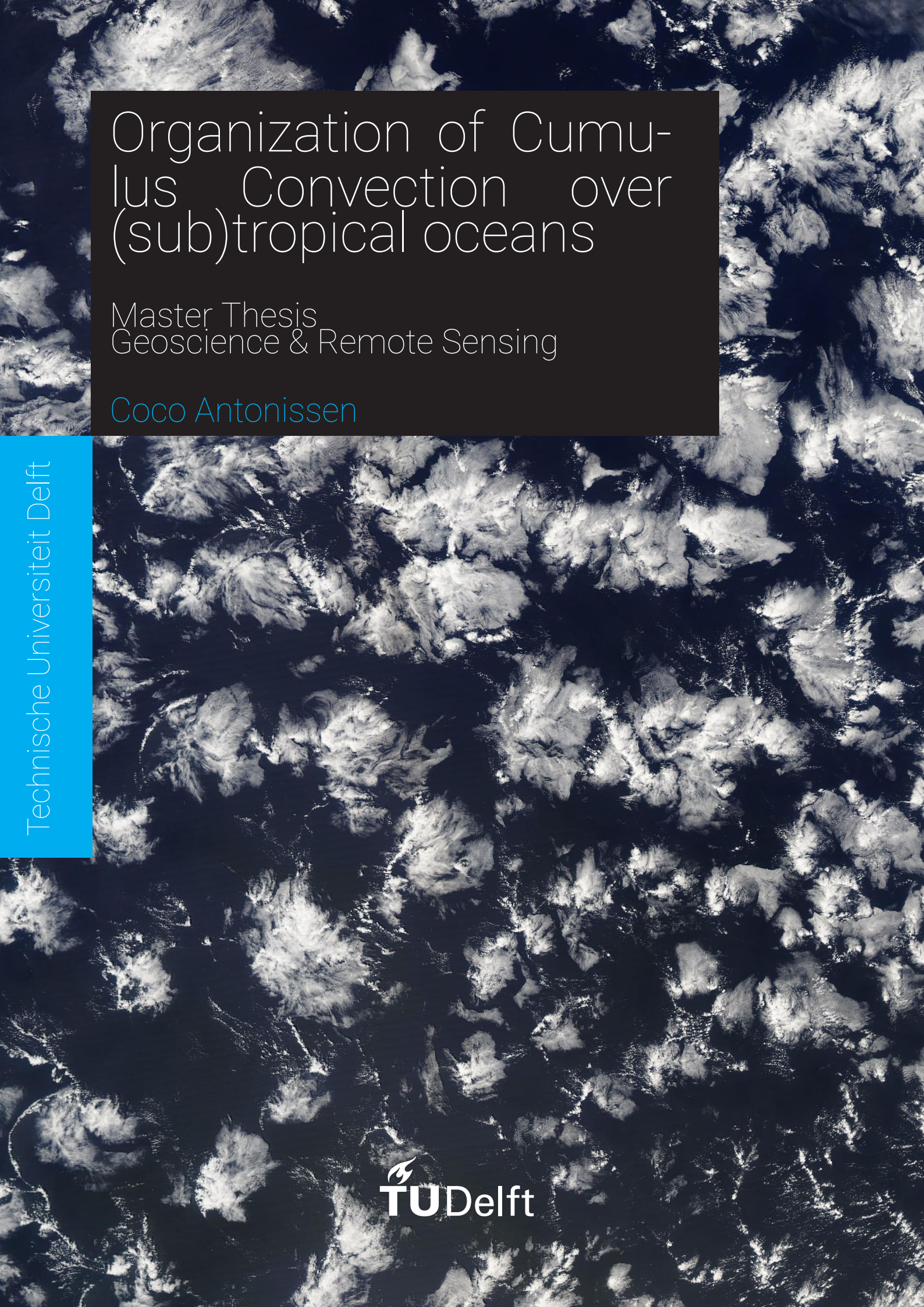


Organization of Cumulus Convection over (sub)tropical oceans

Master Thesis
Geoscience & Remote Sensing

Coco Antonissen

Technische Universiteit Delft

TU Delft

Cover: MODIS (AQUA) corrected reflectance (true colour) image of a cloud field spanning approximately 10° – 23°*N and* 49° – 59°*W (East of Barbados) on February 9, 2017.* (`https://worldview.earthdata.nasa.gov`)

Organization of Cumulus Convection over (sub)tropical oceans

by

Coco Antonissen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday April 24, 2019 at 10:00 AM.

Student number:	4219376	
Project duration:	May 1, 2018 – April 24, 2019	
Thesis committee:	Prof. dr. A.P. Siebesma,	TU Delft
	Dr. L. Nuijens,	TU Delft
	Dr. S.L.M. Lhermitte,	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Preface

Working on this thesis has been a great pleasure. I would like to thank my graduation committee for introducing me to the topic, providing useful feedback and having interesting discussions. Working on my thesis brought many ups, but also downs. Feedback, discussions and motivation from you was of great help and above all very inspiring. All of this made it possible for me to not only finish my thesis, but also to finish it with a positive feeling. Thank you for that!

Furthermore I would like to thank my parents, for supporting me not only during my thesis but also throughout the years. I feel I am lucky to have you not just as parents, but also as friends and sparring partners. To me being *"home"* means having fun, having interesting conversations and getting some rest. My years in university would have been very different without your support and I am very glad I could enjoy the opportunities I got because of that.

Last but not least I would like to thank Jens, my friends and fellow students for all the great times we have had so far. Studying is nice but having good friends around you makes it great.

Coco Antonissen
Delft, April 2019

Abstract

The climate on Earth is changing rapidly. In order to take the right measures it is important to know with which magnitude and rate the warming of the climate continues in the future. However, climate models currently predict different scenarios. The differing in response of shallow cumulus clouds to a warming climate explains the largest part of the spread of climate sensitivity in these models. It is known that clustered clouds have a different climate feedback than randomly distributed clouds and that we can expect more clustered cloud fields in a warmer climate. It is however unknown how these clouds respond and what the exact difference in climate feedback is.

Future research must provide us with more insights and eventually a better understanding of the response of shallow cumulus clouds to warming. However, in order to investigate this an objective measure of the degree of cloud organization is necessary, something that is currently lacking.

In this thesis a dimensionless combined organization measure is introduced that can be applied to a large range of cloud field organizations on different field sizes. This is a combined measure, as it makes use of the existing organization index I_{org} and the size of clouds and clear sky areas in the cloud field. A different approach on calculating I_{org} was introduced taking into account the sizes of the clouds, resulting in an useful and more realistic values from this parameter.

The results of the combined organization measure were compared to a visual inspection of 557 cloud fields with dimensions of 10° × 10°, which showed promising results. Sub-fields of 5° × 5° and 2.5° × 2.5° were also analysed, which showed that the combined organization measure could often be applied on smaller scales as well but yielded some problems when areas contained either no or little clouds or some very large ones.

It was concluded that the newly developed method is an improvement of the already existing method, providing scientists with a better and more reliable index to quantify the degree of cloud organization.

Contents

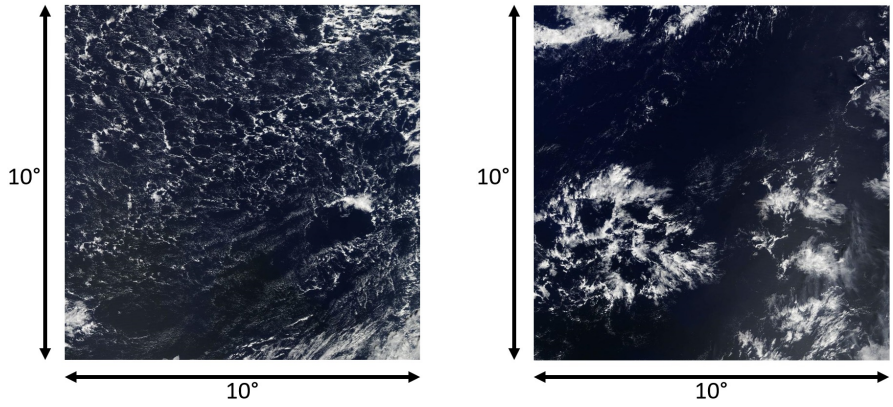
List of Figures

List of Tables

1 Introduction

All over the world, clouds appear in various shapes and sizes. Probably everyone can acknowledge that, as sometimes the sky is dominated by these typical "woolly" cumulus clouds and other days, when stratus clouds dominate the sky, it looks like the earth is covered by a blanket. Individual clouds can easily be distinguished from the earth. However, it is challenging to say something about the cloud *organization*, where we look at the distribution of clouds in a large area rather than identifying individual clouds. Using satellites we can observe large areas from above, and thus have a better overview of what a cloud field looks like. Imagery from satellites therefore offers great opportunities for science.

Figure 1.1 shows two cloud fields with different organizations. Both fields have the same cloud cover (24%), but the spatial distribution of the clouds is different. The cloud field on the left looks rather randomly organized: clouds are distributed over the entire area. This is not the case for the cloud field on the right, where clouds are clustered and large areas containing little clouds can be identified.



(a) Example of approximately evenly distributed clouds.

(b) Example of a clustered cloud field.

Figure 1.1: Satellite true colour images (MODIS) showing two cloud fields with cloud cover = 24%.

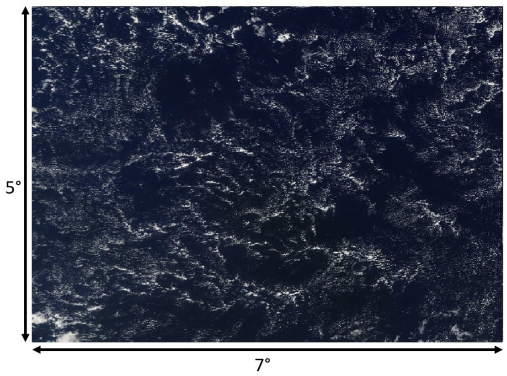
We are interested in studying the organization of a cloud field, as this greatly improves our understanding of our current and future climate. Climate change has been a so-called *hot topic* for a period of time now and modelling it is important for many reasons, but is also very challenging. Advancements in computer power have greatly improved our modelling capabilities in terms of sophistication and comprehensiveness. However, different models still show a large spread in the degree of warming as a response to a doubling of CO_2. This uncertainty in climate sensitivity can largely be traced back to the response of especially marine low level clouds to a changing environment (Bony and Dufresne, 2005; Vial et al., 2013). In particular, our understanding of the variability of shallow cumulus clouds in the trade wind region is limited and most likely because of that, low-level cloud feedback shows a large spread for different models in this region (Brueck et al., 2015). Increasing computer power gives us the opportunity to create sophisticated climate models, but

in order to improve them we need to understand our current climate and in particular what the role of clouds in our climate is.

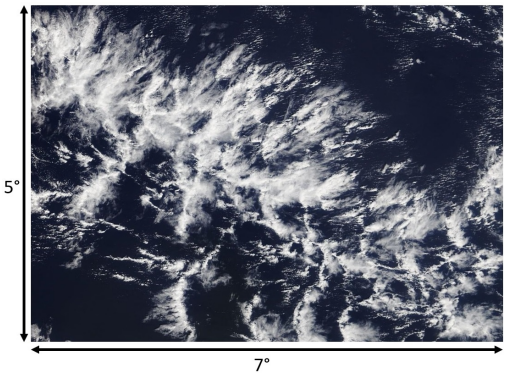
Our limited understanding of how clouds, circulation and climate interact is discussed by Bony et al. (2015). This paper states that four main questions should be answered in future research in order to get a better understanding of the impact of clouds on climate. One of these questions is *what role convective aggregation plays in climate*. Insights from both field and numerical studies showed that aggregation of clouds, also referred to as clustering, has a significant different feedback on the climate compared to a randomly organized cloud field (Bretherton et al., 2005; Cahalan et al., 1994; Tobin et al., 2012). Therefore, the climate feedback of these clouds highly depends on the way they are organized, which illustrates the interest in cloud organization.

The cloud fields in Figure 1.1 not only look different from space, they also have a different climate climate feedback. The images are true colour images created with satellite data from the MODIS instrument. Satellites provide us with frequent and high-resolution imagery and are therefore very useful for assessing cloud organization. MODIS imagery is available free of charge, provides data twice a day and contains a cloud mask with 1 km resolution. It therefore offers a great opportunity to assess the organization of cloud fields on a twice a day basis. The question however is how to assess the degree of cloud organization.

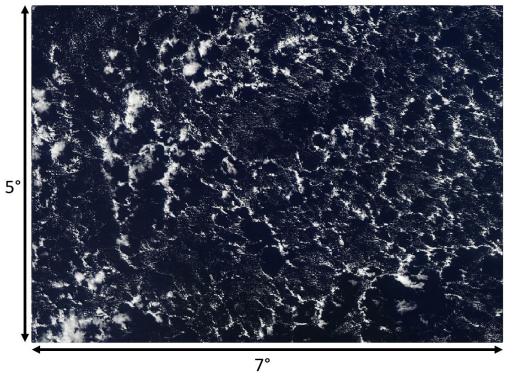
Recently an effort has been made, hosted by the International Space Science Institute, to *subjectively* investigate the patterns of mesoscale organization in the trade wind region. Ten years of MODIS imagery was analysed and resulted in four recurring patterns named Sugar, Fish, Gravel and Flowers, displayed in Figures 1.2a to 1.2d. The captions contain a definition as formulated by the group of scientists working together on this project (Stevens et al., 2019).



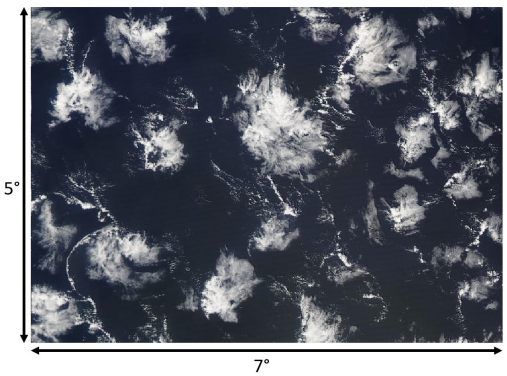
(a) Sugar: Dusting of very fine scale clouds, perhaps patterned by a larger scale flow, but with little evidence of self-organization (cold-pools or gust fronts).



(b) Fish: Large-scale (100s of km) skeletal like networks of clouds separated from each other, or other cloud forms, by well defined cloud free areas and sometimes accompanied by a stratiform cloud shield.



(c) Gravel: Brighter clouds along meso-β (20 km to 100 km) lines or arcs defining randomly interacting cells with intermediate granularity and little evidence of accompanying stratiform cloud veils.



(d) Flowers: Large-scale (50 km to 200 km) stratiform cloud features, typically with reduced reflectivity and appearing in bunches (bouquets) well separated from one another by regions devoid of clouds.

Figure 1.2: Subjective cloud organization patterns. All cloud fields span a region of 5° × 7°. Definitions of these patterns as stated by the group of scientists are provided in the caption.

Although they successfully and relatively consistently identified different cloud patterns, the results are likely to be biased as the patterns were identified in a subjective way. To obtain consistent results, it is therefore of uttermost importance to objectively quantify the degree of organization of a cloud field (Seifert and Heus, 2013).

Objective methodologies, like $SCAI$ (Simple Convective Aggregation Index) (Tobin et al., 2012) and the Organization Index I_{org} (Seifert and Heus, 2013; Tompkins and Semie, 2017; Weger et al., 1992), are formulated and applied in earlier studies (see Sections 2.2.3 and 2.2.4). However, the results of these methods are not as desired. Therefore this thesis focusses on finding a novel approach to assess the cloud organization successfully in an objective way.

1.1. Research Questions

As argued in this chapter, the largest uncertainty in climate models nowadays is the response of shallow cumulus clouds to a changing environment. As it is important for policy makers to know with high confidence levels what our future climate will look like and thus to what extent measures should be taken, we need a better understanding of the climate feedback of these cloud fields. Understanding the evolution of cloud organization is important for this. Satellite imagery provides us with large amounts of useful data. In order to analyse this imagery in a systematic and objective way, a measure must be found which quantifies the degree of cloud organization. The main focus of this thesis therefore lies in finding this measure. Subsequently, this measure is applied to multiple years of satellite data to investigate the temporal and spatial scales of shallow cumulus cloud organization.

Two main questions are formulated. Sub-questions are used to find the answers to the main questions.

1. **How can shallow cumulus cloud organization objectively be characterized using satellite imagery?**
   (a) How is shallow cumulus cloud organization defined?
   (b) Which parameters have been used in the past to characterize shallow cumulus cloud organization?
   (c) What are the most important parameters to measure shallow cumulus cloud organization?
   (d) Can these parameters be linked to a subjective evaluation of the organization?
   (e) How robust are these parameters when different thresholds for data selection are applied?
   (f) How robust are these parameters when applied to different field sizes?
2. **What are the temporal and spatial scales of shallow cumulus cloud organization?**
   (a) On which temporal scales can trends be detected?
   (b) How fast does organization change?
   (c) On what spatial scales do parameters vary?

1.2. Methodology & Outline

To answer these questions satellite data from the MODIS instrument aboard two different satellites (Terra and Aqua) will be used. An appropriate area of 10° × 10° is selected (Fig. 1.3) and multiple years of data (the months December, January and February between December 2010 and February 2018) are analysed. This area is appropriate because shallow cumulus clouds are abundant here, which will be explained further in Section 2.1. More specifically also because a large field campaign, EUREC4A, will be held in this region in February 2020 (Bony et al., 2017). Only the months December, January and February are analysed because the area of interest is dominated by shallow cumulus convection in stead of deep convection during this period. This is due to a shift of the ITCZ (Inter Tropical Convergence Zone) to the southern hemisphere (King et al., 2013). The measure to assess the degree of cloud organization is created from theory and a visual assessment of the success of using certain parameters. Eventually, based on this visual judgement, the most effective measure is selected.

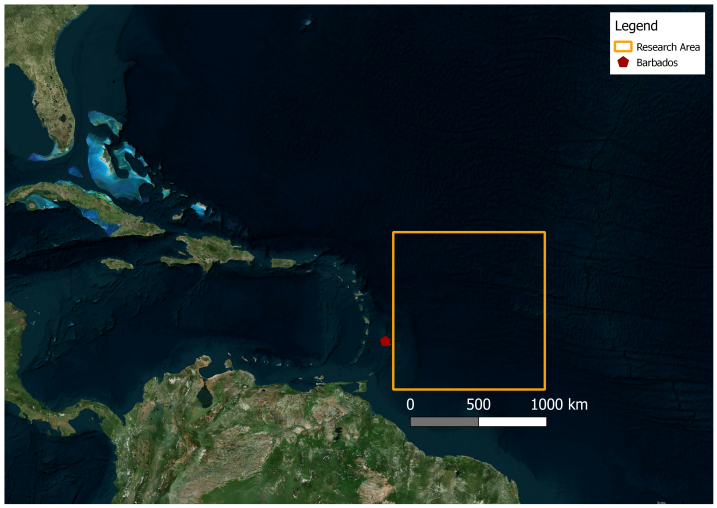


Figure 1.3: The area of interest, indicated by the orange box, is located close to Barbados and spans 10° × 10° between 10° – 20°N and 49° – 59°W.

This thesis contains six chapters. After this introduction Chapter, theory on shallow cumulus clouds, cloud fields and cloud characterization is discussed in Chapter 2. Chapter 3 discusses the MODIS instrument, its cloud mask and the pre-processing of the data. The methodology is explained in Chaper 4. The results and discussion are subsequently discussed in Chapter 5 and finally the conclusions are presented in Chapter 6.

2

Theory

This chapter is split up in two parts. First a short theoretical background on shallow cumulus clouds and the trade wind region is provided. The second part discusses existing methods to characterize cloud organization.

2.1. Shallow Cumulus Clouds

As mentioned already in the introduction the main goal is to be able to find a quantitative measure of the organization of shallow cumulus clouds in the trade wind region. This section discusses the definition of shallow cumulus clouds and why they are so abundant in the trade wind region.

Shallow Cumulus Clouds

The World meteorological Organization (WMO) has created an International Cloud Atlas, which describes the classification system for clouds (Howard, 1803). An example of cumulus clouds is shown in Figure 2.1. The Cloud Atlas defines shallow cumulus clouds as described here:

"Detached clouds, generally dense and with sharp outlines, developing vertically in the form of rising mounds, domes or towers, ofwhich the bulging upper part often resembles a cauliflower. The sunlit parts of these clouds are mostly brilliant white; their base is relatively dark and nearly horizontal" (Howard, 1803).

Figure 2.1: Cumulus Clouds (Burt, 2009).

Cumulus clouds are formed via atmospheric convection, which means that air warmed by the surface starts to rise due to instability. The temperature of the rising air drops, which results in a larger relative humidity. Once the relative humidity reaches 100% water vapour condenses (under the condition that nuclei like dust are present in the atmosphere) and forms cumulus clouds. Depending on the temperature profile of the atmosphere and the presence of any inversions, cumulus clouds vary in depth. Shallow cumulus clouds are cumulus clouds with a depth up to approximately 5 km. A warm ocean surface in combination with a moderate or weak large-scale subsiding motion results in the majority of marine low-level clouds being shallow cumulus (Nuijens et al., 2014).

The climate feedback of shallow cumulus clouds depends on different properties of the cloud field. An

important climate feedback is radiation, as a changing cloud cover affects the reflectance of shortwave radiation from the sun. However, reflectance does not only depend on cloud cover fraction but is also affected by the horizontal distribution of clouds (Cahalan et al., 1994). This can also be referred to as *cloud organization*.

As mentioned before, shallow cumulus clouds are ubiquitous over tropical oceans and therefore their radiative properties have a significant impact on the Earth's radiation budget. It has been shown that their differing in response to a warming climate explains the largest part of the spread of climate sensitivity in climate models (Bony and Dufresne, 2005; Bony et al., 2004; Boucher et al., 2013; Medeiros et al., 2008, 2015; Vial et al., 2013; Webb et al., 2006), as was already discussed in the Introduction. Shallow cumulus clouds span large parts of the world's oceans and are common over land during the day in periods when fair weather prevails. Particularly, shallow cumulus clouds are abundant in the trade wind region.

Trade Wind Region

The trade wind region is defined as the region roughly between the Inter Tropical Convergence Zone close to the equator and delimited by the Tropic of Cancer in the north and the Tropic of Capricorn south of the equator. Trade winds are the prevailing winds at the surface near the equator, directing from northeast (southeast) to West in the northern (southern) hemisphere. The winds strengthen during the boreal winter and act as the steering flow for tropical storms.

Atmospheric Circulation

The trade winds are a result of the atmospheric circulation, which is a global movement of air mass. As more sunlight reaches the earth at the equator than at the poles, energy is unevenly divided over our planet. Driven by the temperature gradient at the surface, air mass flows from the equator towards the poles by rising at the equator and subsequently flowing polewards at a height of approximately 17 km because it cannot ascent further than the tropopause. However, due to the rotation of the Earth air can not flow directly from the equator towards to poles. The atmospheric circulation is instead split up in three cells on each hemisphere: the Hadley cell, the Ferrel cell and the Polar cell. Instead of flowing all the way to the poles near the tropopause, air descends at the subtropics and flows back towards the equator at the surface. This is schematically visualized in Figure 2.2. Due to the rotation of the earth, the near surface return flows to the equator are deflected towards the West, resulting in north-easterly trade wind in the subtropical Northern Hemisphere and south-easterly winds at the subtropical Southern Hemisphere. These winds are denoted with green arrows in the Figure. For this reason the region is referred to as the *trade wind region*.

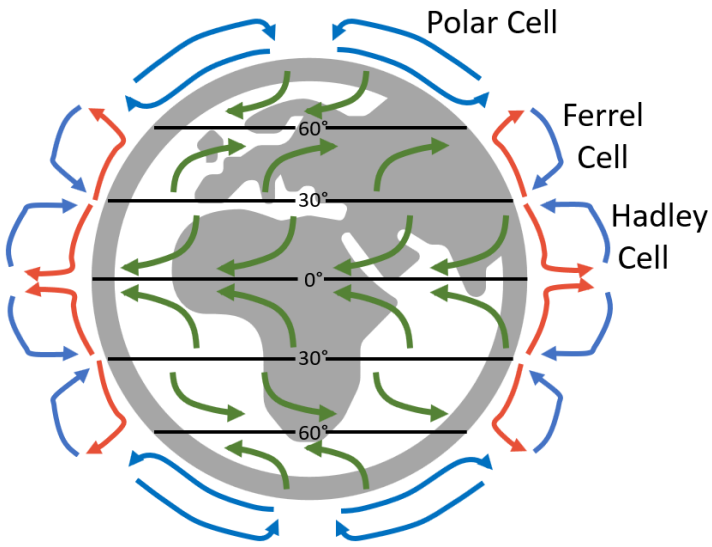


Figure 2.2: The atmospheric circulation model showing the different cells on each hemisphere. Red and blue arrows indicate respectively warm and cold air masses. The green arrows denote the surface winds.

2.2. Cloud Characterization

This section describes four basic cloud field parameters and existing methods designed specifically to characterize cloud organization like $SCAI$ and I_{org}. This section also provides a short explanation of their practical implications. The way they are implemented in this thesis is described in Chapter 4.

2.2.1. Basic Cloud Parameters

In this section the basic parameters describing clouds are discussed. These basic parameters are the cloud cover, the average cloud size, the cloud number density and the average nearest neighbour distance. The parameters provide useful and simple method to have an initial idea about the clouds in a field. However, they do not necessarily provide information about the organization of the cloud field.

Due to the fact that the area of interest is close to the equator, often some data is missing= which will be explained in Chapter 3. Furthermore, as a finite area is assessed, clouds will sometimes directly be adjacent to the edge of the area or to the no-data area. This may have consequences for the results of the basic cloud characterization analysis and will be discussed in this section as well.

Figure 2.3 contains an example cloud field.

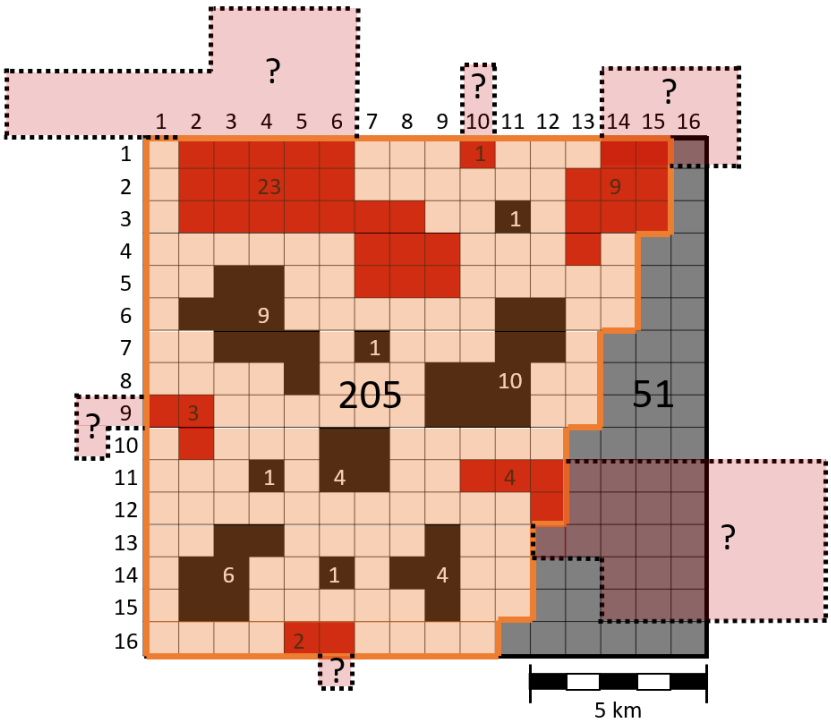


Figure 2.3: Example of a cloud field. Grey pixels denote no-data area, black pixels denote clouds fully within fthe field and red pixels represent clouds adjacent to the no-data area or edge of the field. The red areas outside the area denote possible realizations of the clouds adjacent to the edge or no data area. These cloud however can have any shape.

In this example cloud field the grey pixels denote no-data area and the orange highlighted area does contain data. The red and black pixels denote cloud pixels. In particular, the red pixels show clouds adjacent to the no-data area or the edge of the field. The red semi-transparent areas surrounded by dotted black lines represent possible cloud realizations. However, the values of these pixels are unknown and no assumptions on the sizes of the complete clouds at the edges of the field can be made. This yields some statistical issues, but they reduce when the domain size increases with respect to the cloud size.

The sizes of the two areas (containing data and no-data) and the individual clouds are written down in the Figure. The question marks in the red semi-transparent areas show that these sizes are unknown.

The theory on the basic cloud characterization will be on the basis of this field.

Cloud Cover

One of the first characteristics of a field that is usually discussed is the *cloud cover*, also referred to as the *cloud fraction* (e.g. by Brueck et al. (2015)). This parameter is defined as the percentage of the field that is covered

by clouds, given by:

$$CC = \frac{\sum_{i=1}^{N} a_i}{A} \times 100\% \tag{2.1}$$

Here CC is the cloud cover, a_i is the area representing cloud i and A is the total area of the field.

A completely clear sky field will thus have a cloud cover of 0%, since there are no pixels representing clouds, and a field that is completely covered in clouds has a cloud cover of 100%. However, these extremes are not common and cloud cover usually has a value in between these values, although that also depends on the size of the field.

The field shown in Figure 2.3 is used here as an example to calculate cloud cover. This cloud field consists of 16 x 16 = 256 pixels, but only 205 pixels contain data. The total area of the field A thus equals 205 pixels. The field contains in total 79 cloud pixels (black and red pixels). Under the assumption that all pixels represent equal-size areas (1 km^2 here, although not of importance for the cloud cover) and making use of Equation (2.1), the cloud cover of this field is found to be: $\frac{79}{205} \times 100 = 38.5\%$.

It was mentioned before that the so-called "cut-off" of the clouds at the edges and the no-data area influences the basic parameters. However, for the cloud cover it can be assumed that this effect is negligible, as the cloud cover does not depend on the number of clouds, their sizes or their geometrical positions. Therefore all clouds can always be taken into account when calculating the cloud cover.

Number of Clouds

Another parameter used to describe a cloud field is the number of clouds that a field contains. In order to compare this value to fields with different fractions of missing data, the number of clouds is usually presented as the number of clouds / unit area, or in other words the *cloud density*. The unit area is defined as the complete area that is evaluated, regardless of whether pixels contain data or not. For the example field in Figure 2.3 this means that the unit area equals $16 \times 16 = 256$ km^2. The cloud density is given by the following equation:

$$N_{clouds} = \frac{n}{A} \times A_{unit} \tag{2.2}$$

Here N_{clouds} is the number of clouds per unit area, n is the number of clouds, A is the area of the field that contains data and A_{unit} is the unit area.

If this equation is applied to the example field in Figure 2.3, we find that the number of clouds is 15, the area containing data is 205 km^2 and the unit area is 256 km^2. Therefore the cloud density of this field is equal to $\frac{15}{205} \times 256 = 18.7$ clouds/256 km^2.

However, since it is unknown how large the clouds adjacent to the edges and no-data areas are and thus what their spatial extent outside this area is, these clouds must be removed from the data. Otherwise it is likely that the cloud density gets overestimated, as the number of clouds found in the field actually cover a larger area than that is evaluated. This was sketched with red-semitransparent areas in Figure 2.3. If these possible cloud extensions would be true, the number of clouds would remain the same but the area A would be larger, and thus the cloud density would be smaller.

In order to avoid such biases, the clouds adjacent to the edge or no-data area are removed. Their sizes are subtracted from the total area A as well, otherwise the cloud density would be biased in the opposite direction. Doing so for the example field in Figure 2.3 leaves us with 9 clouds. The total area A is reduced to $205 - 23 - 1 - 9 - 3 - 4 - 2 = 163$ km^2. The unit area remains the same, and thus we obtain a cloud density of $\frac{9}{163} \times 256 = 11.3$ clouds / 256 km^2. This value is much lower than the value found when the red clouds are taking into account, but as argued above it also is a more realistic value.

Average Cloud Size

This parameter gives the average size of the clouds in the field and is given by the following equation:

$$\overline{CloudSize} = \frac{\sum_{i=1}^{N} a_i}{n} \tag{2.3}$$

Where $\overline{CloudSize}$ denotes the average cloud size, $\sum_{i=1}^{N} a_i$ is the sum of all cloud areas and n is the number of clouds.

Taking the field in Figure 2.3 as an example again, in which all pixels represent a similar area of 1 km^2, the average cloud size is 5.20 km^2. However, similar as for the cloud density, the clouds adjacent to the edge and

no-data area should be removed since their size is unknown. Therefore, the number and the sum of all sizes of the clouds denoted by black pixels must be used. In this example, n then equals 9 and the sum of these 9 cloud areas is 37 km^2. This gives an average cloud size of $\frac{37}{9} = 4.11$ km^2. This is smaller than the average cloud size computed from all clouds in the field. This can also be expected from theory as the *chance* of clouds being directly adjacent to the edge or no-data area is larger for larger clouds. Removing these usually relatively large clouds from the dataset will then results in a smaller average cloud size.

Average Cloud Nearest Neighbour

Each cloud has a geometrical centroid. The nearest neighbour of a cloud is defined as the distance to the nearest centroid of another cloud. Therefore the nearest neighbour is found by first finding the geometrical centroids of all cloud, and subsequently calculating its distances to the other geometrical centroids. The shortest distance is the nearest neighbour of this cloud. The average nearest neighbour of a field is found by taking the average of all these nearest neighbours. The geometrical centroids have an x- and a y-coordinate. These coordinates are given by:

$$C_{x,i} = \sum_{j=1}^{n} \frac{(x_j a_j)}{a_i} \tag{2.4}$$

$$C_{y,i} = \sum_{j=1}^{n} \frac{(y_j a_j)}{a_i} \tag{2.5}$$

Here $C_{x,i}$ and $C_{y,i}$ are respectively the x- and y-coordinates of the geometrical centroid of cloud i. Furthermore x_j and y_j are the x and y coordinates of the pixels making up the cloud, a_j is the size for that pixel and a_i is the total area of cloud i. Using these centroids, the distances to all other clouds are determined using the following equation:

$$\overline{NN} = \frac{\sum_{i=1}^{N} min\left\{\sqrt{(C_{x,i} - C_x)^2 + (C_{y,i} - C_y)^2}\right\}}{n} \tag{2.6}$$

Here $\overline{NN}$ stands for the average nearest neighbour distance, $C_{x,i}$ and $C_{y,i}$ are taken from Equations (2.4) and (2.5) and C_x and C_y represent all centroids except for $C_{x,i}$ and $C_{y,i}$. The numerator in this equation defines the sum of the smallest distance to any of the other clouds (thus nearest neighbour) of all clouds. This is divided by the number of clouds n to find the average nearest neighbour. The example field in Figure 2.3 has an average nearest neighbour of 2.90 km when using all clouds. Again however, since the sizes of the clouds adjacent to the edges and no-data area are unknown, and yet there centroids are unknown, these clouds should be removed from the dataset. Doing so, an average nearest neighbour distance of 3.22 km is found.

2.2.2. Cloud Size Distribution

Instead of determining the average cloud size, one can also look at the cloud size distribution. This parameter provides more information on the number of clouds in the field that have a given size.

The first studies on the cloud size distributions of shallow cumulus clouds using satellite data suggested an exponential (Hozumi et al., 1982; Plank, 1969; Wielicki and Welch, 1986) or lognormal (LeMone and Zipser, 1980; López, 1977) decay of cloud sizes. More recent studies suggest a power-law decay, however only over a restricted range of scales. Clouds larger than this scale break less common than expected (Benner and Curry, 1998; Cahalan and Joseph, 1989; Kuo et al., 1993; Machado and Rossow, 1993; Nair et al., 1998; Sengupta et al., 1990; Zhao and Di Girolamo, 2007). The same power-law decay is suggested from numerical models (Dawe and Austin, 2012; Heus and Seifert, 2013; Neggers et al., 2003). A schematic representation of such a typical shallow cumulus cloud size distribution would look like the one sketched in Figure 2.4.

Note that both the x- and y-scale are logarithmic.

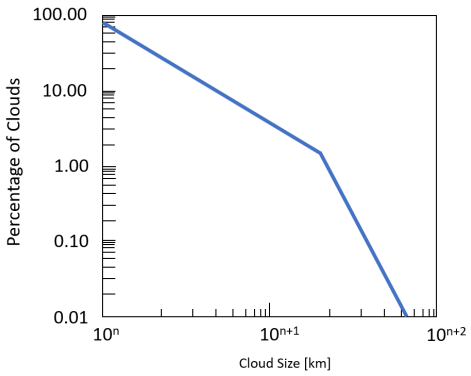


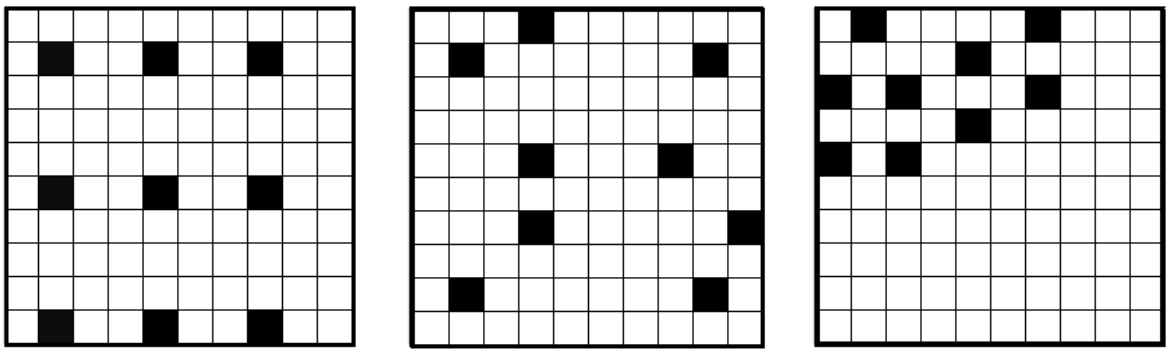
Figure 2.4: Example graph of a typical shallow cumulus cloud size distribution in a log-log plot. The cloud sizes are given on the x-axis and can have different size ranges. The y-axis shows the corresponding percentage of the clouds in the field for a given cloud size. Typically the graph can be divided into a less steep and a steeper part.

2.2.3. Characterizing Cloud Organization

Since the main goal of this thesis is to find a representative measure for cloud organization, different parameters then the aforementioned parameters are required. Although it is likely that the cloud density and average nearest neighbour distances decrease and that the average cloud size increases for a higher degree of organization, organization can not be described using these parameters alone. In this sub-section the organization index I_{org} and the organization parameter $SCAI$ are discussed.

Organization index I_{org}

The organization index discussed in this section is referred to as I_{org} and was applied in different studies (Seifert and Heus, 2013; Tompkins and Semie, 2017; Weger et al., 1992). The result is relatively simple: it classifies an area either as regular (Fig. 2.5a), random (Fig. 2.5b) or clustered (Fig. 2.5c). This is done by comparing the field to a theoretical field with a random organization. Although the method used for this thesis will deviate slightly, but substantially, from the widely used parameter I_{org}, the method to obtain I_{org} will be discussed here as the first steps are similar, and because the widely used method will also be used to explain the improvements of the related parameter proposed in this thesis.



(a) Example of a regular field. (b) Example of a random field. (c) Example of a clustered field.

Figure 2.5: Different organized fields. All fields have dimensions of 10 x 10 pixels with 9 pixels representing clouds. 2.5a shows a regular distributed field, 2.5b a random field and 2.5c shows an example of a clustered field. I_{org} is promising for separating each of these fields, despite their sizes and the total number of clouds being equal.

For the calculation of the I_{org} the nearest neighbour distance of each cloud must be found. For this a part of Equation (2.6) is used, but instead of determining the average nearest neighbour of the field directly, the nearest neighbours of all clouds are stored in a parameter. These nearest neighbours for each cloud are thus defined as the smallest distance from this given cloud to the other clouds in the field. In an equation this is written as:

$$NN_i = min\left\{\sqrt{(C_{x,i} - C_x)^2 + (C_{y,i} - C_y)^2}\right\} \tag{2.7}$$

Here NN_i represents the nearest neighbour distance of a given cloud.

Since the area is finite, periodic boundary conditions should be taken into account. This is only possible for rectangular shaped fields. Doing so, a schematic of the nearest neighbours looks like as depicted in Figure 2.6. Here clouds a and b are each others nearest neighbours. The same holds for clouds d and e, although this is only true because periodic boundary conditions are considered. The nearest neighbour for cloud c is cloud d, but the nearest neighbour of cloud d is cloud e. Clouds are thus not always pairs when looking at nearest neighbours.

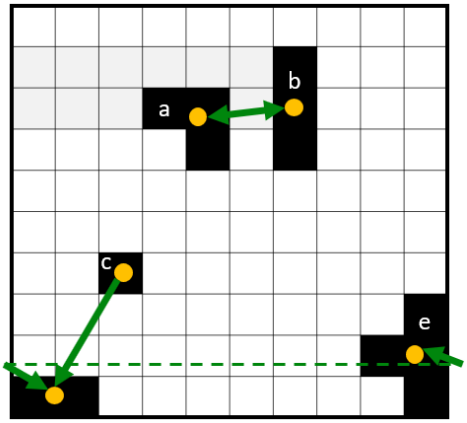


Figure 2.6: Schematic of the Nearest Neighbour calculation with Periodic Boundary Conditions. The black pixels represent clouds, the orange dots their centroids and the green arrows the nearest neighbour of each cloud.

After obtaining all nearest neighbour values, the cumulative distribution function of these nearest neighbour distances is calculated. The cumulative distribution function is given by:

$$F_X(x) = P(X \leq x) \tag{2.8}$$

In this equation the right-hand side gives the probability that the variable X is smaller than or equal to x. In terms of the cumulative distribution function of the nearest neighbours of a field this means that X represents the nearest neighbour distances of all clouds and x is given by the unique values of the nearest neighbour distances. $F_X(x)$ will from here on be referred to as the $NNCDF$ (Nearest Neighbour Cumulative Distribution Function).

The $NNCDF$ will be compared to the cumulative distribution function of the nearest neighbours if clouds in the area were organized randomly. The cloud organization can in that case be considered as a Poisson point process for which the expected distribution for the $NNCDF_{ran}$ is given by a Weibul distribution (Stoyan et al., 1987). In terms of the Nearest Neighbour distances, the Weibul distribution can be written as:

$$NNCDF_{ran} = 1 - exp(-\lambda \pi r^2) \tag{2.9}$$

Here λ is the number of clouds per unit area and r is the nearest neighbour distance, equal to the unique nearest neighbour distances found in the actual data.

After calculating the $NNCDF_{ran}$, the $NNCDF$ of the area is compared to it. This can visually be done by plotting the actual $NNCDF$ against the $NNCDF_{ran}$ and comparing this to the diagonal. Three possible outcomes are schematically sketched in Figure 2.7. The $NNCDF$ is plotted on the y-axis and the $NNCDF_{ran}$ on the x-axis. If the clouds in the field that is being analysed are randomly distributed, the graph of $NNCDF$ against the $NNCDF_{ran}$ will be on the diagonal. If the clouds are clustered the graph will be above the diagonal, because there are more small nearest neighbours in the actual field than that is expected from a theoretical random field. For a regular organization this is the other way around: the graph will be below the diagonal since there are less small nearest neighbours in the actual field compared to the theoretical random field.

The organization index I_{org} is defined as the area under the graph. As both the x and y axes run from 0 to 1, a random field would have value 0.5. A clustered field has a higher value and a regular field has a lower value. This does, however, not necessary mean that organization is the same for all spatial scales. An integrated value of 0.5, indicating a random field, could also be a result from cancellation: if clouds for example

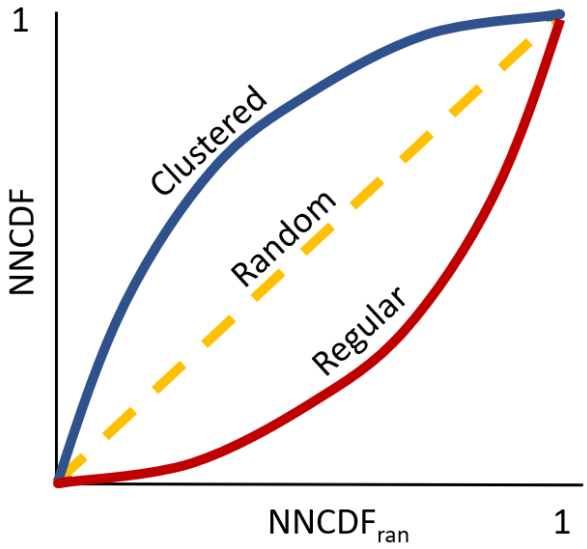


Figure 2.7: Schematic illustration of the determination of the organisation parameter I_{org}. The yellow dashed line is the diagonal and represents a random distribution. If the graph is plotted above this diagonal the clouds are clustered, below the diagonal means that the clouds are regularly distributed.

are regularly organised for small nearest neighbours (graph is below the diagonal) and clustered for larger values (graph is above the diagonal), integration cancels out these statistics and the field may seem random. Therefore, to assess organization for different spatial scales, the graph should still be examined.

Although I_{org} has been recognized as a powerful measure of organization, this parameter lacks one important cloud characteristic: its size. The $NNCDF_{ran}$ found in equation (2.9) does assume all clouds are points, located at the geometrical centroids of these clouds. However, in reality these centroids may represent clouds with sizes ranging from 1 pixel up to hundreds of pixels.

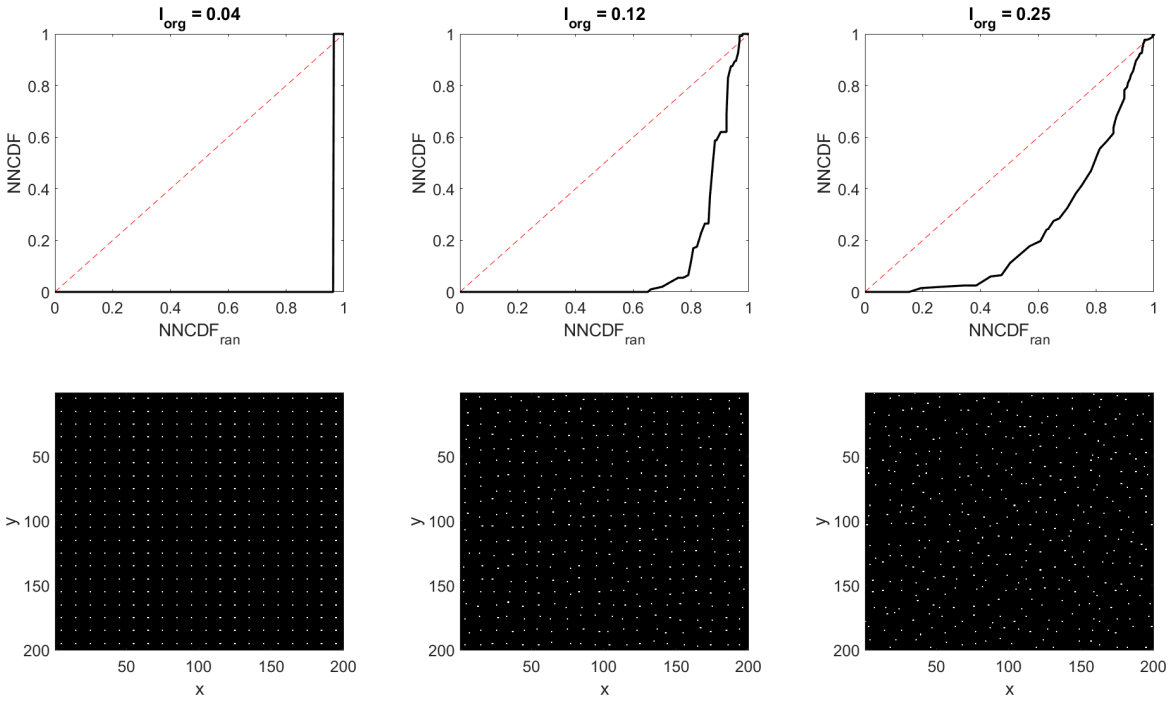
The assumption used for a poisson point process leads to a bias in I_{org} due to the fact that a given size of a cloud precludes the existence of another cloud within a certain distance. In reality, clouds are thus forced to be further apart due to their sizes than when they are considered single points. This results in an apparent regularity at smaller scales and either randomness of clustering at larger scales (Benner and Curry, 1998), because more small, and less large values for nearest neighbours are found. If the graph would be sketched in Figure 2.7, it would initially lie below the diagonal but would be located above the diagonal for larger scales.

Benner and Curry (1998), therefore also applied a related but essentially different method to assess cloud organization. They created a randomly generated inhibition $NNCDF$ that took not only the number of clouds into account, but also their sizes (but not their shapes, all clouds were assumed to be circles). They showed that a plot of the observed $NNCDF$ versus the inhibition $NNCDF$ provided the best means for assessing the organization of a cloud field. The apparent regularity at smaller scales was eliminated and cloud fields appeared to be clustered more strongly.

Here some examples for fields with a certain amount of clustering or regularity with their corresponding values for the altered I_{org} are presented, calculated by applying the method of Benner and Curry (1998) which is discussed further in Chapter 4. However, all clouds are points, as using random sizes would computationally be very complicated. These examples therefore mostly show the implications and results of regular or random fields.

First three examples of regular fields with increasing deviations from their fixed position are shown in Figure 2.8. On the left is a completely regular field shown with dimensions of 200 x 200 pixels with 400 single-pixel clouds placed every 10 pixels between 5 and 195 for both x and y coordinates. The field and plot in Figure 2.8b represent a slightly more random field: each pixel deviates with a maximum of 2 pixels in all x and y directions. In Figure 2.8c this maximum deviation from its regular position is increased to 4 in all directions. The values for I_{org} are increasing for a larger deviation from the regular positions but are still far below the value for a random distribution of clouds of 0.5. Increasing the random deviations even further will eventually

lead to a completely random field with the plot being on the diagonal. However, simulations have shown that an originally regular field with large deviations will never lead to a significant clustered field. Increasing the random signal in the regular field to infinite values will thus lead to an I_{org} of 0.5.



(a) I_{org} of a regular field with cloud centroids every 5 pixels in both directions.

(b) I_{org} of a semi-regular field with deviations of maximum 2 pixels in all directions from its regular position.

(c) I_{org} of a semi-regular field with deviations of maximum 4 pixels in all directions from its regular position.

Figure 2.8: I_{org} of different measures of regularity. (a) is completely regular with centroids every 5 pixels in both directions. Some random deviations from these regular cloud centroids are introduced in (b), where the maximum deviation from the regular position is 2 pixels in all directions. For (c) this maximum deviation is increased to 4 pixels in all directions.

Figure 2.9 shows some clustered fields with the same dimensions and number of clouds as the (semi-)regular fields shown in Figure 2.8.

Their locations are generated from a normal distribution with a fixed mean value μ = 100 for both x and y coordinates and a varying standard deviation σ. The value of σ was increased from 10 to 150 pixels to achieve a random cloud distribution. The fields shown in Figure 2.9 show the simulations of σ = 10 until σ = 90 in steps of 10. These fields were simulated making use of certain restrictions: clouds can not be positioned in a place if another cloud was positioned there before and they must be positioned within the fields' boundaries. If the randomly generated values do not meet these requirements they are generated again, until they are assigned to a legal position. X and y coordinates are generated independently from each other. For each value of σ the simulations is run 100 times. Variations are small but do increase for decreasing cloud numbers. The average, minimum and maximum I_{org} values out of 100 simulations for σ = 10 pixels for example are 0.93, 0.90 and 0.96 respectively. For σ = 90 pixels, these values are respectively 0.51, 0.45 and 0.56.

The average I_{org} for each σ after 100 simulations is plotted in Figure 2.10 indicated by the dashed green line. The results for a field with similar dimensions but only 100 clouds (blue solid line) and the results of a field with 100 x 100 pixels and 100 (black solid line) or 400 clouds (dashed red line) are also shown. The I_{org} values are decreasing for an increasing standard deviation. However, at some points the scenes become random instead of clustered and the values of I_{org} have reached the value for a random distribution: 0.5. Similar to the shift from regular to random fields, fields that are originally clustered do not become regular. Their minimum value is 0.5. Depending on the size of the field, this value is reached at a smaller or larger value of σ. The smaller the field, the sooner the clouds are randomly distributed. Figure 2.10 shows that this point is reached when σ is approximately equal to half of the size of the domain (σ = 50 pixels when the domain is 100 x 100 pixels and σ = 100 pixels when the domain is 200 x 200 pixels). The number of clouds in a field also

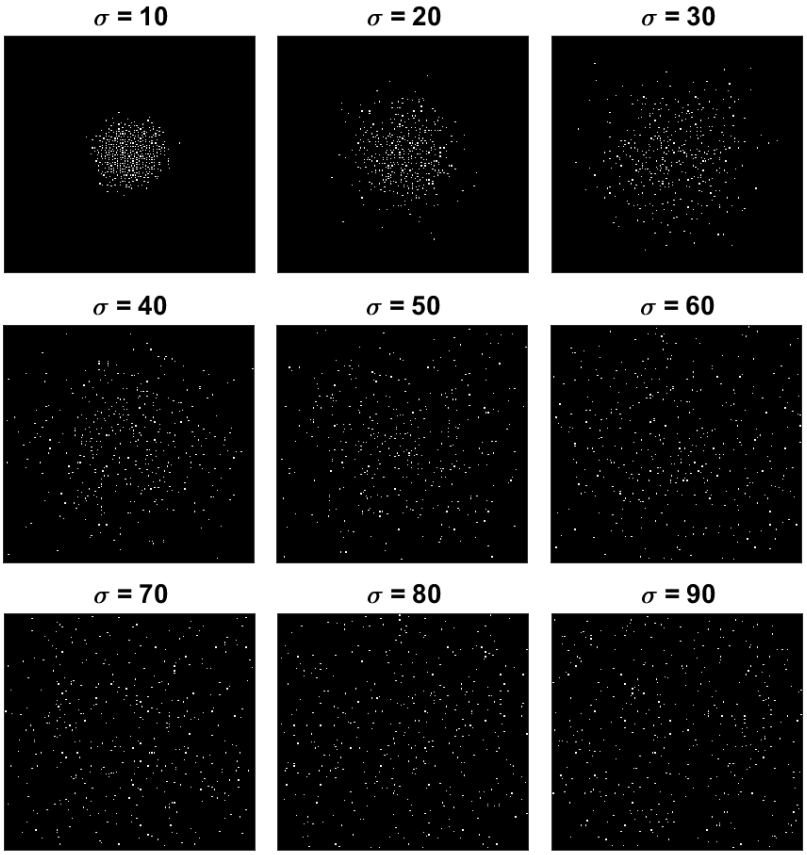


Figure 2.9: Realizations of clustered scenes with varying standard deviations. The fields have dimensions of 200 x 200 pixels and contain 400 single-pixel clouds. A higher standard deviation σ leads to a more random scene.

influences the value of I_{org}: the lower the number of clouds, the more I_{org} varies between individual runs. In practice, fields that are analysed typically contain several thousands of clouds and the value of I_{org} does not change significantly for different runs.

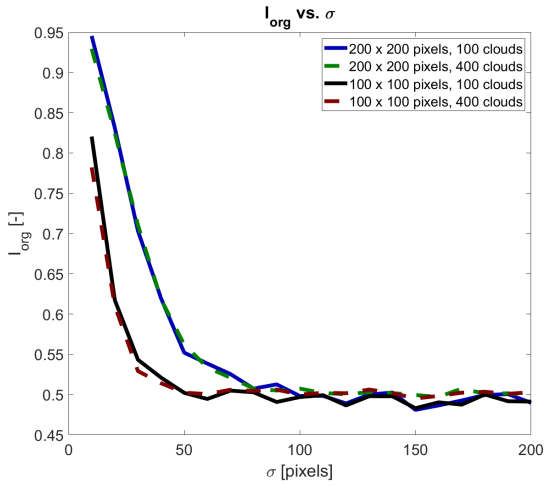


Figure 2.10: I_{org} compared to the standard deviation from a clustered field. The scenes have dimensions of 200 x 200 pixels and contain 400 single-pixel clouds. Cloud locations are generated using a normal distribution function with a mean value μ at 100 for both the x and y coordinates. Increasing σ leads to a lower value of I_{org}.

2.2.4. Cloud Organization using parameter SCAI

The Simple Convective Aggregation Index (SCAI) was introduced by Tobin et al. (2012). The parameter is relatively simple and takes into account the number of clusters in a domain and the average nearest neighbour distance.

This method assumes that the lower the number of cloud clusters is in a domain, the more aggregated these clouds are. Furthermore small nearest neighbours indicate more aggregation. They showed that these parameters are not equivalent but instead provide complementary information on the degree of aggregation. In their studies the method was applied to deep convection, but should in theory also be applicable to shallow cumulus convection as the same assumptions (less clouds and smaller nearest neighbours indicate a higher degree of aggregation) can be assumed for both cloud types.

The parameters can be combined in a Simple Convective Aggregation Index (SCAI) given by:

$$SCAI = \frac{N}{N_{max}} \frac{D_0}{L} \times 1000 \tag{2.10}$$

Here N is the number of clouds, N_{max} is a theoretical maximum number of clouds, D_0 denotes the average nearest neighbour and L is the characteristic length of the field. The number of clouds in a field and the average nearest neighbour are thus normalized by fixed parameters for a given field. The characteristic length L is the squared root of the total area that is being evaluated and N_{max} is proportional to $(L/a)^2$ (Tobin et al., 2012). Here, a is set to 2.

Finding the value of SCAI for the three different example fields of Figure 2.5 requires the values of the normalizing factors N_{max} and L first. Assuming each pixel is 1 km^2, we can say that N_{max} is proportional to L^2. The characteristic length is the average length of the sides of the field, 10 pixels (or 10 km) for these examples. The maximum number of clouds that fit in these fields are $0.25 * 10^2$, since pixels can not be laterally or diagonally adjacent to one another. Therefore N_{max} equals 25. All fields contain the same number of clouds: 9. Using Equation (2.10), we find for the regular field displayed in Figure 2.5a a value of SCAI of 108.0. For the random field in Figure 2.5b we find 85.7 and for the clustered field in Figure 2.5c the SCAI value is 72.9. As a lower value for SCAI indicates more aggregation these results meet the expectations.

3

MODIS and data processing

This chapter discusses which data has been used for this research (Sections 3.1 and 3.2) and how this data was processed (Section 3.3). In Section 3.4 the thresholds which are applied for the data processing are discussed based on some initial results.

3.1. Earth Observation using Satellites

For this project data acquired by the MODIS instrument (more in section 3.2) is used. MODIS is aboard two satellites: Terra and Aqua. The satellites were launched on December 18, 1999 and May 4, 2002, respectively. Both satellites are in low-orbit, at an altitude of 705 km. The satellites are part of NASA's Earth Observing System (EOS) and have sun-synchronous, near-polar (inclination angle of 98.2°, see Figure 3.1) circular orbits.

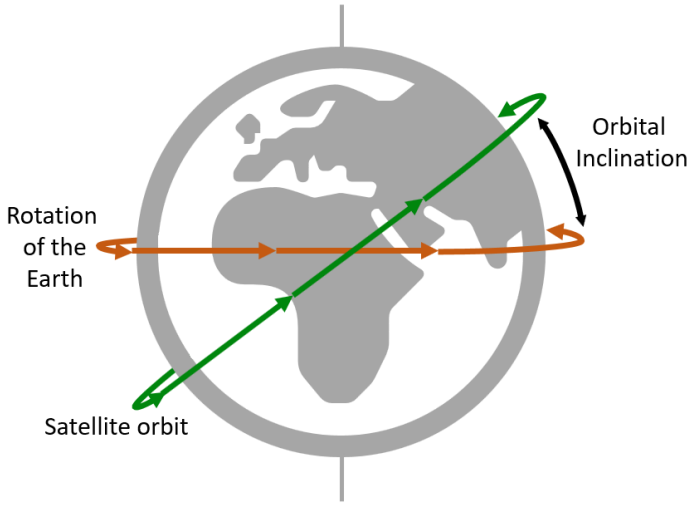


Figure 3.1: Schematic representation of an orbital inclination.

A sun-synchronous orbit means that the orbit precesses around the Earth's polar axis with a rate similar to the Earth's average angular speed around the sun, thus one revolution per year. The big advantage of such an orbit is that it crosses the same latitude at the same local solar time, regardless of the longitude or date. Terra is descending (north to south) and crosses the equator during daytime at approximately 10:30 AM (14:30 UTC), whereas AQUA is ascending and crosses the equator during daytime at approximately 1:30 PM (17:30 UTC). With an orbit period of 99 minutes, the repeat cycle is 16 days, meaning that the swaths are similar every 16 days. The approximate orbit for Terra on December 7, 2015, is shown in Figure 3.2.

A low inclination angle implies that polar areas are not captured, whereas larger inclination mean that some areas closer to the equator won't be captured. With an inclination angle of 98.2° high latitude areas are mostly captured and little area near the equator is not captured on each day. The resulting world view images on December 7, 2015, for Terra and Aqua are shown in Figures 3.3 and 3.4 respectively. From these images it

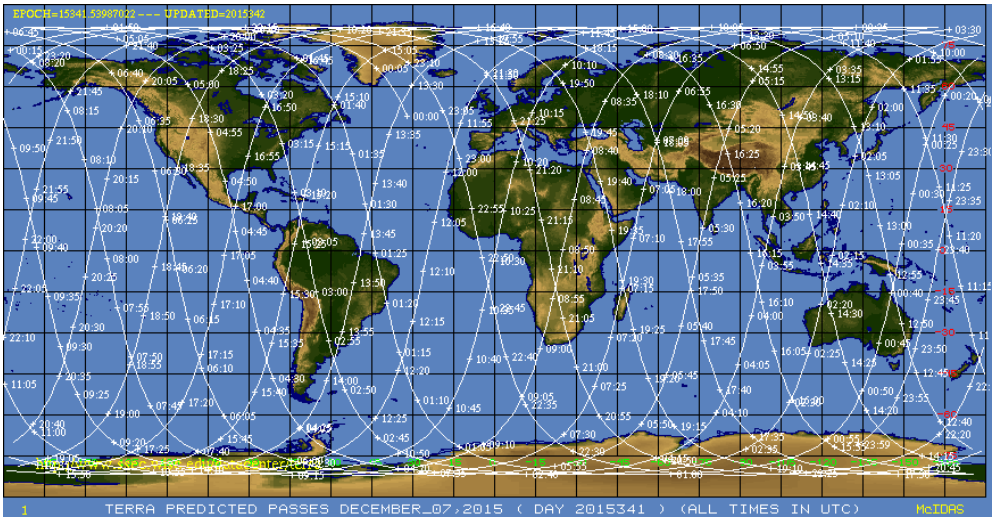


Figure 3.2: Overpass times Terra on December 7, 2015. *Source:* *https://www.ssec.wisc.edu/datacenter/terra/*

can be seen that around the equator some data is missing. As the orbits are slightly different and the repeat cycle is 16 days, the areas of missing data are similar again after 16 days. However, the latitude boundaries do not change, only the longitude boundaries are different.

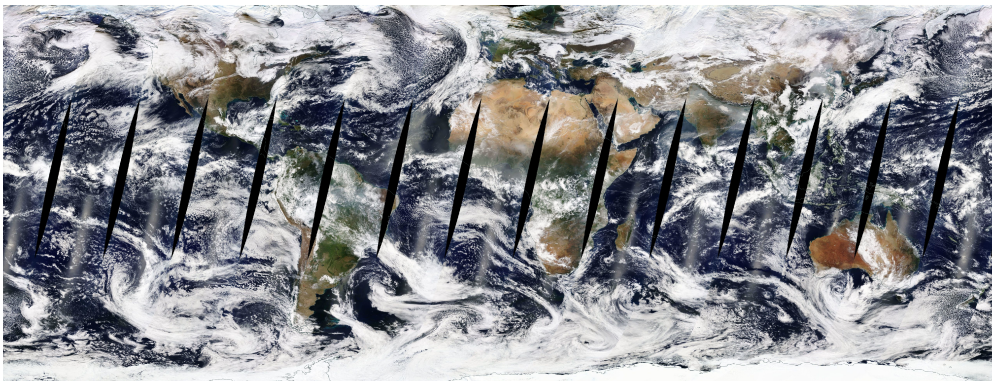


Figure 3.3: Terra swaths (descending) on December 7, 2015. Taken from `https://worldview.earthdata.nasa.gov`

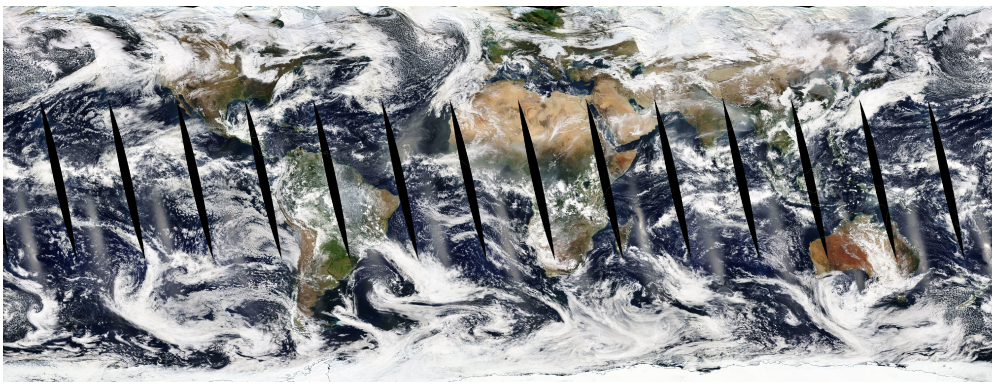


Figure 3.4: Aqua swaths (ascending) on December 7, 2015. Taken from `https://worldview.earthdata.nasa.gov`

3.2. MODIS

MODIS (Moderate Resolution Imaging Spectroradiometer) is one of the five instrument aboard the Terra and Aqua satellites. It acquires data in 36 spectral bands ranging from the short wave visible to the long wave infrared (0.4µm to 14.4µm), or in groups of wavelengths. Data acquired by MODIS is meant to improve our understanding of global dynamics and processes occuring on the land, in the oceans, and in the lower atmosphere. Because MODIS has a scanning pattern of 55° and is at an altitude of 705 km, the viewing swath width is 2330 km. Because of that the entire surface of the earth is viewed every 1-2 days.

The spatial resolution at nadir is 250 m for bands 1 and 2, 500 m for bands 3-7 and 1000 m for bands 8-36. With increasing sensor zenith angle, pixels become larger. This is visualized in Figure 3.5 taken from Sayer et al.(2015).

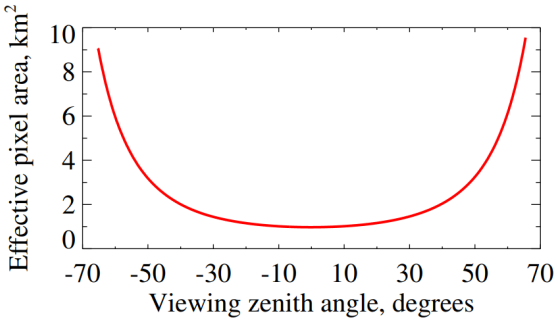


Figure 3.5: The pixel size with increasing sensor zenith angle for a pixel with a resolution of 1km at nadir (M. Sayer et al., 2015).

3.2.1. MODIS Cloud Mask

From this data a cloud mask of 1km resolution at nadir is created. Table 3.1, taken from the MODIS Cloud Mask User's Guide (Strabala, 2005), gives an overview of all bands on the MODIS instrument and states whether these bands are implemented in the Cloud Mask algorithm.

Table 3.1: MODIS Bands (Strabala, 2005).

Band	Central Wavelength (µm)	Used in Cloud Mask	Primary Application		Band	Central Wavelength (µm)	Used in Cloud Mask	Primary Application
1	0.659	Y	Clouds, shadow		19	0.940	Y	Shadows
2	0.865	Y	Low clouds		20	3.750	Y	Thin cirrus
3	0.470	N			21/22	3.959	Y(21)/N(22)	Window
4	0.555	N	Snow		23	4.050	N	
5	1.240	Y	Snow		24	4.465	N	
6	1.640	Y	Snow, shadow		25	4.515	N	
7	2.130	N			26	1.375	Y	Shadows
8	0.415	N			27	6.715	Y	High moisture
9	0.443	N			28	7.325	N	
10	0.490	N			29	8.550	Y	Mid moisture
11	0.531	N			30	9.730	N	
12	0.565	N			31	11.030	Y	Window
13	0.653	N			32	12.020	Y	Low moisture
14	0.681	N			33	13.335	N	
15	0.750	N			34	13.635	N	
16	0.865	N			35	13.935	Y	High cloud
17	0.905	N			36	14.235	N	
18	0.936	Y	Low clouds					

The resulting Cloud Mask actually contains a 48-bit cloud mask product, containing much more information than just a cloud classication. Some examples are:

- Sun glint flag.
- Land/Water flag classifying pixels as Water, Coastal, Desert or Land.
- Non-cloud obstruction flag (heavy aerosol).

The cloud mask information is stored in bit fields 1 and 2, because each pixel is classified as either *Cloudy*, *Uncertain Clear*, *Probably Clear* or *Confident Clear*. Table 3.2 provides a simple overview of these bit field values and the associated classification. The classification of pixels into one of the four classes is based on 14 different bands (Table 3.1).

Table 3.2: MODIS Cloud Classification and bit values (Strabala, 2005).

Classification	**Bit Field 1**	**Bit Field 2**
Cloudy	0	0
Uncertain Clear	0	1
Probably Clear	1	0
Confident Clear	1	1

3.3. Data processing

MODIS data is available in *Hierarchical Data Format (HDF)*, which is designed for the storage of large amounts of data. This format is supported by different software programs, for example *MATLAB*. This section discusses which steps must be taken to extract the desired information from the MODIS data.

The MODIS Cloud Mask data is free of charge and can be downloaded to a computer. The number of files that should be downloaded per day varies because MODIS data is available in tiles, of which the boundaries depend on the overpass and are therefore different every day. Sometimes a single tile covers the entire area of interest and thus only one tile is downloaded on that day. But often more tiles (up to 4) are required to cover the entire area of interest. This requires some extra work before the data can be analysed. The original cloud mask data from Aqua and Terra on December 7, 2015, are displayed in Figure 3.6 and Figure 3.7 respectively. It becomes clear that the cloud mask product, as mentioned in Section 3.2.1, contains more information than just the cloud mask. Land, Water and coastal areas can be distinguished and also the sun glint is visible in dark green colours. For this thesis, only the cloud mask will be used.

In Figure 3.7 the data is displayed in one image, but is stored in three different files. Therefore, the first step is to mosaic (stitch) the files together in the correct way. Then a subset of the area of interest can be extracted and the data can be converted from *Hierarchical Data Format (HDF)* to *Tagged Image File Format (TIFF)*. There are various ways to do this, but a useful and relatively fast tool is called the *HDF-EOS to GeoTIFF Conversion TOOL*, or in short *HEG tool*. This tool can do all the steps from mosaicking to converting in one run. It is developed by the Synergy progam, with the support of NASA's Earth Observing System Program (E. Moghaddam-Taaheri, 2017) and can be downloaded for free via the NASA website. It can be operated via a GUI or from the command line. If you want to use the tool to mosaic/convert data for a multiple days, operating it from the command line is way more convenient. The *subset_stitch_swath* program is part of the *HEG tool* and is used to read one or more files, stitch them together, obtain a subset of the area of interest and convert the data to TIFF format. Again, this program can be used from the GUI or from the command line. In the command line the following code must be used:

```
subset_stitch_swath -p parameter_file.prm
```

The parameter file contains field-value pairs as input for the program. A separate parameter file is created for every day, every platform and every field name. Because information on the Sensor Zenith Angle and Cloud Top Pressure is necessary for later steps, these fields must be processed on top of the Cloud Mask Product. On top of stitching, subsetting and converting their resolution is also scaled up because the original data of the Sensor Zenith Angle and Top Cloud Pressure has a resolution of 5km at nadir.

Creating the parameter files can be done in *MATLAB*. Creating a batch (*.bat*) file at the same time containing these code lines for every day, every platform and every field name, gives the possibility to just run one

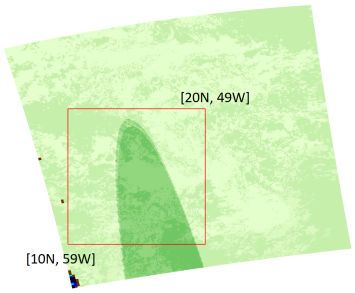


Figure 3.6: An example of the original cloud mask data of Aqua on December 7, 2015. On this day only one tile was required to cover the area of interest, indicated by the red square.

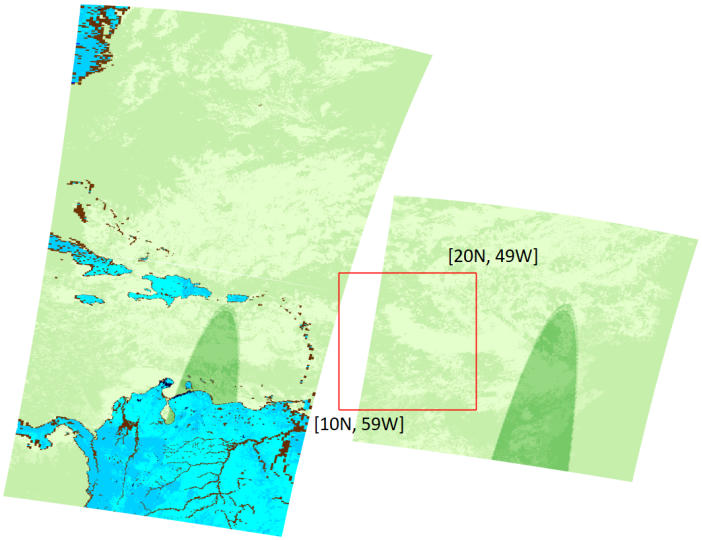


Figure 3.7: An example of the original cloud mask data of Terra on December 7, 2015. On this day three tiles were required to cover the area of interest, indicated by the red square. Two of these tiles were from the same swath, the third one is part of another swath.

file in the command line eventually. For N parameter files, your batch file would look like this:

```
ECHO OFF
CALL subset_stitch_swath -p parameter_file_1_CloudMask.prm
CALL subset_stitch_swath -p parameter_file_1_SensorZenith.prm
CALL subset_stitch_swath -p parameter_file_1_CloudTopPressure.prm
                                .
                                .
                                .
CALL subset_stitch_swath -p parameter_file_N_CloudMask.prm
CALL subset_stitch_swath -p parameter_file_N_SensorZenith.prm
CALL subset_stitch_swath -p parameter_file_N_CloudTopPressure.prm
```

ECHO OFF suppresses the commands to be echoed. Echoing a command can be useful sometimes but is unnecessary here. To execute a command from a batch file, the word *CALL* is added before every command.

The parameter files must contain specific information. The fields, together with a short explanation, are listed here.

- NUM_RUNS: the number of runs in file. This is always equal to one, since each swath field is processed separately.
- BEGIN and END: after the number of runs the keyword BEGIN must be used. After all parameters (except for NUM_RUNS) are listed, the word END is attached at the end of the file.
- NUMBER_INPUTFILES: the number of files that need to be stitched. For this project this number varies between 1 and 4.
- INPUT_FILENAMES: the names of the input file(s). These names must also contain the directory path from the folder where the program will run. If the number of input files is larger than 1, the different files must be separated by a PIPE "|".
- OBJECT_NAME: the name of the swath, which is *mod06* for both AQUA and TERRA.

- FIELD_NAME: the exact name of the swath field. For this project 3 different swath fields are processed: *Cloud_Mask_1km*, *Sensor_Zenith_Day* and *Cloud_Top_Pressure_Day*. The name of the field must end with a PIPE "|", otherwise the parameter file is not read correctly.

- BAND_NUMBDER: this field contains the number for the desired band. Field data can be 3-dimensional and in that case 2-dimensional slices from the data field must be read. For 2-dimensional data the band number will default to 1, which is the case for all fields processed for this project.

- SPATIAL_SUBSET_UL_CORNER and SPATIAL_SUBSET_LR_CORNER: The upper-left and lower-right corners for spatial subsetting. For this project these values are (20 -59) en (10 -49) respectively.

- OUTPUT_OBJECT_NAME: the name of the grid in the output .hdf file. This is *mod06* in this case, although the .hdf file will not be saved: only the .tif file is required.

- OUTGRID_X_PIXELSIZE and OUTGRID_Y_PIXELSIZE: These contain the pixel sizes along the X-axis (longitude) and Y-axis (latitude). The default value of 999.0 is used for both axes, resulting in a grid of 1126x1099 pixels of 999.0 x 999.0 meter per pixel. Since Cloud Top Pressure and Sensor Zenith angle values are only available in 5km resolution, their resolution is scaled up.

- RESAMPLING_TYPE: this can be nearest neighbour, bilinear cubic or cubic convolution. For this project nearest neighbour (NN) is used.

- OUTPUT_PROJECTION_TYPE: there are multiple options but here the Universal Transverse Mercator (UTM) is used.

- ELLIPSOID_CODE: this field is required for processing but at the moment the program always resorts to using the WGS84 Ellipsoid.

- UTM_ZONE: this is an optional field when choosing for output projection type UTM. The area of interest is located in UTM zone 21.

- OUTPUT_PROJECTION_PARAMETERS: An optional field and by default all projection parameter values will be set to zero, which is also the case for this project.

- OUTPUT_FILENAME: a required field that must include to full path directory. Files are named in a systematic way: PLATFORM_field_YYYYMMDD.tif, in which PLATFORM is either TERRA or AQUA, field is clouds, sz or ctp for the cloud mask, sensor zenith and cloud top pressure respectively.

- SAVE_STITCHED_FILENAME: YES or NO, but the user guide asks to use NO. This field will also be removed from the parameter file in the future.

- OUTPUT_STITCHED_FILENAME: the field is not used by the tool and will also be removed in the future. The user guide suggests to use a dummy name here.

- OUTPUT_TYPE: the type of output is specified here, which is GEO for this project. It must be set to HDFEOS for Hdf-Eos grid output, or BIN for raw binary output.

An example of such a parameter file is shown in the box below. This example is the parameter file for the cloud mask of AQUA on December 1, 2010. The values that vary per parameter file are printed in purple.

```
NUM_RUNS = 1
BEGIN
NUMBER_INPUTFILES = 3
INPUT_FILENAMES = 15-16/TERRA/MOD06_L2.A2015341.1340.061.2017324004833.hdf|
15-16/TERRA/MOD06_L2.A2015341.1515.061.2017324004805.hdf|
15-16/TERRA/MOD06_L2.A2015341.1520.061.2017324004533.hdf
OBJECT_NAME = mod06|
FIELD_NAME = Cloud_Mask_1km|
BAND_NUMBER = 1
SPATIAL_SUBSET_UL_CORNER = ( 20.0 -59.0 )
SPATIAL_SUBSET_LR_CORNER = ( 10.0 -49.0 )
OUTPUT_OBJECT_NAME = mod06|
OUTGRID_X_PIXELSIZE = 999.0
OUTGRID_Y_PIXELSIZE = 999.0
RESAMPLING_TYPE = NN
OUTPUT_PROJECTION_TYPE = UTM
ELLIPSOID_CODE = WGS84
UTM_ZONE = 21
OUTPUT_PROJECTION_PARAMETERS = ( 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 )
OUTPUT_FILENAME = F:/Thesis/Data/MODIS/MYD06/TERRA_clouds_20151207.tif
SAVE_STITCHED_FILE = NO
OUTPUT_STITCHED_FILENAME = HELLO
OUTPUT_TYPE = GEO
END
```

After executing the batch file and obtaining all *.tif* files, the next step of data processing can start. From here on *MATLAB* is used.

As mentioned before in Section 3.2.1, the cloud mask is a 48-bit dataset, the bits have to be read individually using *MATLAB* function *bitget*. Bit fields 1 and 2 contain information on the classification (Table 3.2). The bit values are converted from class *int8* to class *double* (floating point) for further evaluation and get a value of 1-4 for cloudy, uncertain clear, probably clear and confident clear respectively. If areas are not covered by the swath on that day, pixels contain no data and get value 0. This happens for example to the pixels located between the swaths in Figure 3.7. Most pixels containing data are classified either as Cloudy or as Confident Clear.

Figure 3.8 shows the cloud classification for MODIS aboard Aqua on December 7, 2015. As can be seen in Figure 3.6 only one tile was required to cover the entire area of interest. Stitching of tiles was therefore not necessary for Aqua on this day. The *HEG tool* solely created a subset and converted the data to *GeoTIFF* format. Since only bit fields 1 and 2 were read in *MATLAB*, the sun glint visible in Figure 3.6 can not be seen anymore in Figure 3.8.

A choice must be made regarding which cloud classification classes will be considered cloudy in the end. Only considering "cloudy" would result in a more conservative estimation, whereas treating both "cloudy" and "uncertain clear" as clouds would most likely result in an overestimation of cloud pixels. Both options will initially be used for calculations such that a quantitative argumentation can be used to decide what the best option is in a later stage.

As mentioned before, the data fields *Top Cloud Pressure* and *Sensor Zenith Angle* are required for further steps and were processed in the *HEG tool* as well. They both are signed 16-bit integer values and are converted to *double* class as well in *MATLAB*. Also both must be multiplied by a scale factor, which is a constant and can be found in the meta data of the original file.

The *HEG tool* works good for the cloud mask product, leaving no "no data" values in between files from the same swath. However for the Top Cloud Pressure and the Sensor Zenith Angle, a thin "line" with a width of approximately 5 pixels of "no data" exists between files from the same swath. This thin "line" is also, when looking carefully, visible in Figure 3.7. Having this data gap is undesirable, because this data is used to mask out high clouds or areas with large sensor zenith angles and with missing data, pixels are possibly not masked out when actually they should have been. This problem can be overcome in *MATLAB*, by identifying these

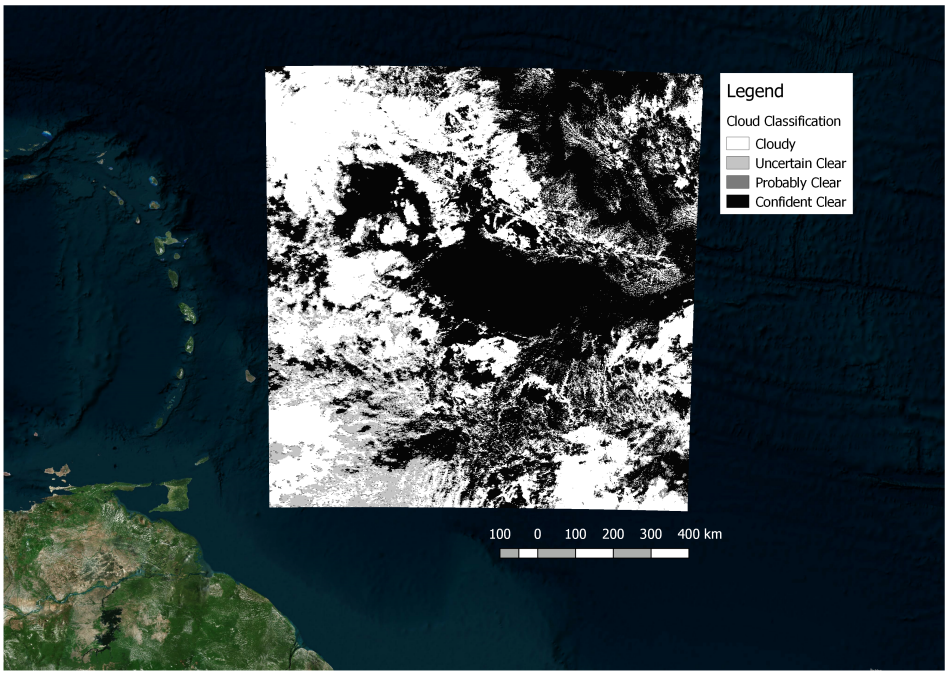


Figure 3.8: Cloud Classification for MODIS aboard Aqua on December 7, 2015. Clouds are classified as Cloudy, Uncertain Clear, Probably Clear of Confident Clear. This image does not contain no-data, as the swath on this day contained the full area of interest (see Fig 3.6).

pixels and interpolating values from the pixels surrounding the line. It might be that this problem occurs because the Top Cloud Pressure and Sensor Zenith Angle must be scaled up to end up with 999x999 meter resolution, whereas the cloud mask is already available in this resolution. For interpolation data gaps are filled with a moving average with a window length of 50 to ensure all gaps get a value. Since the input parameter file for the *HEG tool* set the same pixel resolution for the three data fields, their sizes are the same: 1126x1099 pixels. The resulting matrices for the Cloud Top Pressure and Sensor Zenith Angle are shown in Figures 3.9 and 3.10.

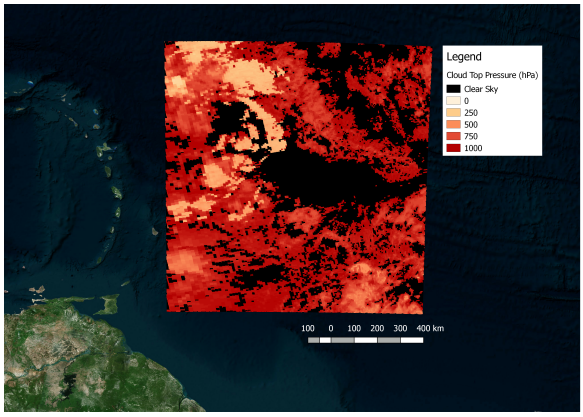


Figure 3.9: The Cloud Top Pressure for MODIS aboard Aqua on December 7, 2015. Data is already scaled by the scale factor.

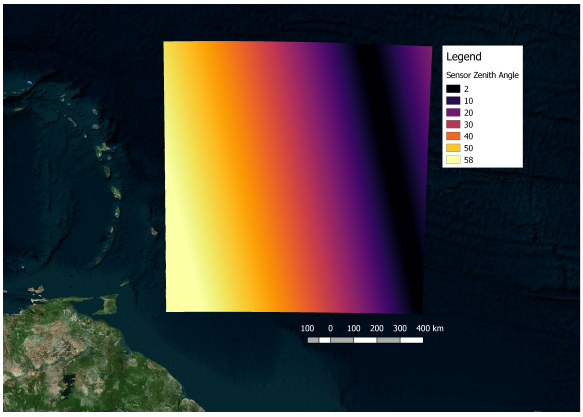


Figure 3.10: The Sensor Zenith Angle for MODIS aboard Aqua on December 7, 2015. Data is already scaled by the scale factor.

These matrices can now be used to mask out Cloud Mask data. A Cloud Top Pressure below 550 hPa means that these pixels represent clouds above approximately 5 km. Since this thesis focuses on shallow cumulus clouds, these pixels are masked out and classified in a separate matrix as high clouds. Choosing the threshold for the Sensor Zenith Angle is less straightforward. Pixels increase in size with an increasing Sensor Zenith angle which may alter the results and on top of that clouds are observed more from the side which makes the cloud classification less reliable, as argued by ... Therefore, analysis will be done using different Sensor Zenith

Angles: 30°, 45° and 60° will be used and final statistics will be compared to choose the best option for this thesis.

The last processing step is explained with an example. Again the results of MODIS aboard Terra on December 7, 2015, are used for this. In this example only pixels classified as *Cloudy* are considered as clouds, pixels classified as *Uncertain Clear* are thus considered being clear sky. For the Sensor Zenith Angle an angle of 45° is used as threshold. Figure 3.11 shows the intermediate results of the cloud classification and the two masks. After applying these two masks to the initial cloud classification the final Shallow Cumulus Cloud Classification is obtained, displayed in Figure 3.12. All cloudy pixels have value 1, all other pixels get value 0. The information on the number of pixels classified as high clouds and the number of pixels masked out due to their high sensor zenith angle is stored in a different parameter.

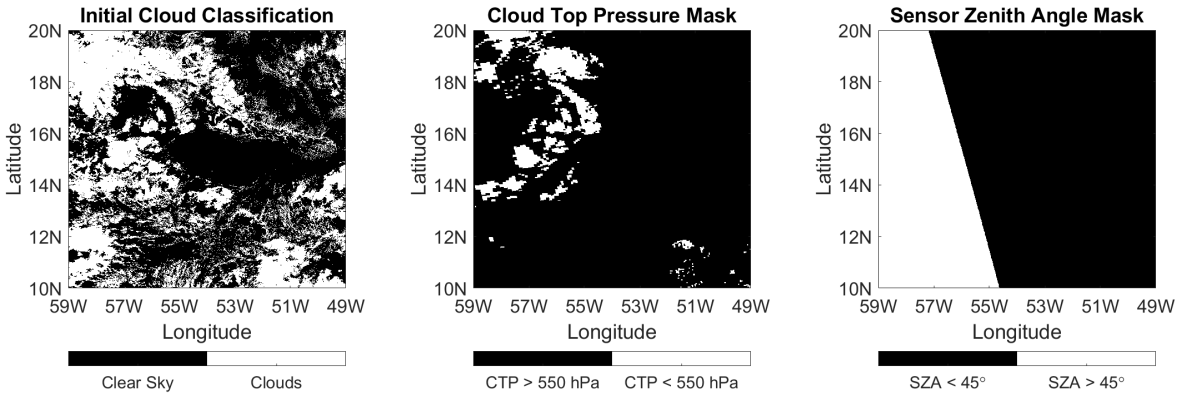


Figure 3.11: Intermediate results of the Cloud Classification, Cloud Top Pressure and Sensor Zenith Angle. The Cloud Top Pressure Mask and the Sensor Zenith Angle Mask are used to mask the Initial Cloud Classification. The result is shown below.

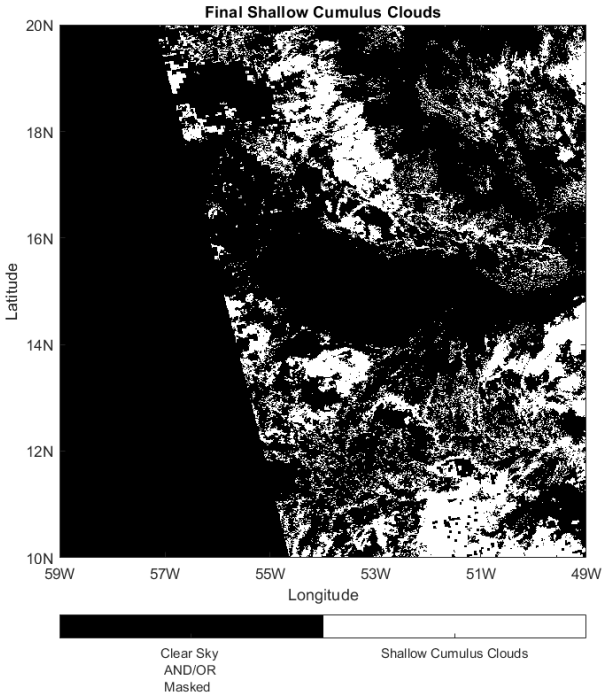


Figure 3.12: Final Shallow Cumulus Clouds of MODIS aboard Terra on December 7, 2015. Pixels not classified as Shallow Cumulus Clouds can represent clear sky, high clouds or an area with a sensor zenith angle above 45°.

After obtaining the final result of each sensor on each day it is decided if the day is used for further analysis or not. Following previously chosen thresholds based on the best agreement between cloud cover derived from MODIS and from ground-based lidar instruments, days/areas are only evaluted further if both of the following statements are true (Brueck et al., 2015):

- The area that can be evaluated is larger than 50% of the total area. A description of the definition is given below.
- High clouds cover is less than 20% of the area that can be evaluated.

The evaluable area is defined as the total area minus the no data area and minus the high sensor zenith angle area.

On top of doing the above analysis for each day separately, the steps are repeated for smaller portions of the total area. By doing so, data is collected to analyse spatial variability of the cloud organization. This will be done for different field sizes as the complete research area will be split in 4 equal fields and in 16 equal fields. Results from smaller fields can eventually be compared to the results from the larger fields to identify correlations.

3.4. Data selection

In the previous section it was explained that masks are applied to the initial data and depending on the amount of high clouds and the size of the area that was left after applying the Sensor Zenith Angle mask, it was decided whether days should be further evaluated or not. The threshold for the maximum amount of high clouds was fixed to 20% of the area that was left after applying the Sensor Zenith Angle mask. This section compares the results for different Sensor Zenith Angles in order to determine the best Sensor Zenith Angle. Furthermore the different platforms (Terra or Aqua) and the different options for Cloud Classification (only *Cloudy* or *Cloudy* and *Uncertain Clear* as clouds) derived from the MODIS Cloud Mask are evaluated.

The different Sensor Zenith Angles are compared in the first subsection. The second subsection discusses the different platforms and the third subsection discusses the different options for Cloud Classification. When not the focus of analysis, the Sensor Zenith Angle is fixed to 45°, an average value of the results of the two platforms is taken and only pixels classified as *Cloudy* by the MODIS Cloud Mask are used to define clouds. Investigation showed that changing these parameters does affect the absolute values, but does not change the positive or negative bias that can be concluded about the parameters that are varied. Therefore the two thresholds/options that are not under discussion in the subsection are fixed.

3.4.1. Sensor Zenith Angle

The threshold for Sensor Zenith Angle was set to different values: 30°, 45° and 60°. The number of days that were left after applying these different masks are shown in Table 3.3. Under the column "sum", the total number of datasets is calculated, whereas "combined" gives the total number of days that can be evaluated. It is then assumed that Terra and Aqua are unbiased, that the physical differences between their overpass times are insignificant and thus that their data can be combined. On days that both Terra and Aqua are evaluated their average is taken. On days that only one of them was evaluated, that value is set as the average value for that day. From the Table it can be derived that more days can be evaluated when the Sensor Zenith Angle threshold is higher. This is not surprising, as with a larger threshold less data is removed and thus there is a higher chance that more than 50% of the area is covered by data points.

Table 3.3: Number of evaluable days for different Sensor Zenith Angles separately for Terra and Aqua, the sum of the evaluable days seperataly and for a combination of the two sensor platforms. CC_{high} stands for the high-level cloud cover. Only pixels classified as *Cloudy* by the MODIS Cloud Mask are used to define cloud pixels.

Mask	**Terra**	**Aqua**	**Sum**	**Combined**
No mask	720	720	1440	720
SZA = 30 °, CC_{high} < 20%	142	135	277	242
SZA = 45 °, CC_{high} < 20%	281	276	557	403
SZA = 60 °, CC_{high} < 20%	378	400	778	507

Not only the number of days that can be used for further evaluation is different. The average values of the "basic" parameters like cloud cover, average nearest neighbour, the number of clouds per unit area and the

average cloud size are also different. Figure 3.13 shows the histograms of these different parameters for the different thresholds.

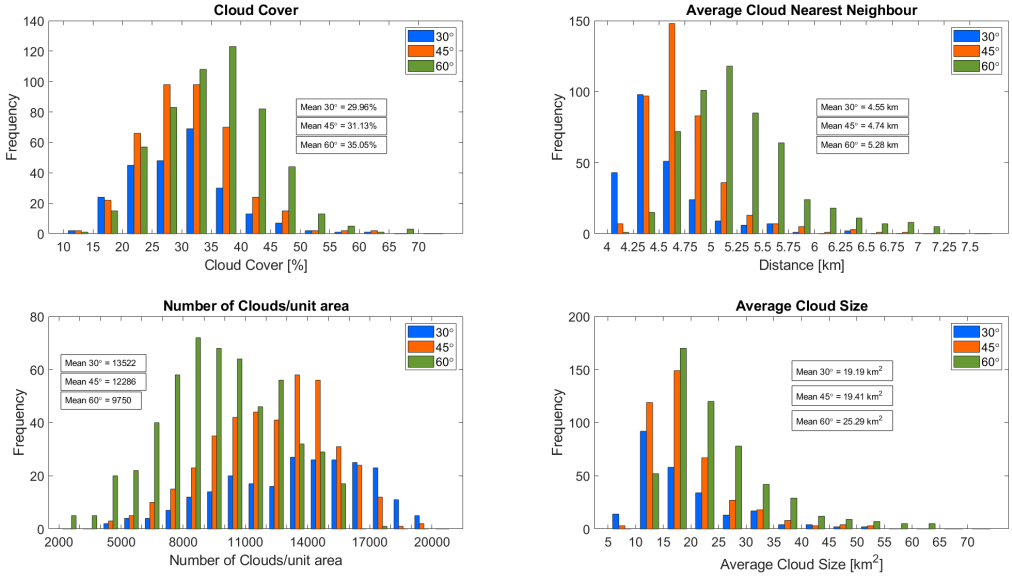


Figure 3.13: Results for different Sensor Zenith Angles.

With larger Sensor Zenith Angle the cloud cover, average nearest neighbour distance and average cloud size increase and the number of clouds per unit area decreases. This can most likely be explained by the increasing ground pixel size with increasing Sensor Zenith Angle, displayed in Figure 3.5 on page 19. From that graph it can be derived that the effective pixel area increases exponentially for larger Sensor Zenith Angles and that it is approximately 1.4 km^2, 2.5 km^2 and 6 km^2 for Sensor Zenith Angles of 30°, 45° and 60° respectively. The difference in effective pixel area is thus particularly large for a Sensor Zenith Angle for 60° and the difference between 45° and 60° is much larger (+3.5 km^2) than the difference between 30° and 45° (+1.1 km^2). The same can be concluded from the results presented in Figure 3.13, as the differences between 60° and 45° are larger than the differences between 45° and 30°.

Not only are the average values different, the shapes of the histograms also vary. In the sub-plot showing the number density, the histogram of 60° is shaped almost like a normal distribution, whereas the histograms of 45° and 30° are both skewed with the highest bin moving to the right. This is more extreme for the Sensor Zenith Angle set to 30° than to 45°.

3.4.2. Platform

The same sort of analysis can be done to show the difference between Terra and Aqua. Table 3.3 shows that for Aqua the number of evaluable days is higher for Sensor Zenith Angles of 30° and 60°. For 45° this is the other way around. The differences in the number of evaluable days become larger with increasing Sensor Zenith Angle. The results of the basic parameters for both Terra, Aqua and their combined values, are displayed in Figure 3.14.

This figure shows that, on average, the cloud cover, average nearest neighbour and average cloud size are higher for Terra. The number of clouds per unit area is larger for Aqua. As mentioned before, the results are similar when the threshold for Sensor Zenith Angle is changed, but some differences become larger with increasing Sensor Zenith Angle. Therefore the results of using different platforms with different Sensor Zenith Angles are summarized in Table 3.4. The differences of the Average Nearest Neighbour and the Number of Clouds per Unit Area are larger when comparing a Sensor Zenith Angle of 30° with 45° than when 45° is compared to 60°. This is different for the cloud cover and the average cloud size: not only are the different results of Terra and Aqua larger when choosing different Sensor Zenith Angles, the differences also increase for larger Sensor Zenith Angles.

Several studies on the results of the cloud cover detected by Aqua and Terra found that the values for Terra are slightly higher in December, January and February than the values found by Aqua. Minnis et al. (2004) for instance found that the general patterns in cloud cover are very similar for both satellites, but that

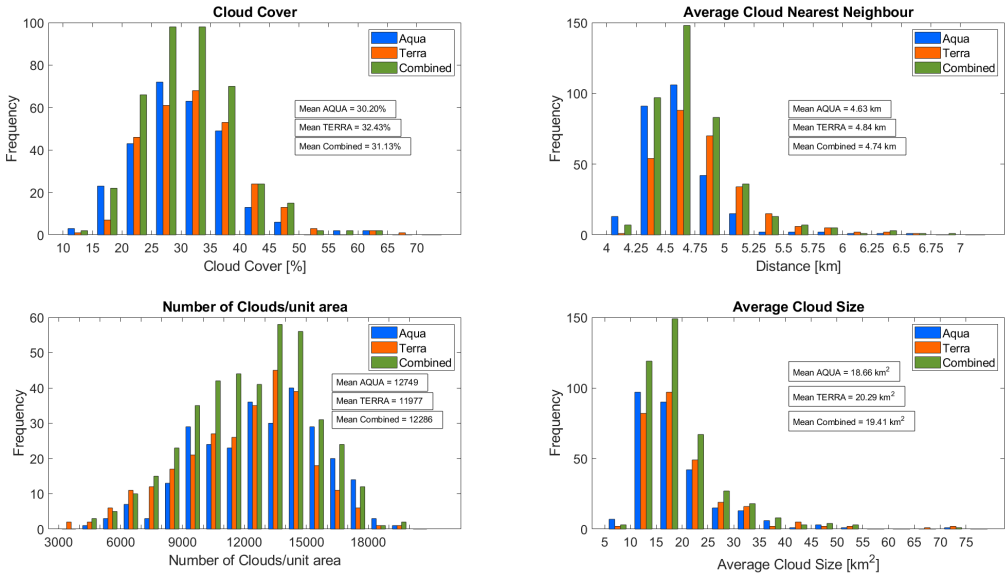


Figure 3.14: Results for Aqua, Terra and combined results. The threshold of the Sensor Zenith Angle is set to 45° here.

regions which are dominated by low-level marine stratocumulus clouds have a smaller cloud cover at 1330 local time (Aqua) than at 1030 local time (Terra). The same study showed that during daytime the cloud cover found by Aqua is generally smaller than the cloud cover detected by Terra. The exception is the inter-tropical convergence zone near 10°N. Since this is exactly where the area of interest is located the results found here do not match the conclusions presented in the paper of Minnis et al. However, a different study (King et al., 2013) shows that the differences between Aqua and Terra are different during different times of the year. From this more detailed study they concluded that the average cloud cover in the months December, January and February in the years 2002 - 2011 is slightly higher for Terra, which is also observed in the results presented here in the upper-left plot in Figure 3.14. Furthermore it was also stated that these and related results are consistent with climatology (Minnis et al., 2004).

These differences in cloud cover between Terra and Aqua can thus be expected and are unlikely to be caused by a bias in the data. Rather it is a physical difference between their times of overpass.

Table 3.4: Results of Terra and Aqua for different Sensor Zenith Angles.

		Terra	**Aqua**	**Percentual difference Terra and Aqua**	**Combined Cloudy**	**Combined Cloudy + Uncertain Clear**
Cloud Cover	**SZA = 30°**	30.33%	29.68%	-2.14 %	29.96%	38.66%
	SZA = 45°	32.43%	30.20%	-6.81 %	31.13%	40.58%
	SZA = 60°	37.01%	33.14%	-10.46 %	35.05%	48.60%
Average Nearest Neighbour	**SZA = 30°**	4.67 km	4.42 km	-5.35 %	4.55 km	4.28 km
	SZA = 45°	4.84 km	4.63 km	-4.34 %	4.74 km	4.46 km
	SZA = 60°	5.29 km	5.12 km	-3.21 %	5.28 km	4.99 km
Number of Clouds / Unit Area	**SZA = 30°**	12856	14371	+11.78 %	13522	16518
	SZA = 45°	11977	12749	+6.45 %	12286	14997
	SZA = 60°	9915	10304	+3.92 %	9750	12136
Average Cloud Size	**SZA = 30°**	20.87 km^2	17.20 km^2	-17.59 %	19.19 km^2	13.61 km^2
	SZA = 45°	20.29 km^2	18.66 km^2	-8.03 %	19.41 km^2	13.85 km^2
	SZA = 60°	24.31 km^2	24.24 km^2	-0.29 %	25.29 km^2	15.57 km^2

3.4.3. Cloud Classification

In this subsection the effect of the choice for cloud classification is investigated. The MODIS Cloud Mask classifies each pixel as *Cloudy*, *Uncertain Clear*, *Probably Clear* or *Confident Clear*, as explained in Chapter 3.2.1. It is likely that only using pixels classified as *Cloudy* results in an underestimation of cloud pixels and that using both *Cloudy* and *Uncertain Clear* pixels would lead to an overestimation of cloud amount. To investigate to what extent this choice influences the final statistics and which choice leads to the most realistic results, the values are compared. The histograms of the basic parameters are displayed in Figure 3.15. The results are also shown in Table 3.4 in the last two columns.

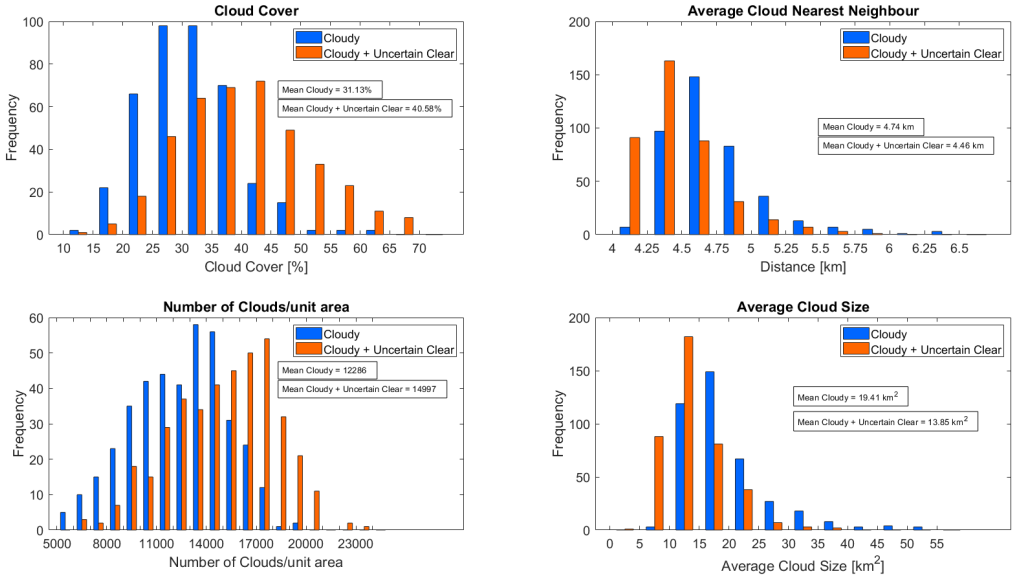
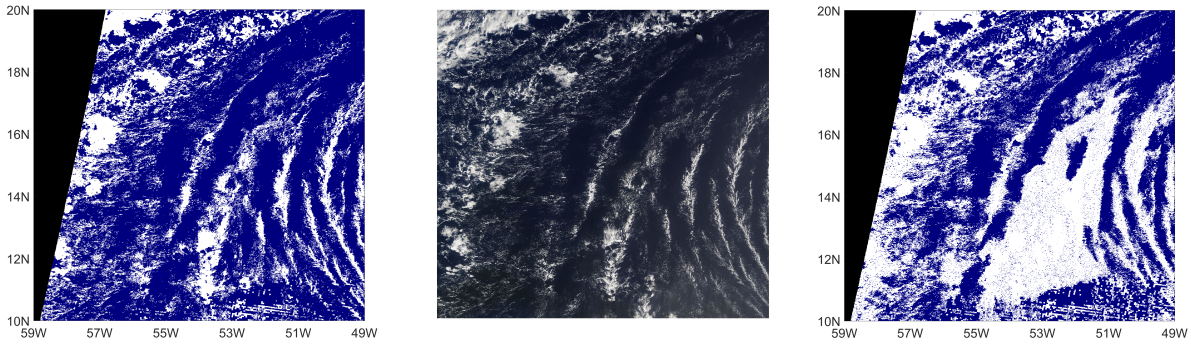


Figure 3.15: Results for different thresholds for Cloud Classification. The Sensor Zenith Angle threshold is set to 45° and a combination of results of Terra and Aqua are used.

The histograms clearly show some differences, which are mostly visible in the plots of the cloud cover and cloud density. For both parameters the histograms showing the results when both *Cloudy* and *Cloudy* + *Uncertain Clear* are considered clouds are shifted to higher values. The average values also are much higher: the cloud cover increases from 31.13% to 40.58% with factor 1.3 and the cloud density increases from 12286 to 28954 clouds / unit area even with factor 2.36. This difference is very large and can be explained (partly) by the way the number of clouds are determined: when clouds adjacent to the masked out area or the edge of the scene, they are removed from the dataset (see Section 4.1.1 and Equation (2.2)). Furthermore, their area is subtracted from the area that is left of the scene. As large clouds are more likely to be situated directly adjacent to the masked out area or edge and will therefore be removed, it is likely that the cloud density increases due to the greatly reduced area that is left. As clouds are often on average larger when pixels classified as *Uncertain Clear* are also considered cloudy, they are also more likely to be adjacent to an edge and subsequently be removed from the dataset. This would also explain the smaller average cloud size: large clouds have a bigger chance of being removed from the dataset and are thus not taken into account for the calculation of the average cloud size. Since the cloud cover increases with approximately a factor of 1.3 the multiplication of the number of clouds / unit area and the average cloud size should also be a factor 1.3 higher if no data was removed from the database. However, the multiplication leads to an increase of factor 1.68 and thus the fact that clouds are removed must have a large influence on the statistics.

When looking at the example of February 2, 2015 in Figure 3.16, the difference between only using the MODIS Cloud Mask *Cloudy* (a) pixels or using both *Cloudy* and *Uncertain Clear* (c) becomes clear. As a reference a snapshot of MODIS Worldview is added (b). In this case there is a clear difference between (a) and (c), and only using MODIS Cloud Mask *Cloudy* as clouds gives a better approximation of (b), which shows the corrected reflectance, also referred to as *true color* because it is similar to what the human eye would see. Comparing other days gave a similar results, with the differences sometimes being more clear than other times, but only using *Cloudy* as cloud pixels always results in an image that visually matches the MODIS worldview image best.



(a) Only MODIS Cloud Mask *Cloudy* pixels as clouds.

(b) Snapshot from MODIS Worldview.

(c) MODIS Cloud Mask *Cloudy* and *Uncertain Clear* pixels as clouds.

Figure 3.16: Differences in cloud classification on February 2, 2015 (TERRA) when only using MODIS Cloud Mask *Cloudy* pixels as clouds (a) or also considering MODIS Cloud Mask *Uncertain Clear* as clouds (c). A snapshot from MODIS Worldview (b) is added to compare.

3.4.4. Conclusions on threshold implications

Based on the results presented above a decision must be made regarding further data evaluation. A trade off must be made between the amount of data that can be used and the representativeness of this data. Using data representing very different effective pixel sizes is not desirable. On the other hand it is important to have as much data as possible for further analysis: a trade off must thus be made. The results further presented in this chapter therefore are processed using **combined data of Terra and Aqua**. The threshold for the **Sensor Zenith Angle is set to 45°** and **only pixels classified as *Cloudy*** by the MODIS Cloud Mask are considered cloudy. Some of these cloudy pixels may be classified as high clouds for which a **threshold of 550 hPa** is used. Using these thresholds results in a total of 403 days that are analysed (see Table 3.3).

4

Methodology

This chapter provides a short overview of how existing methodology is applied in this research, and which new methodologies are applied and incorporated.

4.1. Cloud Characteristics

After all processing steps are completed, the remaining data to be examined is a binary matrix with a value of 1 for shallow cumulus clouds and 0 for clear sky, high clouds or no data pixels. Several methods to characterize clouds were examined. In the first subsection the methodology to obtain the basic parameters of the cloud field is explained. The second subsection explains how the organization index, referred to as OI_3, is obtained.

4.1.1. Basic parameter analysis

The basic parameters were described in Section 2.2.1. The cloud cover (Eq. 2.1), cloud number density (Eq. 2.2), average cloud size (Eq. 2.3) and average nearest neighbour distance (Eq. 2.6) are applied to all evaluable days after applying the high cloud mask and Sensor Zenith Angle mask as was explained in Sections 3.3 and 3.4. They will be applied to the complete area of interest that spans $10° \times 10°$, and on top of that also to smaller sub-grids of $5° \times 5°$ and $2.5° \times 2.5°$. The basic parameters are used to test the correlation between them and the final results of a combined organization index OI_3, which will be explained in the next section.

Determining all the required properties for the basic cloud parameters the *MATLAB* function *regionprops* can be used. This function is designed to return properties for all 8-connected components in an image or matrix. This implies that cloud pixels are assumed to be part of one cloud when they are directly horizontally, vertically or diagonally adjacent to each other. Of each cloud, the *Area*, *Centroids* and *Pixel indices* are obtained using this function. From these parameters, all basic parameters can be determined.

The pixel indices of each cloud are necessary to determine whether a cloud is adjacent to the edge or no-data/masked out area and should thus be removed from the dataset when determining the cloud number density, the average cloud size and the average nearest neighbour. This approach in which the pixel indices are used, is based on the following: each pixel has a number, the most upper left pixel in a matrix gets value 1, the pixel below gets value 2 and the pixel to the right gets the value equal to the sum of the vertical length of the matrix + 1. A database is created with the pixel indices of the pixels at the edges and the pixels directly adjacent to the no-data/masked out area. If one or more pixel indices of a cloud match at least one of these pixels in the database, the cloud is removed from the arrays storing the area and centroids, and its area is subtracted from the entire area containing data. The pixels making up this cloud are thus considered as no-data area. This must be done because the clouds that are "cut-off" do not have representative sizes and centroids for the entire cloud, which was explained in Section 2.2.1.

Determining these basic parameters is relatively simple and provides us with better insights in how they are related to the organization of a cloud field.

4.1.2. Organization Index OI_3

As discussed already in Chapter 2.2 there are various methods to assess cloud organization. For this thesis the method for finding the organization index I_{org} introduced by Benner and Curry, 1998, will be applied in combination with parameters that provide measures for the size of clear sky areas and the largest cloud

size in a field. Determining the I_{org} involves the simulation of a random cloud field, taking into account the number and sizes of the actual cloud field. In this section the methods for finding the I_{org} and the measures for the clear sky areas and the largest cloud size are discussed seperately. The last subsection discusses how these separate parameters are combined in order to find the value of the organization index OI_3.

Calculating I_{org}

Although the procedure of creating a randomly generated inhibition NNCDF is existing and was applied by Benner and Curry 1998, the methodology to create such a random field can not simply be copied, nor was it explained explicitly in literature. Therefore, the applied methodology to obtain such a field is discussed here.

The first step is to find the number of clouds and their sizes of both shallow cumulus clouds and high clouds. On top of that the centroids of the shallow cumulus clouds must be determined as well. The high clouds are also considered in the simulation, as high clouds also imply an inhibition for shallow clouds to be placed in this region. Therefore two binary input matrices are created: one for the shallow cumulus clouds and one for high clouds. The clouds (shallow cumulus or high) are denoted with a 1 and all other pixels get value 0 and can thus represent no-data areas, areas with high sensor zenith angle or high (shallow cumulus) clouds for the input matrix representing shallow cumulus (high) clouds. An example of an input matrix for shallow cumulus clouds is shown in Figure 3.12 on page 25. The input matrix for the high clouds is obtained after applying the sensor zenith angle mask in Figure 3.11 (right) to Cloud Top Pressure mask representing high clouds (middle) on page 25.

The number of clouds, their sizes and for the shallow cumulus cloud matrix their centroids are determined using the *MATLAB* function *regionprops*. The function determines the centroids as in Equations (2.4) and (2.5). The *Pixel indices* are not obtained, as clouds adjacent to the edge or no-data area are not removed to the dataset.

Eventually all clouds (both shallow cumulus and high clouds) must be placed randomly in a field. To get the most realistic simulation of a random field with similar cloud properties (number of clouds and sizes) as the actual field, clouds are not allowed to overlap in the simulated field. Therefore the largest clouds are placed in the field first, which gives the highest chance of placing all clouds in the field. For that, all clouds are stored in one matrix in the order of largest to smallest. Since using the exact shapes of the clouds is computationally complicated the clouds are simplified to circles. Their radii are calculated under the assumption that the total area of the clouds is a circle, they are thus calculated using the following equation:

$$R_i = \sqrt{\frac{A_{cloud,i}}{\pi}} \tag{4.1}$$

In which R_i is the radius of pixel i in pixel dimensions and $A_{cloud,i}$ is the size of this cloud expressed in number of pixels. The radius of each cloud is stored in column 1 of the matrix and an indicator for shallow cumulus (value 1) or high clouds (value -1) is stored in column 2. The input for the simulation is thus one matrix with dimensions of the number of clouds × 2, in which the number of clouds is the sum of the shallow cumulus clouds and the high clouds.

Before they can be placed in a random field the size of this field must be determined. Since it is desired to use periodic boundary conditions the field must be rectangular and can not have the same boundaries as the original data left after applying the sensor zenith angle mask. Therefore the squared root of the entire area left after applying the sensor zenith angle mask is calculated. If this value is within 0.25 km from an integer, the value is rounded up or down to this integer and a square field is created. If the value has a larger deviation from an integer, a rectangle with dimensions of the squared root rounded up × the squared root round down is created. This is also the reason that clouds adjacent to the edge or no-data area do not need to be removed for this calculation: the number of clouds and their sizes in the random field are the same as in the input data, as well is the size of the field.

Looking at the example used before (Aqua December 07, 2015), the area that remains after applying the Sensor Zenith Angle mask with a threshold of 45° is 853907 km^2, and the squared root is 924.0709 km. Since this is only 0.0709 km away from its nearest integer 924, an empty square matrix of 924 × 924 km is created. This results in a field with a total area of 853776 km^2.

Furthermore, two empty databases for the x and y coordinates of the clouds are created. These matrices have dimensions of the number of clouds by 9. This is done to account for the periodic boundary conditions. Not only the simulated location of each cloud in the actual matrix is stored, but also its coordinates of 8 locations on 8 similar matrices around the actual one. This way, 9 positions of the cloud are stored and

periodic boundary conditions can be implemented in further calculations. An example of a simple 5 x 5 field with 3 clouds is shown in Figure 4.1.

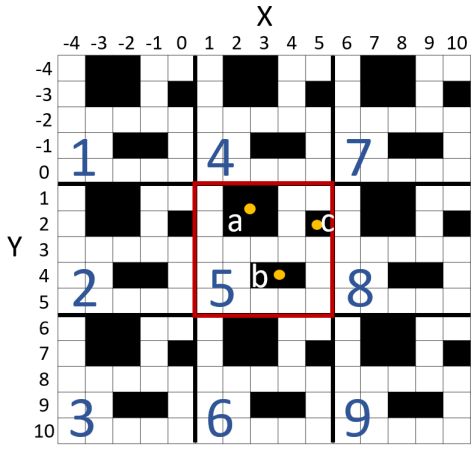


Figure 4.1: The main field is located in the middle, with red outside borders. The orange dots show the centroids. Cloud a is located at (2.5, 1.5), cloud b at (3.5, 4) and cloud c at (5,2). Each field has a number that corresponds to the column of the matrix in which the data is stored.

This field contains 3 clouds with different sizes. The actual field is located in the middle and 8 similar fields are placed around it. To account for periodic boundary conditions, the coordinates of the centroids of each cloud in all fields must be stored. Table 4.1 shows the way of storing each of these coordinates, using the clouds in Figure 4.1 as an example. The clouds are placed and therefore also stored from largest to smallest. To have 9 unique coordinate combinations for each cloud, each column gives the x and y coordinates of one of the fields numbered from 1 to 9. Column 5 thus contains the coordinates of the field in which the simulation takes place.

Table 4.1: Here the coordinates of clouds a, b and c, located at coordinates (2.5, 1.5), (3.5, 4) and (5,2) respectively, are stored in a similar way as will be done to execute the random cloud simulation.

Cloud	**x-coordinates**								
a	-2.5	-2.5	-2.5	2.5	2.5	2.5	7.5	7.5	7.5
b	-1.5	-1.5	-1.5	3.5	3.5	3.5	8.5	8.5	8.5
c	0	0	0	5	5	5	10	10	10
	y-coordinates								
a	-3.5	1.5	6.5	-3.5	1.5	6.5	-3.5	1.5	6.5
b	-1	4	9	-1	4	9	-1	4	9
c	-3	2	7	-3	2	7	-3	2	7

After the creation of these empty matrices the simulation is started. First the largest cloud is placed on a random location using the random value generator of *MATLAB*: *rand*. *Rand* returns a single uniformly distributed number on the interval (0,1). The value is therefore multiplied by the dimension of the matrix in that direction and subsequently rounded to an integer. The x and y coordinates are independently generated. If a coordinate is zero, a new value is generated for this coordinate. The independently and uniformly randomly generated location in the matrix gets the indicator value stored in column 2. The location of this cloud and its locations in the matrices around the actual field are stored in the parameters for the x and y coordinates as described above. The radius of this cloud is stored in another parameter. The combination of the centroids of the cloud in all fields and the radius provides information on the spatial extent of the cloud, given it is a circle.

Now the process of generating random x and y coordinates is now repeated for all clouds in the list. One by one, from largest to smallest, they are placed in the matrix. If the randomly created coordinates assigns the cloud to the exact same location as a previously placed cloud, new x and y coordinates are created. When this is not the case, the distances to all clouds placed before are calculated. This is done by taking the squared

root of the distances in the x and y directions stored in the parameters for the x and y coordinates. Periodic boundary conditions are thus taken into account. If at least on of these distances is smaller than the radius of the new clouds + the radius of that cloud, the clouds would overlap if the new cloud would be placed in this location. This is not allowed and therefore the location is rejected and the process is iterated. If the placement of the cloud was not successful after 10000 iterations, the whole process starts again with placing the first cloud in the matrix. For this, the maximum amount is 10 attempts: after that a message pops us saying it is not possible to create a random simulation, the I_{org} is set to "Not a Number" and the program continues to the next day. The process of the simulation of a random cloud field is summarized in Figure 4.2.

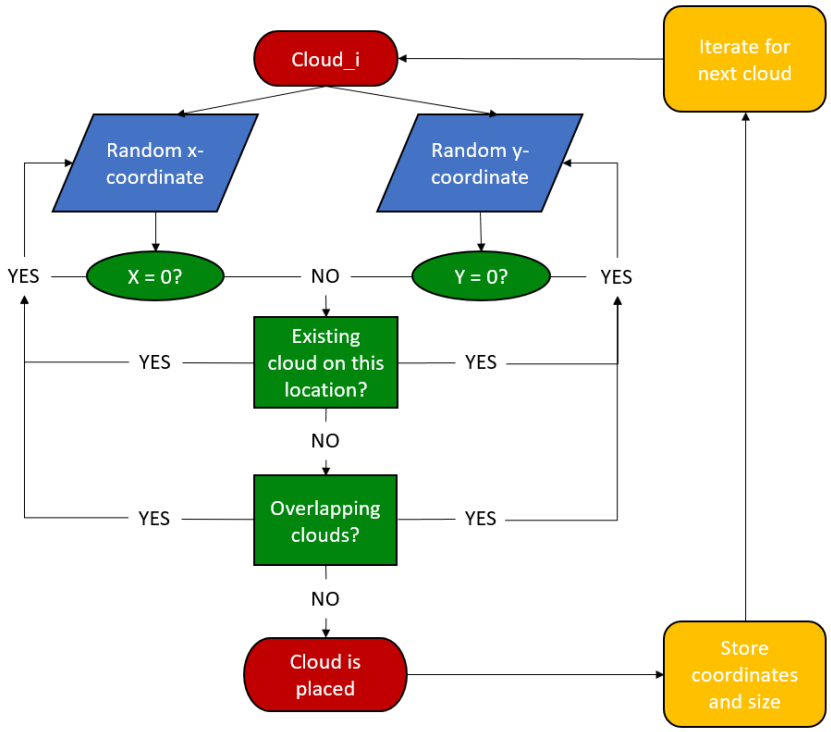


Figure 4.2: Random field simulation flowchart

An example of such a randomly generated field is shown in Figure 4.3. Each white circle represents a shallow cumulus cloud, each gray circle a high cloud and black pixels represent clear sky area. One shallow cumulus cloud is coloured red to highlight the implications of using periodic boundary conditions. This cloud was placed randomly in the upper left corner and since its radius is larger than the distances to both the y and x axes the remaining part of the cloud appear at all other corners as well.

Figure 4.4 shows the matrix with the associated centroids of the shallow cumulus clouds of this random field. These centroids are denoted as white pixels. All other pixels are black and may represent the extent of shallow cumulus or high clouds, or it can be clear sky. Since this is not of important for the algorithm to determine the nearest neighbours, this matrix is used as the input for the nearest neighbour calculation as described in Equation (2.7). Periodic boundary conditions are taken into account.

Once the nearest neighbours are found they are binned in groups of equal nearest neighbour lengths. The cumulative sum of the frequencies in all bins is calculated and divided by the total number of clouds to find the $NNCDF_{ran}$ values in the range [0,1].

The same is done with the original data to find $NNCDF$, although it is not possible here to take periodic boundary conditions into account due to the dimensions of the data that is left after applying the sensor zenith angle. The bin ranges are defined by the unique values of the nearest neighbours of both the original data and the simulated data. The $NNCDF$ and $NNCDF_{ran}$ can now be plotted against each other, which is done in Figure 4.5.

The graph directly visualizes the clustering of the field. In this example the line is clearly plotted above the diagonal and thus the field can be considered as clustered. To what extent the field is clustered is given by the I_{org}. The Organization Index I_{org} is the area under the curve, which can well be approximated by *MATLAB*

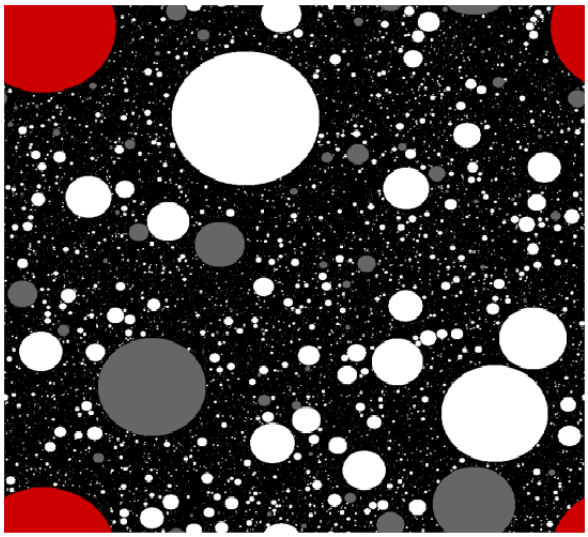


Figure 4.3: Example of a randomly generated field using data from AQUA on December 7, 2015. White (and red) circles represent shallow cumulus clouds, gray circles are high clouds.

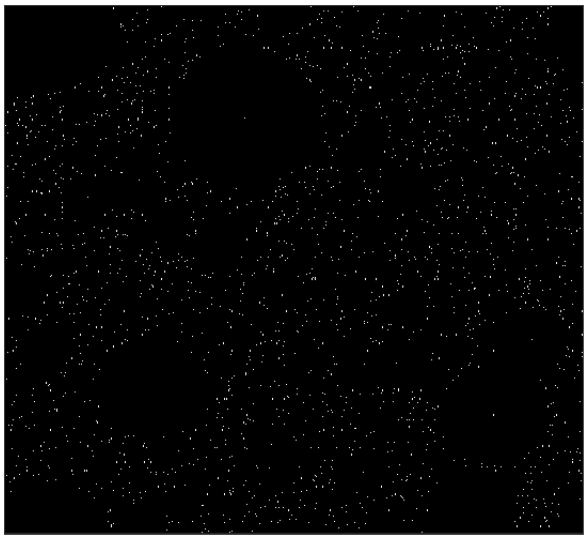


Figure 4.4: The centroids of the randomly generated field in Figure 4.3. White pixels represent centroids of shallow cumulus clouds. Black pixels represent either the extent of a shallow cumulus or high cloud, or clear sky.

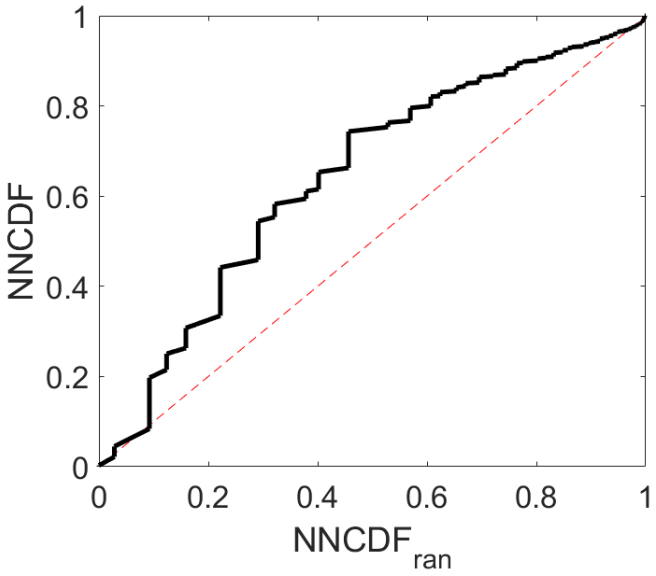


Figure 4.5: NNCDF of December 7, 2015, versus the random inhibition $NNCDF$. The diagonal is denoted by the red dashed line, the black curve shows the result of the $NNCDF$ vs the $NNCDF_{ran}$.

using the function *trapz*. This function computes the approximate integral of the $NNCDF$ with respect to the coordinates of $NNCDF_{ran}$ by making use of the trapezoidal method, which is given by:

$$\int_a^b f(x)\mathrm{d}x \approx \frac{1}{2}\sum_{n=1}^{N}(x_{n+1}-x_n)[f(x_N)+f(x_{n+1})] \tag{4.2}$$

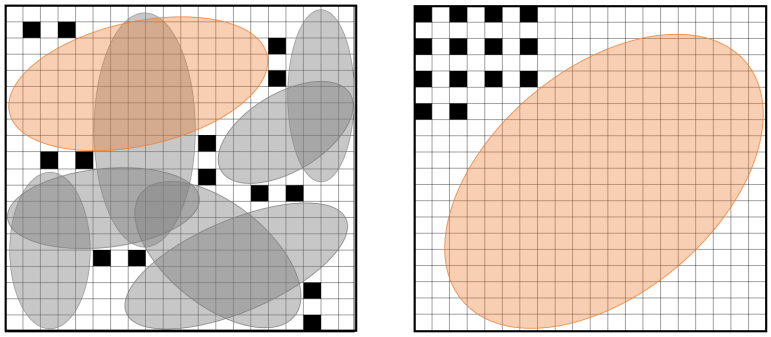
Here $a =< x_1 < x_2 < ... < x_N < x_{N+1} = b$ and $x_{n+1} - x_n$ is the spacing between each consecutive pair of points.

The organization index found for the example here is 0.643.

Clear sky area parameter

A measure for the clear sky area is introduced as solely looking at the cloud size and nearest neighbours

and subsequently calculating the I_{org} does not provide a complete indication of organization. To illustrate this problem, two fields containing the exact same amount of clouds with the exact same size and the exact same nearest neighbour values are shown in Figure 4.6. Theoretically these fields thus have the same value of I_{org}, which intuitively is incorrect. The field on the left (Fig. 4.6a) does not seem extremely organized as clouds are distributed over the entire area, whereas the field on the right (Fig. 4.6b) looks very clustered. Nevertheless, the values of I_{org} are similar, due to the similar size of the field, number and size of clouds and nearest neighbour values. In order to distinguish between these two fields, a measure of the largest clear sky area in the field is defined.



(a) Small clear sky area. Different clear sky areas can be distinguished in this case and these areas may also overlap. The orange coloured shape denotes the approximation of the largest clear sky area.

(b) Large clear sky area. In this case there is only one large clear sky area that is approximated by the orange coloured shape.

Figure 4.6: Examples of very similar fields but with different clear sky areas. (a) contains a field that intuitively is not so clustered. (b) shows an example of a field with the same amount of clouds that have the same size and the same nearest neighbour values, but is intuitively much more clustered than (a). In both figures orange and/or grey coloured shapes show approximations of clear sky areas. The orange shape shows the largest approximate clear sky area of the field, which is larger in (b) and thus this field is considered more clustered.

The clear sky areas are difficult to define, as commonly all non-cloud pixels are connected and should in that sense thus be considered as one clear sky area. However, this is not the value we are interested in, as this is also similar for both fields in Figure 4.6. Therefore a different approach is used to find an approximation of clear sky regions in the field. The largest clear sky region is then used as an additional measure for organization/clustering. Some examples of such clear sky regions are denoted with the orange and/or grey coloured shapes in the two fields in Figure 4.6. From the field on the left it becomes clear that these clear sky regions may overlap. However, the aim is not to connect all non-cloud pixels and thus all these shapes represent separate clear sky regions. The field on the right only contains one large clear sky region, as all clouds are clustered in one area. The orange shapes show approximations of the largest clear sky region in the field.

As a higher value of I_{org} indicates more clustering, a larger clear sky region does the same. The clear sky area parameter is defined as the largest clear sky area divided by the area that is left after applying the Sensor Zenith Angle mask:

$$A_{ClearSky,max} = \frac{\sqrt{max\{A_{ClearSky}\}}}{\sqrt{A_{left}}} \tag{4.3}$$

In which $A_{ClearSky,max}$ is the parameter (ratio) in the range (0,1), $max\{A_{ClearSky}\}$ is the largest clear sky area in the field and A_{left} is the area that is left after applying the Sensor Zenith Angle mask.

It is not possible to simply determine the largest clear sky area by counting all connected clear sky pixel. If that would be done, it would be likely to find only one, very large, clear sky area as most clear sky pixels are connected to each other. In stead we are looking for an approximation of the largest clear sky region, as was illustrated in Figure 4.6. Looking at the field in Figure 3.12 on page 25 we would like to find an approximation for the black area ranging roughly between 14-16N and 51-55W as this can be considered as one region.

To find this approximation the horizontal (East-West) and vertical (North-South) distance from each pixel representing clear sky to the first cloud or no-data pixel is determined. This is done for vertical and horizontal directions separately, but the maximum vertical or horizontal distance is found when a cloud / no data pixel is determined in one direction (either East or West, or North or South). Multiplying these horizontal and

vertical distances then gives the approximation of such a clear sky area. An example cloud field is displayed in Figure 4.7.

Here the green square represent the pixel that is the centre of the largest clear sky region. The blue arrows show the maximum distance in all directions. However, as this maximum distance is restricted to the shortest distance in the horizontal and vertical direction, the actual maximum ranges are smaller and are denoted with red dashed arrows. For the horizontal distance in this example the edge is encountered. Ideally periodic boundary conditions would be applied but since in practice the analysed data always misses some data (the black large area on the left in Figure 3.12) this is in practice not possible for the analysis in this region of the world. For the vertical distance the maximum range is reached because a cloud is encountered. The horizontal maximum range here is $6.5 + 6.5 = 13$ pixels. The vertical maximum range is $4.5 + 4.5 = 9$ pixels. Therefore the rectangle representing the size of this clear sky ara is $13 \times 9 = 117$ pixels. Making use of Equation (4.3), and thus dividing this area by the total area of the field, results in the following calculation:

$$A_{ClearSky,max} = \frac{117}{320} = 0.366.$$

In practice these values are typically somewhat smaller: values between 0.05 and 0.25 are most common. Figure 4.8 shows the results of the calculation in Equation 4.3 for each pixel representing clear sky area.

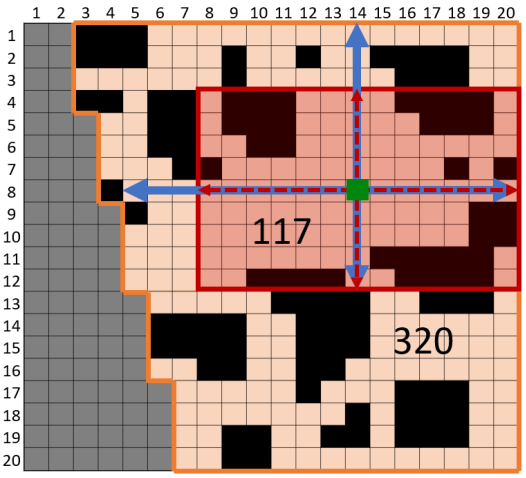


Figure 4.7: Largest clear sky area example. Black pixels represent clouds and white pixels represent clear sky areas. The green square shows the pixel with the largest horizontal and vertical distances to clouds/edges. The blue arrows show the maximum distance in all directions and the red dashed arrows show the maximum distance in vertical and horizontal direction, restricted by the shortest distance calculated separately for the horizontal and vertical directions.

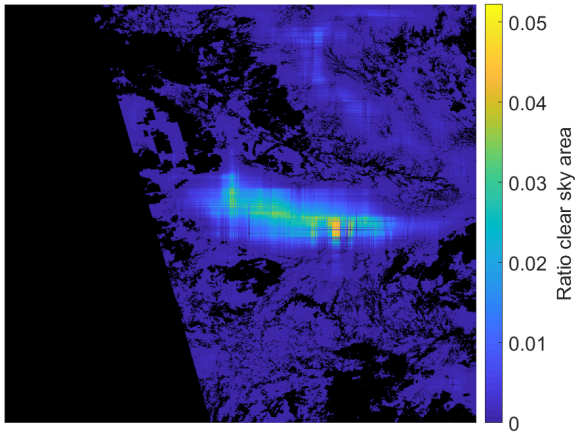


Figure 4.8: Largest clear sky area on December 7, 2015. The black areas represent no-data areas, masked out areas, shallow cumulus clouds or high clouds. For all other pixels the maximum clear sky area is calculated according to the method illustrated in Figure 4.7.

In this Figure shallow cumulus clouds, high clouds and no data / too large Sensor Zenith Angle pixels are coloured black. For all other pixels the analysis is executed. Most clear sky pixels are located in a small clear sky region spanning less than one percent of the entire scene and are thus coloured blue. The clear sky region visible in the final input data product visualized in Figure 3.12 on page 25 can also be detected here, as the ratios are clearly larger here. The largest value found in this scene on December 7, 2015, is 0.2286, which is relatively large.

Largest cloud size parameter

This parameter is introduced as large clouds are associated with clustering of clouds. Furthermore, with the presence of large clouds the freedom of movement for small clouds in the simulation to obtain I_{org} decreases which tempers the value of I_{org}. To correct for this tempering a measure for the largest cloud in the field is taken into account: a larger maximum cloud size relatively increases the value of the final combined measure and in this way reduces the effect of these large clouds on the final organization index value. The largest cloud size is defined as the largest detected cloud in the field, divided by the total area that is left after applying the Sensor Zenith Angle mask:

$$A_{cloud,max} = \frac{\sqrt{max\{A_{cloud}\}}}{\sqrt{A_{left}}} \tag{4.4}$$

In which $A_{cloud,max}$ is the ratio in the range (0,1), $max\{A_{cloud}\}$ is the largest cloud in the field and A_{left} is the area that is left after applying the Sensor Zenith Angle mask.

A fully covered scene would have value 1, however in practice this will not occur for scenes as large as analysed here. In stead the largest cloud will cover a few percent of the total scene and thus have a value of typically 0.1 - 0.5. The ratio of the largest cloud in the scene in Figure 3.12 on page 25 is 0.2261.

The cloud sizes are obtained using the build in *Matlab* function *regionprops*. The *Area* is returned as a scalar and represents the number of pixels of each cloud. For simplicity, it is assumed that all pixels have the same dimensions: their increasing size with increasing *Sensor Zenith Angle* is not accounted for. The largest returned scalar is the largest cloud in the scene.

Combined Measure OI_3

The above described parameters are then combined to one measure. To give them approximately equal weight, they are altered slightly and subsequently multiplied. In theory all three measures range between (0,1), however in reality we are analysing fields of 10° × 10° and therefore values are unlikely to reach the minimum or maximum which only occurs when there are no clouds or when the area is completely covered in clouds. The results, presented in Chapter 5, showed that the I_{org} is never lower than 0.5, but instead ranges between (0.55-0.75). Since the Clear sky parameter and Largest cloud parameter are typically much lower and also have a relatively larger range, I_{org} is rescaled by subtracting 0.5 from the value and then multiplying the result by 2. That way, values typically range between 0.1 - 0.5, which is much closer to the typical values of the other parameters.

The proposed organization index here, taking all three measures into account, is then calculated as follows:

$$OI_3 = 2 \times (I_{org} - 0.5) \times ClearSky \times LargestCloud \tag{4.5}$$

December 7, 2015, was used as an example to explain the individual parameters. These results should be combined to find the combined organization measure OI_3. This then equals:

$$OI_3 = 2(0.643 - 0.5) \times 0.2286 \times 0.2261 = 0.0148$$

Typically, values for OI_3 are within 0.003 and 0.03. Note that whenever a cloud field would be regular and I_{org} would thus have a value between 0 and 0.5, OI_3 would be negative.

4.2. Analysis of the Cloud Characteristics

The obtained data can be analysed in various ways. In order to obtain information on the spatial and temporal variability we make use of the variance parameters within the field and the autocorrelation of the values. The methods to obtain these values are discussed here.

4.2.1. Temporal Variability

To obtain an idea about the temporal variability of a parameter we can make use of the autocorrelation. This is a measure of the correlation between y_t and y_{t+k}, where k represents a lag with respect to t between 0 and K. The autocorrelation for lag k is given by:

$$r_k = \frac{c_k}{c_0} \tag{4.6}$$

In which c_0 is the sample variance and c_k equals:

$$c_k = \frac{1}{T} \sum_{t=1}^{T-k} (y_t - \overline{y})(y_{t+k} - \overline{y}) \tag{4.7}$$

Where T is the effective sample size and $\overline{y}$ is the mean value of the time series. Since not all days can be evaluated, the effective sample size is different for each lag. The sample variance c_0 is defined as:

$$V = \frac{1}{N} \sum_{i=1}^{N} |A_i - \mu|^2 \tag{4.8}$$

Where N is the number of analysed days, A is the vector containing the values of the analysed days and μ is the average of A.

The autocorrelation is calculated separately for each winter for lags $k = 0, ..., 20$. This way an idea of the temporal variability of the cloud characteristics can be obtained. A high autocorrelation for multiple lags indicates that days at lag k are correlated to k days before that day. A negative autocorrelation at lag k indicates a negative correlation with k days before.

4.2.2. Spatial Variability

The spatial variability can be measured by calculating the variance of the cloud characteristics of the field. Because all parameters are calculated for the complete field that spans $10° \times 10°$, but also for smaller individual areas within this complete field of $5° \times 5°$ and $2.5° \times 2.5°$, we can look at the deviations from these smaller areas to the mean value of the complete area. The sample variance is given by Equation (4.8), but in this case N equals the amount of areas that contain data on that day. As was explained in Section 4.1.1 not always all areas contain data, as they do not comply with the requirements. Therefore, N is different for each day.

5

Results and Discussion

The results are presented in this chapter. Data of both Terra and Aqua was analysed for the months December, January and February between December 2010 and February 2018. This means that 90 days were analysed for 8 years and for two platforms. In total this provides us with $90 \times 8 \times 2 = 1440$ possible datasets, divided over 720 days, to use for further evaluation. In Section 3.4 the effects of setting different thresholds for data selection is shown. Based on a short conclusion the most suitable thresholds were selected and applied to the data for further evaluation. The results are presented in Section 5.1 and discussed in Section 5.2.

5.1. Cloud Characteristics and Organization

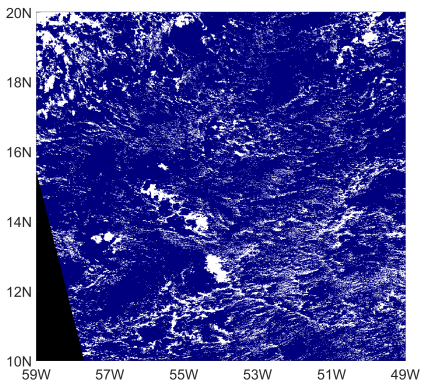
The resulting days after setting the thresholds as discussed in Section 3.4.4 can now be evaluated further. This section presents the results of the different cloud characteristics such as cloud cover, average nearest neighbour distance, number of clouds, average cloud size and the combined organization index OI_3. Since the focus of this thesis is to find an effective cloud organization parameter, the performance of OI_3 is discussed first in a separate section. In the other sections the results on the parameters, their correlations and possible trends are presented. The parameters are determined for the complete area of interest of 10° × 10° and for smaller areas. These results are discussed separately.

5.1.1. Performance OI_3

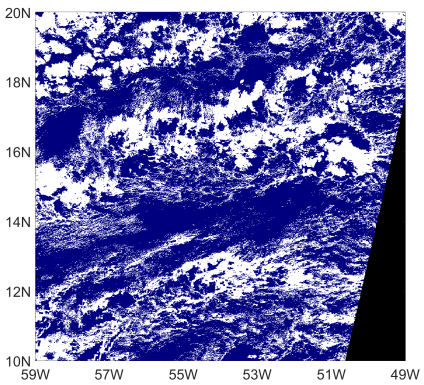
Assessing the performance of OI_3 is a challenge, as the question remains what the definition of a *good* value is to asses the organization. However, a comparison of the obtained results of OI_3 to a subjective classification of the area into the classes *Sugar*, *Gravel*, *Fish* and *Flowers* shows promising results. Appendix A contains all fields (557) that were evaluated after using the Sensor Zenith Angle mask of 45°. The fields are ordered from smallest to largest organization according the the value of OI_3. From these plots a good general idea of the performance of OI_3 can be obtained. When a subjective analysis would be done, it is likely that the order would not be very different (although this is not tested for this specific dataset), because fields with a low OI_3 value have many small clouds scattered throughout the entire area and fields with higher values have less, but larger clouds and larger clear sky areas in between. Based on this judgement, OI_3 can be considered successful in ordering fields from least to most organised. Days would be ordered differently when only I_{org} would be considered or if it would be based on *SCAI*. This comparison is summarized in Appendix B. The values of I_{org} and $SCAI$ correlate to OI_3, which is shown later in this section, but there are some significant differences and in these cases the performance of OI_3 often correlates better with the subjective analysis.

Figure 5.1 shows six examples of cloud fields with different OI_3 values (increasing from (a) to (d), fields (e) and (f) have similar values for OI_3).

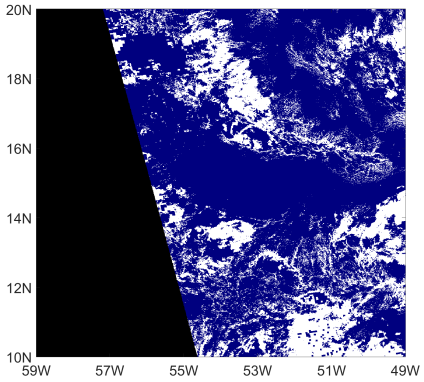
Cloud field (a) is, according to the value of OI_3, one of the least clustered fields that has been analysed. With many small clouds that do not seem clustered when subjectively inspected, no large clouds and no large clear sky regions this is also expected. From the subjective classification this field is classified as a mixture *Sugar* and *Gravel* (Stevens et al., 2019). Cloud field (b) is slightly more organized. On average clouds are larger. This field is also classified as a mix of *Sugar* and *Gravel* (Stevens et al., 2019). Cloud field (c) is the cloud field that has been used in Chapters 3 and 4 to illustrate large parts of the data (pre)processing and the methodology. This field contains some large clouds and little "*Sugar*" clouds. On top of that also a large clear



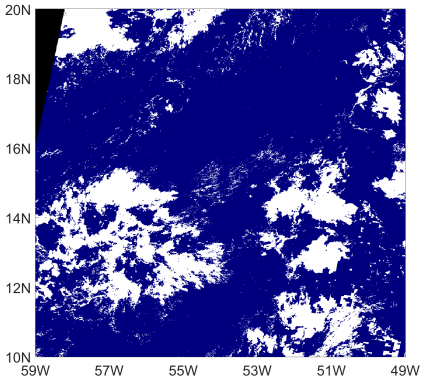
(a) Clouds on February 14, 2013 (AQUA). OI_3 is 0.0014.



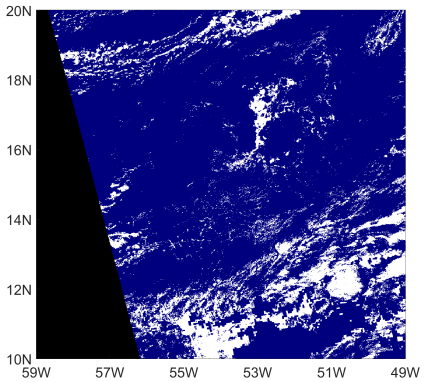
(b) Clouds on December 22, 2011 (TERRA). OI_3 is 0.0051.



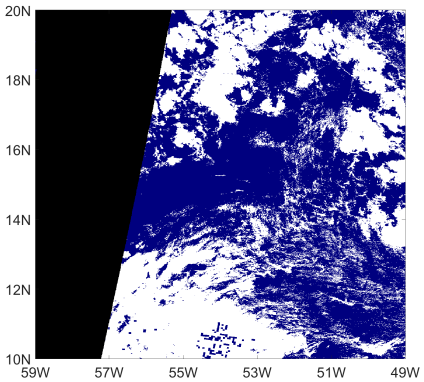
(c) Clouds on December 7, 2015 (AQUA). OI_3 is 0.0148.



(d) Clouds on February 14, 2017 (TERRA). OI_3 is 0.0399.



(e) Clouds on February 23, 2013 (AQUA). OI_3 is 0.0200.



(f) Clouds on February 5, 2013 (TERRA). OI_3 is 0.0201.

Figure 5.1: Cloud fields with different values for the combined organization measure OI_3. Clouds are coloured white, clear sky is coloured blue and black pixels denote either no-data areas or pixels with a sensor zenith angle larger than 45°. The OI_3 values are increasing for the cloud fields (a) to (d), with respectively values of 0.0010, 0.0051, 0.0148 and 0.0399. Cloud fields (e) and (f) respectively have OI_3 values of 0.0200 and 0.0201.

sky area can be detected in the centre. Therefore this cloud field can be considered more clustered than cloud fields (a) and (b). Stevens et al. (2019) classified this field as *Flowers* and *Sugar*. Cloud field (d) shows both larger cloud structures and larger clear sky regions than cloud fields (a)-(c). Also, the amount of very small "*Sugar*" clouds has highly reduced. Therefore, all parameters have high values and as a result this field has a very large value of OI_3. This field has not been analysed by Stevens et al. (2019), but does fit the description of *Fish* best due to the large cloud structures and large clear sky areas.

Fields (e) and (f) have similar values of OI_3 (0.0200 and 0.0201), but look very different. These two exam-

ples show that a similar OI_3 does not necessarily mean that all individual parameters have the same values. In the study of Stevens et al. (2019), five out of six people classified these fields into the class *'Undefined'*, which is for example used when there are too many high clouds or when the field contains too many different organization types.

The individual values of I_{org}, the *Clear Sky Parameter* and the *Largest Cloud Parameter* of all six fields are summarized in Table 5.1.

Table 5.1: OI_3 and individual parameter values for 4 fields

Field	OI_3	I_{org}	Clear Sky Parameter	Largest Cloud Parameter	SCAI
a	0.0014	0.5806	0.1454	0.0592	0.2225
b	0.0051	0.5834	0.1075	0.2826	0.1604
c	0.0148	0.6433	0.2286	0.2261	0.1781
d	0.0399	0.7274	0.3164	0.2774	0.0811
e	0.0200	0.7068	0.2777	0.1743	0.1205
f	0.0201	0.6418	0.1669	0.4251	0.1344

In this Table the fields (a) to (d) are ordered from smallest to largest value of OI_3 with corresponding cell colours (from dark red to yellow). However, this order may be different for the individual parameters. The order of the other parameters becomes clear from the cell colours. The differences between the individual parameters for these fields are discussed here.

For instance, I_{org} of fields (a) and (b) is almost similar, and the *Clear Sky Parameter* is even higher for field (a). However, since the *Largest Cloud Parameter* is much higher for field (b), the value of OI_3 for field (b) is almost 4 times higher than the OI_3 of field (a). Comparing fields (b) and (c) the opposite is true: the *Largest Cloud Parameter* is higher for field (b) than for field (c), but since field (c) contains a large clear sky region and I_{org} is higher, OI_3 of field (c) is significantly higher than field (b). Field (d) has a significant higher OI_3 value than the other fields and also its I_{org} and *Clear Sky Parameter* values are higher. The *Largest Cloud Parameter* however is still smaller than the one of field (b).

The possible differences for individual cloud fields becomes even more clear when comparing fields (e) and (f). The values of I_{org}, the *Clear Sky Parameter* and the *Largest Cloud Parameter* are very different, but multiplying these values (I_{org} is reduced with 0.5) results in similar values for the combined measure OI_3. This implies that OI_3 can be used as a measure for cloud organization/clustering, but that it is difficult to relate this extent of clustering to climate feedbacks of the cloud field.

The results from Stevens et al. (2019) can be compared to the values of OI_3. In total, 65 days could be compared based on the requirement that at least five out of six scientists from the project of Stevens et al. (2019) agreed on the type of cloud field, in combination with the availability of evaluated data from this thesis. Out of these 65 days most fields were classified as *Gravel* (65), then *Fish* (8) and *Sugar* (8) and for *Flowers* only 4 days could be compared. The results are presented in Figure 5.2.

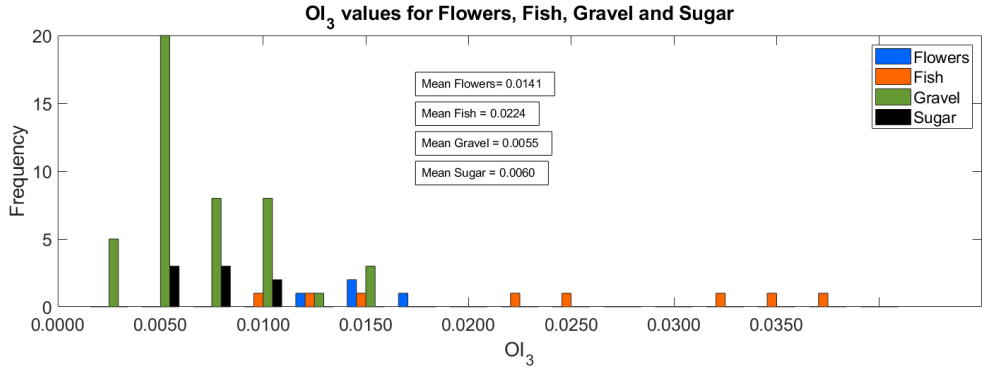


Figure 5.2: OI_3 values compared to subjective analysis results.

This Figure shows that the values of OI_3 are lower for fields which are subjectively classified as *Gravel* and

Sugar. From these results it also seems like the spread in values is smaller for *Sugar* and *Flowers*, but these results may be biased due to the small number of days that can be evaluated. Nevertheless, the spread in results classified as *Fish* by Stevens et al. (2019) is very large. Figure 5.3 shows the fields classified as *Fish* with the smallest (left) and largest (right) values.

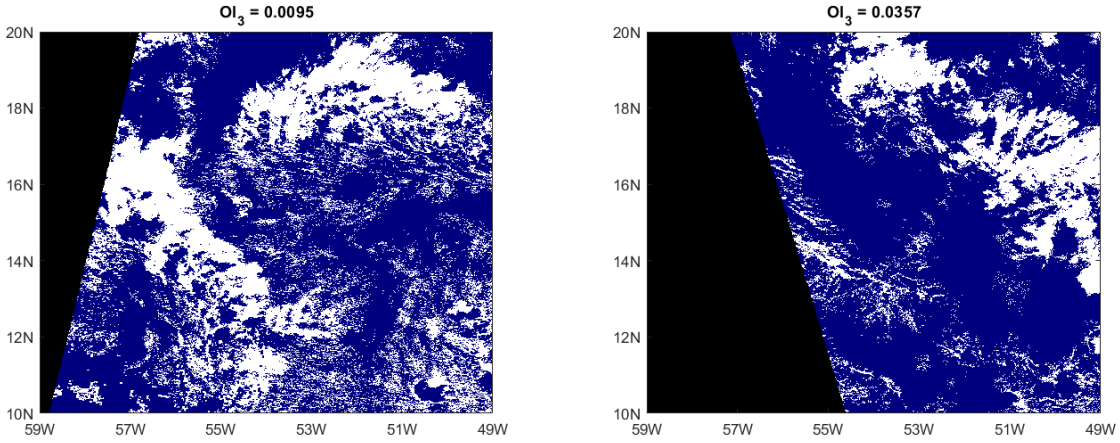


Figure 5.3: Cloud field classified as *Fish* by Stevens et al. (2019) with different OI_3 values.

The probability of such high OI_3 values as seen for some *"Fish"* fields is small. This can be derived from Figure 5.4, in which the blue bins show the probability of OI_3 values to occur and the dark red line represents the approximation of the cumulative density function. The black dashed line marks the point at which the cumulative density function equals 0.5 and thus this Figure shows that 50% of the days have a value of OI_3 smaller than 0.01.

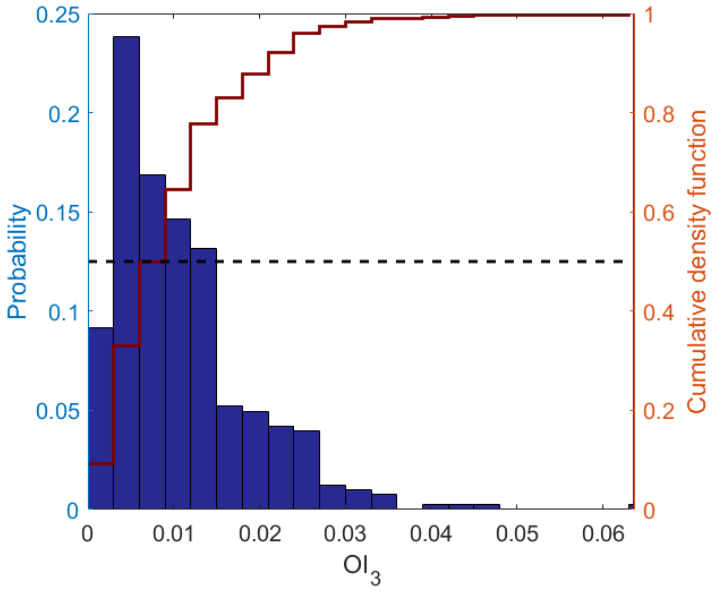


Figure 5.4: Probability of OI_3 values.

As discussed above it remains difficult to judge the performance of OI_3 in an objective way, but a subjective judgement shows that OI_3 performs as expected and cloud fields sorted based on their value of OI_3 are, when analysed subjectively, ordered from least to most clustered. However, OI_3 does not provide information on what drives the clustering. OI_3 can be high because all individual parameters are (relatively) high (for example field (d)), but the clustering can also be mostly driven by the *Clear Sky Parameter* (field (e)) or the

Largest Cloud Parameter (field (f)).

Figure 5.5 contains three different scatter plots to show the correlation between the three individual parameters that are part of OI_3. The markers are coloured by the corresponding value of OI_3. The correlation coefficients are printed above the scatter plots. These values show that the *Clear Sky Parameter* and I_{org} are highly correlated, as their correlation coefficient equals 0.76. The correlation between the *Largest Cloud Parameter* and I_{org} however is low (0.13) and the correlation between the *Largest Cloud Parameter* and *Clear Sky Parameter* is even non-existing (0.01).

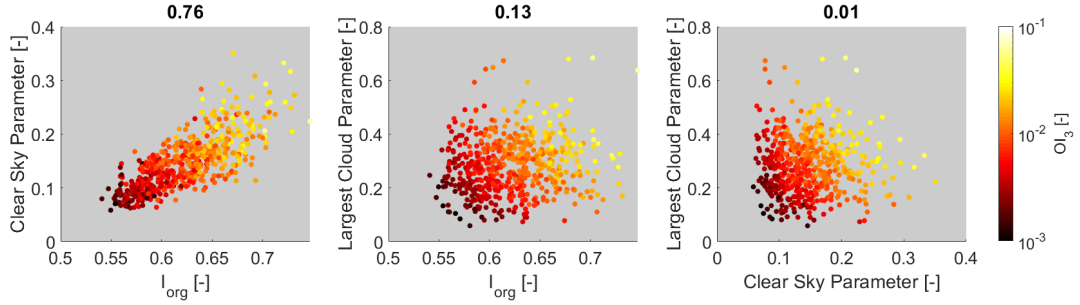


Figure 5.5: Scatter plot of the individual parameters of OI_3 against each other. The circles are coloured on a logarithmic scale after their corresponding values of OI_3.

Figure 5.6 shows the scatter plots of the individual parameters against OI_3 and the corresponding correlation coefficients. Note that the x-axis has a logarithmic scale. Since OI_3 is a multiplication of the three parameters high correlation coefficients are expected. The correlation coefficient with I_{org} is 0.79 and is the highest of the three. The *Clear Sky Parameter* also has a high correlation with OI_3: the correlation coefficient is 0.74. The *Largest Cloud Parameter* also clearly correlates with OI_3 but to a lesser extent as the correlation coefficient between OI_3 and the *Largest Cloud Parameter* is 0.53. This can also be derived from Figure 5.6 as the spread of the values on the y-axis is largest for the *Largest Cloud Parameter*.

Although the correlation between OI_3 and I_{org} is high, they do sort the analysed days in a different order. This is summarized in a Table in Appendix B.

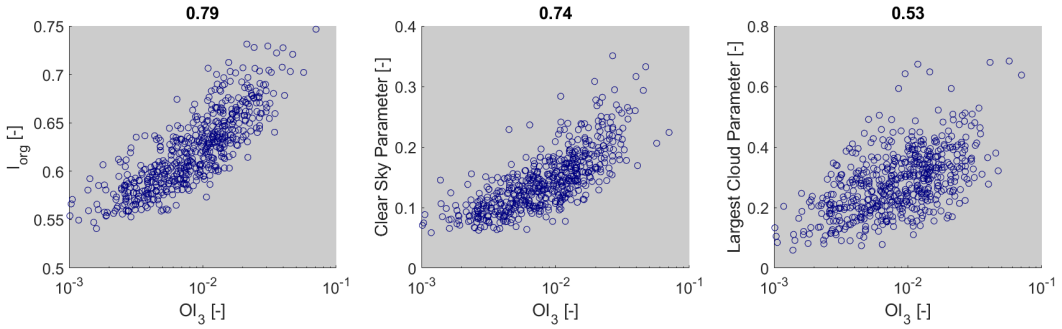


Figure 5.6: Scatter plot of the individual parameters of OI_3 against OI_3.

To get an idea about the correlation between OI_3 and the so-called *Basic Parameters* (*Cloud cover*, *Average nearest neighbour*, *Cloud density* and *Average cloud size*) scatter plots are also created for these parameters. These plots are displayed in Figure 5.7. Again the x-axis has a logarithmic scale. The correlation coefficients are relatively high for the *Cloud Density* (-0.68), and the *Average Nearest Neighbour* (0.44). The correlation with the *Cloud Cover* (0.15) and *Average Cloud Size* (0.04) are relatively low.

The relatively high negative correlation coefficient with the *Cloud Density* is as expected: in general clouds tend to merge together for higher degrees of clustering and thus the *Cloud Density* decreases. The average nearest neighbour is also expected to decrease for a higher degree of clustering, as clouds are expected to be closer together. However, the opposite seems to be true when looking at the relatively large positive correlation coefficient: apparently the average nearest neighbour increases for a higher degree of clustering. This may be due to the fact that clouds are increasing in size: their nearest neighbours are then automatically further away from their centroid as the area that they span is larger. The relatively low correlation with the *Cloud Cover* is as expected, cloud fields can be very clustered for both small (Fig. 5.1e) or high (Fig. 5.1f) cloud

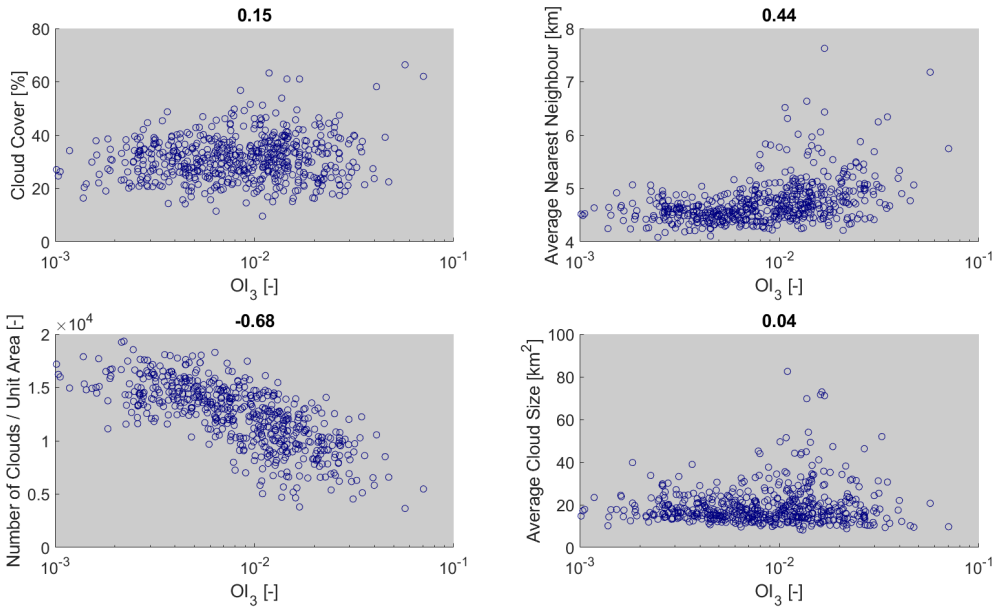


Figure 5.7: Scatter plot of the basic parameters against OI_3.

covers. The correlation coefficient with the *Average Cloud Size* was expected to be higher, since cloud cluster and eventually merge for high cloud organization. Apparently this is not necessarily the case and can fields also be classified as clustered/organized without the presence of large clouds. Another reason for this low correlation may be that clouds directly adjacent to the edge of the field and/or the no-data/masked out area were removed from the dataset. These clouds were taken into account when calculating the OI_3, *Clear Sky Parameter* and *Largest Cloud Parameter*. As the *chance* of *touching* the edge or no-data/masked out area increases with cloud size, these large clouds are more likely to get removed from the dataset which alters the statistics.

The correlation with the earlier applied parameter SCAI was investigated as well. A scatter plot between OI_3 and SCAI is shown in Figure 5.8. Note that the x-axis again has a logarithmic scale. From this plot it can be derived that a strong correlation is present. The correlation coefficient is -0.66 and thus shows that parameter SCAI and OI_3 are indeed relatively strongly correlated. Similar to I_{org}, $SCAI$ does sort the analysed days in a different order than OI_3. This is also summarized in Appendix B.

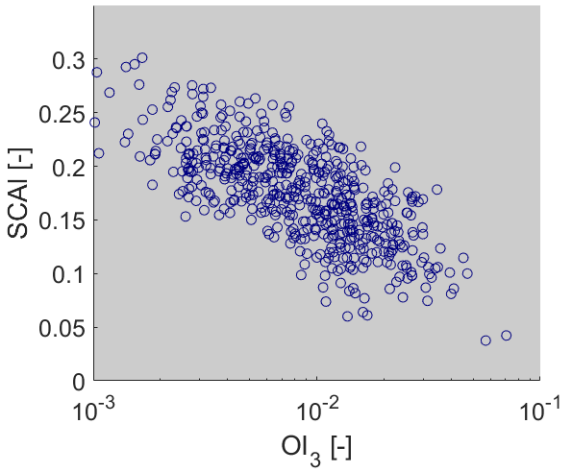


Figure 5.8: Scatter plot SCAI against OI_3.

The results in this section show that there is more confidence in the performance of OI_3 compared to the performances of I_{org} and $SCAI$. Although the correlations between OI_3 and the organization parameters I_{org} and $SCAI$ are high, respectively 0.79 and -0.66, the results from OI_3 are better when compared to a subjective analysis. The order in which OI_3 places the evaluated days is shown in Appendix A. Appendix B gives the differences with I_{org} and $SCAI$.

5.1.2. Temporal variability of OI_3 and basic parameters

In this section time series of OI_3, the cloud cover, the average nearest neighbour, the cloud density and the average cloud size are presented. As data was collected for the months December, January and February between December 2010 and February 2018, the data is presented *per winter*. Furthermore the correlation between OI_3 for Terra and Aqua and the autocorrelation function of OI_3 are presented.

The development of OI_3 during the separate winter months is plotted in Figure 5.9.

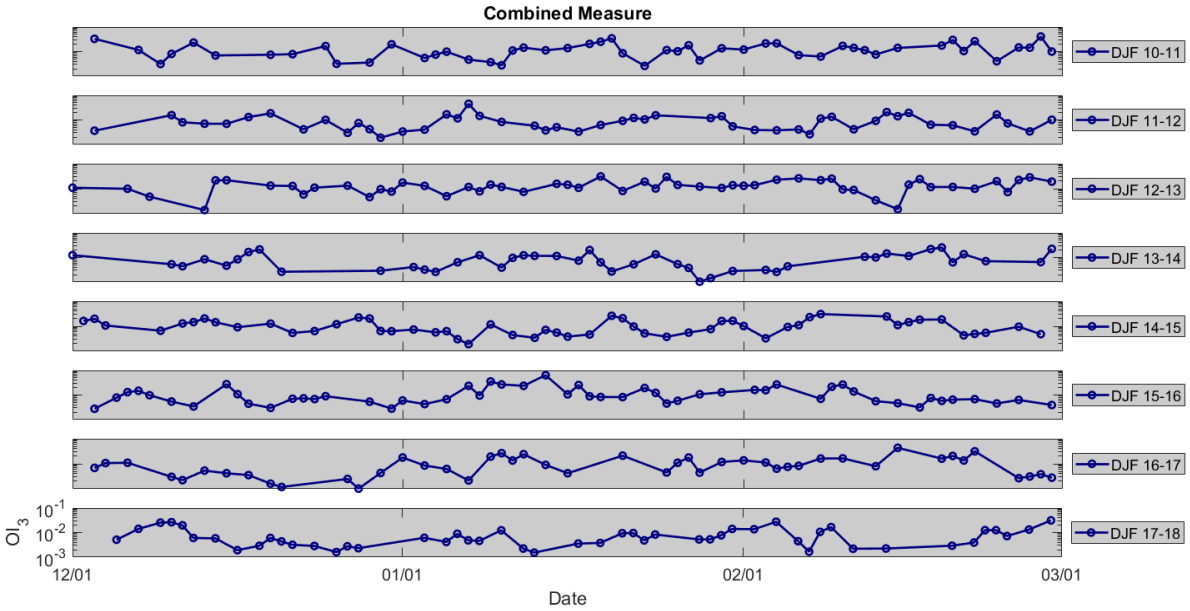


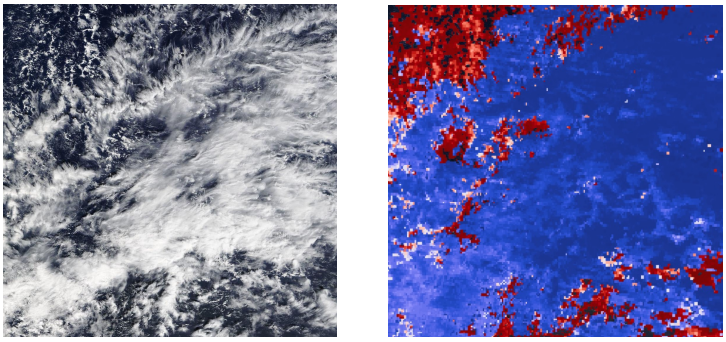
Figure 5.9: Development of OI_3 for different winters. On the y-axis OI_3 on a logarithmic scale. The circles denote the days that were analysed."*DJF*" stands for December, January and February.

As the circles represent data points and the lines just show the connection between those data-containing days, it can be seen that there are multiple longer periods of missing data. This can sometimes be explained by the overpassing location of Terra and Aqua, as days with insufficient data due to the Sensor Zenith Angle are not analysed. However, these long periods of no data can not only be explained by this. Instead, they are most likely caused by a too large amount of high clouds in the area. This is for example the case during the beginning of December 2011, the beginning and end of December 2013 and the beginning of February 2014. An example of such a day is given in Figure 5.10, which shows two snapshots of the same area (area of interest) on December 27, 2013. Figure 5.10a shows the corrected reflectance and Figure 5.10b shows the Cloud Top Pressure. Pixels coloured (light) blue have a Cloud Top Pressure below 550 hPa and are thus considered as high clouds. As this is more than 20% of the entire area, this day can not be analysed. These high cloud fields often remain for multiple consecutive days and therefore cause a gap in the data time line.

From Figure 5.9 it looks like the values of OI_3 are not random. Instead values tend to remain roughly similar for multiple consecutive days or gradually increase or decrease. That the value of OI_3 does usually not change abruptly from day to day is also known from visual inspection. Especially *Flower* structures are often visible for multiple days in a row.

An example of a gradually increasing value of OI_3 can be found between February 11-18, 2014. The corresponding plots of these days are shown in Figure 5.11. These plots show *Flower*-like structures gradually transforming into *Fish* structures. The *SCAI* values of these days are 0.1367, 0.1586, 0.1108, 0.0739,0.1108, 0.1567. According to those values, the cloud field is most organized on February 15, 2014.

Figure 5.12 shows a scatter plot of all OI_3 values collected on a specific day during the 8 years that were analysed accompanied by a 1-day and 10-day average line plot. From these line plots there seems to be some



(a) Snapshot showing the corrected reflectance.

(b) Snapshot showing the Cloud Top Pressure.

Figure 5.10: Snapshots from 2013-12-27 (TERRA), taken from the MODIS Worldview, showing large high cloud areas. (a) shows the corrected reflectance image, whereas (b) shows the Cloud Top Pressure associated to image (a).

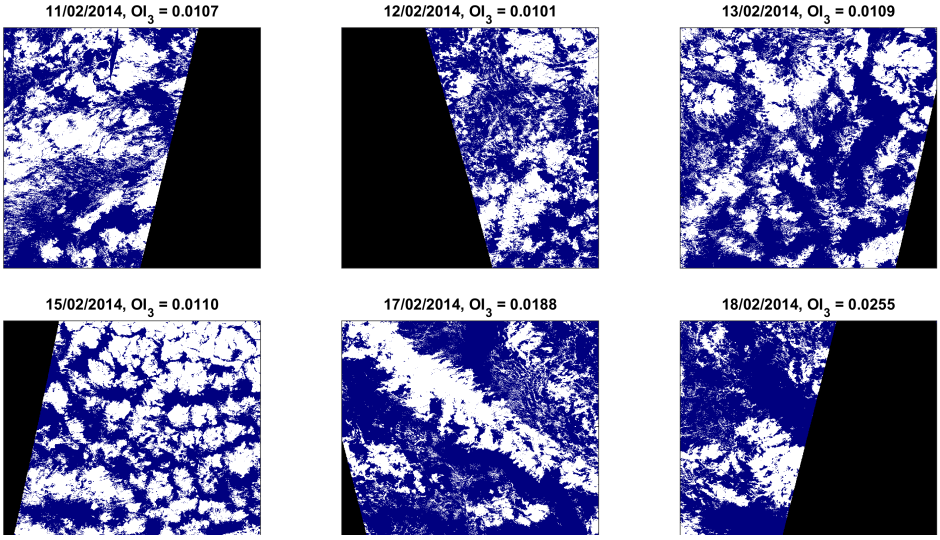


Figure 5.11: Consecutive days with gradually increasing OI_3.

sort of trend with the organization being higher in January and February compared to December. However, as the data record is only 8 years and only 403/720 (56%, see Table 3.3) of the days have been analysed, each day is on average based on 4.5 days and thus prone to biases due to large extremes in the data.

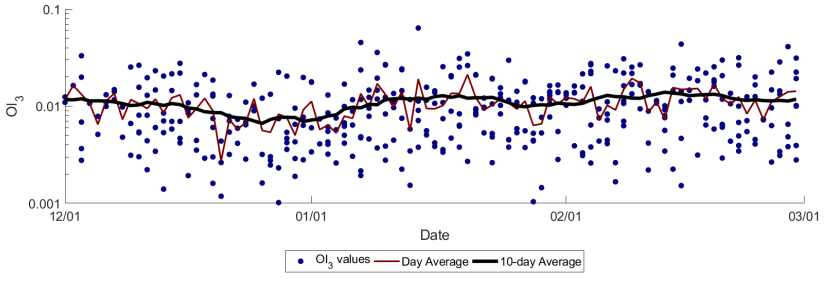
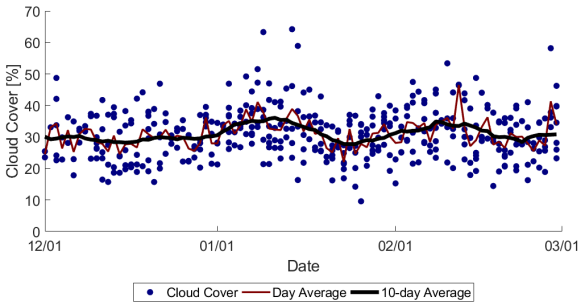
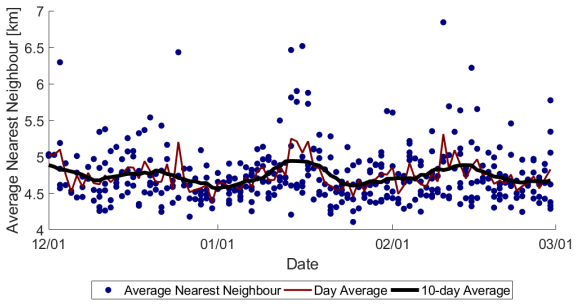


Figure 5.12: Day and 10-day averages for OI_3. The y-axis has a logarithmic scale.

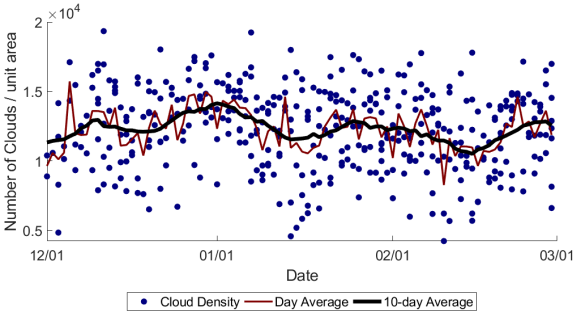
Figure 5.13 shows the results of respectively the cloud cover, average nearest neighbour, number of clouds and cloud size are shown. These Figures are created in a similar way as the graph of OI_3. The minimum, maximum, average and median values per parameter are summarized in Table 5.2. From this Table it can be concluded that there are mainly some large outliers for the average cloud size and average nearest neighbour.



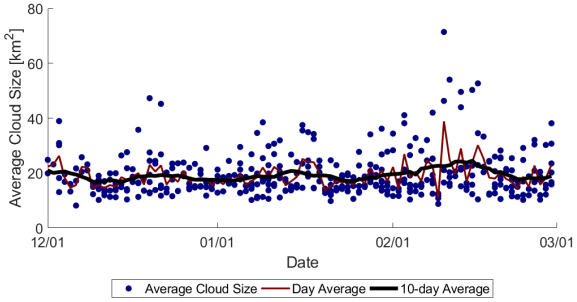
(a) Cloud cover values.



(b) Cloud nearest neighbour values.



(c) Cloud density values.



(d) Average cloud size values.

Figure 5.13: Average values basic parameters.

Table 5.2: Smallest, largest, average and median values of parameters from 8 years of data.

Parameter	Minimum	Maximum	Average	Median
OI_3 [-]	0.0010	0.0637	0.0110	0.0090
Cloud Cover [%]	9.59	64.21	31.13	30.86
Average Cloud Nearest Neighbour [km]	4.11	6.84	4.74	4.65
Number of Clouds / unit area [-]	4224	19342	12286	12541
Average Cloud Size [km^2]	8.17	71.40	19.41	17.01

For all parameters it can be concluded that there seems to be some sort of trend, but it is difficult to prove that due to the limited amount of data. The *Cloud Density* shows a large spread in values but does not show very large extremes. This is different for the *Average nearest neighbour* and the *Average cloud size*, both show some large positive extreme values.

Figure 5.14 shows the median (red line), 25th and 75th percentile (lower and upper box limits), extreme non-outliers (lower and upper whiskers) and outliers ('+' signs) per December-January-February period of OI_3 for Terra and Aqua separately. This way insights in the differences between different years are obtained but also differences between Terra and Aqua can be identified. Some variability between years can be distinguished: the periods DJF 10-11 and DJF 12-13 are more organised than the other winters. The winter period 17-18 is least organised based on OI_3. The differences between Terra and Aqua do not seem significant: 5 times the median value of Terra is higher, versus 3 times for Aqua. The height of the boxes also is equal during most years, except for DJF 12-13 when the spread in values is higher for Terra and for DJF 15-16, when this is the other way around.

The correlation between Terra and Aqua is high: 0.89. This is visualized in the scatter plot in Figure 5.15.

The autocorrelation is calculated for the entire dataset, without combining Terra and Aqua. This way, the autocorrelation can be calculated in hours (as the overpass times of Terra and Aqua respectively are 10:30 AM and 1:30 PM). The result is shown in Figure 5.16. At t - 3, 24, 27, 51, 96 and 120 hours the autocorrelation is significant. However, as the autocorrelation is small at t = 48, 72, 75 and 99 hours it should be concluded that the autocorrelation is significant up until 27 hours.

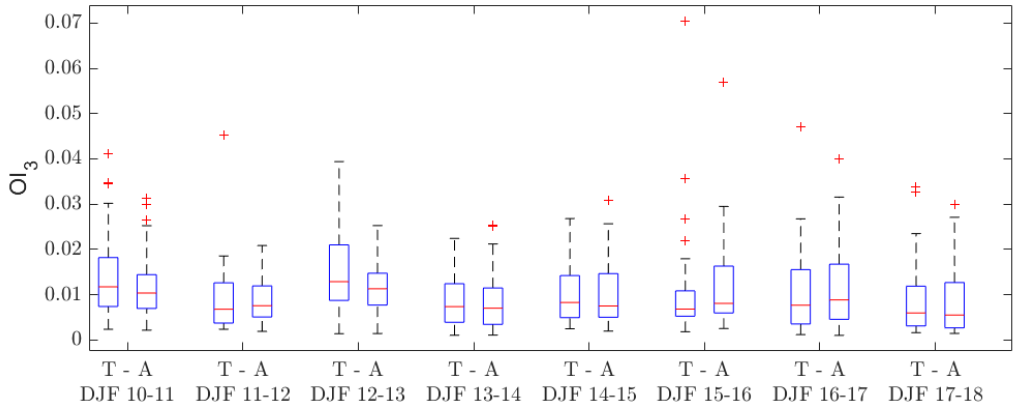


Figure 5.14: OI_3 statistics of Terra and Aqua in box plots. The median value is indicated by the red line, 25th and 75th percentiles by the lower and upper box limits, extreme non-outlier values by the lower and upper whiskers and the '+' signs indicate outliers.

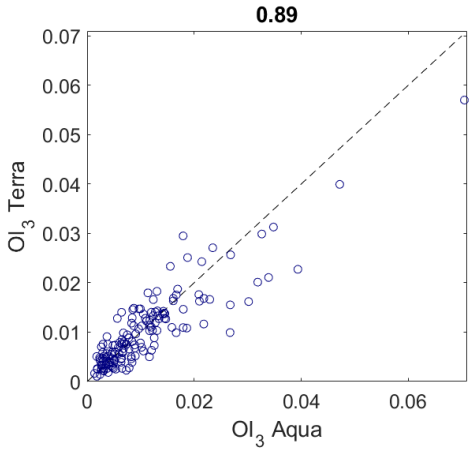


Figure 5.15: Correlation between Terra and Aqua for OI_3

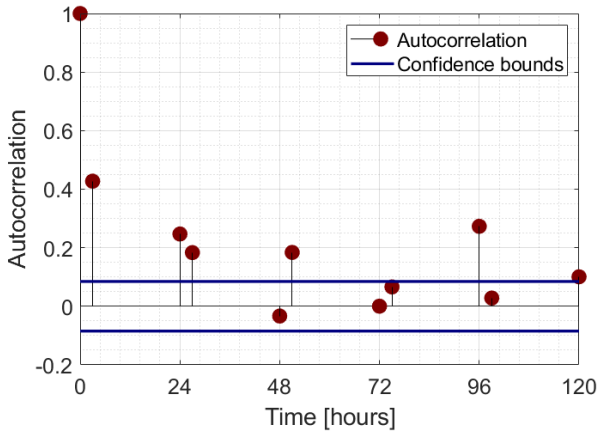


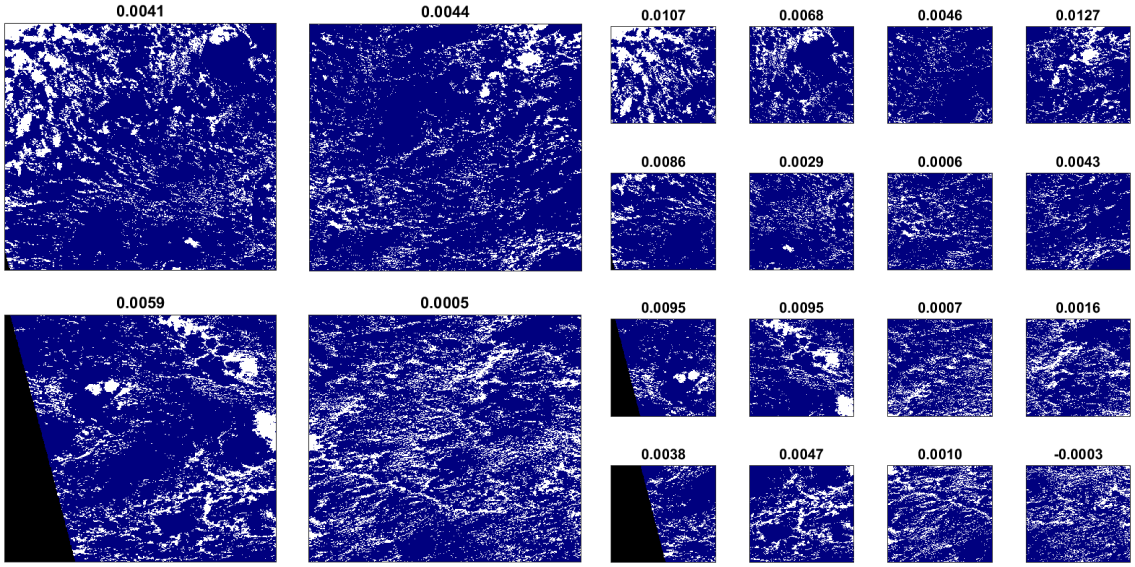
Figure 5.16: Autocorrelation OI_3.

5.1.3. Spatial variability of OI_3

The spatial variability of OI_3 can be investigated by looking at the results of OI_3 when calculated for smaller areas within the complete area of interest. This is done for areas of 5° × 5° (4 smaller areas) and 2.5° × 2.5° (16 smaller areas). Examples from two days are plotted in Figures 5.17 and 5.18. Figure 5.17 shows the smaller fields from the full field that is plotted in Figure 5.1a, which is one of the least clustered fields that has been analysed. Figure 5.18 shows the smaller fields from the full field that is plotted in Figure 5.1f. This field is interesting because the variation in cloud types and organizations within this field is relatively large.

The smaller fields in Figures 5.17a and 5.17b are thus representing a rather random, non-clustered area. The value of OI_3 for the complete field is 0.0014. The largest and smallest values are found for the smallest fields, which can also be expected. Here the largest value is 0.0127 and the smallest value is -0.0003, which implies that I_{org} was below 0.5 and thus classified this field as regular. This does not occur frequently, and only for the smallest fields and for those fields only 16 times out of 8046 small fields that are analysed in total. The variance of OI_3 of this field is relatively small and the difference between the minimum and maximum value is 0.0130. The variance of the four areas of 5°×5° is 3.89×10^{-6}. The variance of the 16 areas of 1.58×10^{-5}.

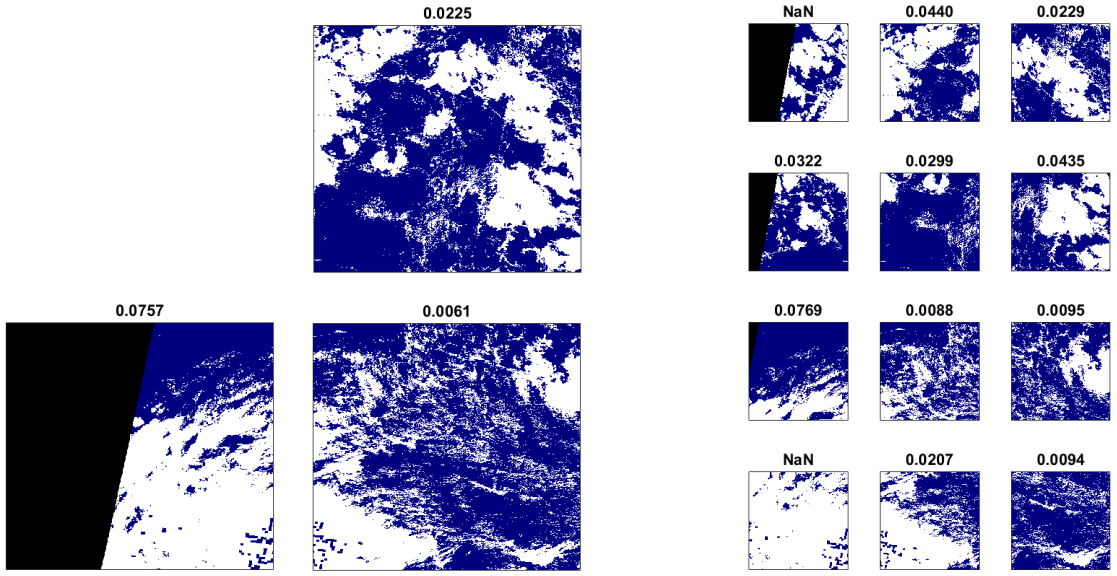
Figures 5.18a and 5.18b show the smaller fields from the complete area shown in Figure 5.1f, which has a value of OI_3 of 0.0201. Some sub-plots are missing, as the data coverage of is sub-plot was smaller than 50%. The variance of this field is larger than the variance of the previous field and equals 6.49×10^{-4} and 3.41×10^{-4} for the 4 and 16 fields respectively. With this example it is shown that it is not always possible to calculate the value of OI_3. In these cases, it was not possible to compute I_{org}, which happens when the cloud cover is very high and the cloud sizes are large. It is then not always possible to randomly place these clouds in a field and therefore the algorithm fails. This is very unlikely to happen when the complete area is analysed, as such large



(a) Fields of 5° × 5°.

(b) Fields of 2.5° × 2.5°.

Figure 5.17: Smaller fields on February 14, 2013 (AQUA).



(a) Fields of 5° × 5° on February 5, 2013 (TERRA).

(b) Fields of 2.5° × 2.5° on February 5, 2013 (TERRA).

Figure 5.18: Smaller fields on February 5, 2013 (AQUA).

shallow cumulus clouds do not exist. However, as the size of the area is reduced, the probability that I_{org} fails increases. This is what happened to two small fields in Figure 5.18b, where the value of OI_3 is denoted as NaN.

The variance of OI_3 was calculated for each analysed day. This variance was then compared to the value of OI_3 of the complete area. A scatter plot of the variance found for the 4 areas of 5° × 5° against the variance found for the 16 areas of 2.5° × 2.5°, coloured by the value of OI_3 for the complete area, is shown in Figure 5.19.

From this Figure it can be derived that there is a positive correlation between the variance of the 4 areas and the 16 areas. Also, from the colours, it can be derived that both variances also exhibit a positive correlation with the value of OI_3 of the entire area: the variance is expected to be higher for a higher degree of organization, and thus the spatial variability of the clouds is larger. In numbers, the correlation between OI_3 and the variance of OI_3 calculated for the 4 and 16 smaller areas is equal to 0.42 and 0.49 respectively. The correlation between the variance for 4 areas of 5° × 5° and the 16 areas of 2.5° × 2.5° is 0.45. What is interest-

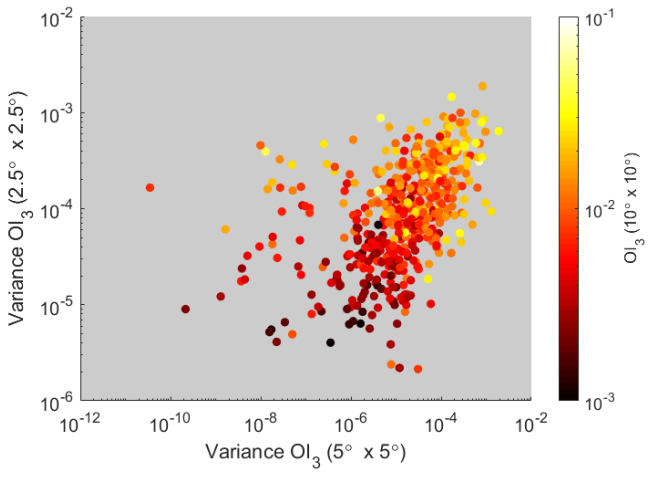


Figure 5.19: Correlation spatial variance OI_3.

ing, is that the variance is significantly higher when looking at the results from Aqua for all three cases. The results of the correlations are summarized in Table 5.3. The values above the diagonal show the correlation coefficients when calculated for both Aqua and Terra simultaneously. Below the diagonal the two values give the correlation of Aqua and Terra respectively. All correlation coefficients are significantly larger for Aqua, and the difference between Aqua and Terra is largest when looking at the correlation coefficients between OI_3 and the variance of OI_3 for the 16 smaller areas.

Table 5.3: Correlation coefficients OI_3 between different spatial scales. The values below the diagonal of the table show the correlation coefficients of Aqua and Terra respectively. The values above the diagonal show the correlation coefficients when calculated for the complete dataset.

	OI_3	**Variance OI_3 4 areas**	**Variance OI_3 16 areas**
OI_3	-	0.42	0.49
Variance OI_3 4 areas	0.46 | 0.36	-	0.45
Variance OI_3 16 areas	0.58 | 0.37	0.49 | 0.41	-

5.2. Discussion

The results presented in this chapter imply interesting insights in quantifying cloud organization. However, some uncertainties remain. For example, the relatively large difference between Aqua and Terra is not well understood. It can partly be explained by their differences in overpass time, but this is unlikely to explain the significant differences in the variance values for Terra and Aqua.

Furthermore, the applied method for the calculations of the basic parameters (except for the cloud cover) where clouds adjacent to the edge or no-data area were removed from the dataset alters the statistics. For example it was expected that the average nearest neighbour as well as the average cloud size would increase when setting both MODIS cloud classifications *Cloudy* and *Uncertain Clear* to clouds. The results show the opposite, most likely because the extremely large (and incorrect, see Figure 3.16) obtained clouds with both classifications had a larger probability of being removed from the dataset, whereas the extra single cloud pixels were mostly kept.

For the calculations of OI_3 however all clouds were kept, and thus it is the question to what extent correlations can be expected between OI_3 and the basic parameters. The correlation with the average cloud size was also expected to be higher, as fields with only small clouds are not often classified as very clustered. This may be due to the fact that mostly large clouds were removed from the dataset which results in a smaller average cloud size.

Furthermore, the correlation with $SCAI$ is higher than expected. Most likely because no clouds were removed in the computation of $SCAI$.

6

Conclusions and Recommendations

To answer the two main questions **How can shallow cumulus cloud organization objectively be characterized using satellite imagery** and **What are the temporal and spatial scales of shallow cumulus cloud organization** several sub-questions were defined. Sections 6.1 and 6.2 are devoted to answering the main questions. In Section 6.3 general conclusions are discussed and Section 6.4 provides some recommendations for future research.

6.1. How can shallow cumulus cloud organization objectively be characterized using satellite imagery?

Shallow cumulus cloud organization can be characterized by making use of MODIS data, in particular the MODIS Cloud Mask. This thesis focussed on finding an improved organization index for shallow cumulus cloud fields, because confidence in existing methods (I_{org} and $SCAI$) lacks. Although the theory behind I_{org} is very useful and effective, the existing methodology did not take the size of clouds into account. The results were therefore not always realistic.

The organization parameter OI_3 is partly based on the existing method I_{org}, but does take into account the size of the clouds. The nearest neighbour distances of the actual field, from calculations based on cloud fields obtained the MODIS Cloud Mask, are compared to a randomly generated cloud field. This random cloud field generator is based on the number of clouds and their sizes. The individual clouds are simplified to circles, so their shapes are not taken into account. The differences between the random generated cloud field and the actual field tell us to what extent a cloud field is clustered, random or regular. OI_3 is calculated from this measure, combined with parameters of the largest cloud and the largest clear sky region in the field.

The results are promising, as comparisons with subjective analyses show high correlations. Based on subjective judgement, OI_3 performs well.

The correlation coefficients with I_{org} and $SCAI$ respectively are 0.79 and -0.66 and thus it can be argued that these parameters should perform good as well. However, crucial differences exist when comparing the results to a subjective analysis. OI_3 performs better overall and is therefore considered an improvement on the existing methods.

OI_3 can be applied to almost all cloud fields. However, as the actual field is compared to the result of a random simulation, fields with less than approximately 50 clouds may show significantly different outcomes for different simulations. Therefore the field should be large enough. In this thesis the smallest fields which were evaluated were $2.5° \times 2.5°$, which was large enough in terms of the number of clouds that were present in the field. However, the cloud cover may also cause the calculations to fail. When the cloud cover is extremely high it may not always be possible to place all clouds in the random field and thus it is not possible to do all calculations.

In general the combined organization index OI_3 performs well. As all parameters are dimensionless it is possible to apply the calculations to different field sizes.

6.2. What are the temporal and spatial scales of shallow cumulus cloud organization?

The results showing the results of OI_3 and the basic parameters on each day (Fig. 5.12 and 5.13d) do show some differences between December, January and February. However, due to the limited amount of data, it is difficult to say whether these differences are significant.

The correlation between the values of OI_3 obtained with Terra and Aqua with overpass times at respectively 10:30AM and 1:30PM is 0.89. As it was concluded that OI_3 performs effectively, this high correlation means that the organization in the morning is not very different from the organization in the afternoon. This can also be derived from visual inspection: the cloud field does usually not change very much over the course of 3 hours. The autocorrelation shows significant values until t = 27 hours. This also shows that usually organization remains the same or gradually increases or decreases over the course of two days. This is also visible in the plot of the evolution of OI_3 (Fig.5.9).

In the previous section it was concluded that, based on a few requirements, OI_3 can be calculated for different field sizes. As expected the variance of the organization values was higher when the large field was split up in smaller areas. From subjective analysis it was known already that *Flower* and *Fish* structures are larger than *Sugar* and *Gravel* fields. However, when evaluating the entire area of interest of 10° × 10° subjectively, often multiple organization structures can be detected. OI_3 therefore rather gives an average value of the entire area, whereas it can be concluded that when smaller areas are evaluated they are more often dominated by only one (subjective) organization structure.

6.3. General Conclusions

In general it can be concluded that the combined organization index OI_3 is an effective measure for cloud organization. It can be applied on different scales and is useful for comparisons between fields due to its dimensionless properties. Existing measures of organization like I_{org} and $SCAI$ have relatively high correlations but lack to be effective for some extreme cases. When visually inspecting the analysed scenes sorted from least to most organized according to OI_3 in Appendix A, it can be concluded that OI_3 performs well. Since a broad range of scenes has been analysed, ranging from 2.5° × 2.5° to 10° × 10° in size, it can be assumed that OI_3 will perform well in almost all cases.

6.4. Recommendations

An interesting follow-up on this research would be to apply the combined organization index OI_3 to larger datasets to better investigate the temporal and spatial variability. This could also be done with data of a geospatial satellite, such as GOES-16. When applying OI_3 to for example hourly data better insights in the development of organization can be obtained. Furthermore, it would be interesting to look at the correlation with atmospheric conditions such as temperature, wind speed and wind inversion. This would improve our insights and understanding of what drives shallow cumulus cloud organization.

Lastly, although this thesis shows that the combined organization index OI_3 performs well, one could also think about further fine-tuning this index by not only taking into account the sizes of the clouds but also their shapes in the calculations of the altered I_{org}.

A

Fields ordered based on OI_3

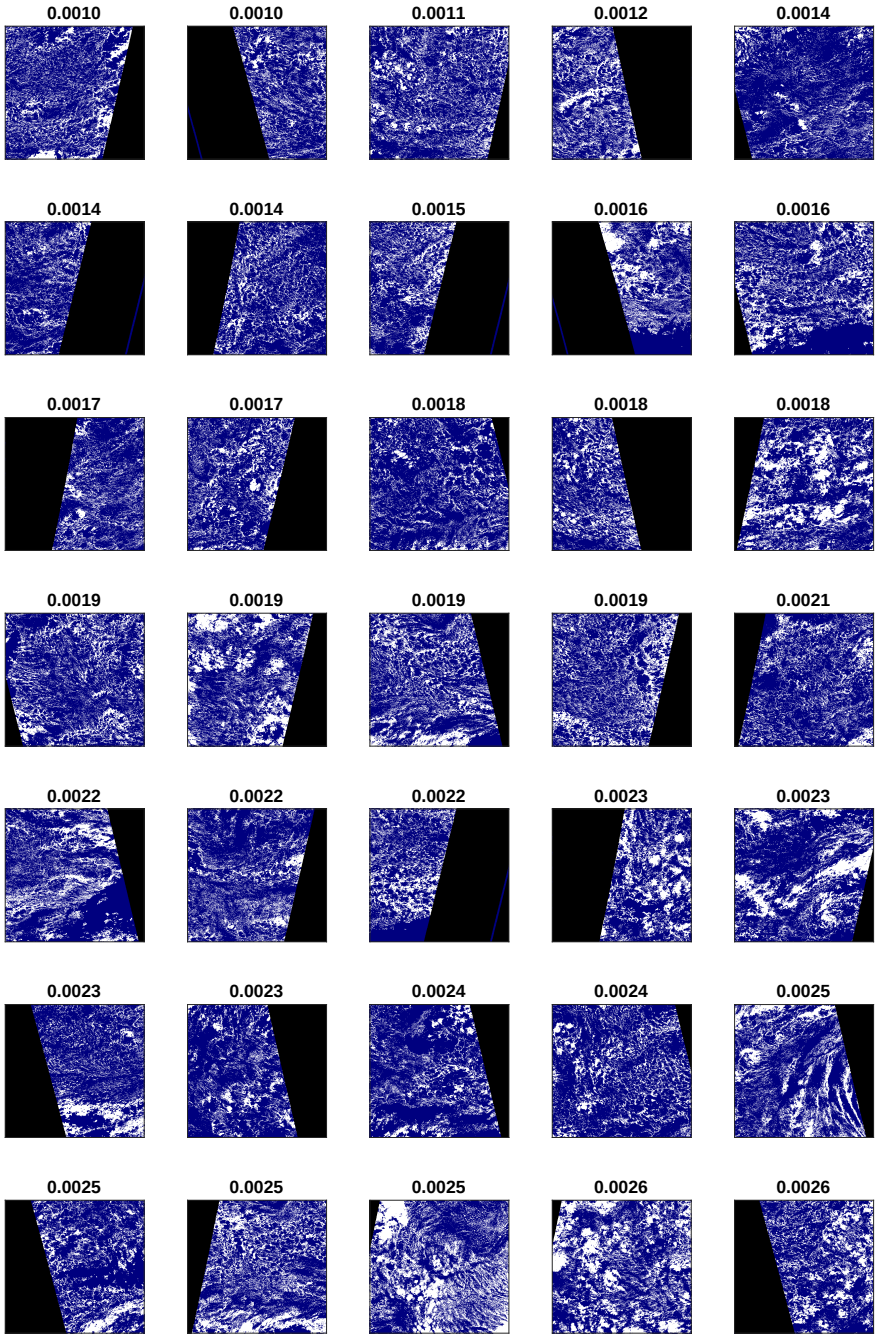


Figure A.1: Ordered fields 1-35

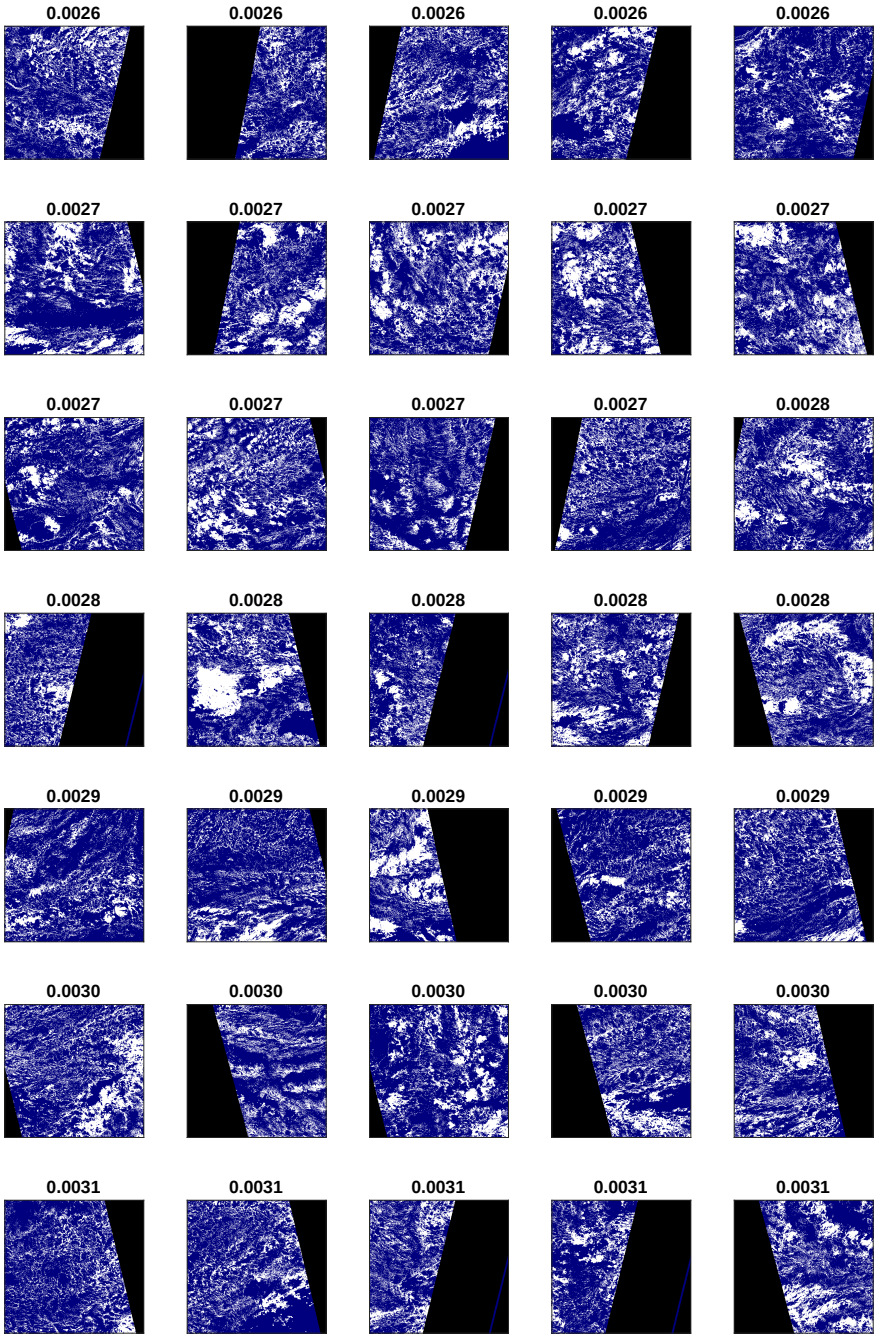


Figure A.2: Ordered fields 36-70

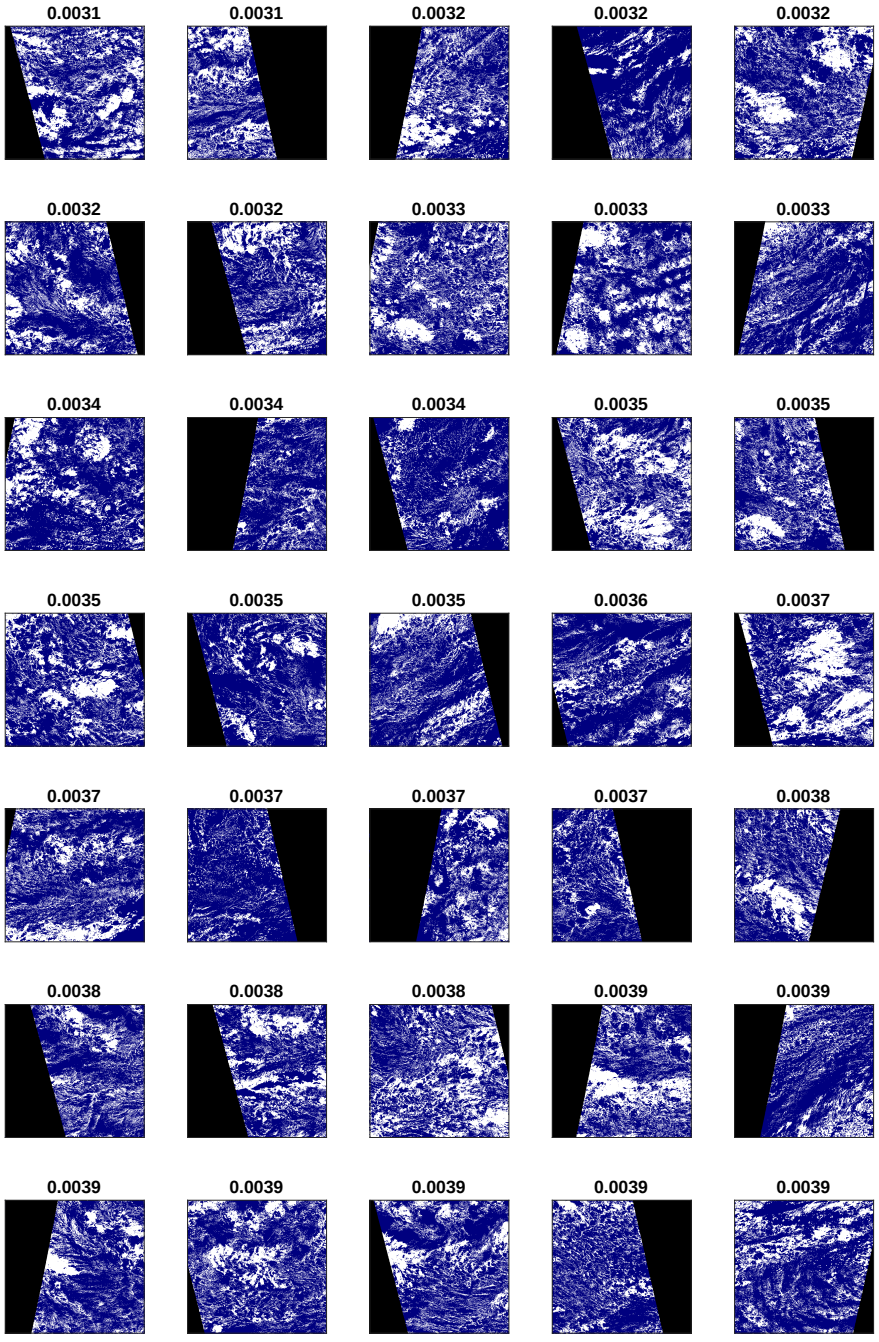


Figure A.3: Ordered fields 71-105

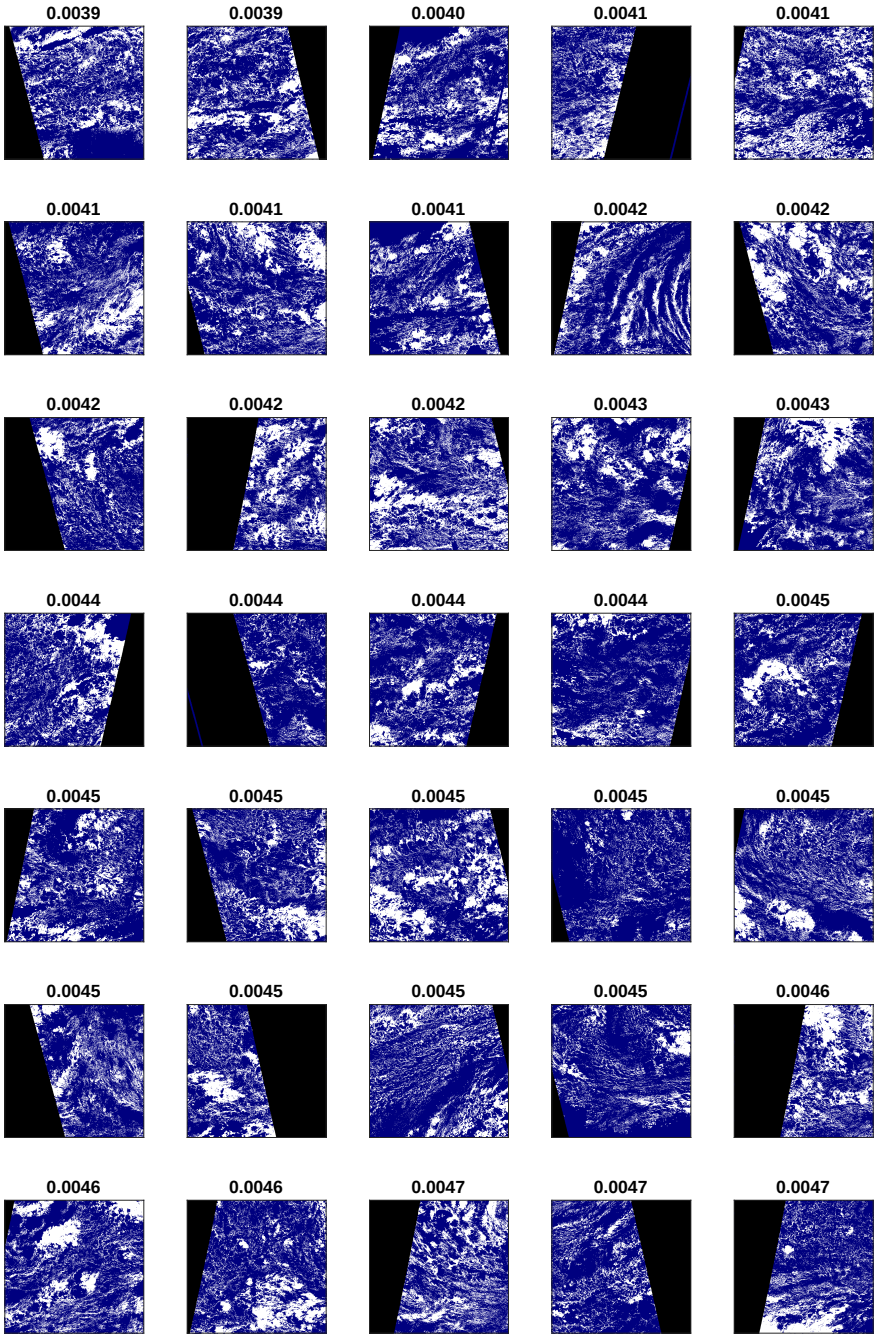


Figure A.4: Ordered fields 106-140

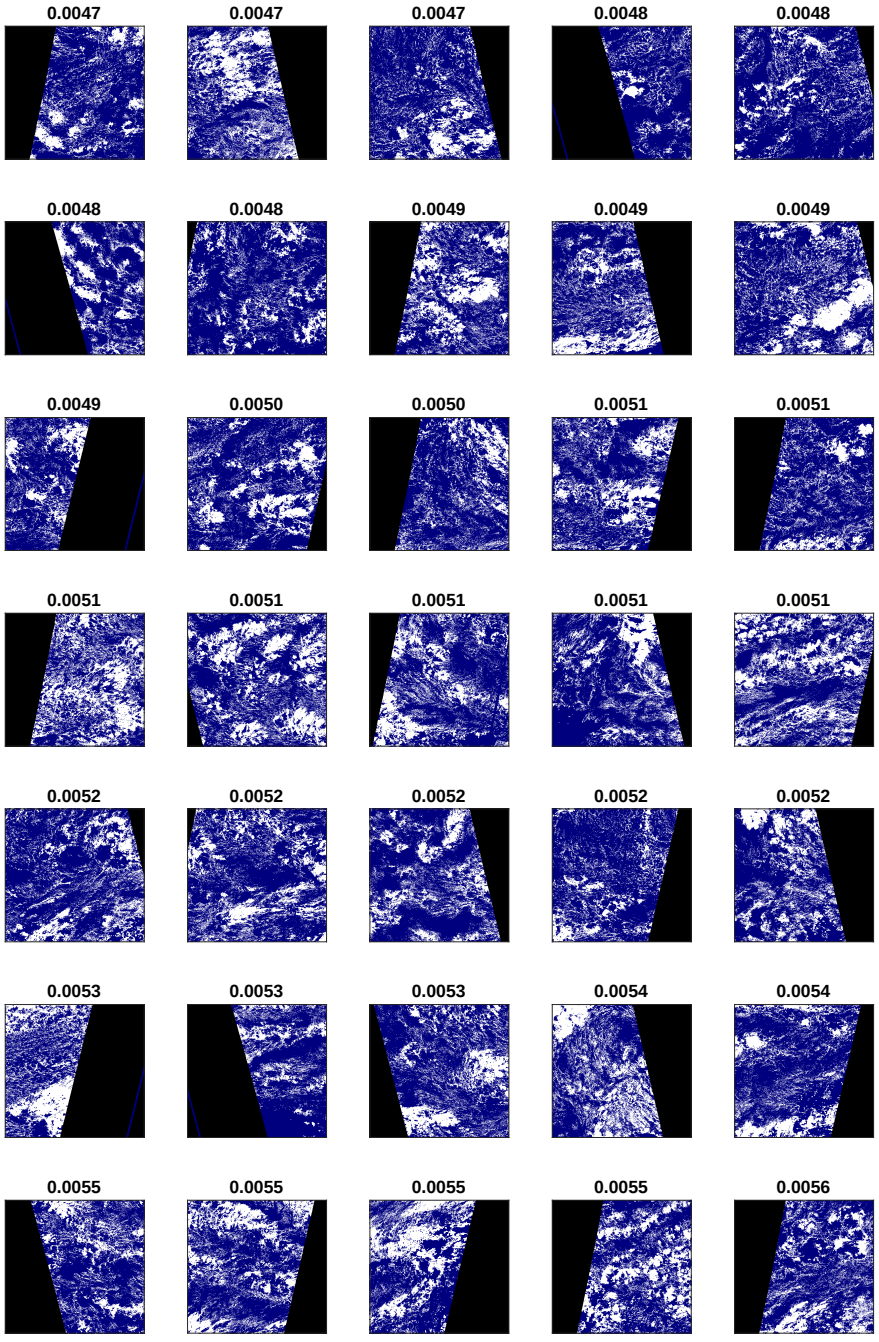


Figure A.5: Ordered fields141-175

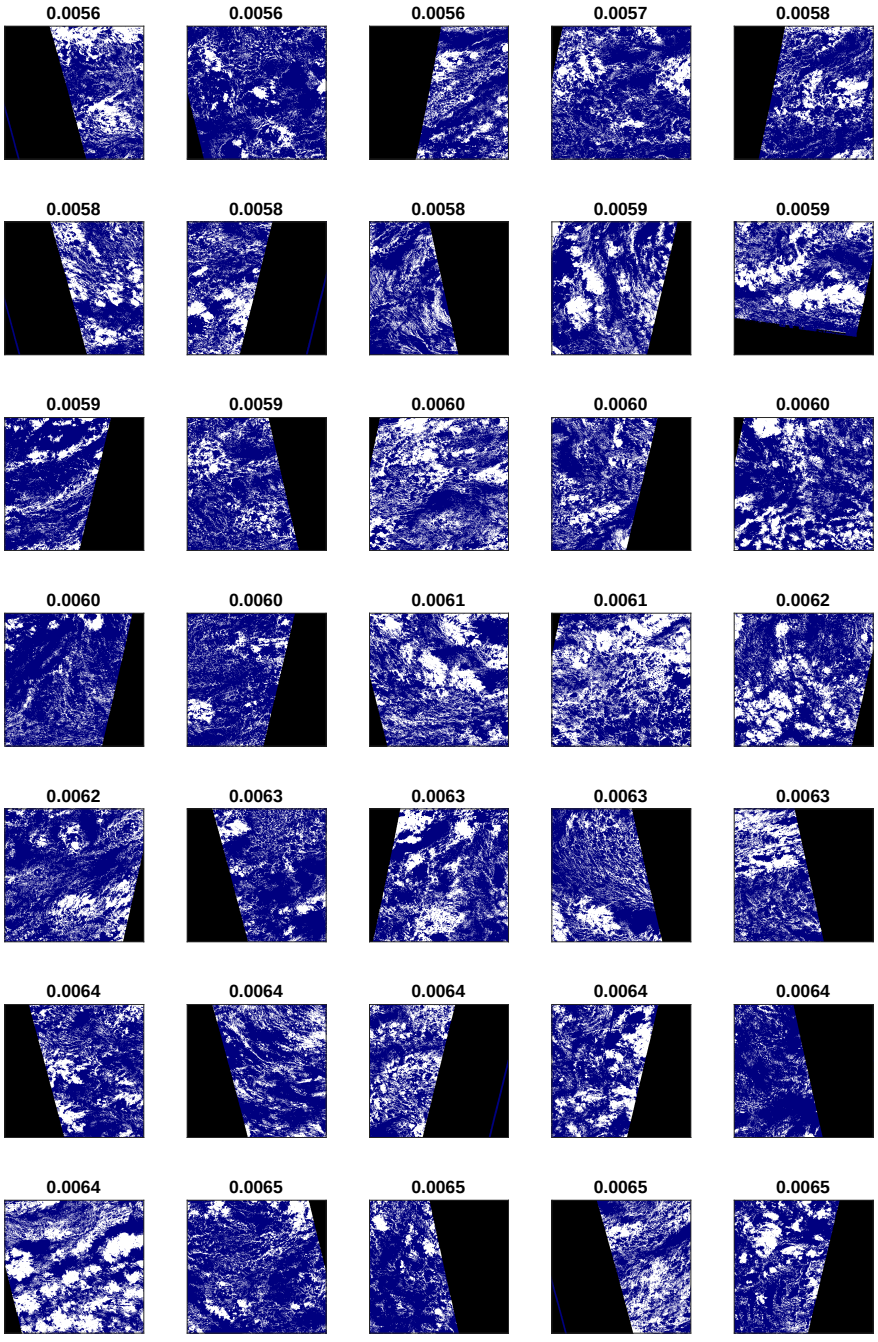


Figure A.6: Ordered fields 176-210

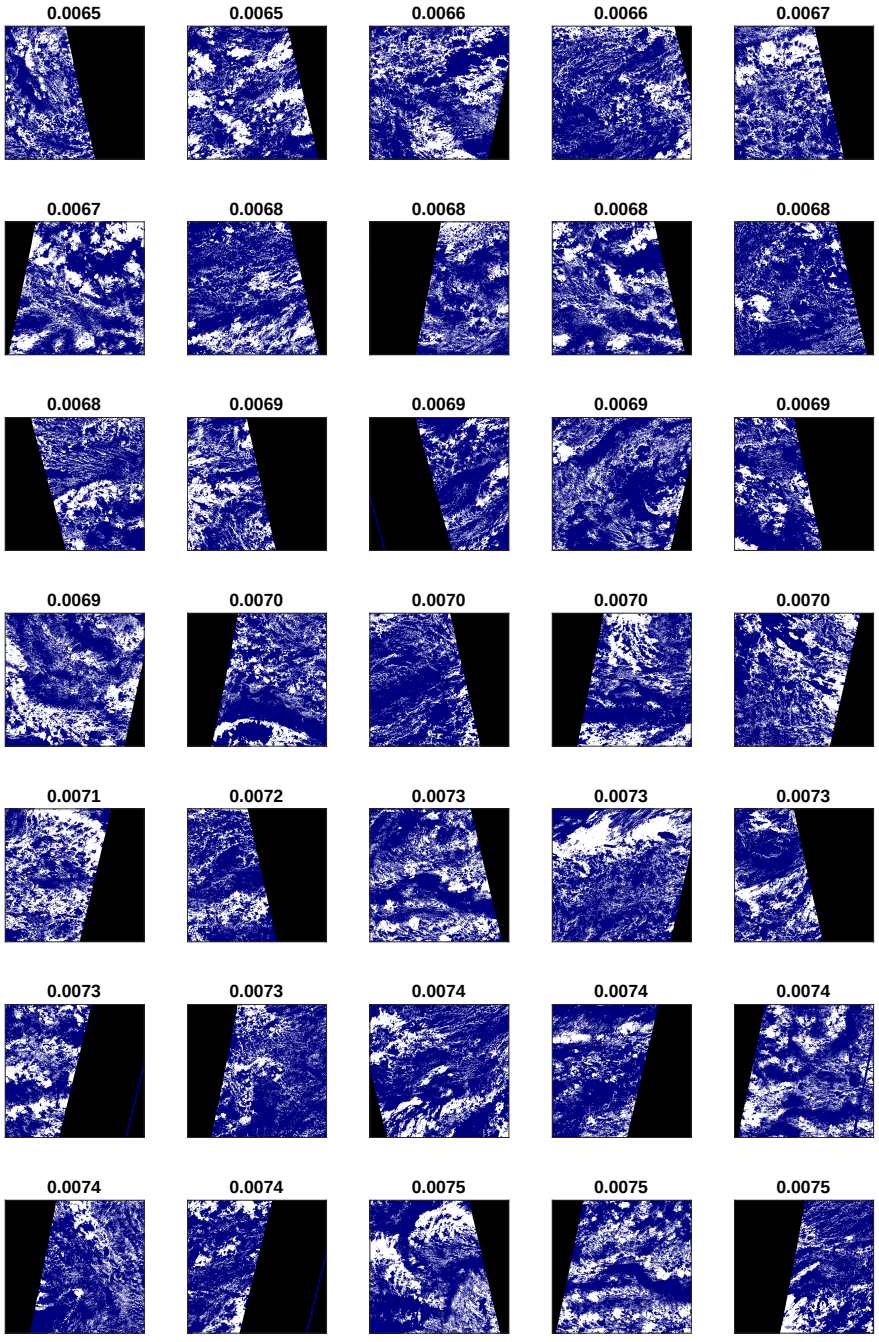


Figure A.7: Ordered fields 211-245

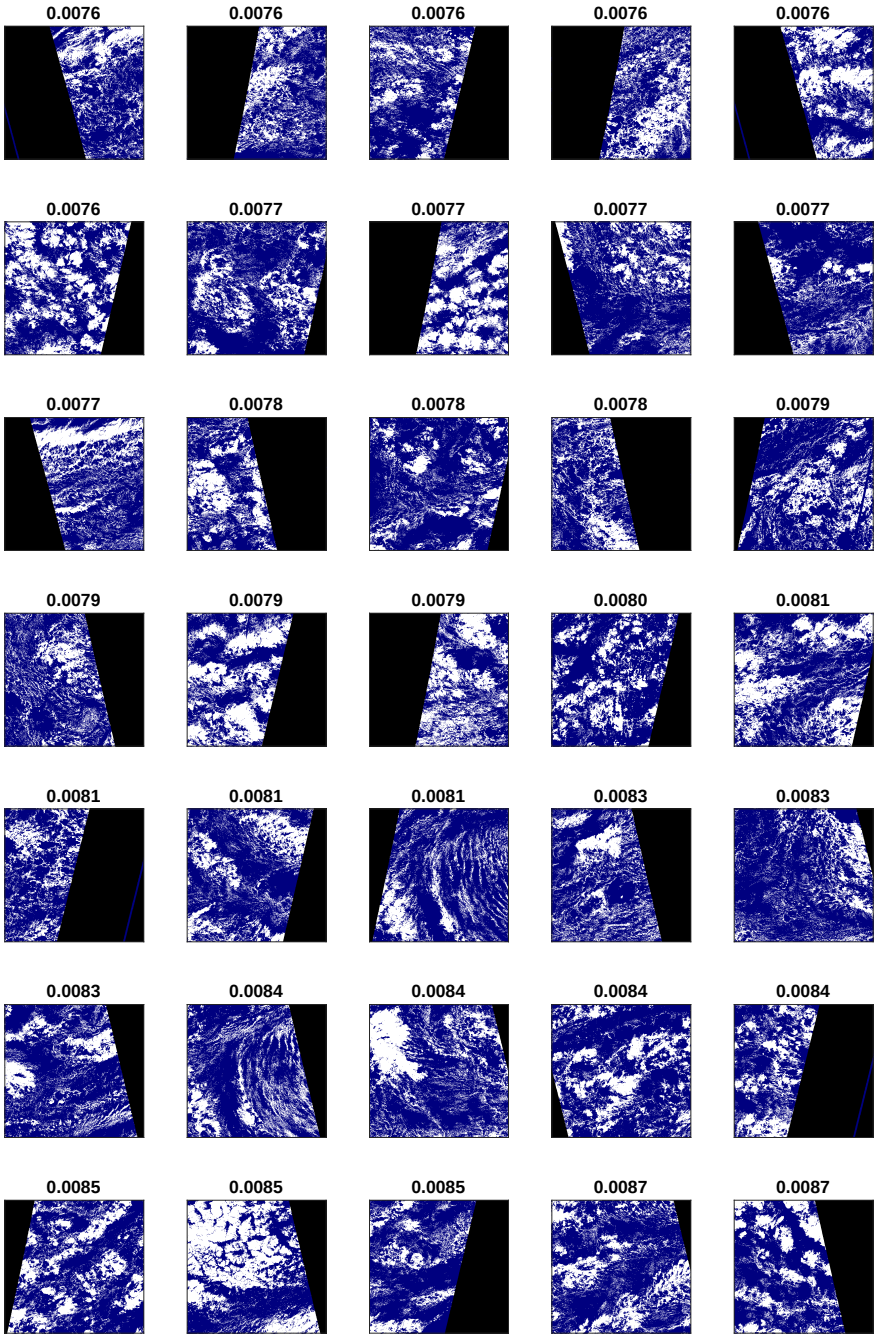


Figure A.8: Ordered fields 246-280

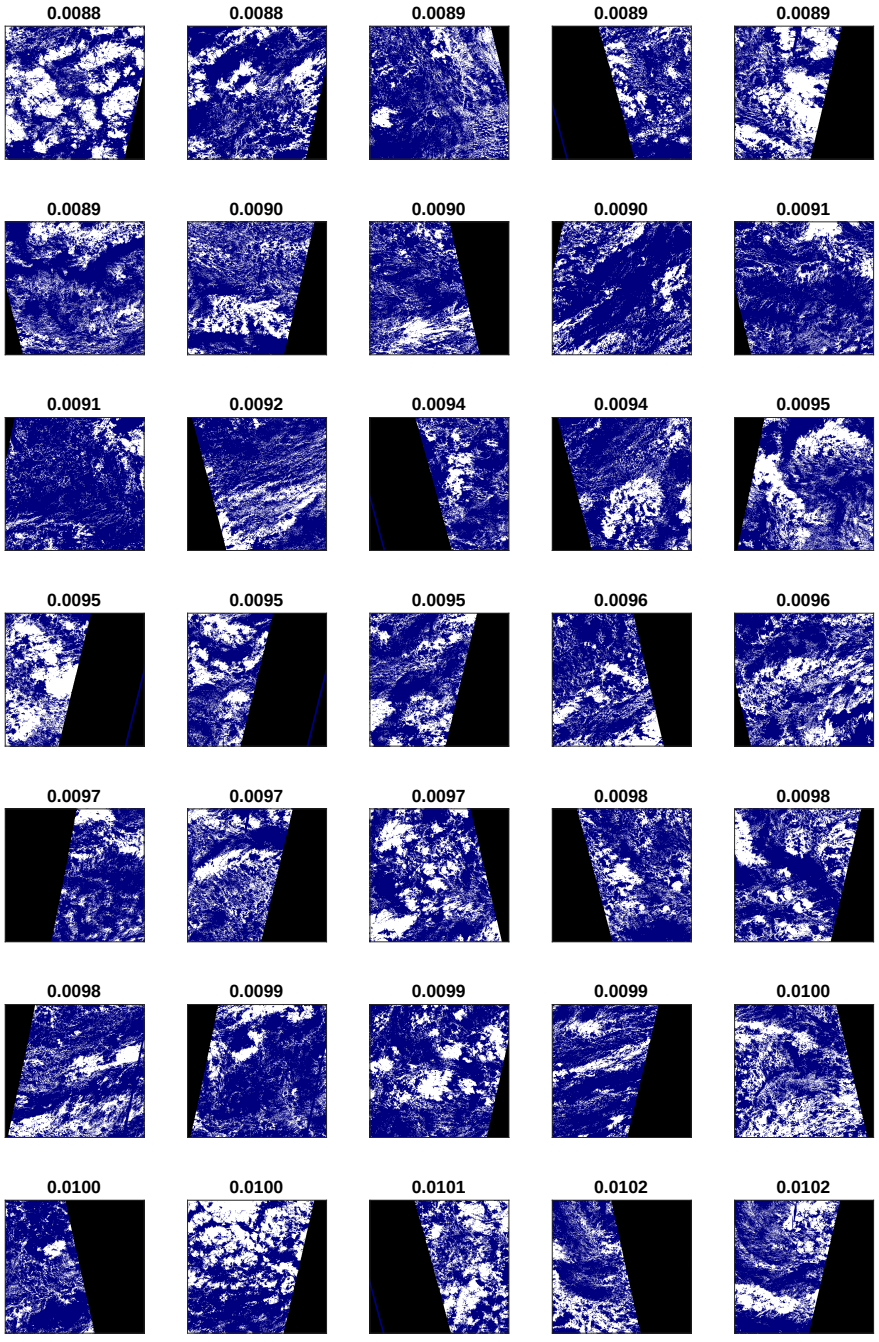


Figure A.9: Ordered fields 281-315

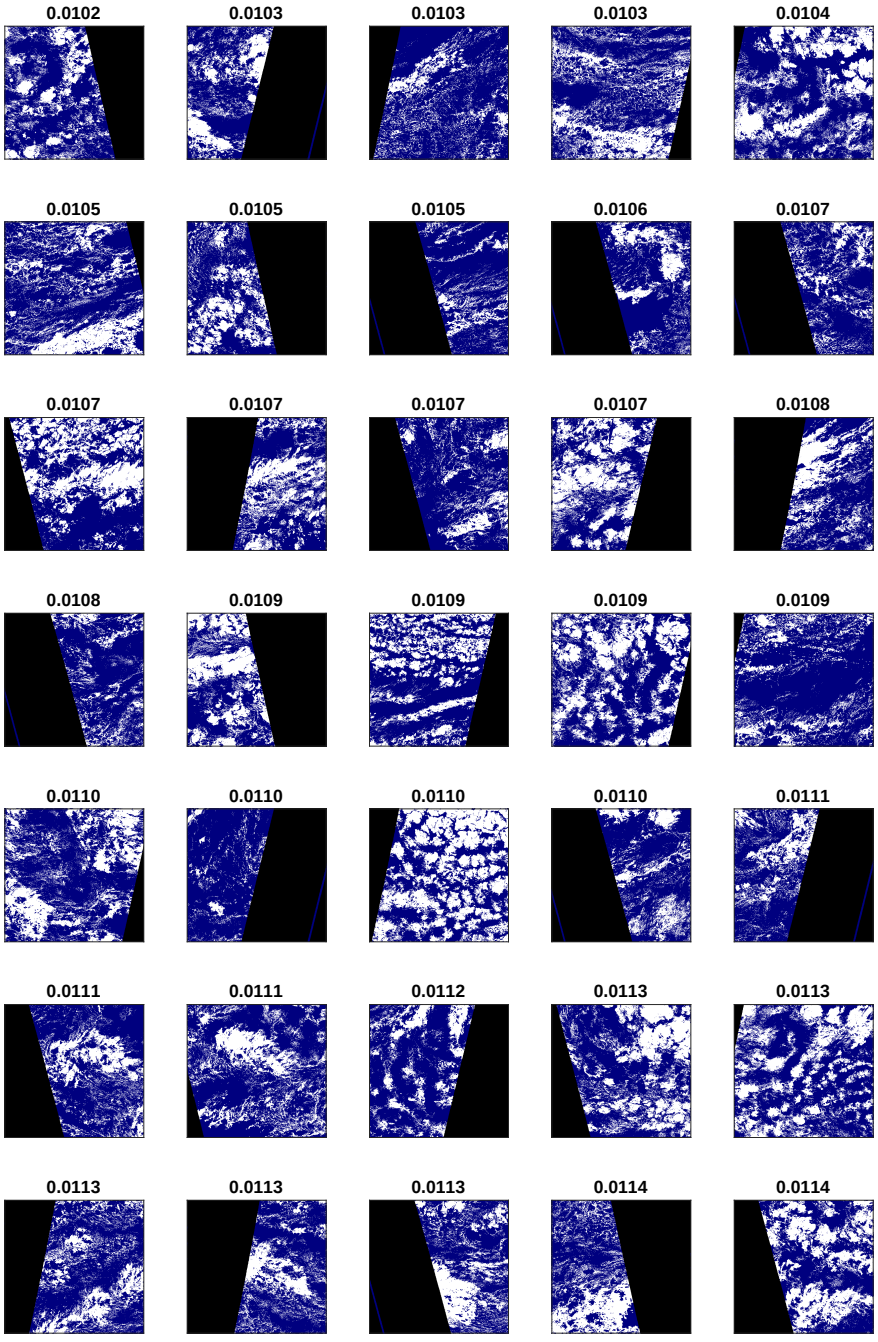


Figure A.10: Ordered fields 316-350

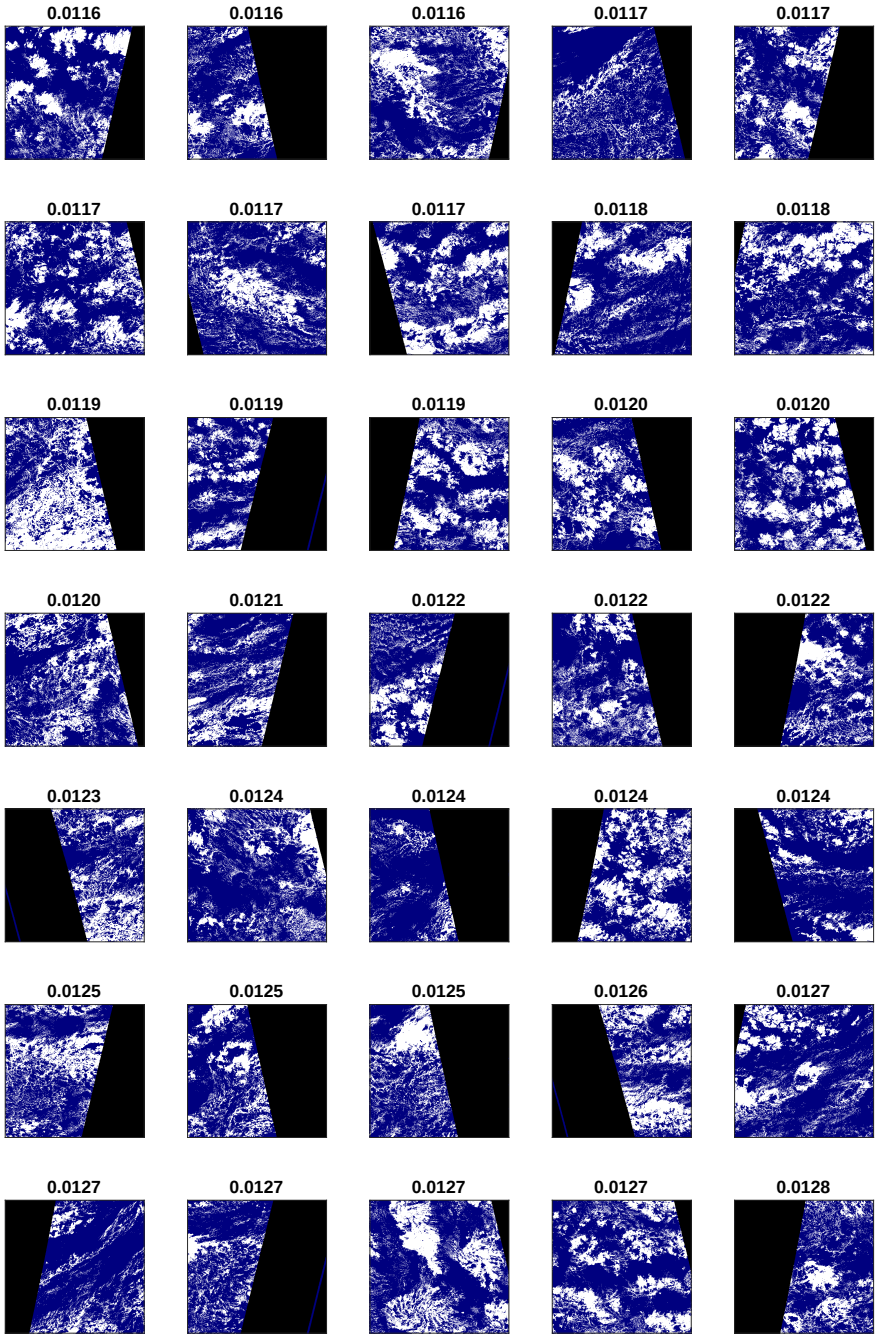


Figure A.11: Ordered fields 351-385

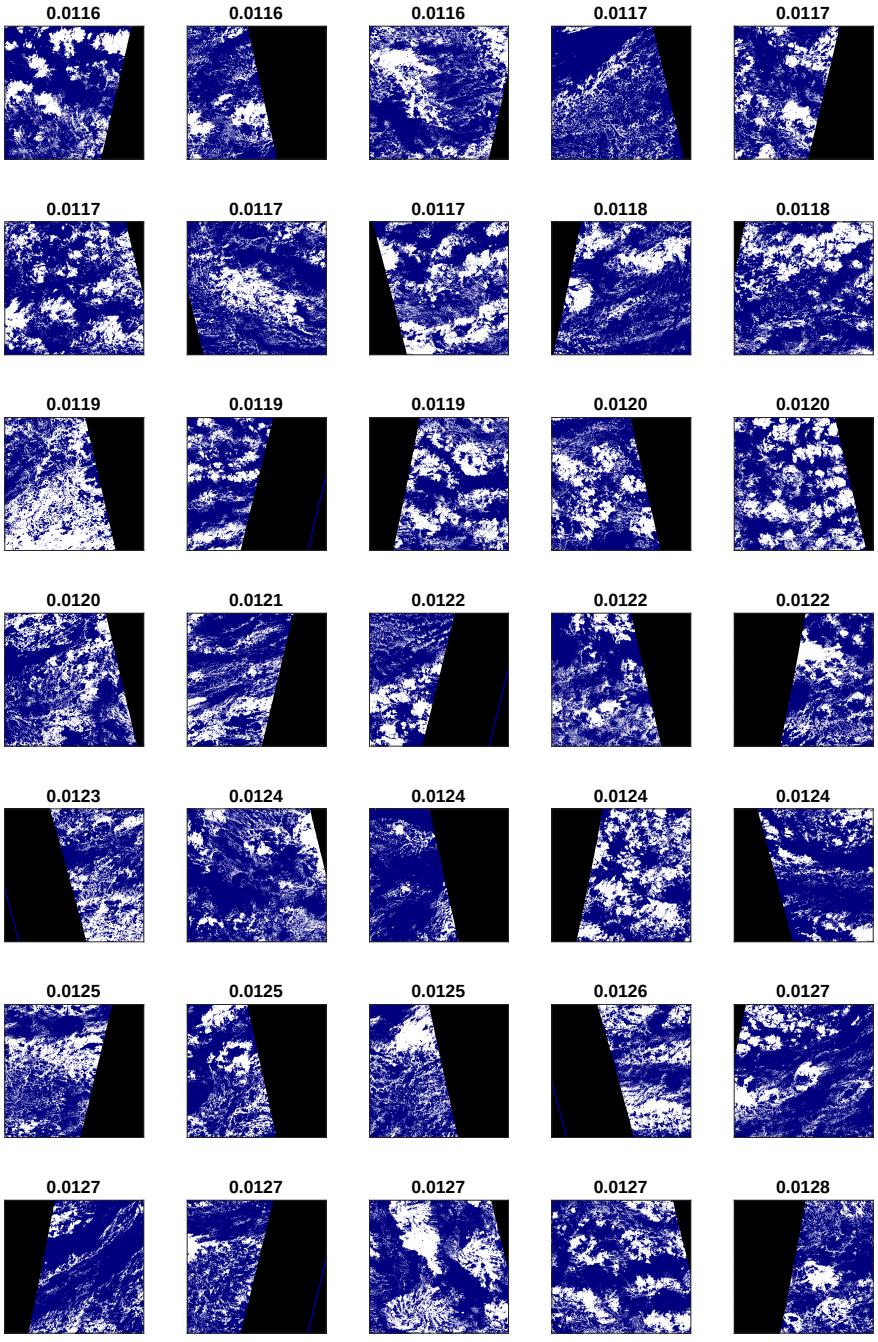


Figure A.12: Ordered fields 351-385

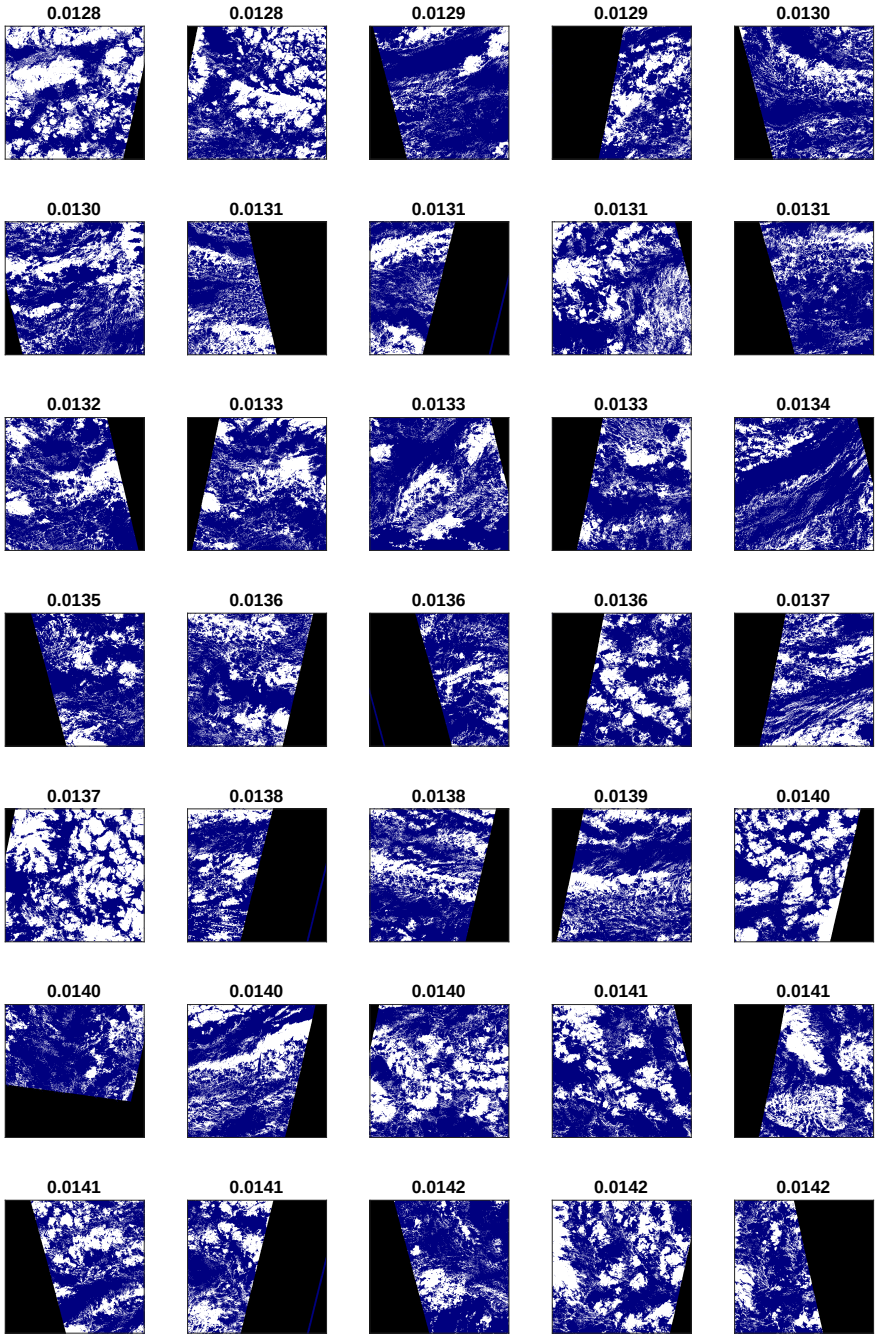


Figure A.13: Ordered fields 386-420

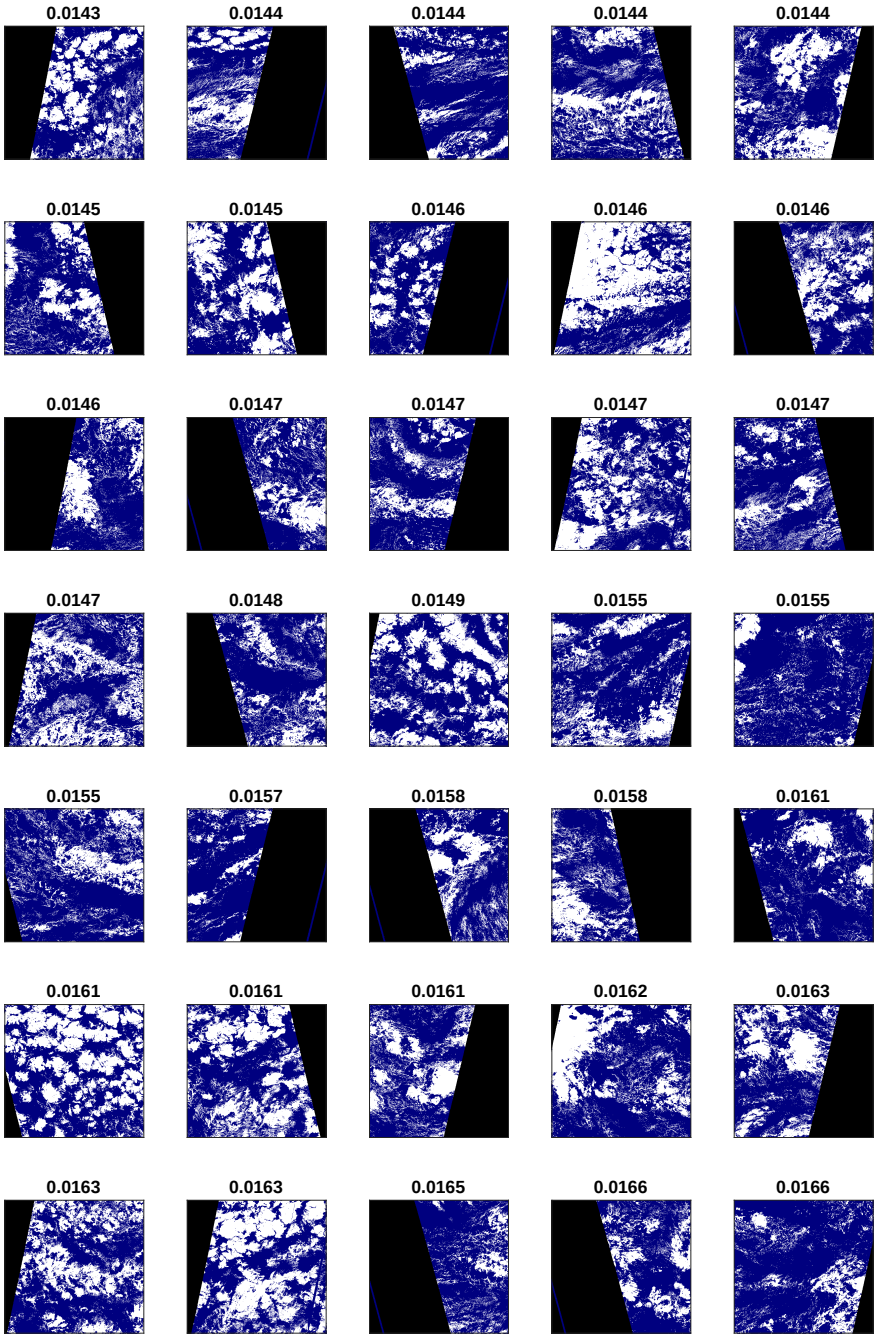


Figure A.14: Ordered fields 421-455

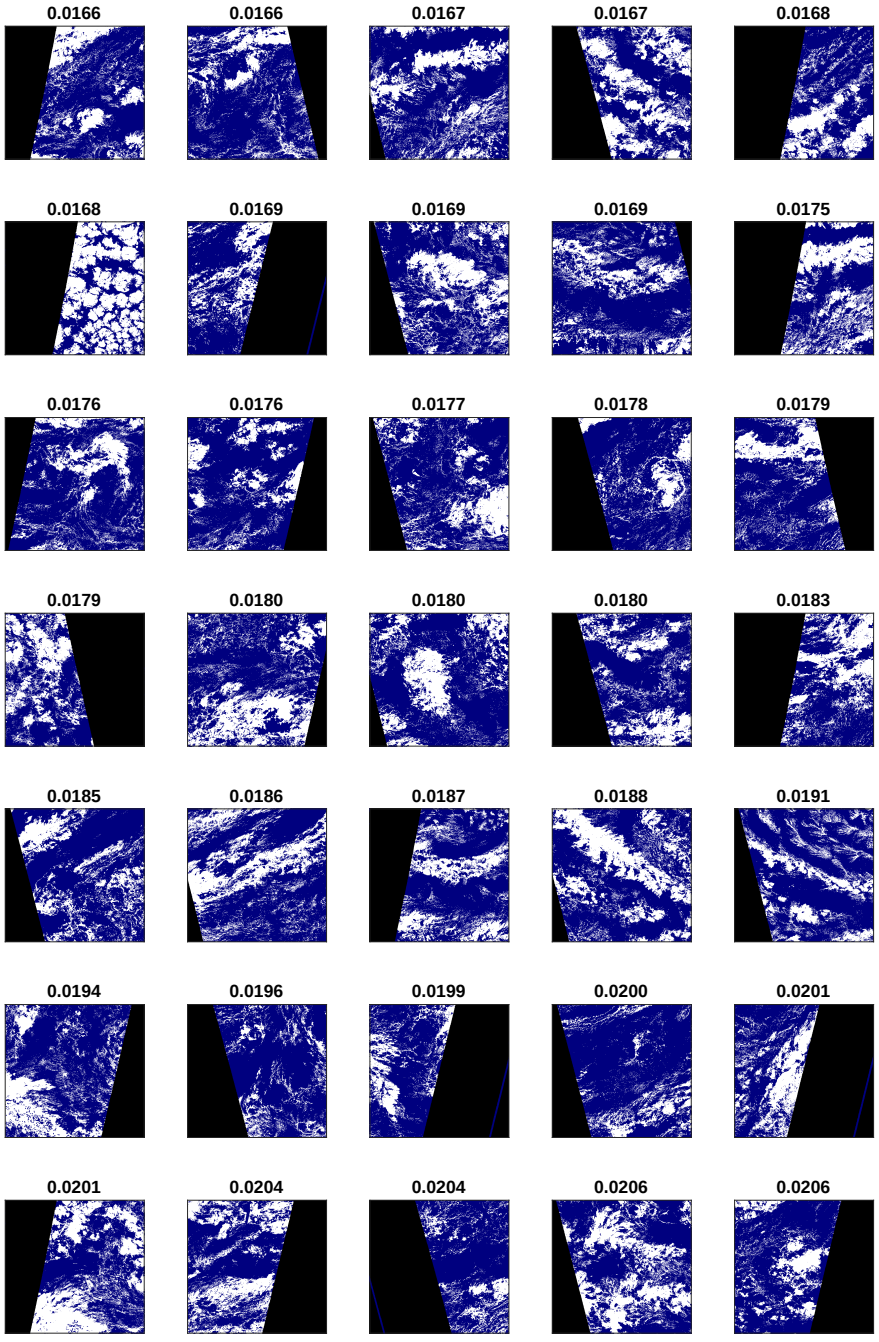


Figure A.15: Ordered fields 456-490

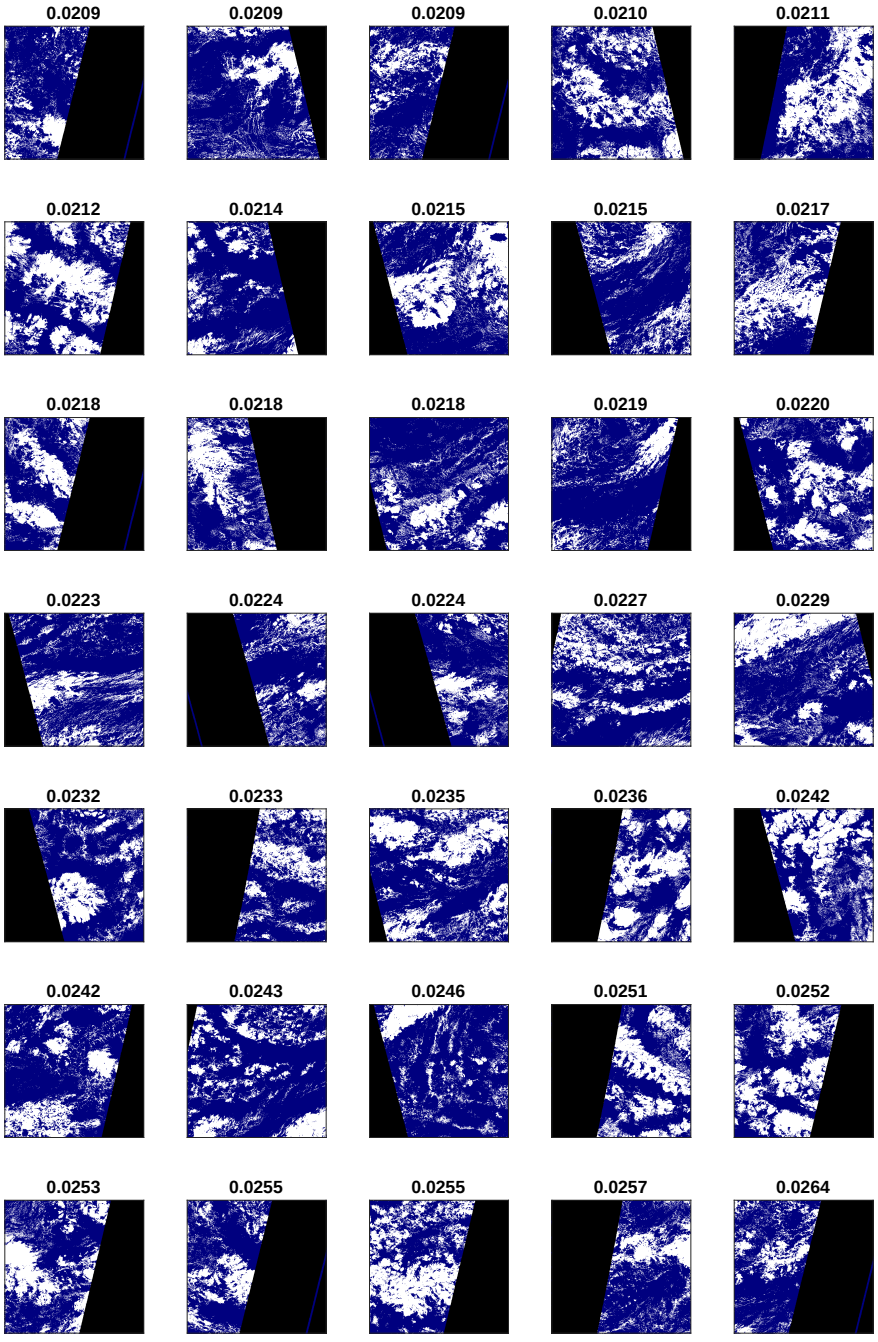


Figure A.16: Ordered fields 491-525

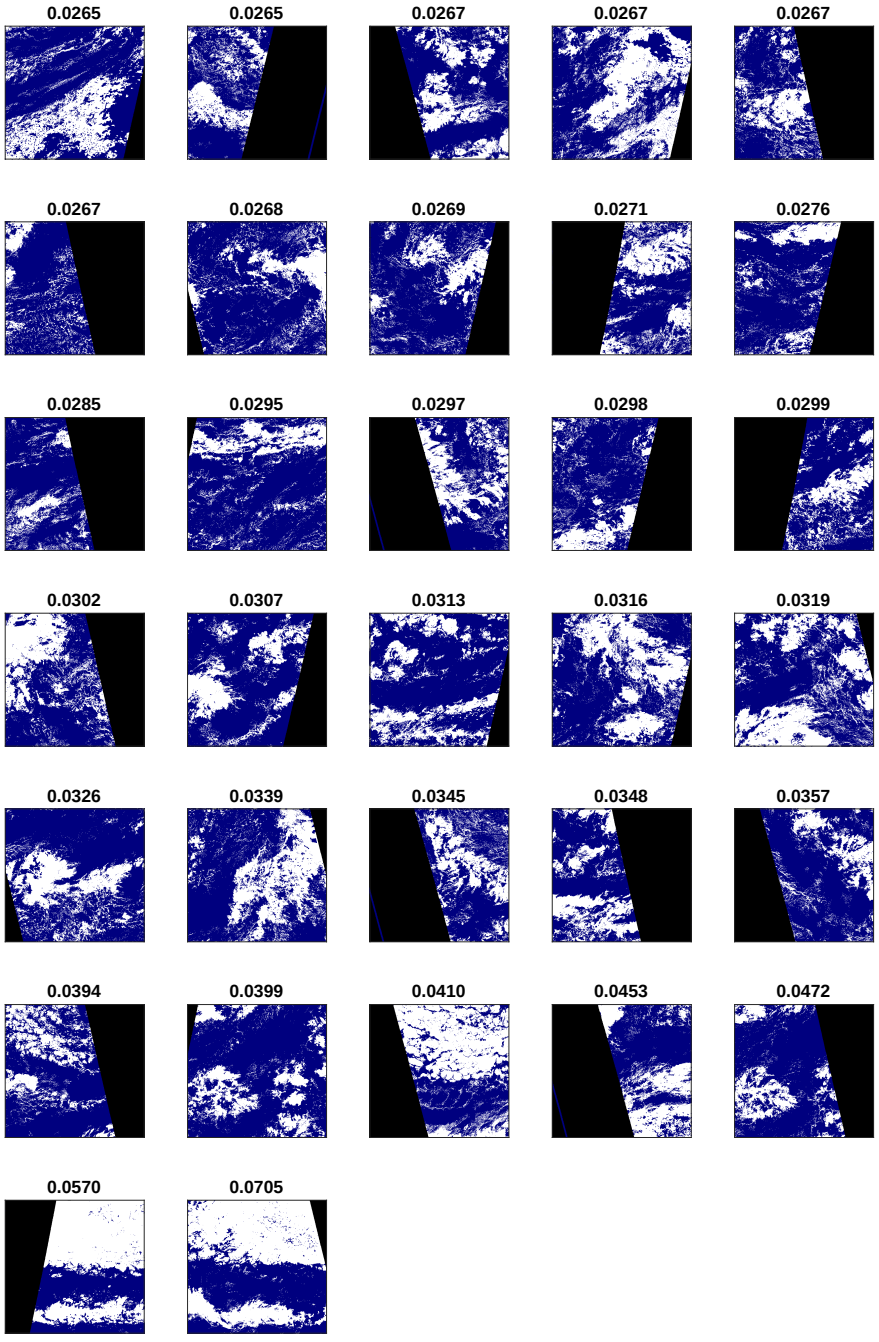


Figure A.17: Ordered fields 525-557

B

Comparison fields ordered between OI_3, I_{org} and $SCAI$

OI_3	I_{org}	$SCAI$	OI_3	I_{org}	$SCAI$	OI_3	I_{org}	$SCAI$	OI_3	I_{org}	$SCAI$
1	6	38	51	12	6	101	134	67	151	77	17
2	39	4	52	18	150	102	89	233	152	240	293
3	50	110	53	76	12	103	126	221	153	200	99
4	3	11	54	59	195	104	152	42	154	20	79
5	86	72	55	16	112	105	53	127	155	253	90
6	17	3	56	283	319	106	101	209	156	63	315
7	155	56	57	102	157	107	66	208	157	183	394
8	2	2	58	4	30	108	160	184	158	120	257
9	1	5	59	304	234	109	38	24	159	118	74
10	35	125	60	148	103	110	131	377	160	100	354
11	22	1	61	28	178	111	69	100	161	173	239
12	110	34	62	74	20	112	217	329	162	165	290
13	236	144	63	129	249	113	137	107	163	179	227
14	113	23	64	32	25	114	185	230	164	230	126
15	11	247	65	73	54	115	97	271	165	242	167
16	87	115	66	130	62	116	171	84	166	82	135
17	5	117	67	44	27	117	70	215	167	192	14
18	7	106	68	45	13	118	46	317	168	95	141
19	10	63	69	105	9	119	190	316	169	24	129
20	65	69	70	23	65	120	51	186	170	78	48
21	9	22	71	156	360	121	177	192	171	206	123
22	15	15	72	37	51	122	380	58	172	158	275
23	49	10	73	33	97	123	229	212	173	19	146
24	58	55	74	369	130	124	237	116	174	124	224
25	25	201	75	71	277	125	132	68	175	195	140
26	27	7	76	98	165	126	157	175	176	123	59
27	259	43	77	42	143	127	80	171	177	411	374
28	163	45	78	48	281	128	83	256	178	141	28
29	247	287	79	61	292	129	352	179	179	321	349
30	99	89	80	142	180	130	114	193	180	273	226
31	92	35	81	75	214	131	104	31	181	208	98
32	29	198	82	138	8	132	79	16	182	139	36
33	8	273	83	231	124	133	167	218	183	256	44
34	88	388	84	40	328	134	189	76	184	297	367
35	109	156	85	21	46	135	47	132	185	93	244
36	103	137	86	181	318	136	188	322	186	294	228
37	112	40	87	175	93	137	186	237	187	223	80
38	34	71	88	81	81	138	149	162	188	64	326
39	67	41	89	198	197	139	364	164	189	146	73
40	219	231	90	55	381	140	62	50	190	340	433
41	52	187	91	43	185	141	249	131	191	374	216
42	108	114	92	381	152	142	36	154	192	288	91
43	60	302	93	84	33	143	225	158	193	133	312
44	41	182	94	203	18	144	308	37	194	164	444
45	14	120	95	30	47	145	313	282	195	241	331
46	257	196	96	68	29	146	295	217	196	136	205
47	54	263	97	56	88	147	346	305	197	367	173
48	135	111	98	72	356	148	125	191	198	222	357
49	172	200	99	26	145	149	13	52	199	106	53
50	122	254	100	147	92	150	128	251	200	31	19

OI_3	I_{org}	$SCAI$	OI_3	I_{org}	$SCAI$	OI_3	I_{org}	$SCAI$	OI_3	I_{org}	$SCAI$
201	345	240	251	279	470	301	280	119	351	520	480
202	245	294	252	275	355	302	238	190	352	268	78
203	176	139	253	246	434	303	412	450	353	202	404
204	305	269	254	302	270	304	329	189	354	475	429
205	502	109	255	376	155	305	368	424	355	355	341
206	178	456	256	96	148	306	312	361	356	486	523
207	335	286	257	250	259	307	404	323	357	254	332
208	420	102	258	414	385	308	390	430	358	392	453
209	90	121	259	174	108	309	342	229	359	378	371
210	153	60	260	332	425	310	228	379	360	325	426
211	111	66	261	226	96	311	443	105	361	264	491
212	314	382	262	218	460	312	293	500	362	419	344
213	194	369	263	85	299	313	266	363	363	370	447
214	258	285	264	363	442	314	162	75	364	388	339
215	145	172	265	251	451	315	166	202	365	461	499
216	278	421	266	255	133	316	362	283	366	351	396
217	213	258	267	193	276	317	261	203	367	263	220
218	150	82	268	292	264	318	424	389	368	389	147
219	252	296	269	248	118	319	151	359	369	382	272
220	357	204	270	282	246	320	379	497	370	339	280
221	184	149	271	344	310	321	286	398	371	220	232
222	154	122	272	260	236	322	296	241	372	331	336
223	410	113	273	243	343	323	415	159	373	523	161
224	285	298	274	474	488	324	299	39	374	395	527
225	356	134	275	244	250	325	234	94	375	482	353
226	205	358	276	430	463	326	398	541	376	196	222
227	232	160	277	107	532	327	57	219	377	431	306
228	350	207	278	303	213	328	435	278	378	227	87
229	91	169	279	318	402	329	276	445	379	216	252
230	94	181	280	489	439	330	310	284	380	479	448
231	121	335	281	406	513	331	451	199	381	538	469
232	143	26	282	449	462	332	287	378	382	311	86
233	211	303	283	289	255	333	499	534	383	375	473
234	116	248	284	384	243	334	438	509	384	484	479
235	127	32	285	204	401	335	393	364	385	341	211
236	214	210	286	215	288	336	327	413	386	436	540
237	290	85	287	161	225	337	547	384	387	408	525
238	324	370	288	140	64	338	354	552	388	532	411
239	338	206	289	442	496	339	316	128	389	387	368
240	159	307	290	337	314	340	212	49	390	333	308
241	197	61	291	450	423	341	326	188	391	397	440
242	284	136	292	180	142	342	269	366	392	199	170
243	182	304	293	334	77	343	509	406	393	191	95
244	221	399	294	348	321	344	322	459	394	426	506
245	119	21	295	235	309	345	330	505	395	365	242
246	170	104	296	169	397	346	291	266	396	416	347
247	144	101	297	224	151	347	233	70	397	358	409
248	210	177	298	201	153	348	168	168	398	300	345
249	117	183	299	239	163	349	115	57	399	301	262
250	265	383	300	187	350	350	307	494	400	522	481

OI_3	I_{org}	$SCAI$	OI_3	I_{org}	$SCAI$	OI_3	I_{org}	$SCAI$	OI_3	I_{org}	$SCAI$
401	472	372	451	439	507	501	503	408	551	549	529
402	298	362	452	500	550	502	270	265	552	553	548
403	464	289	453	477	194	503	537	486	553	513	544
404	518	501	454	459	340	504	490	472	554	542	503
405	306	435	455	446	403	505	524	476	555	551	530
406	481	555	456	425	342	506	465	432	556	540	557
407	281	235	457	423	351	507	485	320	557	557	556
408	359	412	458	456	461	508	429	176			
409	272	391	459	405	508	509	526	539			
410	492	546	460	457	419	510	399	417			
411	478	295	461	360	554	511	515	437			
412	317	431	462	274	330	512	514	466			
413	320	482	463	371	395	513	512	536			
414	527	520	464	432	400	514	445	474			
415	373	465	465	400	452	515	507	478			
416	391	410	466	469	392	516	504	516			
417	418	338	467	539	495	517	554	549			
418	361	138	468	476	390	518	546	477			
419	462	517	469	483	300	519	519	515			
420	471	334	470	455	333	520	494	521			
421	434	458	471	353	352	521	413	376			
422	277	223	472	319	405	522	510	375			
423	536	427	473	448	455	523	467	538			
424	267	348	474	473	416	524	394	279			
425	309	414	475	315	327	525	505	386			
426	427	365	476	506	492	526	511	528			
427	433	526	477	402	464	527	447	311			
428	530	436	478	428	418	528	544	533			
429	209	543	479	470	510	529	421	535			
430	366	387	480	487	467	530	409	301			
431	262	174	481	349	346	531	493	324			
432	422	253	482	534	420	532	517	487			
433	323	238	483	497	380	533	533	438			
434	377	504	484	548	489	534	463	449			
435	417	261	485	372	274	535	460	297			
436	383	457	486	401	454	536	528	325			
437	407	268	487	454	471	537	531	511			
438	498	547	488	501	337	538	488	422			
439	343	446	489	453	512	539	466	313			
440	508	475	490	525	415	540	521	443			
441	403	428	491	437	245	541	480	502			
442	550	485	492	496	373	542	555	542			
443	271	83	493	440	291	543	545	551			
444	207	166	494	468	490	544	444	518			
445	543	468	495	385	498	545	491	531			
446	535	553	496	495	545	546	541	537			
447	441	522	497	556	514	547	516	524			
448	328	260	498	452	493	548	396	267			
449	386	484	499	529	441	549	458	519			
450	347	393	500	336	407	550	552	483			

Bibliography

Timothy C Benner and Judith A Curry. Characteristics of small tropical cumulus clouds and their impact on the environment. *Journal of Geophysical Research: Atmospheres*, 103(D22):28753–28767, 1998.

Sandrine Bony and Jean-Louis Dufresne. Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, 32(20), 2005.

Sandrine Bony, J-L Dufresne, Hervé Le Treut, J-J Morcrette, and Catherine Senior. On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, 22(2-3):71–86, 2004.

Sandrine Bony, Bjorn Stevens, Isaac H. Held, John F. Mitchell, Jean-Louis Dufresne, Kerry A. Emanuel, Pierre Friedlingstein, Stephen Griffies, and Catherine Senior. *Climate Science for Serving Society: Research, Modelling and Prediction Priorities*. Springer, Berlin, 2013.

Sandrine Bony, Bjorn Stevens, Dargan MW Frierson, Christian Jakob, Masa Kageyama, Robert Pincus, Theodore G Shepherd, Steven C Sherwood, A Pier Siebesma, Adam H Sobel, et al. Clouds, circulation and climate sensitivity. *Nature Geoscience*, 8(4):261, 2015.

Sandrine Bony, Bjorn Stevens, Felix Ament, Sebastien Bigorre, Patrick Chazette, Susanne Crewell, Julien Delanoë, Kerry Emanuel, David Farrell, Cyrille Flamant, et al. Eurec4a: a field campaign to elucidate the couplings between clouds, convection and circulation. *Surveys in Geophysics*, 38(6):1529–1568, 2017.

Sandrine Bony, Addisu Semie, Tristan L'Ecuyer, Alyson Douglas, Bjorn Stevens, and EUREC4A ISSI Science Team. Variability of the mesoscale organization of shallow convection over the tropical atlantic. Presentation at EGU 2018, 2018.

Olivier Boucher, David Randall, Paulo Artaxo, Christopher Bretherton, Gragam Feingold, Piers Forster, V-M Kerminen, Yutaka Kondo, Hong Liao, Ulrike Lohmann, et al. Clouds and aerosols. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 571–657. Cambridge University Press, 2013.

Christopher S Bretherton, Peter N Blossey, and Marat Khairoutdinov. An energy-balance analysis of deep convective self-aggregation above uniform sst. *Journal of the atmospheric sciences*, 62(12):4273–4292, 2005.

Matthias Brueck, Louise Nuijens, and Bjorn Stevens. On the seasonal and synoptic time-scale variability of the north atlantic trade wind region and its low-level clouds. *Journal of the Atmospheric Sciences*, 72(4): 1428–1446, 2015.

Stephen Burt. Cumulus clouds, 2009. URL `https://cloudatlas.wmo.int/species-cumulus-humilis-cu-hum.html`. Last visited 2018-10-10.

Robert F Cahalan and Joachim H Joseph. Fractal statistics of cloud fields. *Monthly weather review*, 117(2): 261–272, 1989.

Robert F Cahalan, William Ridgway, Warren J Wiscombe, Thomas L Bell, and Jack B Snider. The albedo of fractal stratocumulus clouds. *Journal of the Atmospheric Sciences*, 51(16):2434–2455, 1994.

Brenda G Cohen and George C Craig. The response time of a convective cloud ensemble to a change in forcing. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 130(598):933–944, 2004.

George C Craig and Brenda G Cohen. Fluctuations in an equilibrium convective ensemble. part i: Theoretical formulation. *Journal of the atmospheric sciences*, 63(8):1996–2004, 2006.

JT Dawe and PH Austin. Statistical analysis of an les shallow cumulus cloud ensemble using a cloud tracking algorithm. *Atmospheric Chemistry and Physics*, 12(2):1101–1119, 2012.

Kristene Rodriques E. Moghaddam-Taaheri. *HDF-EOS to GeoTIFF Conversion TOOL (HEG) Stand-alone User's Guide*, 2017.

Thijs Heus and Axel Seifert. Automated tracking of shallow cumulus clouds in large domain, long duration large eddy simulations. *Geoscientific Model Development*, 6(4):1261, 2013.

Laura M Hinkelman, K Franklin Evans, Eugene E Clothiaux, Thomas P Ackerman, and Paul W Stackhouse Jr. The effect of cumulus cloud field anisotropy on domain-averaged solar fluxes and atmospheric heating rates. *Journal of the Atmospheric Sciences*, 64(10):3499–3520, 2007.

L. Howard. Definition of cumulus, 1803. Last visited 2018-08-31.

Kunihiko Hozumi, Toshio Harimaya, and Choji Magono. The size distribution of cumulus clouds as a function of cloud amount. *Journal of the Meteorological Society of Japan. Ser. II*, 60(2):691–699, 1982.

IPCC. Global warming of 1.5°c, summary for policymakers., 2018.

Nadir Jeevanjee and David M Romps. Convective self-aggregation, cold pools, and domain size. *Geophysical Research Letters*, 40(5):994–998, 2013.

Michael D King, Steven Platnick, W Paul Menzel, Steven A Ackerman, and Paul A Hubanks. Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7):3826–3852, 2013.

Kwo-Sen Kuo, Ronald M Welch, Ronald C Weger, Mark A Engelstad, and SK Sengupta. The three-dimensional structure of cumulus clouds over the ocean: 1. structural analysis. *Journal of Geophysical Research: Atmospheres*, 98(D11):20685–20711, 1993.

Margaret A LeMone and Edward J Zipser. Cumulonimbus vertical velocity events in gate. part i: Diameter, intensity and mass flux. *Journal of the Atmospheric Sciences*, 37(11):2444–2457, 1980.

Raúl Erlando López. The lognormal distribution and cumulus cloud populations. *Monthly Weather Review*, 105(7):865–872, 1977.

Peter Lynch. The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, 227(7):3431 – 3444, 2008. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2007.02.034. URL `http://www.sciencedirect.com/science/article/pii/S0021999107000952`. Predicting weather, climate and extreme events.

A M. Sayer, N C. Hsu, and C Bettenhausen. Implications of modis bow-tie distortion on aerosol optical depth retrievals, and techniques for mitigation. *Atmospheric Measurement Techniques*, 8, 12 2015. doi: 10.5194/amt-8-5277-2015.

LAT Machado and WB Rossow. Structural characteristics and radiative properties of tropical cloud clusters. *Monthly Weather Review*, 121(12):3234–3260, 1993.

Andrew J Majda and Boualem Khouider. Stochastic and mesoscopic models for tropical convection. *Proceedings of the National Academy of Sciences*, 99(3):1123–1128, 2002.

Guillaume S Mauger and Joel R Norris. Assessing the impact of meteorological history on subtropical cloud fraction. *Journal of Climate*, 23(11):2926–2940, 2010.

Brian Medeiros, Bjorn Stevens, Isaac M Held, Ming Zhao, David L Williamson, Jerry G Olson, and Christopher S Bretherton. Aquaplanets, climate sensitivity, and low clouds. *Journal of Climate*, 21(19):4974–4991, 2008.

Brian Medeiros, Bjorn Stevens, and Sandrine Bony. Using aquaplanets to understand the robust responses of comprehensive climate models to forcing. *Climate Dynamics*, 44(7-8):1957–1977, 2015.

Patrick Minnis, David F Young, Sunny Sun-Mack, Patrick W Heck, David R Doelling, and Qing Z Trepte. Ceres cloud property retrievals from imagers on trmm, terra, and aqua. In *Remote Sensing of Clouds and the Atmosphere VIII*, volume 5235, pages 37–49. International Society for Optics and Photonics, 2004.

Caroline J Muller and Isaac M Held. Detailed investigation of the self-aggregation of convection in cloud-resolving simulations. *Journal of the Atmospheric Sciences*, 69(8):2551–2565, 2012.

US Nair, RC Weger, KS Kuo, and RM Welch. Clustering, randomness, and regularity in cloud fields: 5. the nature of regular cumulus cloud fields. *Journal of Geophysical Research: Atmospheres*, 103(D10):11363–11380, 1998.

RAJ Neggers. Attributing the behavior of low-level clouds in large-scale models to subgrid-scale parameterizations. *Journal of Advances in Modeling Earth Systems*, 7(4):2029–2043, 2015.

RAJ Neggers, HJJ Jonker, and AP Siebesma. Size statistics of cumulus cloud populations in large-eddy simulations. *Journal of the atmospheric sciences*, 60(8):1060–1074, 2003.

Roel AJ Neggers, J David Neelin, and Bjorn Stevens. Impact mechanisms of shallow cumulus convection on tropical climate dynamics. *Journal of climate*, 20(11):2623–2642, 2007.

NOAA. Goes-r series satellites, 2018. URL `https://www.ncdc.noaa.gov/data-access/satellite-data/goes-r-series-satellites`. Last visited 2018-07-04.

Louise Nuijens and Bjorn Stevens. The influence of wind speed on shallow marine cumulus convection. *Journal of the Atmospheric Sciences*, 69(1):168–184, 2012.

Louise Nuijens, Ilya Serikov, Lutz Hirsch, Katrin Lonitz, and Bjorn Stevens. The distribution and variability of low-level cloud in the north atlantic trades. *Quarterly Journal of the Royal Meteorological Society*, 140(684):2364–2374, 2014.

Vernon G Plank. The size distribution of cumulus clouds in representative florida populations. *Journal of Applied Meteorology*, 8(1):46–67, 1969.

Xin Qu, Alex Hall, Stephen A Klein, and Peter M Caldwell. On the spread of changes in marine low cloud cover in climate model simulations of the 21st century. *Climate dynamics*, 42(9-10):2603–2626, 2014.

Johannes Quaas. Evaluating the "critical relative humidity" as a measure of subgrid-scale variability of humidity in general circulation model cloud cover parameterizations using satellite data. *Journal of Geophysical Research: Atmospheres*, 117(D9), 2012.

A Robert. Cloud clusters and large-scale vertical motions in the tropics. *Journal of the Meteorological Society of Japan*, 60(1), 1982.

Axel Seifert and Thijs Heus. Large-eddy simulation of organized precipitating trade wind cumulus clouds. *Atmos. Chem. Phys*, 13(11):5631–5645, 2013.

SK Sengupta, RM Welch, MS Navar, TA Berendes, and DW Chen. Cumulus cloud field morphology and spatial patterns derived from high spatial resolution landsat imagery. *Journal of Applied Meteorology*, 29(12):1245–1267, 1990.

Bjorn Stevens, David Farrell, Lutz Hirsch, Friedhelm Jansen, Louise Nuijens, Ilya Serikov, Björn Brügmann, Marvin Forde, Holger Linne, Katrin Lonitz, et al. The barbados cloud observatory: Anchoring investigations of clouds and circulation on the edge of the itcz. *Bulletin of the American Meteorological Society*, 97(5):787–801, 2016.

Bjorn Stevens, Sandrine Bony, Helene Brogniez, Laureline Hentge, Cathy Hohenegger, Christoph Kiemle, Tristan S. L'Ecuyer, Ann Kristin Naumann, Hauke Schulz, Pier Siebesma, Jessica Vial, Dave M. Winker, and Paquite Zuidema. Sugar, gravel, fish and flowers. Unpublished paper., 2019.

Dietrich Stoyan, Wilfrid S Kendall, and Joseph Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 1987.

Kathleen I Strabala. *MODIS cloud mask user's guide*. University of Wisconsin–Madison, 2005.

Isabelle Tobin, Sandrine Bony, and Remy Roca. Observational evidence for relationships between the degree of aggregation of deep convection, water vapor, surface fluxes, and radiation. *Journal of Climate*, 25(20):6885–6904, 2012.

Adrian M Tompkins. Organization of tropical convection in low vertical wind shears: The role of cold pools. *Journal of the atmospheric sciences*, 58(13):1650–1672, 2001.

Adrian M Tompkins and Addisu G Semie. Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *Journal of Advances in Modeling Earth Systems*, 9(2):1046–1068, 2017.

Jessica Vial, Jean-Louis Dufresne, and Sandrine Bony. On the interpretation of inter-model spread in cmip5 climate sensitivity estimates. *Climate Dynamics*, 41(11-12):3339–3362, 2013.

Jessica Vial, Sandrine Bony, Jean-Louis Dufresne, and Romain Roehrig. Coupling between lower-tropospheric convective mixing and low-level clouds: Physical mechanisms and dependence on convection scheme. *Journal of advances in modeling earth systems*, 8(4):1892–1911, 2016.

Jessica Vial, Sandrine Bony, Bjorn Stevens, and Raphaela Vogel. Mechanisms and model diversity of trade-wind shallow cumulus cloud feedbacks: a review. *Surveys in geophysics*, 38(6):1331–1353, 2017.

Mark J Webb, CA Senior, DMH Sexton, WJ Ingram, KD Williams, MA Ringer, BJ McAvaney, R Colman, BJ Soden, R Gudgel, et al. On the contribution of local feedback mechanisms to the range of climate sensitivity in two gcm ensembles. *Climate dynamics*, 27(1):17–38, 2006.

RC Weger, J Lee, Tianri Zhu, and RM Welch. Clustering, randomness and regularity in cloud fields: 1. theoretical considerations. *Journal of Geophysical Research: Atmospheres*, 97(D18):20519–20536, 1992.

Bruce A Wielicki and Ronald M Welch. Cumulus cloud properties derived using landsat satellite data. *Journal of Climate and Applied Meteorology*, 25(3):261–276, 1986.

Allison A Wing and Kerry A Emanuel. Physical mechanisms controlling self-aggregation of convection in idealized numerical modeling simulations. *Journal of Advances in Modeling Earth Systems*, 6(1):59–74, 2014.

Minghua Zhang, Christopher S Bretherton, Peter N Blossey, Phillip H Austin, Julio T Bacmeister, Sandrine Bony, Florent Brient, Suvarchal K Cheedela, Anning Cheng, Anthony D Genio, et al. Cgils: Results from the first phase of an international project to understand the physical mechanisms of low cloud feedbacks in single column models. *Journal of Advances in Modeling Earth Systems*, 5(4):826–842, 2013.

Guangyu Zhao and Larry Di Girolamo. Statistics on the macrophysical properties of trade wind cumuli over the tropical western atlantic. *Journal of Geophysical Research: Atmospheres*, 112(D10), 2007.