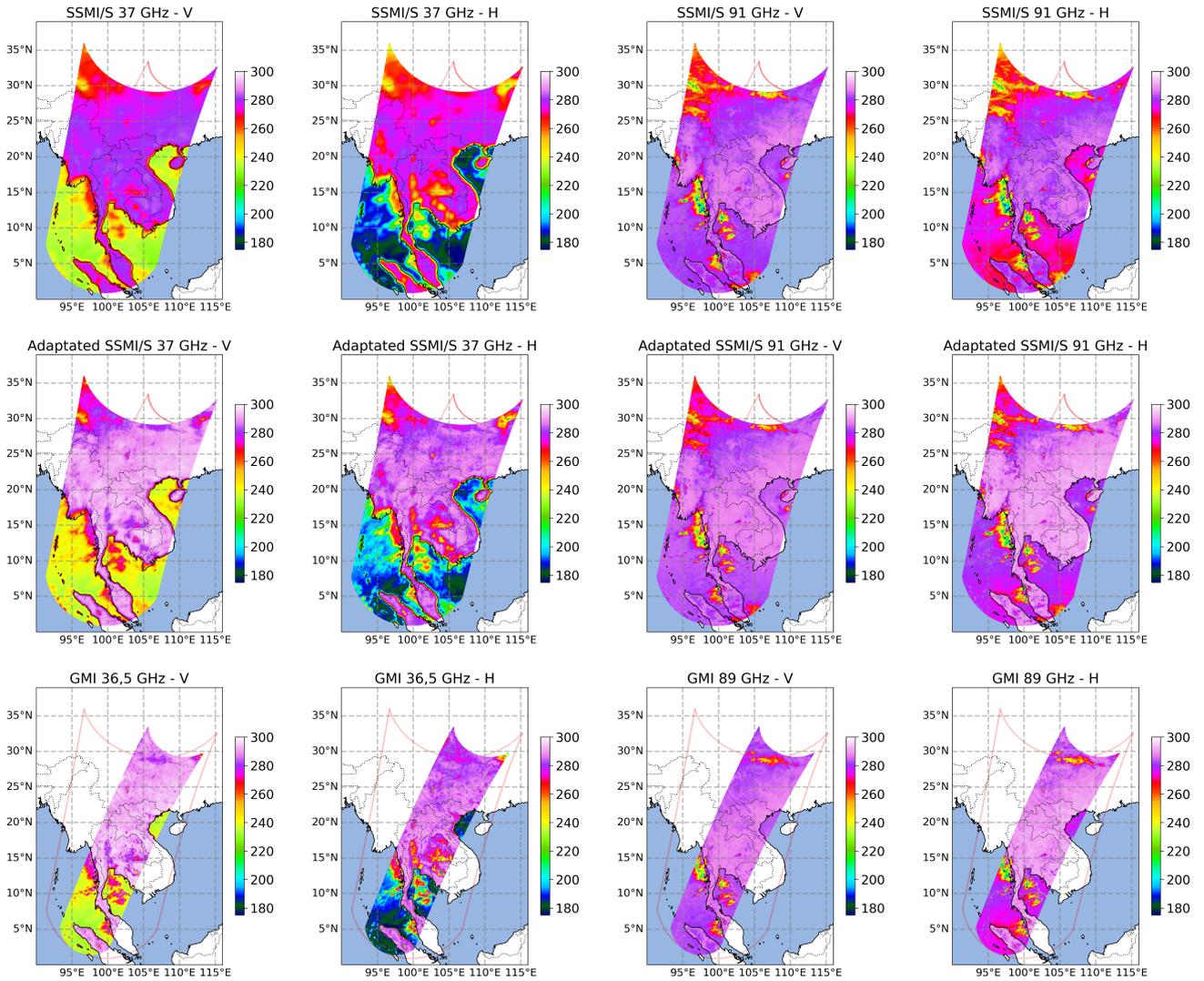


Fig. 5. Brightness temperatures in Kelvin in the same observation as figure 2 but middle row shows the adapted SSMI/S channels (SSMI/S transformed into GMI).

quantifies the dissimilarity between an image and its complete cycle adaptation. That means that the dissimilarity between SSMI/S adapted to GMI, then adapted back to SSMI/S, and the dissimilarity between GMI adapted to SSMI/S, then adapted back to GMI, are minimized during training. Identity loss, on the other hand, is the difference between an image and its identity transformation, i.e., SSMI/S adapted to SSMI/S. The first step in evaluating the generated images is to confirm that the structures are preserved throughout the cycle of transformations. To monitor that there is no transfiguration of the structure, we monitored the loss and displayed the pixel difference between the images and their transformations to ensure that there is no structural high error. The quantification of these losses during the validation process further confirms that they are also minimized when considering unseen images.

Secondly, the visual evaluation is done in complement to the monitoring of the training and validation loss. A few

overpasses of the two radiometers GMI and SSMI/S over the same area are used to qualitatively analyze the TB fields. As mentioned earlier, and as shown in Figure 1, the two radiometers do not have the same swath width or the same angle of view. It is, therefore, not expected to obtain precisely the same TB field. The model is supposed to correct for systematic differences such as resolution and frequency but not for differences due to different viewing angles of the same system since this is not an image translation problem. During its flight, each satellite encounters different precipitation systems, which it sees from different angles depending on the relative positions of the latters and the satellites.

Figure 5 shows the adaptation results of SSMI/S to GMI on the overpass case presented before. The first row is the original images of the four SSMI/S channels. The second row shows the adapted images, and the third row shows the GMI channels. The most prominent feature is the small increase

TABLE IV
SUMMARY OF TEST DATASET

	Number of pixels
Overpass between GMI and SSMI/S (section V-A)	$\sim 88 \times 10^5$
Overpass between DPR and SSMI/S (section V-B)	$\sim 36 \times 10^5$
Colocation between Météo-France and SSMI/S (section V-C)	$\sim 25 \times 10^5$

in the perceived resolution of the adapted images, particularly the 37-GHz channel, as some details become visible in the adapted images. This validates the SSMI/S-to-GMI generator’s capacity to increase the lower resolution of the SSMI/S to match the higher resolution of the GMI. The adapted images also conserve all precipitation structures, even if they appear somewhat deformed because of the difference in view angle.

V. EVALUATION ON QUANTITATIVE PRECIPITATION ESTIMATION TASK IN THE TARGET DOMAIN

The previous section shows an evaluation of the domain adaptation between F18 and GPM observations. Since the goal of the adaptation is ultimately rain retrieval, an in-depth assessment of the latter is required to evaluate the transfer learning method’s potential. The rain retrieval step takes the adapted SSMI/S data from the CycleGAN generator directly as inputs to the DRAIN model. The complete details of the training and performance of DRAIN on GPM can be found in [3].

Quantitative comparisons are performed on three different data sets (Table 4). The 847 passes of the two satellites between latitudes 70° N and 70° S during the whole year 2019 are considered. A first test set is constituted from the intersection of the swaths of the two radiometers. A subset containing only the observation area covered by the narrower swath of the DPR radar constitutes the second test set. The third test is a co-location in time and space over the whole of 2019 with the Météo-France five-minute rain product at a resolution of 1 km, which is a good reference for mid-latitude QPEs.

For these three test sets, a pixel-by-pixel comparison is performed. The ability to detect rain, which is to distinguish rain and no-rain pixels, is analyzed through the contingency table. As the rainy/non-rainy classes are not balanced, we have a majority of non-rain pixels, which completely distorts the computation of accuracy. For example, due to the rare cases of rain pixels, if a model predicts non-rain for all pixels, it still has a relatively high accuracy score. Therefore, the F1-score that appears more appropriate than the latter is calculated to facilitate comparison. F1-score resolves the unbalanced data by taking into consideration both recall and precision scores. The recall score is the ratio of the True Positives to all the positive samples in the dataset. In contrast, the precision score is the ratio of True Positives to all the predicted positives. Combining these two metrics can deal with the rarity of positive data points. The F1-score formula is given by,

$$F1 = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Probability of Detection (POD) and False Alarm Rate (FAR) are other metrics popular in classification. POD is another

term for recall, which indicates the model’s ability to identify positive cases out of all the actual positive cases. FAR, on the contrary, focuses on the False Positive by measuring the proportion of negative cases that were identified incorrectly as positive. It is the ratio of the False Positive to all the actual negative cases.

The bias and standard deviation calculated on all pixels are dominated by zero values, which correspond to the majority of true non-rain pixels. To emphasize the improvement in rain intensities, the true positives (pixels considered as rainy simultaneously by DRAIN and by the used rain reference) are then retained. The bias and the standard deviation of the rainfall rate differences are finally calculated on these pixels.

A. Comparison between overpasses

As described in section II, the resolution and swath of the two satellites are different. In order to compare the DRAIN estimation on GMI, non-adapted, and adapted SSMI/S, it is essential to colocate the pixels of GMI and SSMI/S. A KD-Tree algorithm [29] (implemented with SciPy library in Python) is used to calculate the nearest SSMI/S pixels of a given GMI pixel. The rain rate computed for that nearest pixel from the SSMI/S TB (adapted or not) is then affected to the GMI pixel in question to allow a direct comparison.

Figure 6 shows the result for the Southeast Asian case presented in Figure 1. Pixels with no rain (RR=0) are not plotted. The left-hand side image is the reference DRAIN applied directly to GMI data. The center image is the fully adapted, and the right-hand-side image is the non-adapted result. In this example, dominated by tropical convection, the non-adapted result does not perform too poorly in terms of the general structure of the rain regions. Locally, however, some rain pixels are retrieved as non-rainy, and the rain intensities are substantially off, with rain rates no greater than 20 mm/hr. On the fully adapted image, both the structure and intensities appear much closer to the reference, showing that the adaptation indeed succeeded in transforming the SSMI/S TB image into a GMI TB image. Two more overpasses are shown in the Annex.

Table V presents the contingency and base statistics of the adapted and original SSMI/S TB against DRAIN-GMI used as a reference for the 847 overpasses mentioned above. One must note that most of these overpasses occur at high latitudes where the GPM-core reaches its maximum latitude (65°). As expected in both cases, nearly 95 % of the pixels are non-rainy. However, adaptation clearly improves the fraction of true positives, which goes from 0.66 % before adaptation to 2.61 % after. This improvement comes mainly from the bad detections, which go from 3.12 % before adaptation down to 1.17 % after. Table V also shows that the adaptation clearly

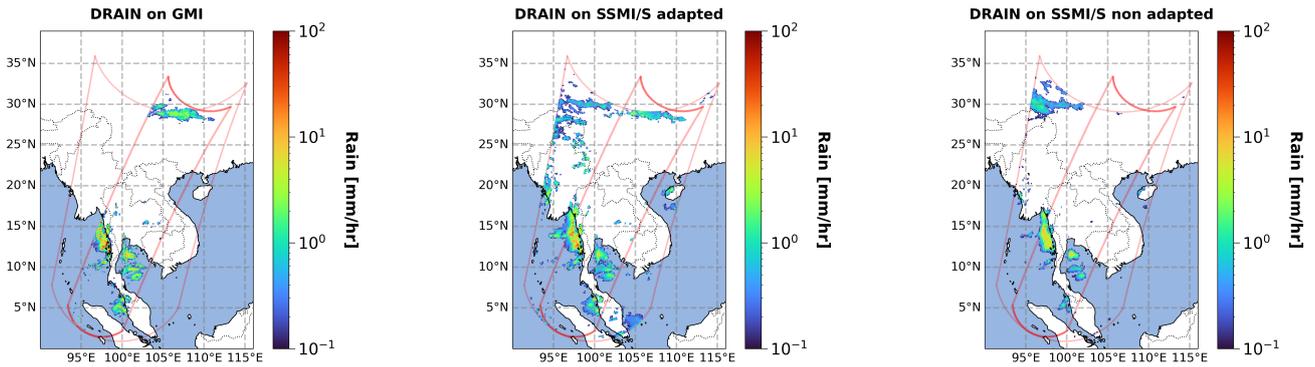


Fig. 6. Same as figure 1 but for the rain intensities retrieved by DRAIN. The left-hand side is DRAIN on GMI data. The Center is DRAIN on adapted SSMI/S data. The right-hand side is DRAIN on unadapted data. All rain rates in mm/hr. Both SSMI/S-inferred rain rates are co-located with the GMI pixels for comparisons.

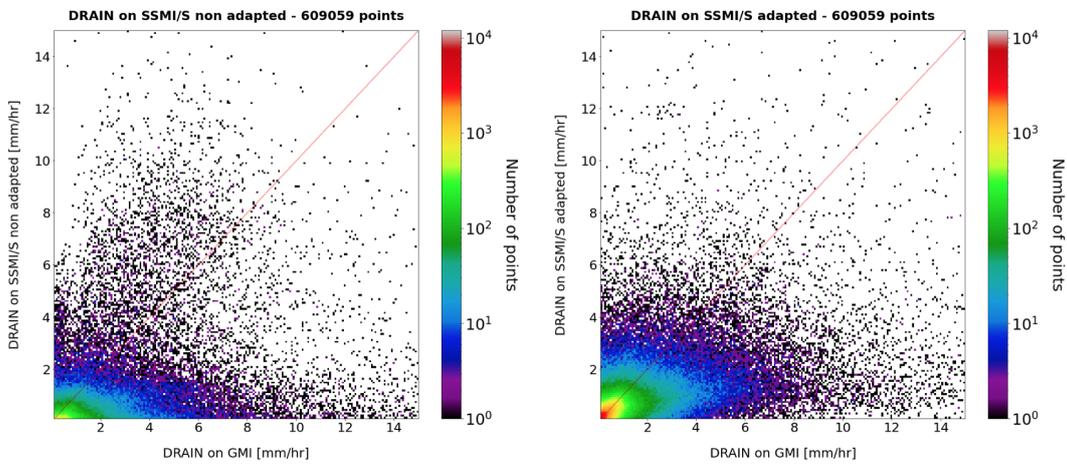


Fig. 7. Scatter plot for the 847 overpasses between GMI and SSMI/S data, after retrieval with DRAIN and co-location to the GMI pixels. The left-hand side plot is for non-adapted SSMI/S data, and the right-hand side is for adapted ones. The number of points above each plot indicates the points where at least one of the three datasets (DRAIN on GMI, on original SSMI/S, and on adapted SSMI/S) has non-zero rain rates.

improves the score on both rain and non-rain predictions. The F1-score increases from 0.20 for the non-adapted images to 0.66 for the adapted data. Both the bias and standard deviation of errors are significantly reduced from 1.64 to 0.75 mm/hr and 3.05 to 1.74 mm/hr, respectively, while the number of rain pixels has increased considerably for the adapted data. The reduction in overall error may be largely due to the number of low rain-intensity pixels that the adapted images are able to capture. The improvement in intensity is comparable between adapted and non-adapted for the rain intensity between 1-3 mm/hr. The bias reduced from 1.24 mm/hr (mean target rain rates at 1.75 mm/hr) to 0.90 mm/hr (mean target rain rates at 1.66 mm/hr).

Figure 7 shows the scatter plots corresponding to the statistics of Table V. The improvement brought by the adaptation is clear, although, above 3 to 4 mm/hr, an underestimation of the intensities can still be seen. However, these overpasses comparisons must be taken with some degree of caution because first, the co-location of the SSMI/S pixels with the GMI pixels might affect the statistics, and second, most of these overpasses happen at high latitudes where rain intensity

is predominantly light, and DRAIN might not be giving its best performances (see [3]). In addition, the difference between the three cases is the ability to capture low rain rates. DRAIN and the adapted case are better at estimating low rain, which leads to lower average rain rates for the collocated pixels.

B. Comparison with DPR

In this section, DRAIN estimation on collocated data between the GMI and the SSMI/S is compared to the DPR data. As in the previous section, data of the 847 overpasses between F18 and GPM-Core are considered. The SSMI/S data are used as an input in DRAIN and co-located afterward with the GMI pixels. The co-location of DPR pixels was carried out in a previous study ([2]) as follows: depending on the location of the pixels in the swath and in an orbit, 1 to 5 DPR pixels are averaged for each GMI pixel within the DPR swath. Due to the smaller swath of DPR, there are less collocated data than in Table V. As expected, the performance in Table VI is slightly worse than in Table V. The precipitation rates estimated by DRAIN in the adapted target domain (adapted SSMI/S) are

TABLE V
RAIN RETRIEVAL RESULTS WITH DRAIN ON GMI AS THE REFERENCE.

DRAIN on GMI	non-adapted SSMI/S		adapted SSMI/S	
	Rain	Non-Rain	Rain	Non-Rain
Rain	0.66%	3.12%	2.61%	1.17%
Non-Rain	1.85%	94.35%	1.45%	94.75%
F1 score	0.20		0.66	
Bias on rain (mm/hr) (TP)	1.64		0.75	
Standard deviation of error (TP)	3.05		1.74	
Number of rain pixels (TP)	58799		232164	
Mean of rain rates in mm/hr (TP)	2.19		1.21	

TABLE VI
RAIN RETRIEVAL RESULTS WITH DPR AS THE REFERENCE.

DPR	non-adapted SSMI/S		adapted SSMI/S		GMI	
	Rain	Non-Rain	Rain	Non-Rain	Rain	Non-Rain
Rain	0.67%	5.27%	2.66%	3.28%	4.25%	1.95%
Non-Rain	1.81%	92.23%	1.16%	92.88%	0.76%	93.02%
F1 score	0.16		0.55		0.75	
Bias on rain (mm/hr) (TP)	2.47		1.13		0.46	
Standard deviation of error (TP)	6.36		3.65		1.76	
Number of rain pixels	24584		96740		286884	
Mean of rain rates in mm/hr (TP)	2.98		1.61		1.17	

closer to those estimated in the source domain (GMI) than to the labels (DPR) learned by DRAIN.

However, performance improvements are brought by the adaptation and are clearly shown in Table VI. The True Positives go from 0.67 % for the non-adapted data to 2.66 % for the adapted ones. This result is to be compared with the 4.25 % of the GMI data. True Negatives are similar for all three configurations, and, once again, it's the False Alarm rate that changes the most between non-adapted at 5.27 % and adapted at 3.28 %. These values remain higher than the optimal 1.95 % for GMI data. A substantial improvement is also seen in the F1 score, the bias, and the standard deviation of the error, but the performances of DRAIN on the adapted data cannot quite match the performances on the original GMI ones. Similar to the previous evaluation, the lower mean target rain rates in Table VI are due to the increase in the capability of capturing low rain rates. However, for rain rates between 1-3 mm/hr, the non-adapted error is 1.33 mm/hr (for mean target rain rates of 1.79 mm/hr), while the adapted model error is 1.00 mm/hr (for mean target rain rates of 1.69 mm/hr). For this interval of rain rates, the error for DRAIN is 0.55 mm/hr with the mean target rain of 1.63 mm/hr.

The same caution should be used when looking at these results as for the previous ones, but they give an idea of the potential of the domain adaptation and the margin of progression left.

C. Comparison with Météo-France Mosaic product

This section presents an evaluation of the rain retrieval results of the adapted images over an independent set of validation data. As in section V-A and V-B, the results will be evaluated over the whole year 2019, which was not part of the training of either DRAIN or CycleGAN. The Météo-France mosaic [24], hereafter mosaic, provides a precipitation product at 5-minute intervals over a 1536x1536 pixel grid

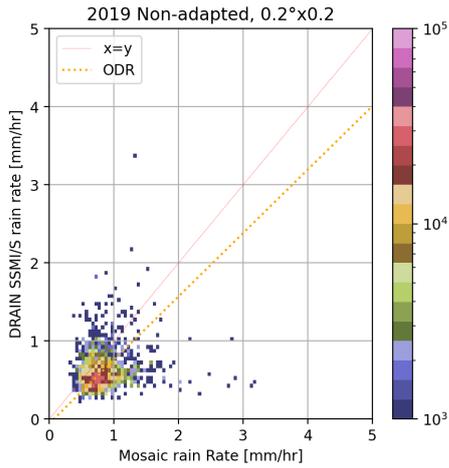
with a resolution of 1 km. Due to the difference in data structure between the mosaic and the satellite observation, the comparison requires a co-localization process again. All satellite and mosaic data are averaged in a $0.2^\circ \times 0.2^\circ$ grid. Each satellite overpasses France at a different time and on different days over the course of the year 2019. In particular, DMSP-F18 is a geosynchronous platform, while GPM-Core is not. We can, therefore, expect their rain rate distributions to be significantly different. However, the errors between the mosaic and the three satellite-based estimates, SSMI/S non-adapted, SSMI/S adapted, and GMI, should give us robust estimates of the adaptation performances.

Table VII shows the performances of the three satellite estimators against the mosaic. As in Table V, the performances are better for the adapted data when compared to the non-adapted and are very similar to the GMI. Interestingly, the True Positive for SSMI/S adapted is slightly better than that of GMI (4.94 % vs 4.85 % resp.). It is also noticeable that the imbalance between False Alarm and Bad Detection is also similar for the three estimators. The Bad Detection starts at 13.34 % for the non-adapted and goes down to 10.80 % for the adapted and 9.81 % for the GMI. This shows that the mosaic detects rain more often than the satellite estimators about 10% of the time, and it comes as no surprise since one would expect ground-based radars to be more efficient at detecting very light rains near the surface.

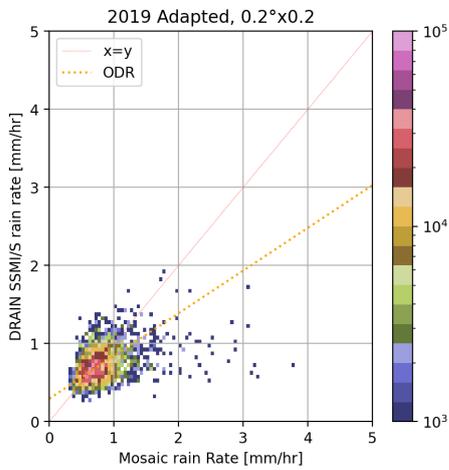
Figure 8.a and .b show the scatter plots of non-adapted and adapted SSMI/S, respectively, vs the mosaic rain rates. The shades of color show the density of points from dark blue, low density, to purple, high density. For the non-adapted case, the densest area of points is clearly off the $x = y$ line with an underestimation of about 0.3 mm/hr. For the adapted case, the points are definitely closer to the $x = y$ line and are less scattered than for the non-adapted case. After the adaptation, there are also fewer outliers; even if above 1.5 mm/hr, there

TABLE VII
RAIN RETRIEVAL RESULTS WITH DRAIN ON THE MOSAIC AS THE REFERENCE.

Mosaic	non-adapted SSMI/S		adapted SSMI/S		GMI	
	Rain	Non-Rain	Rain	Non-Rain	Rain %	Non-Rain %
Rain	2.40%	13.34%	4.94%	10.80%	4.85 %	9.81 %
Non-Rain	2.14%	82.13%	1.49%	82.78%	0.36% %	84.98 %
F1 score on rain	0.24		0.45		0.49	
Bias on rain (mm/hr) (TP)	0.21		0.07		0.02	
Standard deviation on rain (TP)	0.33		0.21		0.33	
Number of rain pixels	2556036		2556036		6645997	



(a) Non-adapted

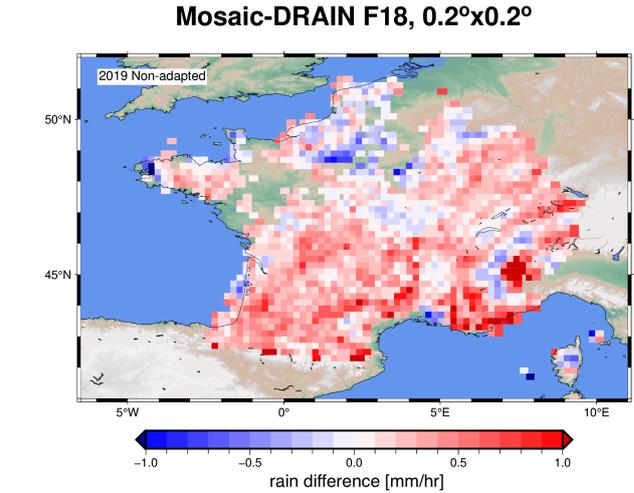


(b) Adapted

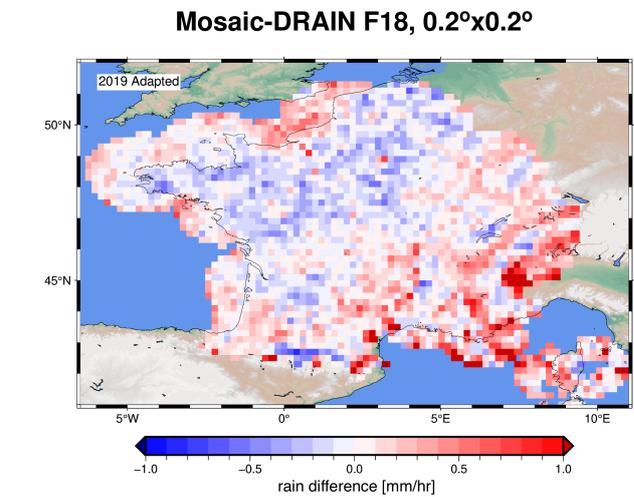
Fig. 8. (a) Scatter plot of retrieved rain from non-adapted SSMI/S vs. Météo-France Mosaic. (b) same as (a) but for adapted SSMI/S.

is still a clear signature of underestimation.

To further analyze the scatter plot, the Orthogonal Distance Regression (ODR), as proposed by [30], is used to represent the points. This approach is chosen over the simple linear regression due to the fact that both the Mosaic rain rate and the retrieved rain rates are affected by uncertainties. The regression line for the non-adapted may appear to be closer to the identity line ($y = x$) in figure 8. However, this is misleading as the explained variance by each of these linear



(a) Non-adapted



(b) Adapted

Fig. 9. Maps of the difference Mosaic-DRAIN for the True Positives. (a) for the non-adapted et (b) for the adapted. Data of all the overpasses are averaged in 0.2x0.2 boxes over the whole year 2019.

regressions is 0.14 before adaptation and 0.32 after. Even though the regression slope on the non-adapted case (0.81) seems closer to the $x=y$ line, the barycentre of the histogram is off-center, and the regression is poor. The position of the regression line being closer to the identity line may be due to the wrong estimates that are coincidentally almost symmetric for both the low and high rain intensity. After adaptation, the

regression slope is further (0.55) from the $x=y$ line, but the barycentre of the scattered points is spot on the $x=y$ line, and the explained variance was multiplied by a factor of two. In addition, lower rain rates are better estimated in the adapted case.

Figure 9 show respectively the difference Mosaic-DRAIN averaged in 0.2×0.2 boxes for all the DMSP overpasses of 2019. In order to enhance the differences, only True Positive has been kept here. In the non-adapted case (figure 9.a), almost no rain is retrieved over the ocean, a lot of rain is missed along the Loire Valley, and south of the latter, a clear underestimation of the rain intensities can be seen. Over the Alps and the Pyrennees mountains, the Météo-France mosaic might be affected by larger errors due to imperfectly filtered ground clutter despite the use of the quality flag provided in the Météo-France product: only rain estimates with a confidence greater or equal 80% were kept here. The adapted case (figure 9.b), on the other hand, shows a much more balanced distribution of the differences, even if oceanic regions seem to be slightly underestimated. The northern part of France (where rain is usually dominated by frontal systems) seems to be overestimated, while the southern part of the country appears to be more underestimated, maybe due to the difference in the nature of the rain regimes as mentioned above.

VI. DISCUSSIONS

The validation process is one of the main challenges in training CycleGAN since it is meant to be used when no paired data are available. In order to get the best performance, it is important to monitor the training loss and save the best generator. Without any paired data for comparison, cycle and identity loss on a validation set determine the best epoch to select the optimal generators. In numerous trainings, the generators that ultimately give the best F1-score on the rain retrieval step are the ones in the earlier epochs (around epoch 14) rather than the later epochs. This may be due to the different properties of the two datasets, in particular, one containing images that were selected on rain-based criteria (GMI), while in contrast, the other dataset contains images that were only filtered for latitudes (SSMI/S). It is very likely that the discriminator working with the first dataset forces its generator to generate rain systematically. Data selection for a more balanced dataset requires an algorithm, either rule-based or with a Machine Learning approach, to identify rain pixels from TB alone in the SSMI/S dataset. An in-depth study of such an algorithm is important for the future application of CycleGAN or other unsupervised domain adaptation methods.

Another challenge in optimizing a CycleGAN model is that it is very sensitive to initialization. Indeed, a common weakness of generative models is that performance can vary greatly from one learning experiment to another for a given set of hyperparameters. Although choosing the best model for the domain adaptation between GMI and SSMI/S involved the model with the lowest validation score, as described above, the performances in terms of low rain retrieval errors also had to be taken into account. Moreover, each modification of CycleGAN's hyper-parameters leads to different performances,

which means that CycleGAN has to be trained several times to select the best model.

CycleGAN is a domain-to-domain adaptation. In its application on rain retrieval for the GPM constellation, the domain of each satellite microwave imager needs to be adapted to that of the GMI as it is the only satellite that allows for the supervised training of the DRAIN algorithm. As a result, a different CycleGAN is required for each satellite in the constellation. This can lead to different rain retrieval errors for each satellite, as CycleGAN works better when the two domains are close. Nevertheless, the results presented validate the machine learning approach to domain adaptation for rain retrieval applications. In future applications, a multi-source domain adaptation approach where one model can be used for all satellites in the constellation might be a more sustainable method. For instance, an approach of domain generalization proposed by [31] using an auxiliary vision bridge domain with mappings from multiple domains. Combined with self-supervision learning techniques, it can train an effective and efficient model for generalization across multi-domain tasks. Instead of cross-domain adaptation, StarGAN is a GAN-based model capable of learning multiple domains with a single generator [32]. StarGAN comprises a discriminator and a generator. Its discriminator is trained to distinguish between real and fake images as well as predict the label of identified real images. The training of the generators contains more steps. The generator first tries to transform an input image to a target domain. Next, it transforms the fake image and its concatenated original label back to the original domain. The generator tries to fool the discriminator with fake images to train the model with adversarial loss. After optimization, the generator will have learned mapping from multiple domains.

VII. CONCLUSION

In this paper, CycleGAN was used to adapt the passive microwave brightness temperatures from the SSMI/S domain to the GMI domain. Ultimately, the goal is to use the adapted data as input in a deep-learning rain retrieval algorithm, DRAIN, trained on GMI data. The adaptation process is an unsupervised approach that does not rely on any paired images. One of the unique characteristics of the present study is the use of a technique originally developed for computer vision and its application to abstract pseudo-images made of geophysical measurements. Unlike RGB images, our task deals with four channels of brightness temperatures describing the condition of the atmosphere over a given scene. Multiple tests regarding hyper-parameters have thus carefully been examined to find the best configuration. This process has shown that using CycleGAN with the intention of a downstream task for the adapted images is very challenging in terms of optimization of the hyper-parameters.

However, the adapted SSMI/S data significantly increase DRAIN performances over the unadapted data. The retrieved rain field is improved both in terms of structure and intensity on the test case of coastal convection in Southeast Asia. An in-depth analysis using the whole year 2019 over three different successive references, DRAIN on GMI, DPR, and

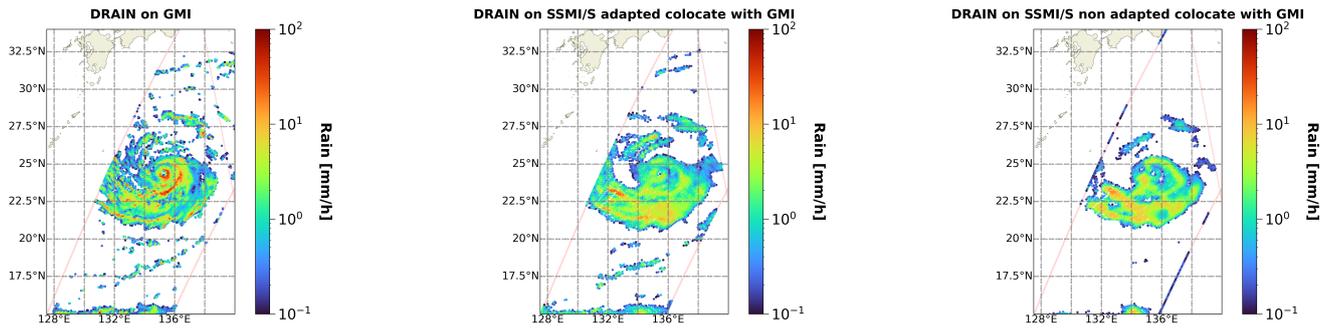


Fig. 10. Rain retrieval of an overpass on 16/09/2022. From left to right: DRAIN on GMI data, DRAIN on adapted SSMI/S, DRAIN on original SSMI/S. The rain rates are in mm/hr

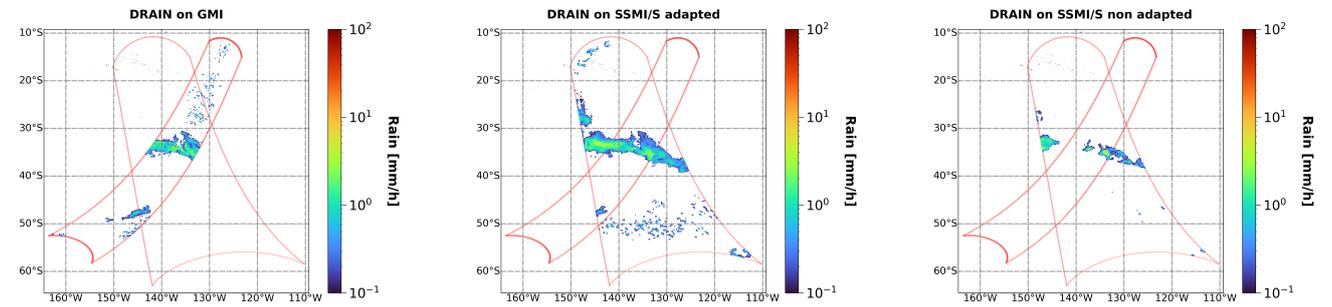


Fig. 11. Rain retrieval of an overpass on 01/02/2019. From left to right: DRAIN on GMI data, DRAIN on adapted SSMI/S, DRAIN on original SSMI/S. The rain rates are in mm/hr

Meteo-France mosaic, also shows an improvement in rain detection and rain intensity estimation. The most noticeable improvement is reached in the comparison with the Météo-France mosaic. Although spatially limited to mainland France, the total number of overpasses is much higher than between the two satellites, providing a more robust estimate for the considered region.

The results in this paper validate the Deep Learning approach to rain retrieval from passive microwave brightness temperatures from the GPM constellation. The success of Deep Learning Unsupervised Domain Adaptation to mitigate the discrepancy in the distribution of radiometric images observed by the different satellites proves that DRAIN can be applied to the whole constellation. However, with the difficulties in training CycleGAN described in this paper, as well as its limitation within the one-to-one domain, we believe that future methods for unsupervised domain adaptation of brightness temperatures should consider a domain generalization approach or a multi-source domain adaptation approach.

ACKNOWLEDGEMENT

This work was performed with funding from both Programme National de Télédétection Spatiale and CNES-TOSCA. The HAL GPU cluster from IPSL-ESPRI was used for the training of the CycleGAN. The GMI, DPR, and SSMI/S datasets were downloaded from NASA’s Precipitation Processing System (PPS). The Météo-France datasets were downloaded from SEDOO-AERIS.

APPENDIX

As in section V-A, figure 10 and 11 show the rain retrieval by DRAIN on GMI image, adapted SSMI/S image, and original SSMI/S image on two different overpass cases. The orbit of the SSMI/S can be recognized by the larger swath. The first case, figure 10, is a cyclone between 0N-40N and 120E to 150E on the 16th of September 2022. It should be noted that there is approximately a one-hour delay between the passage of GMI and SSMI/S. The second case, figure 11, shows another rain structure in the Pacific Ocean on the 1st of February 2019 with a 4-minute delay. In both cases, the estimated rain structure on the adapted images shares more similarity to that of the GMI images. The difference in the first case could be due to the delay between the two passages.

REFERENCES

- [1] E. A. Smith, G. Asrar, Y. Furuhashi, A. Ginati, A. Mugnai, K. Nakamura, R. F. Adler, M.-D. Chou, M. Desbois, J. F. Durning, J. K. Entin, F. Einaudi, R. R. Ferraro, R. Guzzi, P. R. Houser, P. H. Hwang, T. Iguchi, P. Joe, R. Kakar, J. A. Kaye, M. Kojima, C. Kummerow, K.-S. Kuo, D. P. Lettenmaier, V. Levizzani, N. Lu, A. V. Mehta, C. Morales, P. Morel, T. Nakazawa, S. P. Neeck, K. Okamoto, R. Oki, G. Raju, J. M. Shepherd, J. Simpson, B.-J. Sohn, E. F. Stocker, W.-K. Tao, J. Testud, G. J. Tripoli, E. F. Wood, S. Yang, and W. Zhang, *International Global Precipitation Measurement (GPM) Program and Mission: An Overview*. Dordrecht: Springer Netherlands, 2007, pp. 611–653. [Online]. Available: https://doi.org/10.1007/978-1-4020-5835-6_48
- [2] N. Viltard, P. Lepetit, C. Mallet, L. Barthès, and A. Martini, “Retrieving rain rates from space borne microwave sensors using u-nets,” in *Climate Informatics 2020. 10th International Conference*, Sep. 2020. [Online]. Available: <https://hal-insu.archives-ouvertes.fr/insu-02894942>

- [3] N. Viltard, V. Sambath, P. Lepetit, A. Martini, L. Barthès, and C. Mallet, "Evaluation of drain, a deep-learning approach to rain retrieval from gpm passive microwave radiometer." *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [4] C. D. Kummerow, D. L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. J. Munchak, and V. Petkovic, "The evolution of the goddard profiling algorithm to a fully parametric scheme." *Journal of Atmospheric and Oceanic Technology*, vol. 32, no. 12, pp. 2265 – 2280, 2015. [Online]. Available: https://journals.ametsoc.org/view/journals/atot/32/12/jtech-d-15-0039_1.xml
- [5] S. Pfreundschuh, P. J. Brown, C. D. Kummerow, P. Eriksson, and T. Norrestad, "Gprof-nn: a neural-network-based implementation of the goddard profiling algorithm." *Atmospheric Measurement Techniques*, vol. 15, no. 17, pp. 5033–5060, 2022. [Online]. Available: <https://amt.copernicus.org/articles/15/5033/2022/>
- [6] S. J. Pan and Q. Yang, "A survey on transfer learning." *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [7] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, p. 1180–1189.
- [8] X. Liu, C. Yoo, F. Xing, H. Oh, G. El Fakhri, J.-W. Kang, J. Woo *et al.*, "Deep unsupervised domain adaptation: A review of recent advances and perspectives." *APSIPA Transactions on Signal and Information Processing*, vol. 11, no. 1, 2022.
- [9] M. Xu, M. Wu, K. Chen, C. Zhang, and J. Guo, "The eyes of the gods: A survey of unsupervised domain adaptation methods based on remote sensing data," *Remote Sensing*, vol. 14, no. 17, 2022. [Online]. Available: <https://www.mdpi.com/2072-4292/14/17/4380>
- [10] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [11] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2066–2073.
- [12] V. Sambath, N. Viltard, L. Barthès, A. Martini, and C. Mallet, "Unsupervised domain adaptation for global precipitation measurement satellite constellation using cycle generative adversarial nets," *Environmental Data Science*, vol. 1, p. e24, 2022.
- [13] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [14] OSCAR. (2022) Satellite: Gpm core observatory. World Meteorological Organization. [Online]. Available: https://space.oscar.wmo.int/satellites/view/gpm_core_observatory
- [15] ——. (2022) Instrument: Ssmis. World Meteorological Organization. [Online]. Available: <https://space.oscar.wmo.int/instruments/view/ssmis>
- [16] P. Bauer, L. Schanz, and L. Roberti, "Correction of three-dimensional effects for passive microwave remote sensing of convective clouds," *Journal of Applied Meteorology and Climatology*, vol. 37, no. 12, pp. 1619–1632, 1998.
- [17] Y. Hong, J. L. Haferman, W. S. Olson, and C. D. Kummerow, "Microwave brightness temperatures from tilted convective systems," *Journal of Applied Meteorology*, vol. 39, no. 7, pp. 983–998, 2000.
- [18] S.-W. Kim, D.-B. Shin, and Y. Choi, "Effects of the three-dimensional hydrometeor distributions of precipitating clouds on passive microwave rainfall estimations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 1957–1966, 2016.
- [19] C. Guilloteau and E. Foufoula-Georgiou, "Beyond the pixel: Using patterns and multiscale spatial information to improve the retrieval of precipitation from spaceborne passive microwave imagers," *Journal of Atmospheric and Oceanic Technology*, vol. 37, no. 9, pp. 1571 – 1591, 2020. [Online]. Available: <https://journals.ametsoc.org/view/journals/atot/37/9/jtechD190067.xml>
- [20] G. Skofronick-Jackson, W. Berg, C. Kidd, D. B. Kirschbaum, W. A. Petersen, G. J. Huffman, and Y. N. Takayabu, "Global precipitation measurement (gpm): Unified precipitation estimation from space," in *Remote Sensing of Clouds and Precipitation*. Springer, 2018, pp. 175–193.
- [21] C. Mallet, C. Klapisz, and N. Viltard, "Effects of heterogeneous precipitating atmospheres on simulated brightness temperatures," in *Microwave Radiometry and Remote Sensing of the Earth's Surface and Atmosphere*. CRC Press, 1999, pp. 291–298.
- [22] D. Harris and E. Foufoula-Georgiou, "Subgrid variability and stochastic downscaling of modeled clouds: Effects on radiative transfer computations for rainfall retrieval," *Journal of Geophysical Research: Atmospheres*, vol. 106, no. D10, pp. 10 349–10 362, 2001.
- [23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.
- [24] J. Figueras i Ventura and P. Tabary, "The new french operational polarimetric radar rainfall rate product," *Journal of Applied Meteorology and Climatology*, vol. 52, no. 8, pp. 1817–1835, 2013.
- [25] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [26] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems*, vol. 30, 2017.
- [27] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," *arXiv preprint arXiv:1802.05957*, 2018.
- [28] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*. PMLR, 2019, pp. 7354–7363.
- [29] J. L. Bentley, "Multidimensional binary search trees used for associative searching," vol. 18, no. 9, 1975. [Online]. Available: <https://doi.org/10.1145/361002.361007>
- [30] P. T. Boggs and J. E. Rogers, "Orthogonal distance regression," *Contemporary mathematics*, vol. 112, pp. 183–194, 1990.
- [31] S. Harary, E. Schwartz, A. Arbelle, P. Staar, S. Abu-Hussein, E. Amrani, R. Herzig, A. Alfassy, R. Giryes, H. Kuehne *et al.*, "Unsupervised domain generalization by learning a bridge across domains," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5280–5290.
- [32] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789–8797.

6.2 . References

Reference of the attached article :

V. Sambath, N. Dubois-Quilici, N. Viltard, A. Martini, and C. Mallet. Un-supervised Domain Adaptation to mitigate out-of-distribution problem of spatial radiometer images : Application to Quantitative Precipitation Estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

Additional references for this chapter :

- J. Figueras i Ventura and P. Tabary. The new French operational polarimetric radar rainfall rate product. *Journal of Applied Meteorology and Climatology*, 52(8) :1817–1835, 2013.
- D. Harris and E. Foufoula-Georgiou. Subgrid variability and stochastic downscaling of modeled clouds : Effects on radiative transfer computations for rainfall retrieval. *Journal of Geophysical Research : Atmospheres*, 106(D10) :10349–10362, 2001.
- A. Y. Hou, R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi. The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, 95(5) :701 – 722, 2014. doi : 10.1175/BAMS-D-13-00164.1. URL <https://journals.ametsoc.org/view/journals/bams/95/5/bams-d-13-00164.1.xml>.
- C. Mallet, C. Klapisz, and N. Viltard. Effects of heterogeneous precipitating atmospheres on simulated brightness temperatures. In *Microwave Radiometry and Remote Sensing of the Earth's Surface and Atmosphere*, pages 291–298. CRC Press, 2023.

Conclusion and Perspectives

The water cycle is a life-sustaining system that demands our utmost attention under the current pressure of climate change. The modification of the water cycle due to climate change is a complex issue as each component in the water cycle interacts with and influences each other. **Chapter 1** summarizes some of the current observed and projected changes in the water cycle, including regional and global precipitation patterns, such as the monsoons and the ITCZ, and extreme hydrological events, such as floods and droughts. The changes differ greatly from one region to another.

Remote sensing missions in the past few decades have accumulated an unprecedented amount of data on Earth observation. Among these missions, the Global Precipitation Measurement (GPM) mission is the center of this study. GPM offers a constellation of satellites equipped with microwave imagers. These instruments measure the brightness temperature (TB) at a wide range of frequencies, which allows for a retrieval of multiple geophysical quantities related to the water cycle. The present study uses data from the GPM Core Observatory (GPM-CO), the reference satellite of the constellation, which carries a Microwave Imager (GMI) and a precipitation radar (DPR).

This thesis aims to develop a method to extract information from microwave TB images without requiring any ancillary data. Using Computer Vision, I explored unsupervised image segmentation models to extract information from the TB images and how to validate and interpret the resulting segmentation. As a complement to the unsupervised approach, I also present a supervised learning approach for rain retrieval from TB images. Finally, this thesis offers a feasibility study on an unsupervised domain adaptation as a transfer learning technique to apply these approaches to multiple microwave radiometers.

In line with treating TB measurements as images, **Chapter 2** presents a supervised model for retrieving rain rates from TB images using U-Net models and quantile regression. Using only TB images and DPR rain rates, DRAIN achieves a performance similar to that of the state-of-the-art reference algorithm of the GPM algorithm. Additionally, its performance does not differ between land and ocean. The supervised approach poses less challenge for the evaluation and interpretation of the results. However, DRAIN can only be trained on the GPM-CO, as it carries precipitation radar, observing precipitation events simultaneously with the GMI, which can be used as the target during training. In contrast, unsupervised segmentation can be trained on any satellite as it does not require a target.

Chapter 3 offers an overview of machine-learning approaches for image segmentation that require little to no labels using unsupervised learning and self-supervised learning. First, for unsupervised algorithms, two major branches of algorithms are discussed : the pixel-wise distance-based approaches (K-means, DBSCAN, Super-pixels) and non-pixel-wise approaches (convolutional layers and vision transformer). The non-pixel-wise approaches, which consider the spatial structure within an image, are categorized based on their training framework. In unsupervised training, the training frameworks revolve around the creation of the objective function to segment the image without any labels. This chapter gives example algorithms for three methods for defining the objective function : pseudo-labeling, reconstruction of the input image, and contrasting images within the dataset. Secondly, self-supervised learning differs from unsupervised learning as it aims to first learn information from the input before segmenting the image in a downstream task. Similarly, **Chapter 3** presents some training frameworks for self-supervised learning.

The summary of methods provides some pros and cons of the presented example algorithms as well as some metrics for evaluating the segmentation. The evaluation metrics include indexes based on intra-class and inter-class dispersion and traditional segmentation scores such as accuracy, precision, recall, and F1-score. The selected models are implemented and compared in **Chapter 4**.

Chapter 4 details the implementation and evaluation of the selected unsupervised segmentation models. The approaches tested include a K-means model as well as several models based on fully convolutional networks of different levels of depth. A challenge that arises when using an unsupervised approach is the evaluation, validation, and interpretation of the results. In this study, these models are evaluated based on two external variables : precipitation and SST. All the optimal models for each approach achieved very similar scores. The choice of an optimal model is based on its overall performance on multiple orbits and its number of segmentation classes. Before studying the evolution of these classes on a longer dataset, it is necessary that the labeled precipitation and SST classes can capture the regional and global patterns of precipitation and SST. On a regional level, the identified rain classes are shown to be able to follow the seasonality in monsoon precipitation. They can recover the same patterns and number of rain pixels as the DRAIN algorithm. For global ocean precipitation, the count of rain classes is able to indicate the pattern of areas where rainfall occurs more frequently and the seasonality of precipitation. Similarly, the SST classes are representative of the general structure of SST and are able to follow the seasonality in SST.

The main limitations of current optimal models include their difficulty in distinguishing between snow-covered surfaces and rain in high latitudes. In addition, there is room for improvement in terms of the recall score for precipitation classes. Future efforts to improve the segmentation model for TB data should focus on

these aspects.

After presenting and evaluating the supervised model in **Chapter 2** and the unsupervised model in **Chapter 4**, trend analysis of the ten years of data of the GMI (2014-2024) using the Mann-Kendall test is discussed in **Chapter 5**. First, the trend estimation is done directly on the TB data, using the 89 GHz channel due to its interpretability in terms of convective rain. There is a non-significant decreasing trend in the coldest T_{B89H} . The interpretation of this trend is difficult. Nonetheless, some regional changes are clearly observed, even if it is impossible to attribute a cause to these changes. Next, the DRAIN result for the GMI dataset shows an increase in the global number of rain pixels and a decrease in rain intensity. Lastly, the unsupervised segmentation classes show an upward trend in the number of rain classes for three IPCC regions : WAF (Non-significative), BOB (Significative), and SEA (Significative). The trends observed in DRAIN and in the unsupervised segmentation are coherent with the continuity of precipitation trend presented in the IPCC AR6 for data between 1985-2014. However, the conclusion of the nature of the trend observed using GMI data alone is impossible. With only GMI, it is only possible to observe instantaneous precipitation. The trend could be due to multiple factors, such as the life cycle, the size, and the number of occurrences of precipitation events.

For supervised rain retrieval DRAIN, the application on other satellites requires a transfer learning step, as the performance of a Machine Learning algorithm degrades significantly if the input data distribution is different from the one it was trained on. An unsupervised domain adaptation model is presented in **Chapter 6**. In the presented feasibility study on unsupervised domain adaptation, the model transforms the TB from the SSMI/S to the GMI domain. The adapted TB allows DRAIN to retrieve better precipitation structure and intensity. Using overpasses between SSMI/S and GMI, the scores on rain intensity significantly improve when the adapted images are used. The efficiency of the domain adaptation proves that deep-learning algorithms can be applied to any satellites in the constellation. Nonetheless, the domain adaptation with a generative model such as CycleGAN poses many challenges in obtaining convergence during training. In addition, it is more effective to simultaneously transfer multiple satellites to the GMI domain instead of transferring them one by one.

The large amount of data collected by remote sensing Earth observations offers an unprecedented opportunity for the application of the high-performance and cutting-edge machine learning algorithm. However, the extraction of information from these climatic records is not straightforward, as no labels are readily available for training. Human annotation for each pixel comes at a high cost or is simply impossible. Unsupervised learning has great potential. However, as this deep-learning is basically a black box, it is important to thoroughly analyze the resulting segmentation. The results of the unsupervised segmentation allow us to conclude that

an unsupervised segmentation of TB images is able to provide valid and geophysically meaningful segmentation that can capture the pattern and characteristics of geophysical variables. The trend found using the unsupervised approach is also compared to the more traditional approaches to verify its validity.

As the geophysical properties of the result are verified, this unsupervised approach can be applied to any microwave radiometer. Thus, the next step is to apply this method on multiple satellites to expand the extraction of information from TB images and solve the challenge of harmonizing the segmentation. Next, I will discuss the perspectives for future studies.

Perspectives

Merging segmentations from several models

The results from Chapter 4 show that the algorithms tested provide similar segmentation that has very similar scores. Improvement of the segmentation could be achieved by creating an ensemble model by merging the segmentations from multiple models. It can be advantageous to verify if any of the current models actually specializes in any geophysical variables or if it's possible to train a segmentation model that specializes in a particular quantity. If this is possible, it is advantageous to create an ensemble model with different coefficients for different specializations. This could potentially solve the problem of the current optimal models not being able to distinguish between rain and snow/ice-covered surfaces if it is possible to train an unsupervised model specialized in recognizing snow/ice-covered surfaces.

Application on the whole GPM constellation

Using GMI data from 2014 to 2024, there is a statistically significant increase in the total number of rain pixels through both the supervised and unsupervised methods. However, relying on only GMI, the dataset comprises only an instantaneous observation of precipitation events. Consequently, the nature of the increase in total rain pixels is not known. The GPM constellation consists of multiple satellites carrying radiometers. Using data from the whole constellation will allow us to have more observations and better coverage due to less revisit time.

For the supervised approach to retrieving the rain intensity, DRAIN, the application on other satellites in the constellation requires a transfer learning step. A feasibility study is presented in Chapter 6 for a domain adaptation effort in transforming the observations of other satellites to the domain of the GPM Core Observatory.

For the unsupervised approach, the model can be retrained on all of the satellites in the constellation. The resulting segmentations will have different label numbers but will again represent geophysical variables. Using their overpass, it is

possible to regroup the classes that represent the same geophysical quantity. Additionally, the domain adaptation approach could be used as an intermediate step to harmonize the resulting segmentations.

Improvement of the unsupervised segmentation algorithm

Computer vision is a field that is constantly evolving, with numerous new algorithms emerging. Now that the evaluation tactics for the application in TB are established, it is essential to test innovative new algorithms. In addition, it is interesting to test a self-supervised learning approach where the information extracted from the output can be used to explore the evolution of the water cycle directly without first creating a segmentation.

Irregularity due to instrument deterioration or adjustment

The analysis of the evolution of geophysical quantities based on the satellite dataset requires the distribution of the observed data to stay the same. It is essential to verify if there are any changes due to the degradation of the instrument or simply an adjustment of the instrument.

A - Appendix : Feasibility study on unsupervised domain adaptation

The conference paper presented in this Appendix is the precedent feasibility study for the article in **Chapter 6**. To test the transferability of DRAIN using CycleGAN as an unsupervised domain adaptation model, the 89 GHz in horizontal polarization (89H) is transformed to the domain of the 89 GHz in vertical polarization (89V), and the 89V is transformed into the domain of the 89H. The differences between the horizontal and the vertical polarization are due to the surface emissivity. The TB at 89V is almost always higher than 89H since the surface emissivity at 89V is generally higher than at 89H, except for some rare cases. However, this adaptation is notably less complicated than an adaptation between two satellites due to the fact that these are two channels of the same instruments. Therefore, there are no differences in orbitography.

A modified version of DRAIN that only takes the 89 GHz in both polarization as inputs is trained. Then, we evaluated the performance of this modified DRAIN on the adapted 89H and 89V data. The motivation behind applying domain adaptation to different polarizations of the same satellites is that it allows us to have pixel correspondence for the evaluation, which is the key to deciding if this feasibility study could lead to a real application. Results on rain retrieval show that domain adaptation significantly increases the performance of DRAIN.

The hyper-parameters, model choice, and implementation techniques from this feasibility study serve as the starting point for the actual application of domain adaptation between two different satellites described in **Chapter 6**. In particular, this feasibility study shows the importance of identity loss in CycleGAN as well as the importance in keeping the training loss balanced between the two domains.

Reference of the attached article :

V. Sambath, N. Viltard, L. Barthès, A. Martini, and C. Mallet. Unsupervised domain adaptation for global precipitation measurement satellite constellation using cycle generative adversarial nets. *Environmental Data Science*, 2022.

RESEARCH ARTICLE

Unsupervised domain adaptation for GPM satellite constellation using CycleGAN

Vibolroth Sambath¹, Nicolas Viltard^{1*}, Laurent Barthès¹, Audrey Martini¹ and Cécile Mallet¹

¹Laboratoire ATmosphères, Milieux et Observations Spatiales (LATMOS), Guyancourt, 78280, France

*Corresponding author. Email: nicolas.viltard@latmos.ipsl.fr

Received: 10 January 2022

Keywords: Unsupervised domain adaptation, transfer learning, rain retrieval, microwave brightness temperature

Abstract

Artificial Intelligence has provided many breakthroughs in the field of computer vision. The fully convolutional networks U-Net in particular have provided very promising results in the problem of retrieving rain rates from space-borne observations, a challenge that has persisted over the past few decades. The rain intensity is estimated from the measurement of the brightness temperatures on different microwave channels. However, these channels are slightly different depending on the satellite. In the case where a retrieval model has been developed from a single satellite, it may be advantageous to use domain adaptation methods in order to make this model compatible for all the satellites of the constellation. In this proposed feasibility study, a CycleGAN model is used for adapting one set of brightness temperature channels to another set. Results on a toy experiment show that this method is able to provide qualitatively good precipitation structure but still could be improved in terms of precision.

Impact Statement

Early supervised and end-to-end deep learning approaches in climate studies, especially in satellite observations, are very limited in application due to the lack of available annotated samples, the high demand for pixel accuracy and the non-conventional nature of the data. The present feasibility study on unsupervised domain adaptation will allow for the increase in re-usability of a deep learning model pre-trained on one satellite to many more with similar physical characteristics. While previous unsupervised image domain adaptation approaches focus on qualitative aspect and on classification task, the present objective involves a regression task and non-RGB image data. The results obtained on the unsupervised domain adaptation with CycleGAN on microwave imagery have a significant impact on the practical perspectives of applying deep learning models to the spatial observation of the earth. In terms of climate studies, this unsupervised transfer learning approach will improve the knowledge of the precipitation evolution over the last 30 years. Hence, the study has an important scientific impact in the analysis of the relation between precipitation and climate change as well as a societal impacts on the forecast and preparation for climate change.

1. Introduction

The estimation of precipitation for a given date and location is a very challenging task because rain is very intermittent in time and space. Ground-based observations alone could be very lacking due to the difficulty in obtaining a uniformly calibrated observation with a good spatial resolution over a

large region, especially over the oceans (Hou et al., 2014). On the other hand, global satellite coverage offers a great advantage in estimating a uniform global precipitation. For this exact purpose, the Global Precipitation Measurement (GPM, 2014-present) mission, the successor of the Tropical Rainfall Measuring Mission (TRMM 1997-2015 (Kummerow et al., 1998)), has launched a mother satellite called the GPM Core Observatory and a constellation of daughters. Aboard the GPM Core Observatory, a passive microwave radiometer (GPM Microwave Imager or GMI) provides the brightness temperatures while a Dual-frequency Precipitation Radar (DPR) provides a more direct measurement of the precipitation. The main purpose of the Core Observatory is to serve as the reference for unifying the precipitation estimates from the other satellites in the constellation. The co-located data of brightness temperatures and surface rain rates also open up the opportunity to develop a supervised deep learning model for rain retrieval. Numerous studies have been done on the subject of rain retrieval with a list of literature available in Viltard et al. (2020).

Viltard et al. (2020) developed a deep learning model using U-Net for rain retrieval (DRAIN) on the co-located data of the GPM Core Observatory. U-Net is a fully convolutional neural network containing a contraction path, an expansion path and skip connections (Ronneberger et al., 2015). In DRAIN, U-Net is trained to estimate quantiles of rain products with the brightness temperature from the GMI as inputs and the rain rates of the DPR as targets, with further details available in (Viltard et al., 2020). The next step is to take full advantage of the GPM constellation with this deep learning approach. The GPM constellation, made up of a network of international satellites, can provide up to 80% of the global coverage in less than 3 hours (Hou et al., 2014). Successfully utilizing the whole constellation of GPM will offer a uniform global precipitation map.

GPM official radiometer algorithm is based upon a Bayesian approach in which the GPM core satellite is used to generate an a-priori database of observed cloud and precipitation profiles (Passive Microwave Algorithm Team Facility, 2017). One year (sept 2014-August 2015) of matched GMI/hydrometeor observations is used to construct the a-priori database. The combined product is built with a forward radiative transfer model calculation to compute brightness temperatures sets for the different radiometers, that is to say with different frequency channels and viewing angles. As a description of temperature and water vapor profiles and surface emissivity is needed to perform the simulations, ancillary data coming from GCM's reanalysis are associated with each pixel. The variability of the spatial resolution of the different PMRs (Passive Microwave Radiometers) of the constellation is neglected in GMP-V5. The use of simulated brightness temperatures to develop the retrieval algorithm is an important source of uncertainty especially in the presence of scattering by hydrometeors. The differences in PMR's field of view that are not taken into account can also introduce significant errors (Kidd et al., 2016).

As DRAIN is trained on the co-located brightness temperatures and rain rates from the GMI and the DPR respectively, this model only works for the brightness temperatures of the GPM Core Observatory. This model could not be applied directly to the constellation because between different satellites, there are differences in viewing angle, frequency band, and spatial resolution. In addition, these other satellites are only equipped with microwave imagers and therefore, there is no co-located data available for a supervised learning approach. In order to benefit from the frequent revisit time and better coverage of the whole constellation, a method to transfer the knowledge gained from the GPM Core Observatory is required.

Transfer learning differs from the traditional machine learning algorithms as it relies on a previously trained knowledge. In traditional machine learning algorithm, a model is trained on the source domain and is applied to the source domain. Current Machine Learning techniques, which have given very good results in the field of computer vision, require a large database to train on and are only valid for this domain. As a result, a new model needs to be trained each time the data are from a different feature space or distribution. On the other hand, in transfer learning, a model is first trained with the source domain. Next, the knowledge is transferred in order to create a model for a new task. This is particular advantageous in many cases where there is a lack of data for the targeted task but an abundance of data in a similar domain.

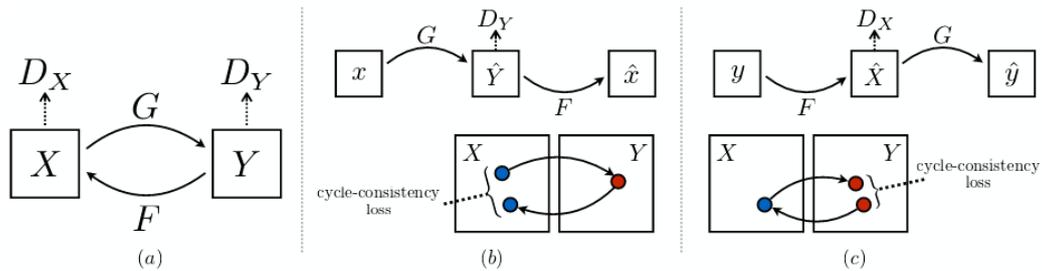


Figure 1. (a) The architecture of CycleGAN. (b) and (c) The illustration of cycle-consistency loss. (Zhu et al., 2017a).

The first challenge in transferring the knowledge is, of course, the lack of training data in the target domain. It is possible to co-locate various satellite observations and GPM Core Observatory data. However, due to the highly intermittent nature of rain and the difference in orbit of the satellites, co-located data are very scarce and certainly inefficient for relying on fine-tuning as a method of transfer learning. In order to use as many satellites data as possible, we have to turn to the transductive transfer learning method where the target domain labels are unavailable while the source domain labels are available (Pan and Yang, 2010). With some unsupervised methods, the domain adaptation and regression could be achieved with one model, for example, the Unsupervised Domain Adaptation by Back-propagation (UDA) (Ganin and Lempitsky, 2015). In UDA, there are three components: the feature extractor, the label predictor, and the domain classifier (Ganin and Lempitsky, 2015). By integrating the domain classifier within the model, UDA is able to make the two domains as common as possible while providing the predictions at the same time. Though with several modifications tested, this model could not be successfully implemented to this experiment. We have then turned to a domain adaptation method done outside of the rain retrieval model. In order to perform this domain adaptation task, the Cycle Generative Adversarial Nets (CycleGAN) was used.

The end goal is to adapt any satellite scans in the GPM Constellation (target domain) to that of the GPM Core Observatory (source domain) in order to be able to use the U-Net for rain retrieval previously trained on the source domain. The feasibility study presented here, the GMI 89 GHz channels in horizontal and vertical polarisation (hereafter 89H and 89V) are the two domains to be adapted using CycleGAN. In the following text, section 2 describes the domain adaptation method. Next, section 3 provides details about the data used for training and testing. Section 4 shows the evaluation of the method in terms of the similarity between original and adapted domain as well as its performance in rain retrieval. Finally, in section 5, next steps and possible improvements will be discussed.

2. Method

The method used in this feasibility study consists of applying an unsupervised domain adaptation on satellite images using a Generative Adversarial Nets (GAN, (Goodfellow et al., 2014)) based approach called CycleGAN (Zhu et al., 2017a). Then, the newly transformed images are tested on a rain retrieval model that was previously trained with original images.

2.1. CycleGAN

CycleGAN is a an image transformation technique that does not require paired images (Zhu et al., 2017a). It consists of two GANs working together, each containing a generator and a discriminator (figure 1). The first generator G takes an image from the source domain X and transforms it into the target domain Y . The second generator F works the other way around, by transforming an image in Y into X . The discriminators D_X and D_Y try to correctly label if a sample is from its respective domain.

The weights of the two generators are updated by their combine loss. Each generator has three terms of loss: cycle-consistency loss, identity loss and adversarial loss. Cycle-consistency is based on the idea that a complete image translation cycle (from X to Y and back to X) as shown in figure 1 (a) should be able to bring back a close enough image to the original (Zhu et al., 2017a). The identity loss is included since it has previously showed the ability to preserve color when transforming between photos and painting (Zhu et al., 2017a). It is the difference between an image and the transformation to its own domain, for example, $\| F(x) - x \|_1$. In the presented experiment, without identity loss, the generator is able to reconstruct the form of the structure but unable to reproduce the value of the brightness temperature. The discriminator loss, on the other hand, is calculated upon its ability to distinguish the real images and the fake images generated by the generators. The complete objective function could be found in (Zhu et al., 2017a).

Zhu et al. (2017a) concluded that the CycleGAN worked best for color and texture changes, for example transforming between different painting styles. However, it is less successful when geometric transformation is involved, for example cat to dog transfiguration. Furthermore, de Bézenac et al. (2019) emphasizes CycleGAN capacity to only perform well for distributions that are close to one another. In the experiment here, this is not really a problem because the rain cells have similar geometries whatever the selected channel. Zhu et al. (2017a) also highlighted the impossibility to achieve as good performance as in the case of paired data approach and the failure when the training characteristic distribution is not representative enough of the test data.

2.2. Training details

Several architectures of the generator proposed by Zhu et al. (2017a) were tested. With previous success shown by Viltard et al. (2020), U-Net seems to be a very good candidate for working with satellite images, particularly the brightness temperatures. Using U-Net as generator, the generated images are able to better imitate the structure native to the targeted domain. The second challenge is the imbalance in the training loss between the two domains. In the first few attempts, with the learning rate schedulers for both generators evolving the same way as the training progresses, the losses remain very imbalance. Hence, different learning rates are set for each generator with higher learning rate for the generator that seems to struggles more. This results in similar losses for both domains, which may be due to the fact that the two generators work together to establish the cycle-consistency loss. It should also be noted that the batch size has an important impact. In this toy experiment, batch of sizes 1, 4, 8 and 16 were tested. After several initialisation of the network, eight images per batch gave the best results in terms of structure within the satellite observation. Smaller or bigger batch size seems to degrade the results. Batch size effect is an empirical remark made equally on the official GitHub depository of CycleGAN (Zhu et al., 2017b). Therefore, it should warrant a careful testing for future application.

3. Toy Experiment Data

The GMI is a multi-channel conically scanning radiometer with a swath of 904 km and with channels ranging from 10 GHz to 183 GHz. These channels are measured in both Horizontal (H) and Vertical (V) polarization. In this toy experiment, only the 89 GHz-channel is used. This channel has the resolution of 4.4 km by 7.3 km. We aim to transform between the horizontal and vertical polarization of the GMI 89 GHz channel using CycleGAN. In this case, X in figure 1 represents 89V while Y represents 89H. This choice of toy data set will later allow comparisons of the adapted image and the targeted satellite scan. The training and validation data sets for each domain do not contain overlapping events to properly test the unpaired image domain adaptation method. The difference between 89H and V is mostly due to the surface emissivity difference between the two channels. The V surface emissivity is almost always higher than the H surface emissivity leading to a generally higher V brightness temperature. However,

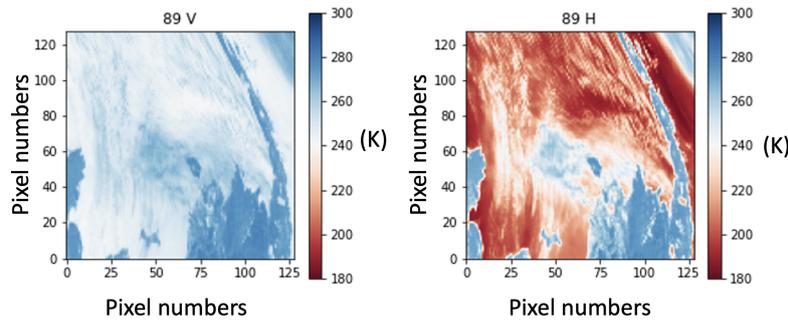


Figure 2. Vertical (89V) and horizontal (89H) polarisation of the 89-GHz channel brightness temperature in Kelvin from the GMI. The image is of 128 by 128 pixels representing roughly (1024 km by 1024 km).

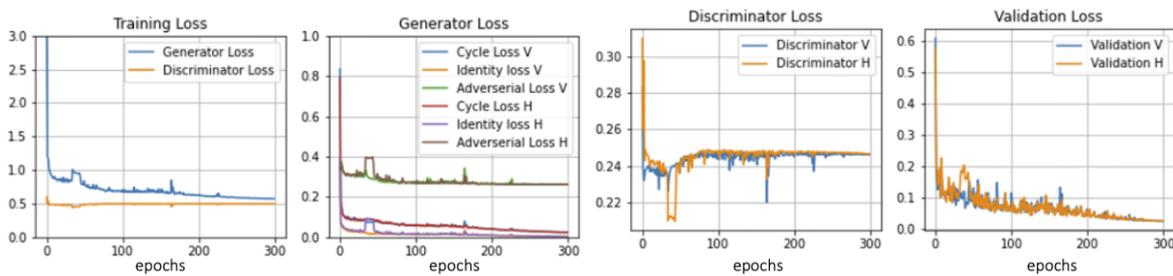


Figure 3. Training and validation losses for different components of the CycleGAN. The training loss (first plot) plot shows the generator and discriminator loss throughout training. Then, generator loss plot (second plot) and discriminator loss plot (third plot) show the details of each component. Finally, validation loss plot (fourth plot) shows the cycle-consistency and identity loss of each domain on validation dataset.

polarisation due to scattering by ice might occur in (rare) cases of oriented particles leading to a H brightness temperature higher than V. This is true for both land and ocean situations.

The training data set consists of 24 000 images for each domain, making 48 000 images in total of different observations taken between 2015 and 2017. The validation set is made up of 4 000 images for each domain taken from the same period. These images contain 221×256 pixels and are chosen with the conditions that they either have at least 100 pixels with more than 10 mm/h rain or at least 10 pixels with more than 100 mm/h rain rates. The selection process is essential to obtain enough images with a precipitation event. Next, the training and validation sets undergo the pre-processing step: data normalisation, random crop (to 128×128 pixels) and random rotation. The random crop (cropping randomly within the image) and random rotation (choosing an angle at random to rotate the image) are added as a data augmentation method and to increase the difficulty of the task for the CycleGAN model. An example of training data is given in figure 2. Note that, though the example given here is of the same event, neither train nor validation images of each domain correspond to each other.

4. Results

Figure 3 shows the training and validation losses. As discussed previously, different learning rate schedules for each domain allows the losses of both domains to evolve in the same way and without a gap between them. According to experiments, this could not be achieved if the learning rates for all networks involved have the same learning rate schedules. To evaluate the performance of CycleGAN on domain adaptation between 89V and 89V, its ability to reconstruct complex rain structure as well as its accuracy in terms of brightness temperatures are discussed.

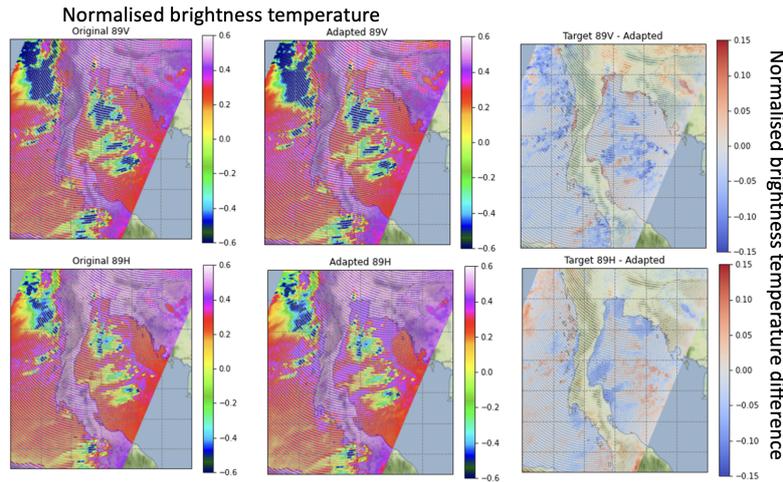


Figure 4. Original, adapted, and their difference of the 89-GHz channel observation from GMI on the 29th May 2017 with latitude $5\text{-}15^\circ\text{ N}$ and longitude $96\text{-}105^\circ\text{ E}$ (over parts of Thailand and Cambodia).

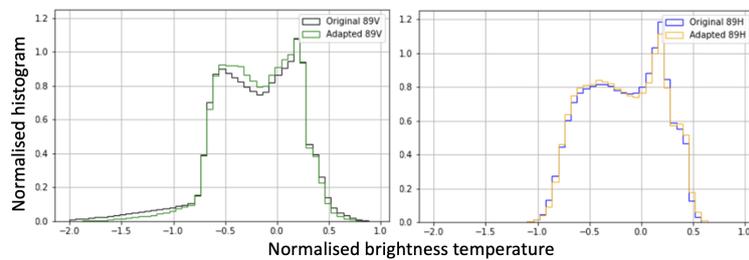


Figure 5. Comparison of the histogram of original 89V data and the adapted 89V data (left) and original 89H data and the adapted 89H (right).

4.1. Results on adapted satellite images

Figure 4 shows the original satellites scans and the adapted images (89V adapted from 89H and 89H adapted from 89V) on a case study. CycleGAN can reproduce very well all the precipitation structures in the original images. However, in terms of values, there are differences of brightness temperature between the original and the adapted images that could not be picked up on qualitatively. Hence, after confirming the re-created structure of brightness temperature for a complex precipitation event, the next step is to investigate the accuracy. With the test data consisting of two months of observation including December 2018 and May 2020, figure 5 shows the comparison between the original normalised data and the adapted results. The original and adapted histograms are almost superimposed though with some inaccuracy. Calculating the Kullback-Leiber divergence (Bishop et al., 1995) also confirms that there is more similarity between the original and adapted domain than without the transformation.

4.2. Results on rain retrieval

Relying on the previous work by Viltard et al. (2020), a U-Net model is used to train a rain retrieval model with the GMI 89 GHz channel as inputs. The U-Net model takes two layers as inputs, the first one is the 89V and the second one being 89H. The target data is the spatially and temporally co-located data from the DPR of the GPM Core Observatory. The DPR surface rain product is the result from the merged of Ka- (13.4 GHz) and Ku-band (35.5 GHz) radars. It has a horizontal resolution of 5 km and a swath of 245 km. Further information about the treatment of this data is available in Viltard et al. (2020).

The rain retrieval is evaluated on three criteria, its ability to distinguish rain and no-rain cases (tables 1 and 2), the mean absolute error when compared to the DPR rain rates, and its ability to reconstruct the

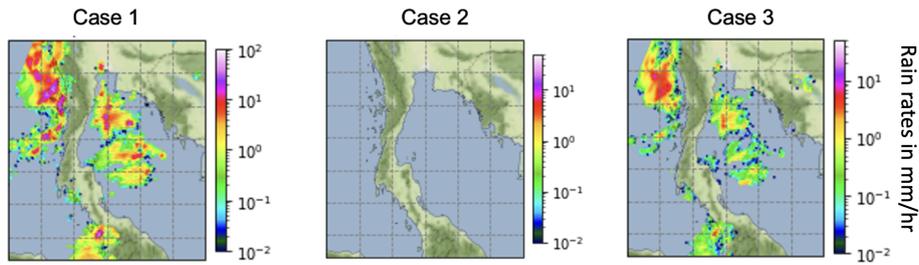


Figure 6. (Same observation as figure 4) Comparison of retrieved surface rain rates in mm/h in Case 1 (left), Case 2 (middle) and Case 3 (right).

Table 1. Classification score for rain vs no-rain cases using the two-month test data.

Score	Case 1	Case 2	Case 3
Recall	0.74	0.002	0.38
Precision	0.81	0.007	0.27
f1-score	0.77	0.003	0.31

Table 2. Mean absolute errors using the two-month test data.

Rain rates (mm/h)	Case 1	Case 2	Case 3
0.1 to 3	0.28	0.57	0.52
3 to 10	2.13	5.01	4.60
Above 10	19.85	20.26	19.0

rain structure. Three cases are compared. Case 1 is the best case scenario, where the rain retrieval model is tested on [89V, 89H] input data, the correct order on which it was train. Case 2, on the other hand, is the worst case scenario, which is using a model trained on [89V, 89H] inputs and tested on [89H, 89V] inputs. The last one, Case 3, the rain retrieval model is tested on the adapted brightness temperature, that is to say, the pair [89V (adapted from 89H), 89 H (adapted from 89V)]. The objective is to situate ideally as close to Case 1 as possible. Figure 6 shows the rain retrieval on the same case study as in the above sub-section. Comparing to the ideal case (Case 1), the retrieved rain intensity with the adapted data as inputs (Case 3) is much weaker. In other words, Case 3 has approximately the same structure but very weak in terms of intensity. Case 2, on the other hand, could not reproduce precipitation at all. Next, the classification score in table 1 shows the ability to distinguish between rain and no rain cases. No-rain cases refer to the prediction below 0.1 mm/h. With domain adaptation, clear improvement was observed across all scores. We could also observe a small improvement in terms of mean absolute error score in table 2.

5. Conclusions and Perspectives

Although this domain adaptation method performs very well in terms of qualitative assessment, there is still room for improvement for its application in rain retrieval. Based on the results presented in section 4, the pre-trained U-Net model is very sensitive to both the structure, and in turns the gradient, as well as the value of the brightness temperature itself. Nevertheless, the results from this toy experiment present a promising proof of concept. The next step would be to study the real application with GPM core observatory as the source domain and one of the satellite in the constellation, for

example SSMI/S, as the target domain. This feasibility study also highlights the importance of quantitative assessment in image domain adaptation for a regression task. In order to give better estimation, a more elegant approach consists of constraining the CycleGAN training process with the loss from rain retrieval model. Prior works related to this approach include the Conditional Generative Adversarial Nets (Conditional GAN) (Mirza and Osindero, 2014) and one of its many variations, Red-GAN (Qasim et al., 2020). In Conditional GAN, additional information, for example class label, is added to the training of the GANs. Red-GAN is built upon the concept of Conditional GAN with a third-player integrated into the two-player Conditional GAN (generator and discriminator) in order to better favor the final objective. In future application on rain retrieval, the U-Net could become an extension of the CycleGAN model. The error in rain rate could then be integrated into the training loss in order to update the generators and discriminators weights. As a consequence, the domain adaptation process is constrained to best work on the rain retrieval application.

Funding Statement. This research was supported by the EUR IPSL and CNES-TOSCA.

Competing Interests. None.

Data Availability Statement. The 1C-R.GPM.GMI.XCAL2016-C, 2A.GPM.DPR.V8 and 2A.GPM.GMI.GPROF2017v1 data sets were provided by the NASA/Goddard Space Flight Center and PPS, which develop and compute these data sets as a contribution to GPM, and archived at the NASA GES DISC.

Ethical Standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Author Contributions. Conceptualization: all authors. Methodology: V.S; C.M; L.B. Data curation: N.V. Data visualisation: V.S; N.V. Writing original draft: V.S. All authors approved the final submitted draft.

References

- Bishop, C. M. et al. (1995). *Neural networks for pattern recognition*. Oxford university press.
- de Bézenac, E., Ayed, I., and Gallinari, P. (2019). Optimal unsupervised domain translation. *arXiv preprint arXiv:1906.01292*.
- Ganin, Y. and Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15*, page 1180–1189. JMLR.org.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.
- Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T. (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 95(5):701 – 722.
- Kidd, C., Matsui, T., Chern, J., Mohr, K., Kummerow, C., and Randel, D. (2016). Global precipitation estimates from cross-track passive microwave observations using a physically based retrieval scheme. *Journal of Hydrometeorology*, 17(1):383 – 400.
- Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J. (1998). The tropical rainfall measuring mission (trmm) sensor package. *Journal of Atmospheric and Oceanic Technology*, 15(3):809 – 817.
- Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.
- Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.
- Passive Microwave Algorithm Team Facility (2017). Gprof 2017, version 1 (used in gpm v5 processing) global precipitation measurement (gpm) mission algorithm theoretical basis doc. https://pmm.nasa.gov/sites/default/files/document_files/ATBD_GPM_GPROF_June1_2017.pdf.
- Qasim, A. B., Ezhov, I., Shit, S., Schoppe, O., Paetzold, J. C., Sekuboyina, A., Kofler, F., Lipkova, J., Li, H., and Menze, B. (2020). Red-gan: Attacking class imbalance via conditioned generation. yet another medical imaging perspective. In *Medical Imaging with Deep Learning*, pages 655–668. PMLR.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Viltard, N., Lepetit, P., Mallet, C., Barthès, L., and Martini, A. (2020). Retrieving rain rates from space borne microwave sensors using u-nets. In *Climate Informatics 2020. 10th International Conference*.
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017a). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.
- Zhu, J.-Y., Park, T., and Tongzhou, W. (2017b). Github depository: pytorch-cycleGAN-and-pix2pix. <https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>. Accessed: 2010-09-30.

B - Appendix : Résumé de Thèse en français

Le cycle de l'eau joue un rôle essentiel pour toute vie sur Terre. L'observation et la quantification de ses divers composantes sous la pression du changement climatique sont devenues de plus en plus importantes pour la société et la science. Toutefois, les réponses de chaque composante du cycle de l'eau au réchauffement climatique diffèrent fortement d'une région à l'autre. Certaines régions ont ainsi connu une augmentation des précipitations, tandis que d'autres ont connu une diminution. Il est donc urgent d'observer le cycle de l'eau à l'échelle régionale mais de façon globale et les mesures spatiales sont particulièrement adaptées pour cela, à condition de surmonter les difficultés méthodologiques associées à la mesure par télédétection.

La télédétection a accumulé une grande quantité de données d'observation spatiale. Concernant les précipitations, la mission Global Precipitation Measurement (GPM) (2014-présent) offre une constellation de satellites équipés de radiomètres micro-ondes permettant d'obtenir une vaste base de données de températures de brillance (TB) à diverses fréquences et polarisations. L'utilisation de plusieurs canaux est essentielle, chacun ayant une sensibilité propre aux différents constituants de l'atmosphère, notamment la vapeur d'eau ainsi que l'eau sous forme liquide ou glacée. Ainsi les canaux aux fréquences les plus basses, beaucoup moins sensibles aux constituants atmosphériques, seront affectés par la surface comme par exemple la température de la surface de la mer (SST), alors que les canaux aux fréquences les plus hautes seront plutôt sensibles à la partie supérieure de l'atmosphère notamment à la température des sommets des nuages de glace.

De façon classique, dans les produits standards, ces températures de brillance sont utilisées en entrée de modèles d'inversion afin d'estimer un certain nombre de variables d'intérêt telles que l'intensité de pluie ou le contenu en eau nuageuse ou des profils d'humidité. Ces modèles sont généralement développés soit à partir de modèles physiques de transfert radiatif permettant de simuler les températures de brillance à partir de profils verticaux atmosphérique, soit à partir de bases de données de télédétection spatiales associées à des modèles de type Bayesian Neural Networks (BNNs). L'approche utilisée dans cette thèse est différente, et peut être assimilée à une approche par objet. Plus précisément, il s'agit sans l'introduction d'aucun a priori ou de variables auxiliaires d'isoler des objets définis par les seules températures de brillance. Ces objets (ou une partie) sont ensuite caractérisés à posteriori à l'aide de variables auxiliaires. L'intérêt d'une telle approche est de ne faire intervenir dans la définition des objets qu'un seul type d'instrument (radiomètre) et qu'elle supprime ainsi les sources d'erreurs liées soit à la modélisation de la physique soit aux problèmes liés à l'observation d'une même scène par deux ins-

truments différents, comme par exemple, la colocalisation des pixels et la résolution différente.

La présente étude utilise le radiomètre micro-onde (GMI) du GPM Core Observatory (GPM-CO), qui est le satellite de référence pour unifier les satellites de la constellation. En effet, ce satellite et contrairement aux autres satellites embarque également un radar précipitation permettant une estimation plus directe des précipitations. En considérant les différents canaux de TB comme des images non conventionnelles, je propose une nouvelle méthode de segmentation sémantique non-supervisée pour observer le cycle de l'eau à partir de ces dernières. Il s'agit d'extraire des informations sous forme d'objet sans l'aide d'une annotation ou de données auxiliaires. Outre les intérêts cités au paragraphe précédent, une approche non-supervisée est applicable à n'importe quel satellite de la constellation indépendamment de sa configuration spécifique (sélection de canaux, incidence de visée, résolution des pixels). Elle est également exempte des incertitudes liées aux données auxiliaires. Enfin, elle est indépendante d'un algorithme d'inversion forcément imparfait.

En résumé, deux axes principaux sont explorés : une approche plus classique d'inversion par apprentissage supervisé et l'approche non-supervisée mentionnée ci-dessus. L'approche supervisée, DRAIN, est un algorithme d'apprentissage profond de l'inversion de pluie à partir des images de température de brillance issues de quatre canaux de GMI : 36.5 GHz en polarisation horizontale et verticale et 89 GHz également en polarisation horizontale et verticale. L'approche non-supervisée explorée dans cette étude est une segmentation non-supervisée des images (sur les mêmes canaux). Les principaux défis abordés pour la méthode non-supervisée sont la labellisation, l'interprétation et la validation de la segmentation par rapport aux variables géophysiques.

Une fois démontré que la partition obtenue est bien interprétable d'un point de vue géophysique., l'étude de l'évolution temporelle des différents segments au cours du temps permet d'analyser leur sensibilité à l'évolution du climat. L'objectif est ainsi de pouvoir identifier, sans a priori, uniquement à partir des observations spatiales, les régions dans lesquelles certaines composantes du cycle de l'eau sont plus particulièrement sensibles au réchauffement climatique.

B.1 . Approche supervisée : DRAIN

En premier lieu, en traitant les mesures de TB comme des images, nous présentons un algorithme d'inversion de pluie supervisé (DRAIN). DRAIN est une régression quantile basée sur l'algorithme U-Net qui contient un encodeur, un décodeur, et une étape de « skip connections » qui sert à conserver les détails et les structures de fine échelle dans une image, ce qui est essentiel dans le cas de la

pluie. Sur le GPM-CO, nous disposons non seulement de mesures de températures de brillance dans les différents canaux évoqués précédemment mais aussi d'un produit de pluie obtenu à partir des mesures du Dual-frequency Precipitation Radar (DPR) qui couvre les 250 km du centre de la fauchée du GMI. En utilisant uniquement des images TB en entrée et des taux de pluie de DPR en sortie, DRAIN atteint une performance similaire à l'algorithme de référence de GPM. Ces très bons résultats sont liés au fait que DRAIN cherche à caractériser la fonction de transfert de l'image des TB à l'image de la pluie en prenant donc en compte la topologie des deux champs, contrairement aux algorithmes classiques qui fonctionnent en pixel-par-pixel. D'autre part, DRAIN a une performance équivalente entre les continents et les océans car l'algorithme s'appuie sur les mesures elles-mêmes pour détecter le type de surface et n'est donc pas affecté par une information a-priori qui peut biaiser les zones de transition.

Cette approche supervisée pose moins de problèmes pour l'évaluation et l'interprétation des résultats. Cependant, DRAIN ne peut être entraîné que sur le GPM-CO, car c'est le seul satellite qui permet de colocaliser la cible pluie issue du DPR et les images de TB. L'application de DRAIN aux autres satellites de la constellation nécessite une étape de l'apprentissage par transfert. En revanche, la segmentation non supervisée peut être entraînée sur n'importe quel satellite de la constellation car les données de cible ne sont pas nécessaires.

Dans la suite, DRAIN est utilisée comme une référence pour évaluer l'approche non-supervisée. Ce choix est motivé par le fait que les données en entrée pour les deux approches sont les images de TB de GMI.

B.2 . Approche non-supervisée

B.2.1 . Les méthodes non-supervisées

La famille des méthodes de segmentation non-supervisée est catégorisée dans deux branches principales (Figure B.1). Les approches « pixel-wise » classifient les pixels un par un sans prendre en compte les structures dans l'image. Un exemple de cette approche est K-means. Les approches « non-pixel-wise » utilisent des techniques de l'apprentissage qui peuvent prendre en compte l'information apportée par la topologie contenue dans les images. Ces approches sont ensuite regroupées par leur fonction de coût. Les modèles qui s'appuient sur des pseudo-labels créent des labels à partir des cartes de caractéristiques obtenues à partir de données en entrée qui servent de supervision pendant leur entraînement. Dans la deuxième approche, la fonction de coût est guidée par la différence de l'image en entrée et sa reconstruction par le réseau. Finalement, la fonction de coût peut se baser sur la comparaison des images en entrée contre l'ensemble de l'image dans la base d'apprentissage.

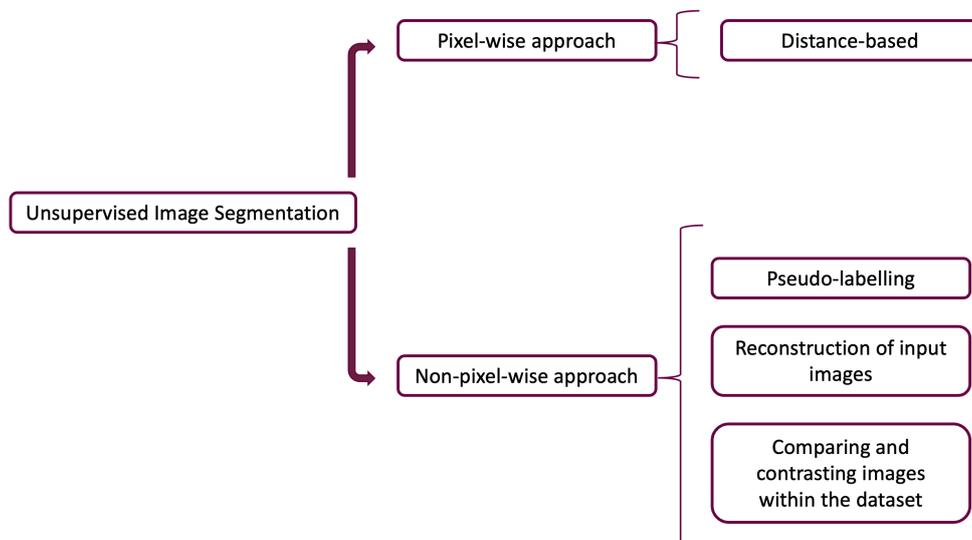


Figure B.1 – Catégories des méthodes de segmentation non-supervisée.

Dans cette étude, des modèles de segmentation non-supervisée de différentes complexités sont explorés. Nous avons testé les approches qui s'appuient sur des réseaux entièrement convolutifs (FCN) pour extraire des cartes de caractéristiques à partir de l'entrée. Premièrement, le modèle FCN le plus simple se compose de plusieurs couches de convolution pour extraire des caractéristiques (modèles M1 et M2 dans la Figure B.2). Ce modèle fait partie des modèles de type « Pseudo-labels ». La fonction de coût est basée sur les labels créés à partir des cartes caractéristiques obtenues par les couches de convolution. Pour le deuxième modèle FCN, nous approfondissons le réseau en utilisant un U-Net comme extracteur de caractéristiques afin d'augmenter la complexité (modèles U1 et U2 dans la Figure B.2). Enfin, le dernier modèle FCN utilisé comprend un réseau encodeur-décodeur avec un mécanisme d'attention. Ce dernier est un modèle de type « reconstruction des données en entrée » car la fonction de coût est construite à partir de la différence entre la sortie de décodeur et les images en entrée (modèles L1 et L2 dans la Figure B.2). En outre, afin d'avoir un élément de comparaison avec une approche par pixel, j'ai testé dans cette étude le Mini-batch K-means qui est efficace pour les grandes quantités de données (modèles K1 et K2 dans la Figure B.2).

Pour l'approche non-supervisée, l'évaluation de la performance des modèles n'est pas une tâche simple car la fonction de coût n'est pas indicative de la qualité de la segmentation. Un bon modèle de segmentation réalise une partition qui est interprétable d'un point de vue géophysique. En conséquence, pour trouver un modèle optimal, les modèles de chaque algorithme sont d'abord entraînés avec plusieurs choix d'hyperparamètres. Ensuite, l'évaluation se base sur les variables externes qui n'ont pas servi pour l'entraînement des modèles.

B.2.2 . La comparaison des modèles

Pour comparer les performances des modèles, nous proposons une évaluation externe où les variables géophysiques pertinentes sont utilisées pour le calcul des métriques de segmentation. Les classes sont attribuées aux variables géophysiques. En convertissant ces variables quantitatives (précipitation et SST) en catégories, elles peuvent être comparées à la segmentation à l'aide de différentes métriques.

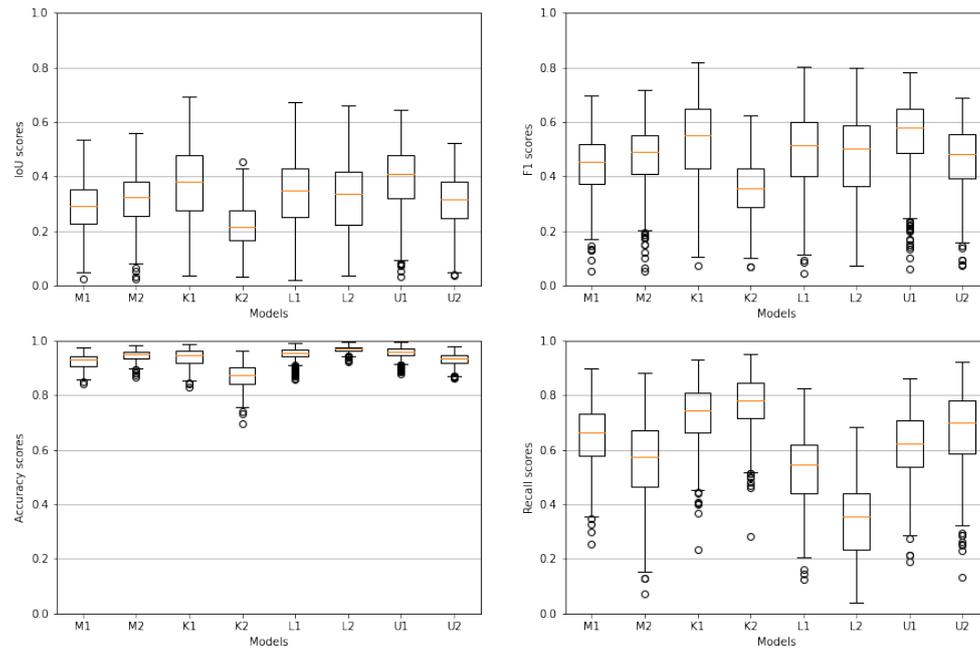


Figure B.2 – Comparaison de modèle utilisant la précipitation comme référence. Les scores ont été calculés avec 450 orbites en 2019. M1 et M2 : modèles FCN simple; K1 et K2 : modèles de K-means; L1 et L2 : modèles avec mécanisme d'attention, U1 et U2 : modèles avec U-Net (voir section B.2.1 pour plus de détail).

L'intensité de pluie de référence utilisée pour la comparaison est la sortie de DRAIN. La première étape est la labélisation des classes « pluie » pour la sortie de la segmentation non-supervisée. En raison du nombre de modèles à évaluer, la labélisation doit être automatisée. Dans un ensemble de données séparé, les pixels de la segmentation sont comparés avec les pixels correspondants de la donnée de précipitation. Ensuite, la classe pluie est définie comme une classe qui partage au moins 15% de pixels de pluie. Par la suite, les intensités de pluie sont discrétisés en pluie (1) et non-pluie (0). Finalement, la segmentation et l'intensité de pluie de référence peuvent être comparées avec les métriques d'évaluation de segmentation.

Utilisant cette démarche, deux modèles optimaux de chaque approche non-supervisée sont sélectionnés pour une comparaison complémentaire. Ces modèles

optimaux ont une performance très similaire. Figure B.2 montre les boîtes à moustache des scores des modèles sur 450 orbites. Un modèle optimal doit avoir un bon score et une faible variance de scores entre orbites. En outre, en matière de structure des précipitations, on remarque que la segmentation de K-Means est beaucoup plus bruitée que les modèles FCN.

La deuxième variable géophysique utilisée pour l'évaluation des modèles est la SST. Le produit de SST NASA (GHRSSST) obtenu à partir de GMI est la référence de la comparaison. De manière similaire à la comparaison précédente, les classes de SST sont labelisées automatiquement. Ces classes sont simplement définies comme des classes de non-pluie sur les océans. L'évaluation de modèle se base sur la compacité et la séparation de leurs classes de SST. L'indice de Calinski-Harabasz est utilisé pour évaluer si des classes de SST sont bien définies (Figure B.3). Les modèles optimaux de chaque approche ont une performance similaire.

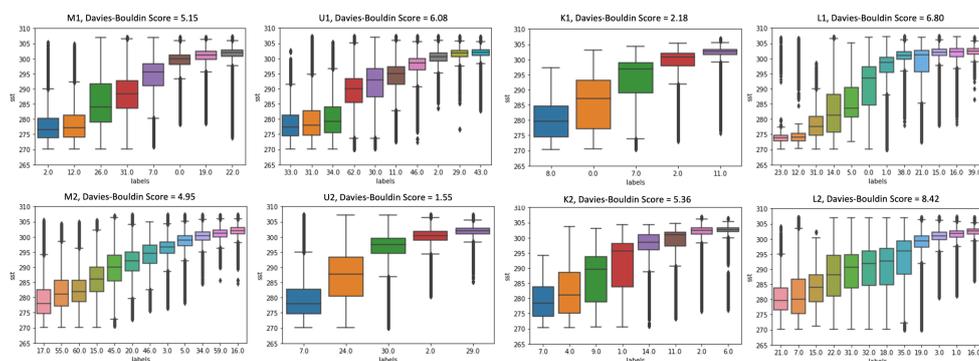


Figure B.3 – Les températures de surface de la mer (K) des classes océaniques non pluvieuses pour les meilleurs modèles. Le résultat de chaque modèle est obtenu à partir des mêmes 100 orbites.

La comparaison quantitative et qualitative avec les précipitations et la SST montre que tous les modèles optimaux de chaque approche ont une performance très similaire. Ils sont capables de segmenter les structures géophysiques dans les images de TB. Le choix d'un modèle optimal pour l'application se pose sur sa capacité à obtenir une partition sensible aux précipitations (bon score et bonne structure de précipitation), sous condition qu'elle discrimine également en fonction de la SST. Figure B.4 montre la segmentation d'un cyclone par le modèle optimal choisi. Le nombre de classes qui représente la pluie a aussi un rôle important dans le choix final. En outre, le temps d'exécution de différentes approches n'est pas de même ordre de grandeur. Pour une orbite, les modèles de mini-batch K-means prends 10 fois plus de temps (0.5 secondes) que les modèles FCN (0.05 secondes).

B.2.3 . Interprétation

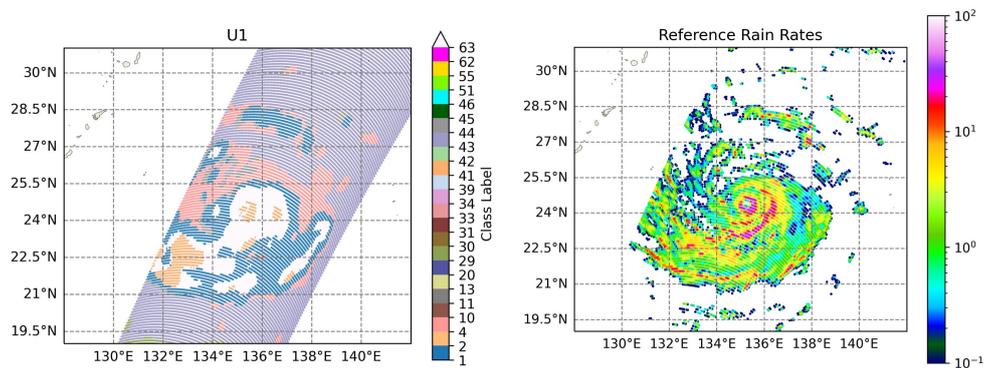


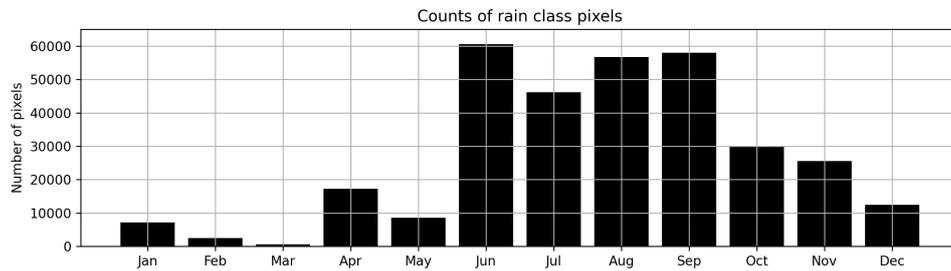
Figure B.4 – Comparaison entre la segmentation d'un modèle optimal et le taux de pluie de référence pour le typhon Namadol observé par le GMI le 16 septembre 2022.

Dans l'utilisation d'un modèle non supervisé pour étudier l'évolution des composantes du cycle de l'eau atmosphérique dans le contexte de l'évolution du climat, il est essentiel que chaque classe obtenue représente les variables géophysiques ainsi que leurs caractéristiques. Les classes de pluie identifiées sont capables de montrer la variabilité intra-annuelles de la précipitation régionale. À plus grande échelle, les classes de pluie montrent les zones de convergence intertropicale (ITCZ), tandis que les classes d'océan peuvent représenter la saisonnalité de la SST. La segmentation préserve également les structures de précipitations telles que les cyclones.

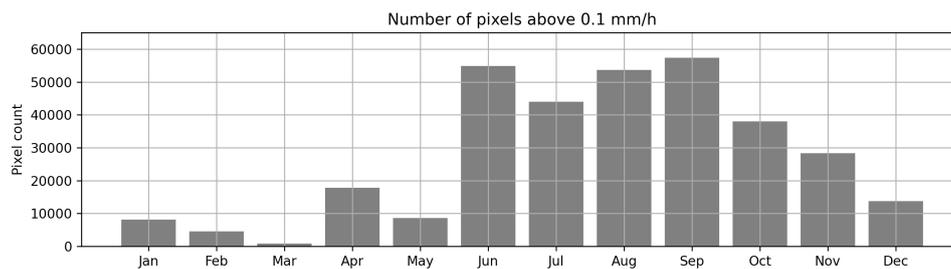
Pour l'analyse régionale, deux régions de référence définies dans le rapport du GIEC et situées entre 40°S et 40°N sont étudiées : BOB (Bay of Bengal) and WAF (Western Africa). BOB contient en majorité de l'océan tandis que WAF contient en majorité des surfaces continentales.

Les précipitations dans BOB sont principalement influencées par la mousson d'été indienne. La mousson commence vers le sud de l'Inde et s'avance vers le nord du pays. A la fin de la mousson, le mouvement s'inverse et la pluie se retire du nord vers le sud du pays. La date de l'arrivée et du départ de la mousson est essentielle pour l'agriculture et l'économie indienne. Avec les données d'un an de GMI (2019), un diagramme de Hovmöller (temps-latitudes) peut être construit à partir de l'occurrence de pixel de pluie de la segmentation. Ce diagramme montre la date approximative de l'arrivée et du départ de la mousson qui sont les mêmes que dans le rapport annuel du département météorologique indien. La Figure B.5 vérifie de plus que le nombre total de pixels de pluie par mois vu par la segmentation et par DRAIN sont comparables et que les variations annuelles sont comparables.

WAF est une région sous l'influence de la mousson de l'Afrique de l'ouest. Les précipitations se produisent entre juin et septembre. Similairement à BOB, le nombre total de pixels de pluie par mois donné par la segmentation et par



(a) Pixels de pluie par la segmentation



(b) Pixels de pluie (supérieurs à 0.1 mm/h) par DRAIN

Figure B.5 – Nombre de pixels de pluie par mois dans la région BOB en 2019. (a) Le nombre de pixels de la classe pluie a été obtenu à partir de la segmentation. (b) Les pixels de pluie au-dessus de 0,1 mm/h récupérés par DRAIN.

DRAIN se ressemble. Par contre, il y a plus de différences entre les deux pour WAF car le modèle de segmentation a plus de mal à détecter les précipitations sur les surfaces continentales que sur les océans. Le début de la mousson est signalé par le changement brusque des latitudes de l'ITCZ, passant de sa position initiale le long de la latitude 5°N à environ 10°N entre mi-juin et fin juin. Ces comportements de la mousson de l'Afrique de l'ouest peuvent être observés avec les occurrences des pixels de class de pluie dans WAF par la segmentation.

Par la suite, l'occurrence de pixels de classe de pluie sur les océans est comptée pour une grille de 0.5°x0.5°. Le comptage est normalisé par le nombre de passages du satellite par bandes de latitude car la distribution de ces derniers n'est pas du tout homogène (orbite inclinée à 65° par rapport à l'équateur). Les résultats montrent que les classes de pluie peuvent bien représenter les régions humides et sèches dans le monde. De plus, les classes de pluie de la segmentation non supervisée suivent les changements saisonniers. Par exemple, la largeur de l'ITCZ est plus grande en JJA qu'en DJF (Figure B.6). La zone de convergence de Pacifique Sud SPCZ voit plus de précipitations en DJF qu'en JJA. Dans l'océan Atlantique, la bande de précipitations se déplace vers le nord de DJF à JJA. Dans l'océan Indien, il y a plus de précipitations en JJA. Les zones humides se déplacent également vers le nord en JJA.

Le comptage de pixels de pluie régionale et globale montre une structure et un comportement saisonnier cohérents pour les classes de pluie. L'analyse globale est ici limitée aux régions océaniques à cause du fait que les modèles de segmentation développés ont du mal à distinguer les précipitations des surfaces recouvertes de glace. Celles-ci ont en effet des signatures très similaires dans les TB et malgré de nombreuses tentatives, aucune segmentation non-supervisée n'a réussi à filtrer les surfaces glacées.

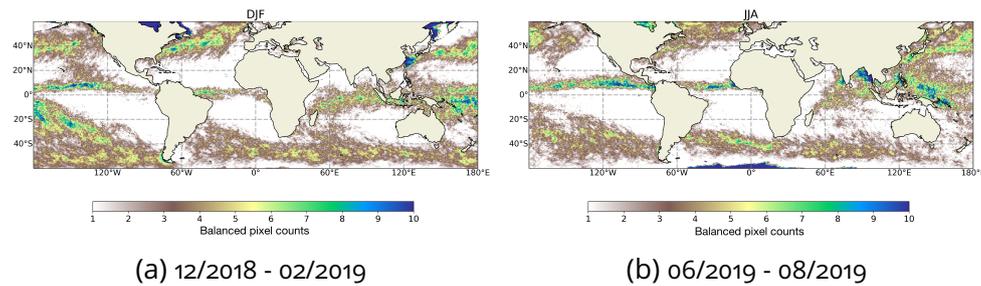


Figure B.6 – Précipitations sur l'océan en utilisant les classes de précipitation. La figure (a) montre les données de DJF (décembre 2018 et janvier-février 2019). La figure (b) montre les données de JJA (juin-juillet-août 2019).

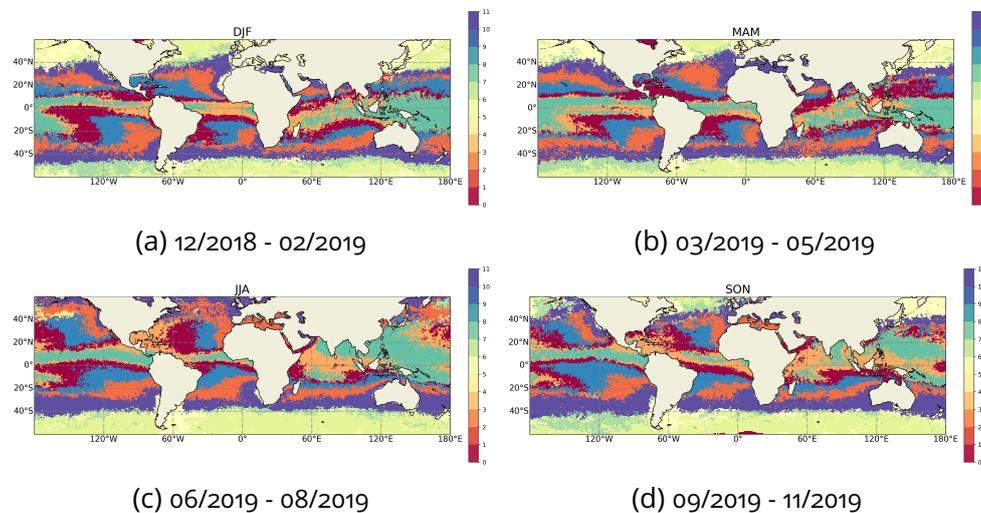


Figure B.7 – La SST représentée par les classes océaniques de segmentation pour (a) DJF, (b) MAM, (c) JJA, et (d) SON. L'étiquette de chaque pixel est obtenue par vote majoritaire des occurrences de chaque classe.

La carte saisonnière de classes d'océan est obtenue à partir de vote majoritaire dans chaque grille sur la période de DJF, MAM, JJA, et SON. Pour valider ces classes, on les compare à certaines caractéristiques saisonnières de la SST. Par exemple, La SST dans le golfe du Mexique est uniformément chaude en JJA, tandis que pendant les autres saisons, la SST est moins chaude avec un léger gradient

nord-sud. Pendant DJF et MAM, cette région est représentée par les classes de SST les moins chaudes. On observe également des gradients de température selon les saisons. Concernant la structure générale, les classes d'océan montrent les gradients par latitudes qui varie selon la quantité de chaleur reçue. Pour toutes les saisons, le bassin de l'est de l'océan Atlantique est plus chaud que le bassin à l'ouest. La bordure ouest de la plupart des océans est plus chaude en raison des effets des gyres océaniques qui transportent de l'eau plus chaude des tropiques vers des latitudes plus hautes.

Ces résultats montrent que les classes correspondant à la SST sont capables de capturer sa structure globale et sa saisonnalité. Chaque classe regroupe des SST à des niveaux similaires. Par conséquent, en fonction de l'objectif, un plus grand nombre de classes de SST permettra de représenter une gamme de températures plus étendue.

B.3 . Application sur une série chronologique de GPM

Après la validation des résultats de la segmentation, les classes obtenues sont utilisées pour analyser les données de GMI sur la période 2014-2024 et la significativité des tendances est évaluée avec le test de Mann-Kendall. En parallèle, l'analyse sur les tendances est aussi faite avec DRAIN et directement sur les températures de Brillance observées dans le canal 89 GHz en polarisation horizontale (TB 89H). Ces analyses sont réalisées à l'échelle globale entre 70°S et 70°N, sur la période 2014-2024 avec une résolution spatiale de 5°.

Notons que DRAIN dans sa version actuelle ne donne pas accès aux cumuls de pluie mais uniquement aux intensités observées. Ceci est dû au fait que GPM-CO ne passe qu'une à deux fois par jour sur la majorité des régions du globe et il est donc impossible de calculer une lame d'eau. On peut toutefois utiliser le nombre de pixels total normalisé par le nombre de pixel par latitude afin de réaliser une analyse statistique.

La TB 89H est un bon indicateur de la nature de convection de la précipitation car elle est presque proportionnelle à la quantité de glace produite dans la partie convective des nuages. Entre 2014 et 2024, on observe une tendance décroissante non significative dans les TB89H les plus froides. L'interprétation de cette tendance reste difficile. Néanmoins, certains changements régionaux sont clairement observés même s'il est impossible d'attribuer une cause. Par exemple, il y a une tendance décroissante significative au-dessus de la forêt d'Amazonie, de la région sud de l'Amérique du Nord, et du Golfe de Guinée. Avec peu de points de grille présentant des tendances significatives, on observe également ce qui semble être une diminution de l'activité convective autour de l'ITCZ.

Ensuite, le résultat DRAIN sur 10 ans de GMI montre une tendance signifi-

cative de l'augmentation du nombre total de pixels de pluie et la diminution de l'intensité de la pluie à l'échelle globale. Concernant l'intensité de pluie, au niveau régional, la majorité des points de grille présentant une tendance avec une tendance négative se trouve dans la région est de l'océan Pacifique. La région est de l'océan Atlantique, la mer d'Arabie, et l'océan Indien voit quelques points de grilles avec une tendance positive. En termes de nombre de pixel pluvieux, une augmentation de l'occurrence de pluie se trouve en majorité dans les régions humides comme le continent maritime, les trajectoires des tempêtes dans le nord-ouest du Pacifique et Atlantique, l'Amazonie, et le SPCZ, avec l'exception de l'ITCZ. Il existe de nombreuses explications possibles pour l'augmentation du nombre total de pixels de pluie. Par exemple, les cellules de pluie pourraient devenir plus grandes, durer plus longtemps ou se produire plus fréquemment. Avec uniquement les données GMI, il n'est toutefois pas possible de déterminer la cause exacte pour les raisons indiquées ci-dessus.

Pour l'approche non-supervisée, les résultats montrent une tendance croissante du nombre de pixels de classes de pluie pour trois régions du GIEC : WAF (non significatif), BOB (significatif) et SEA (significatif). Au-delà de l'analyse « par pixels », on peut étudier les objets de pluie dans la segmentation en définissant un objet comme un groupe de pixels de classe pluvieux qui sont adjacents dans l'image. Il y a globalement une augmentation non-significative de nombre d'objets pluvieux. Les tailles des objets ne montrent pas de tendance pendant cette période. Les tendances observées dans la segmentation non supervisée et dans DRAIN sont cohérentes avec la continuité de la tendance des précipitations présentée dans le sixième rapport du GIEC pour les données entre 1985-2014.

La conclusion de la nature de la tendance observée en utilisant uniquement les données GMI est difficile. Avec GMI seule, nous n'observons que des précipitations instantanées. La tendance pourrait être due à de multiples facteurs, tels que le cycle de vie, la taille et le nombre d'occurrences des événements de précipitations.

B.4 . Adaptation de domaine

Comme mentionné en partie A.3, DRAIN ne fonctionne que pour GPM-CO qui est le seul satellite de la constellation à proposer un radar qui permet d'avoir une mesure partiellement colocalisée avec le radiomètre. Un travail effectué en parallèle a consisté à tenter d'adapter DRAIN pour les autres satellites de la constellation GPM dans le but d'augmenter significativement le nombre de passages en un point du Globe donné. Chaque satellite contient un radiomètre qui mesure la température de brillance dans différents canaux hyperfréquences avec des fréquences allant généralement de 6 à 183 GHz. Cependant, il existe des différences de configuration dans ces instruments qui font varier la distribution des données d'un satellite à l'autre (choix des canaux, incidence des mesures, résolution des instruments).

Pour une approche par apprentissage statistique, ce changement de distribution des données détériore significativement la performance des modèles. Par ailleurs, les orbitographies des différents instruments sont différentes et le nombre d'observations colocalisées est trop faible pour une approche supervisée. Une étude de faisabilité sur une adaptation de domaine pour DRAIN utilisant CycleGAN est présentée dans cette étude. L'objectif est d'augmenter la performance de DRAIN sur l'instrument Special Sensor Microwave Imager/Sounder (SSM/I) sur le Defense Meteorological Satellite Program F18, sans réentraînement. Le SSM/I est assez proche du GMI en choix de canaux, incidence à la surface et résolution mais les différences sont suffisantes pour que les résultats de DRAIN soient fortement dégradés si l'on ne procède pas à une adaptation de domaine des TB SSM/I vers les TB GMI.

CycleGAN est une transformation image-à-image qui contient deux GANs (Generative Adversarial Networks) qui adaptent les images venant d'un domaine à un autre domaine et vice versa. Grâce à sa performance supérieure dans l'algorithme d'inversion DRAIN, U-Net est utilisé comme générateur pour les deux GANs.

Il y a peu d'images colocalisées entre GMI et SSM/I qui contiennent de la pluie. De plus, ces images sont mal distribuées en latitude avec des fortes occurrences aux hautes latitudes où les précipitations ne sont pas de nature très variée. Le cycle GAN permet de transformer une image SSM/I en image GMI puis de retransformer l'image GMI en image SSM/I.

Afin d'évaluer les données adaptées, l'intensité de pluie de DRAIN sur les TBs adaptées sont d'abord comparées avec le radar de précipitation DPR et la sortie de DRAIN à partir de données GMI en utilisant les intersections des orbites de GMI et SSM/I. Les résultats montrent qu'il y a une réduction significative de l'erreur quand les TBs de SSM/I sont adaptées. En outre, la structure de système de précipitation s'améliore avec l'adaptation (Figure B.8). Ensuite, l'intensité de pluie à partir de la donnée adaptée est comparé avec le produit de pluie de Météo-France. Les résultats montrent encore une augmentation de performance de DRAIN sur SSM/I grâce à l'adaptation de domaine.

Cette étude sur l'adaptation de domaine entre SSM/I et GMI montrent le potentiel d'un modèle de l'apprentissage profond pour une inversion de précipitation à partir d'observation satellitaire de TBs. Cependant, CycleGAN est un algorithme difficile à entraîner et faire converger. En outre, CycleGAN ne peut que transformer les données entre deux domaines. Il existe désormais des algorithmes plus innovants qui permettent d'adapter entre plusieurs satellites simultanément.

L'approche de l'adaptation de domaine non-supervisée est avantageuse car elle ne dépend pas de données colocalisées entre satellites et ne nécessite pas de données auxiliaires. Par conséquent, l'approche non-supervisée peut être entraînée sur n'importe quel satellite. Cette étape d'adaptation de domaine pourrait être utile

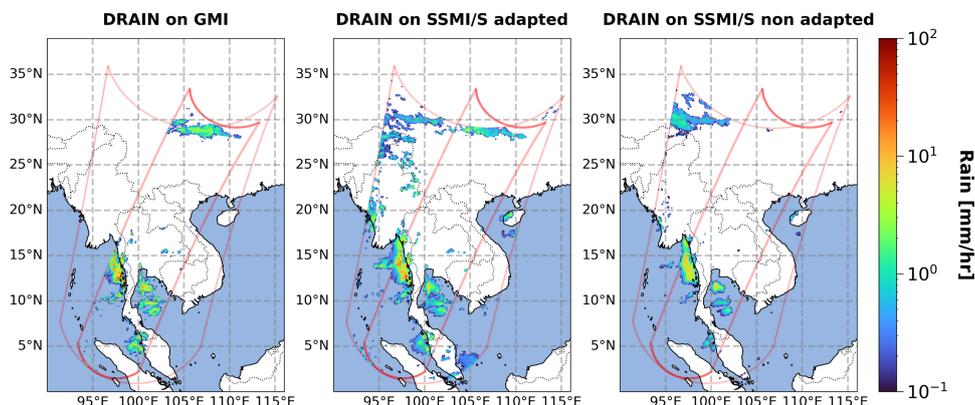


Figure B.8 – L'intensité de pluie par DRAIN à partir de TB de GMI, TB de SSMI/S adaptée, et TB de SSMI/S non-adaptée

pour homogénéiser les segmentations non-supervisées. En outre, la performance de domaine adaptation sur DRAIN est un bon indicateur de l'efficacité sur l'approche non-supervisé.

B.5 . Conclusion et perspectives

Ces résultats montrent que la segmentation non-supervisée est un outil potentiel pour extraire des informations sur le cycle de l'eau à partir d'une immense base de données d'observation spatial. La labélisation et l'interprétation des résultats avec les données de précipitation et de SST montrent que la segmentation peut retrouver la structure de variable géophysique et peut suivre la saisonnalité de chaque variable. De plus, la segmentation permet de discriminer des classes correspondant à d'autres variables géophysiques qui ne sont pas encore explorées dans cette étude.

Sur des données multi-décennales, l'évolution des classes de segmentation pourrait contenir des informations importantes sur les changements dans le cycle de l'eau. L'application de la méthode supervisée et non-supervisée sur les données de GMI entre 2014 et 2024 montre les résultats cohérents avec la continuité de la tendance de précipitation dans le rapport du GIEC. Les résultats de DRAIN montrent une tendance de l'augmentation de nombre de pixels de pluie et une diminution de l'intensité de pluie.

Les perspectives concernant le futur de ce travail consistent à améliorer la performance de la segmentation non-supervisée et l'application sur l'ensemble des satellites de la constellation de GPM.

Les modèles optimaux obtiennent des scores très similaires. L'amélioration de la

segmentation pourrait être réalisée en créant un modèle d'ensemble en fusionnant les segmentations de plusieurs modèles. Il peut être avantageux de vérifier si l'un des modèles actuels se spécialise dans une variable géophysique particulière. Par exemple, une solution possible pour aider le modèle à distinguer la pluie et la glace sur les continents est de créer un ensemble de modèles avec un spécialisé dans la pluie sur continent.

Par ailleurs, la vision par ordinateur est un domaine en constante évolution, et de nouveaux algorithmes plus performants sont régulièrement mis à disposition de la communauté scientifique. Le travail dans cette thèse a exploré en profondeur le processus d'évaluation de la segmentation non supervisée sans vérité terrain. La prochaine étape consiste à tester des algorithmes innovants potentiels en appuyant sur cette démarche d'évaluation pour trouver des améliorations supplémentaires.

En utilisant les données GMI de 2014 à 2024, on observe une augmentation significative du nombre total de pixels de pluie à la fois par les méthodes supervisées et non supervisées. Cependant, en se basant uniquement sur les données GMI, l'ensemble de données ne comprend que des observations instantanées des précipitations. Par conséquent, la nature de l'augmentation du nombre total de pixels de pluie n'est pas connue. L'utilisation des données de la constellation GPM permettra une meilleure fréquence de revisite. Pour l'approche supervisée, l'application sur d'autres satellites de la constellation nécessite une étape d'adaptation. L'approche non supervisée, en revanche, peut être entraînée sur n'importe quel satellite. Cependant, la segmentation qui en résulte aura des étiquettes différentes pour les variables géophysiques. En utilisant leurs passages, nous pourrions regrouper les classes qui représentent la même variable géophysique. De plus, l'approche d'adaptation de domaine pourrait être utilisée comme une étape intermédiaire pour harmoniser les segmentations résultantes. Il convient également de noter que l'analyse de l'évolution des quantités géophysiques basée sur l'ensemble de données satellite nécessite que la distribution des données observées reste la même. Il est essentiel de vérifier s'il y a des changements dus à la dégradation ou à l'ajustement de l'instrument.

C - Appendix : Acronyms and abbreviations

AR6	The Sixth Assessment Report of the IPCC
BNN	Bayesian Neural Networks
BOB	Bay of Bengal
CNN	Convolutional Neural Networks
CycleGAN ...	Cycle Generative Adversarial Nets
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DJF	December-January-February
DPR	Dual-frequency Precipitation Radar
DRAIN	Deep-RAIN
ENSO	El Niño-Southern Oscillation
FCN	Fully Convolutional Networks
GAN	Generative Adversarial Nets
GHRSSST ...	Group for High Resolution Sea Surface Temperature
GHz	Gigahertz
GIEC	Groupe d'experts intergouvernemental sur l'évolution du climat
GMI	GPM Microwave Imager
GPCC	Global Precipitation Climatology Centre
GPCP	Global Precipitation Climatology Project
GPM	Global Precipitation Measurement
GPM-CO	GPM Core Observatory
GPROF	Goddard profiling algorithm
IPCC	Intergovernmental Panel on Climate Change
ISM	Indian Summer Monsoon
ITCZ	Inter-tropical Convergence Zone
JJA	June-July-August
LATMOS	Laboratoire Atmosphères, Observations Spatiales
MAM	March-April-May
MLP	Multilayer Perceptron
MRF	Markov Random Field
NASA	National Aeronautics and Space Administration
NLP	Natural Language Processing
NOAA	National Oceanic and Atmospheric Administration
ReLU	Rectifier Linear Unit
SEA	Southeast Asia

SLIC Simple linear iterative clustering
SON September-October-November
SPCZ South Pacific Convergence Zone
SSMIS Special Sensor Microwave Imager/Sounder
SST Sea Surface Temperature
TB Brightness Temperature
TMI TRMM Microwave Imager
TRMM Tropical Rainfall Measuring Mission
ViT Vision Transformer
WAF Western Africa

Table des figures

1.1	The water cycle elements with storages and fluxes (Oki et al., 1999). The boxes represent the storage, while the arrows represent the fluxes. The values are estimated using data from 1989-1992. . . .	10
1.2	Linear trends in annual mean precipitation per decade (mm/day per decade) based on observational dataset (a,e), CMIP6 multi-model ensemble simulations driven by : all radiative forcings (b,f), GHG-only (c,g), and aerosol only (d,h). The cross over each grid represents a non-significant trend (Douville et al., 2023)	14
1.3	Drought hazard between 1901 and 2020 estimated by the probability of surpassing the median of global severe precipitation deficits. The data for this estimation is the precipitation data from the GPCC data from 1901 to 2010 (Caretta et al., 2022).	19
1.4	Historical (1961-2005) and projected (2051-2070) fluvial flood depth from a combination of land surface model and inundation model in (meters). The simulation of the model is based on the reanalysis of the CMIP5 GCMs.	20
2.1	Scanning geometry of GMI (NESDIS STAR, 2021)	32
2.2	GMI's pixel details.	33
2.3	GPM-CO observation of Cyclone Amphan on 16/05/2020.	34
3.1	Different types of image segmentation (Jung et al., 2022).	52
3.2	An example of a 2-D convolution kernel Li et al. (2021)	54
3.3	The application of 2D convolutional layers on an image with three layers.	54
3.4	Subscale Pixel for an image (Menick and Kalchbrenner, 2018)	55
3.5	Transformer architecture as presented in (Vaswani et al., 2017)	56
3.6	The calculation of attention in attention mechanism.	57
3.7	Architecture of the Vision Transformer (ViT) (Dosovitskiy et al., 2020)	58
3.8	The categories of unsupervised image segmentation	59

3.9	Examples of possible issues in K-means clustering. Figure obtained using the Scikit-Learn Python Machine Learning library (Pedregosa et al., 2011)	60
3.10	SLIC algorithm on medical image	62
3.11	Categorization of the non-pixel-wise approach.	63
3.12	Architecture with CNN and K-Means for unsupervised image segmentation (Barthakur and Sarma, 2019).	64
3.13	The architecture for the unsupervised image segmentation by back-propagation using SLIC (Kanezaki, 2018).	65
3.14	The architecture proposed by Kim et al. (2020) as an improvement to Figure 3.13	65
3.15	W-Net architecture (Xia and Kulis, 2017).	66
3.16	ReDO architecture (Chen et al., 2019)	67
3.17	Autoregressive Unsupervised Image Segmentation (Ouali et al., 2020)	68
3.18	STEGO architecture and prediction scheme (Hamilton et al., 2022)	69
3.19	DINO framework (Caron et al., 2021).	71
3.20	U-Net with attention mechanism (Oktay et al., 2018).	72
3.21	Preliminary result of DINO.	74
3.22	Visual representation of intersection and Union.	78
4.1	Training example from cropped satellite observation on 13/01/2017.	88
4.2	The proposed architecture. The choice of input channels $t = 6$, i.e 19H, 19V, 37H, 37V, 89H, 89V, is shown as an example.	92
4.3	The process of external comparison. The first step is to label the classes obtained from the segmentation. Those that represent precipitation and SST are compared to reference data using traditional segmentation metrics such as accuracy, recall, and F1-score. The classes that represent other geophysical variables are not used for model evaluation in this study.	94
4.4	Comparison of two of the best models from each approach. IoU, F1-score, Accuracy, and Recall score are calculated on 450 orbits observed in 2019. M1 and M2 are models from Kim et al. (2020) with simple convolution layers as feature extractors. K1 and K2 are K-Means models. L1 and L2 are our proposed models. U1 and U2 are models from Kim et al. (2020) with U-Nets as feature extractors.	97

4.5	Comparison between segmentation and reference rain rate for Typhoon Nanmadol observed by GMI on 16th September 2022.	99
4.6	Comparison between segmentation and reference rain rate for a rain structure observed by GMI on 29th May 2017.	100
4.7	Comparison between segmentation and reference rain rate for a frontal system observed by GMI on 18th August 2018.	101
4.8	An example of SST obtained from GMI brightness temperatures orbits by the GHRSSST.	102
4.9	The sea surface temperatures (K) of the non-rain ocean classes for the top models described in Section 4.3.1. The result of each model is obtained from the same 100 orbits.	103
4.10	IPCC reference region <i>Iturbide et al. (2020)</i> . The two chosen regions, BOB and WAF, are highlighted.	105
4.11	Cyclone Amphan observed by the GPM Core Observatory on May 16th, 2020. (a) shows the segmentation results. (b) shows the rain intensity retrieved by DRAIN.	106
4.12	Count of rain pixels per month inside the BOB region in 2019. (a) The count of pixels of the rain class was obtained from the segmentation. (b) The pixel of rain above 0.1 mm/h retrieved by DRAIN.	106
4.13	A Hovmöller diagram of the advancing and retreating Monsoon precipitation over India in 2019. Each date on the x-axis represents the start of the week. The count is the sum of rain class occurrences from the first day of the week to the seventh day.	107
4.14	Using data from 2019, the number of rain class occurrences for each 0.5°x0.5° gridpoint is counted for each month. The area shown is between 5°N-27°N and 6°E to 100°E. The blue line indicates the BOB region.	108
4.15	Tropical storm Fred observed by the GPM Core Observatory on August 30th, 2015. (a) shows the segmentation results, (b) shows the rain intensity obtained by DRAIN on the brightness temperature data.	109
4.16	Using data from 2019, the number of rain class occurrences for each 0.5°x0.5° gridpoint is counted for each month. The area shown is between 5°S to 20°N and 15°W to 15°E. The blue line indicates the WAF region.	110

4.17	Rain pixel counts per month in 2019 inside the WAF region. Top Figure (a) shows the counts obtained by the segmentation. Bottom Figure (b) shows the count from rain pixels above 0.1 mm/h retrieved by DRAIN.	110
4.18	Annual Precipitation over the Ocean in 2019 using pixel counts of the same rain classes as the previous sections. As there are more satellite passages near the poles than in the tropics, the pixel counts are adjusted by the total number of pixels per latitude.	112
4.19	Rainfall on the ocean using the precipitation classes. Figure (a) is the data from DJF (December 2018 and January-February 2019). Figure (b) shows data from JJA (June- July- August 2019).	113
4.20	Seasonal SST comparison. (a)-(d) Seasonal average SST using data from the ESA SST CCI and C3S reprocessed sea surface temperature analyses. (e)-(h) SST represented by the ocean classes from the segmentation. The label for each pixel is obtained by the majority vote from the occurrences of each class.	115
5.1	Mean of 1% of the coldest 89 GHz (T_{BH89}) in horizontal polarization for $5^\circ \times 5^\circ$ grids using rain pixels between 2014 and 2024.	124
5.2	Trend of the seasonal average of the 1% coldest T_{BH89} of rain pixels between 2014 and 2024. Figure 5.2a shows the trend for each $5^\circ \times 5^\circ$ grid box. Figure 5.2b shows the global trend.	125
5.3	Mean DRAIN rain intensity in a $5^\circ \times 5^\circ$ grid for GMI data from 2014 to 2024.	126
5.4	Trend in rain intensity using DRAIN data from 2014-2024. (a) shows the trend for each grid box, while (b) shows the global average rain intensity trend.	127
5.5	Average occurrences of rain pixels seen by DRAIN from 2014-2024 for grid boxes of $5^\circ \times 5^\circ$. The mean number of rain pixels in each grid is adjusted for the difference in satellite passage over each latitude.	128
5.6	Trend in rain pixel counts using DRAIN. (a) shows the trend per $5^\circ \times 5^\circ$ grid boxes. (b) shows the trend in the global seasonal average annually.	129
5.7	The IPCC reference region (Iturbide et al., 2020).	130

5.8	Monthly rain pixel counts from 2015 to 2023. The pixels of the rain classes are counted for the IPCC regions : (a) WAF, (b) SEA, and (c) BOB. The dashed lines show the trend in the data. 'No trend' represents a non-significant trend, while 'increasing' represents a significative increasing trend.	132
5.9	Annual rain pixel counts from 2015 to 2023. The pixels of the rain classes are counted for the IPCC regions : (a) WAF, (b) SEA, and (c) BOB. The Mann-Kendall test shows no significant trends. . . .	133
5.10	Rain intensity per class by comparison to DRAIN data. The order of the range of rain intensity in each class from highest to lowest is class 4, 1, 63, and 41 for both regions.	133
5.11	Monthly pixel counts for each rain class from 2015 to 2023 : (a) WAF, (b) SEA, and (c) BOB. The order of rain intensity from lowest to highest is Class 4, 1, 63, 41. The Mann-Kendall test shows a significant increasing trend in SEA for Class 4.	134
5.12	Monthly rain object counts per class (a) and monthly average object size for BOB (b). There is no significant trend based on the Mann-Kendall test.	136
5.13	Monthly rain object counts per class (a) and monthly average object size for BOB (b). There is no significant trend based on the Mann-Kendall test.	137
5.14	Mann-Kendall per each grid box of 5° by 5° for each rain class. In order of increasing rain intensity representation are Class 4, Class 1, Class 63, and Class 41. The solid-colored grid points are statistically significant based on the Mann-Kendall test.	138
5.15	Seasonal mean of SST class pixel count for GMI data from 2015-2023.. Figure (a) shows the average pixel count for DJF, MAM, JJA, and SON for each class. Figure (b) shows the ranges in SST per class when compared to GHRSSST data (Remote Sensing Systems, 2017). Figure (c) shows the sum of regrouped SST classes based on their ranges of temperatures. The Mann-Kendall test does not show any significant trend.	139
B.1	Catégories des méthodes de segmentation non-supervisée.	182
B.2	Comparaison de modèle utilisant la précipitation comme référence. Les scores ont été calculés avec 450 orbites en 2019. M1 et M2 : modèles FCN simple ; K1 et K2 : modèles de K-means ; L1 et L2 : modèles avec mécanisme d'attention, U1 et U2 : modèles avec U-Net (voir section B.2.1 pour plus de détail).	183

B.3	Les températures de surface de la mer (K) des classes océaniques non pluvieuses pour les meilleurs modèles. Le résultat de chaque modèle est obtenu à partir des mêmes 100 orbites.	184
B.4	Comparaison entre la segmentation d'un modèle optimal et le taux de pluie de référence pour le typhon Nanmadol observé par le GMI le 16 septembre 2022.	185
B.5	Nombre de pixels de pluie par mois dans la région BOB en 2019. (a) Le nombre de pixels de la classe pluie a été obtenu à partir de la segmentation. (b) Les pixels de pluie au-dessus de 0,1 mm/h récupérés par DRAIN.	186
B.6	Précipitations sur l'océan en utilisant les classes de précipitation. La figure (a) montre les données de DJF (décembre 2018 et janvier-février 2019). La figure (b) montre les données de JJA (juin-juillet-août 2019).	187
B.7	La SST représentée par les classes océaniques de segmentation pour (a) DJF, (b) MAM, (c) JJA, et (d) SON. L'étiquette de chaque pixel est obtenue par vote majoritaire des occurrences de chaque classe.	187
B.8	L'intensité de pluie par DRAIN à partir de TB de GMI, TB de SSMI/S adaptée, et TB de SSMI/S non-adaptée	191

Liste des tableaux

3.1	Summary of unsupervised approaches presented	73
4.1	Hyper-parameters for the simple FCN model proposed by Kim et al. (2020). The total number of models tested is 12. These models are labeled M_i	90
4.2	Hyper-parameters for the FCN model proposed by Kim et al. (2020) with U-Net as the feature extractor. The total number of models tested is 20. These models are labeled U_i	90
4.3	Our proposed architecture. Conv2d(K) means 2D convolution with a KxK kernel size.	92
4.4	List of hyper-parameters tested for our proposed model. The total number of models tested is 162. These models are labeled L_i	93
4.5	Kmeans hyper-parameters tested. The total number of models is 18. These models are labeled K_i	93
4.6	Top two models for the algorithm proposed by Kim et al. (2020). M1 and M2 are the models with simple feature extractors. U1 and U2 have U-Net as their feature extractor.	96
4.7	The hyper-parameters of the top two models of our proposed architecture.	97
4.8	The hyper-parameters for the top two K-Mean models.	97
4.9	Average and standard deviation of F1-scores for the 450 orbits shown in Figure 4.4.	98