Universitetet i Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

BACHELOR THESIS

Study program/specialization: Bachelor of engineering / Data technology	Spring semester 2021 Open or Confidential
Author: Joachim Andreassen	
Subject manager: Erlend Tøssebro Supervisor: Steve Jothen	
Title of bachelor thesis: World-wide cloud data compiled from satellite imagery	
Credits: 20	
Keywords: clouds, satellites, wavelengths	Number of pages: + attachments/other: Stavanger - May 15, 2021

Contents

Summary

Denne rapporten fungerer som en mal for din egen rapport.

Chapter 1

Introduction

1.1 Problem

Multiple governmental entities throughout the world provide publicly available images captured by weather satellites at regular intervals. Sourcing these images, processing them, and storing them in an accessible format allows querying for cloud coverage for specific locations on Earth. The goal of this project is to be able to render a visualisation of world-wide cloud coverage by combining and normalising partial satellite imagery provided by different sources.

The produced world-wide visualisation should contain the most recent satellite data available. It is therefore critical that the programs execution time is kept low.

1.2 About the Company

Time and Date AS is a company based just outside Stavanger, Norway. The company operates timeanddate.com, which is the world's top ranked website for time and time zones. More than a million users access the website every day. In addition to time and time zones, timeanddate.com also provide services within weather, astronomy, calculators and much more. [13]

1.3 Structure

This thesis starts off introducing the theory and decisions made in chapter 2. The chapter includes assessments and explanations of why decisions where made, as well as calculations supporting the different conclusions. A lot of the work in this chapter is based on previous work from multiple sources.

The design and construction of software is presented in chapter 3. This chapter shows how the theory explained in chapter 2 is implemented.

Chapter 4 shows the results, and discusses what can be done to further improve this project. Lastly, a short conclusion is presented in chapter 5; following with the bibliography and attachments.

1.4 Technologies

1.4.1 Python

The programming language that is used in this project is Python. Python was chosen because there exists many useful libraries for handling satellite images, such as Satpy (see subsection 1.4.2 below). This reduces the complexity and development time of the project. Speed is a concern though when using Python, as it is a relatively slow language. Many of the libraries that executes the heavy computations in this project reduces this concern, as they uses underlying libraries that is implemented in faster languages like C.

1.4.2 Satpy

A central technology in this project is Satpy. Satpy is a Python library for reading, manipulating, and writing data from satellites. Satpy reduces the concern of different file formats, as multiple file readers is implemented. This allows reading satellite data from many different file formats into python objects called Scene. The satellite data in Scene objects can be altered in multiple ways; including changing projection, image resolution and much more. Combinations of satellite frequencies can also be made easily with Satpy. This is useful when extracting necessary data, based on the differences between the different satellite frequencies (see 2.2 at page 12). [7]

Chapter 2

Theory

2.1 Used satellites

Satellite imagery is used in this project to gather information about the cloud coverage. To cover the whole earth, multiple satellites is needed. The satellites utilized need to be designed for weather monitoring, so the cloud cover can be extracted from the satellite images. Because of these statements, it is important to choose the right satellites for this project's purpose.

2.1.1 Geostationary weather satellites

There is multiple satellites circling the earth in many different orbits. The orbit is chosen based on the satellites' intended applications. Figure 2.1 below contains two illustrations of orbits used by satellites monitoring the earth's weather:

(a) Geostationary orbit (b) Polar orbit

Figure 2.1: Illustration of two different orbits obtained from NOAA [8]. The red circles represents the orbit, while the yellow area on the earth represents the satellite's coverage.

Satellites in geostationary orbits is used in this project. Geostationary satellites circle the earth at the exact same rate as the earth. An illustration of a geostationary orbit is shown above in figure 2.1a. This makes the satellites stay above one particular location on earth. Because of this feature, geostationary satellites is useful for obtaining weather data, and analysing weather changes. [4]

One problem with geostationary satellites is the lack of the ability of monitoring the poles. Physics makes it only possible to have geostationary satellites above the equator. Due to the angle of the satellites, they do only cover from 81.3° south to 81.3° north [12]. This problem is solved by using polar orbiting satellites to monitor the weather at the poles. An example of a polar orbit is illustrated in figure 2.1b above. The weather data from the polar orbiting satellites is updated relatively slow compared to the geostationary weather satellites, due to their orbit [8]. Because of this and the fact that there will be few queries on the weather at the poles, this project will exclude the poles.

2.1.2 Earth coverage

As the geostationary satellites only cover one part of the earth, multiple satellites must be used. Because of earth's curvature, the resolution of the satellite images will be lower near the edges. Figure 2.2 below show this phenomena, where the resolution is defined as square meters per pixel:

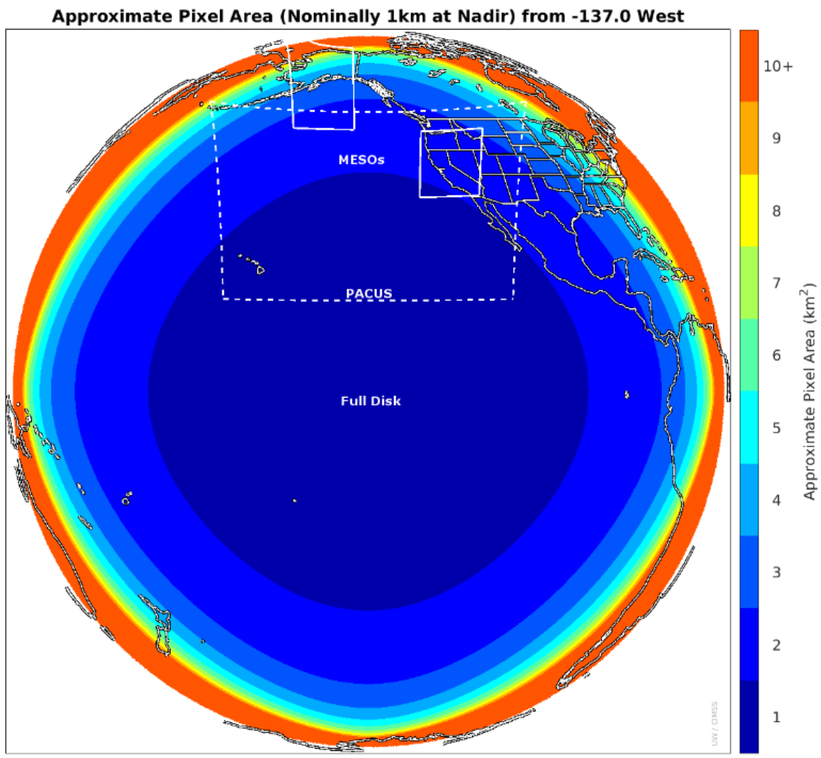

Figure 2.2: Illustration from NOAA showing the falling resolution of satellite images near the edge [9]

.

As a consequence of the lower resolution near the edge of the satellite images (shown in figure 2.2 above), it's beneficial to utilize other satellites for these areas. The more satellites being used, the higher the mean resolution will be. A downside of using more satellites, is that many of the satellite images has to be downloaded from different places. There is also some differences between the data the satellites provide. These differences includes file format, resolution and other image attributes. Because of these facts, each satellite implemented adds complexity and development time. The number of satellites is therefore kept at a reasonable level. The satellites utilized in this project is shown in table 2.1 below:

Name	**Longitude**	**Operator**	**Lowest band resolution**
GOES 16	$75.2°W$	NASA/NOAA	2000m
GOES 17	$137.2°W$	NASA/NOAA	2000m
Himawari 8	$140.7°E$	JMA	2000m
Meteosat 8	$41.5°E$	EUMETSAT/ESA	3000m
Meteosat 11	$0°$	EUMETSAT/ESA	3000m

Table 2.1: Overview over the satellites utilized in this project. Data for GOES 16 and GOES 17 is obtained from NOAA [9]. Continuously the data for the Himawari 8 satellite is obtained from JMA [10], while the EUMETSAT provided the data for Meteosat 8 and Meteosat 11 [6]. The resolutions is received from the University of Twente [11]

The chosen satellites shown in table 2.1 above is spread around the globe at different longitudes. By drawing the tangents from each satellite to the earth, it can be shown if all places at equator is reached. To illustrate this, both earth's radius and the satellites' orbit altitude is needed.

Earth's radius is gotten from NASA Planetary Science Division [3]:

$$earth_{radius} = 6371km \tag{2.1}$$

Altitude of geostationary satellites is gotten from ESA [4]:

$$satellite_{altitude} = 35786km \tag{2.2}$$

An illustration of the satellites from table 2.1 above with the correct measurements of the earth and the satellites is shown in the figure below 2.3:

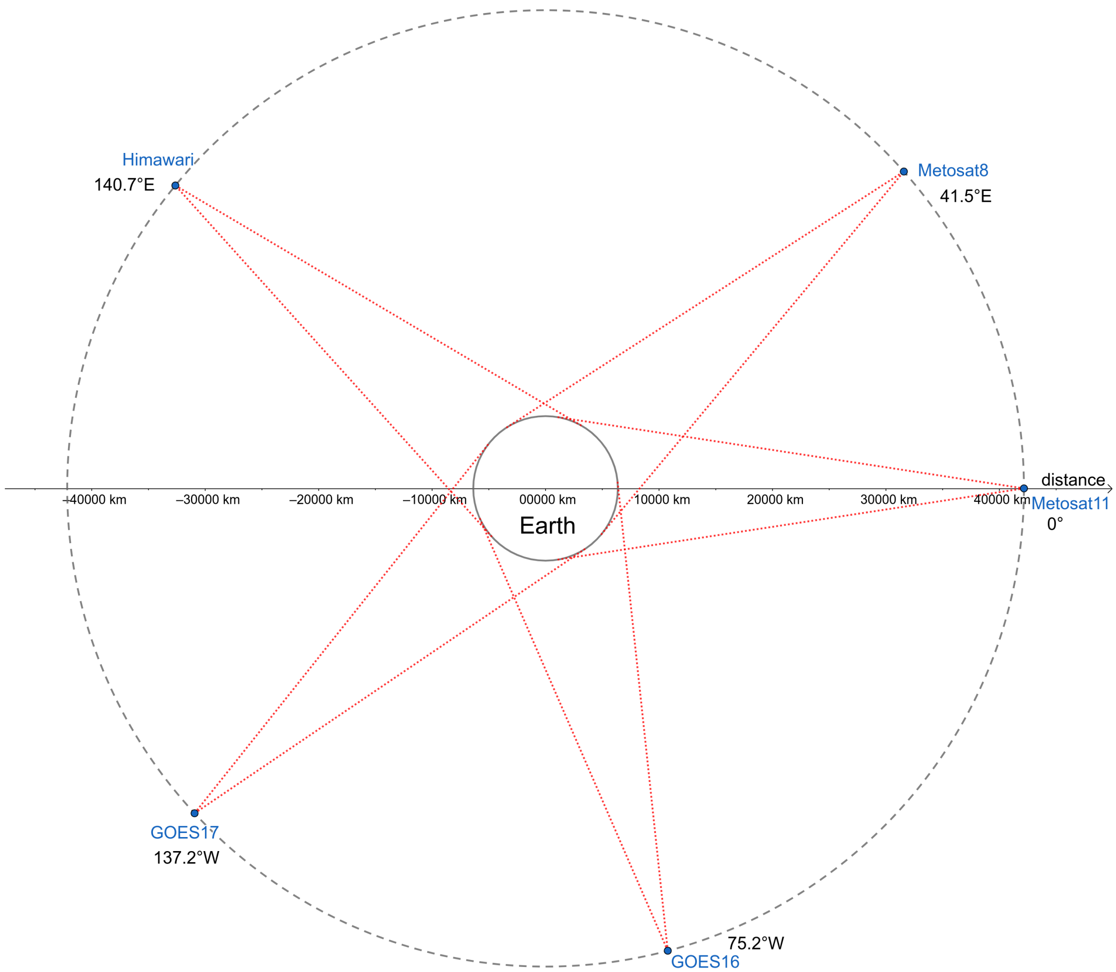

Figure 2.3: Illustration made with GeoGebra showing earth with the chosen satellites. The inner circle represents earth, while the outer represents the geostationary orbit. The red dotted lines shows the outermost view of earth for each satellite.

Figure 2.3 above reveal that every point on earth is covered using the five chosen satellites. As illustrated in figure 2.2 at page 6 the resolution at the outermost view each satellite have of the earth is greatly reduced. By calculating the biggest longitude difference between two adjacent satellites, the point on earth with the lowest resolution can be found. The biggest longitude difference is calculated in the equations below:

$$A = |metosat8_{long} - metosat11_{long}| = |41.5°E - 0°| = 41.5° \tag{2.3}$$

$$B = |himawari_{long} - metosat8_{long}| = |140.7°E - 41.5°E| = 99.2° \tag{2.4}$$

$$\begin{aligned} C &= |goes17_{long} - himawari_{long}| = |137.2°W - 140.7°E| \\ &= 222.8°E - 140.7°E = 82.1° \end{aligned} \tag{2.5}$$

$$D = |goes16_{long} - goes17_{long}| = |75.2°W - 137.2°W| = 62.0° \tag{2.6}$$

$$E = |metosat11_{long} - goes16_{long}| = |0° - 75.2°W| = 75.2° \tag{2.7}$$

By inspecting the equations above (2.3 - 2.7), it's clear that B in equation 2.4 has the biggest angel. This is the longitude difference between the Himawari and Meteosat 8 satellites. Table 2.1 at page 7 also shows that the Meteosat satellites has 3000 metres as the lowest resolution (at a perfect angle). This means that the part where the Meteosat 8 satellite overlaps the Himawari coverage is the covered location that has the lowest resolution. It is beneficial to calculate the resolution difference at the worst point, and verify that it is good enough. A comparison between the resolution at a perfect angle versus the worst angle is illustrated in figure 2.4 below:

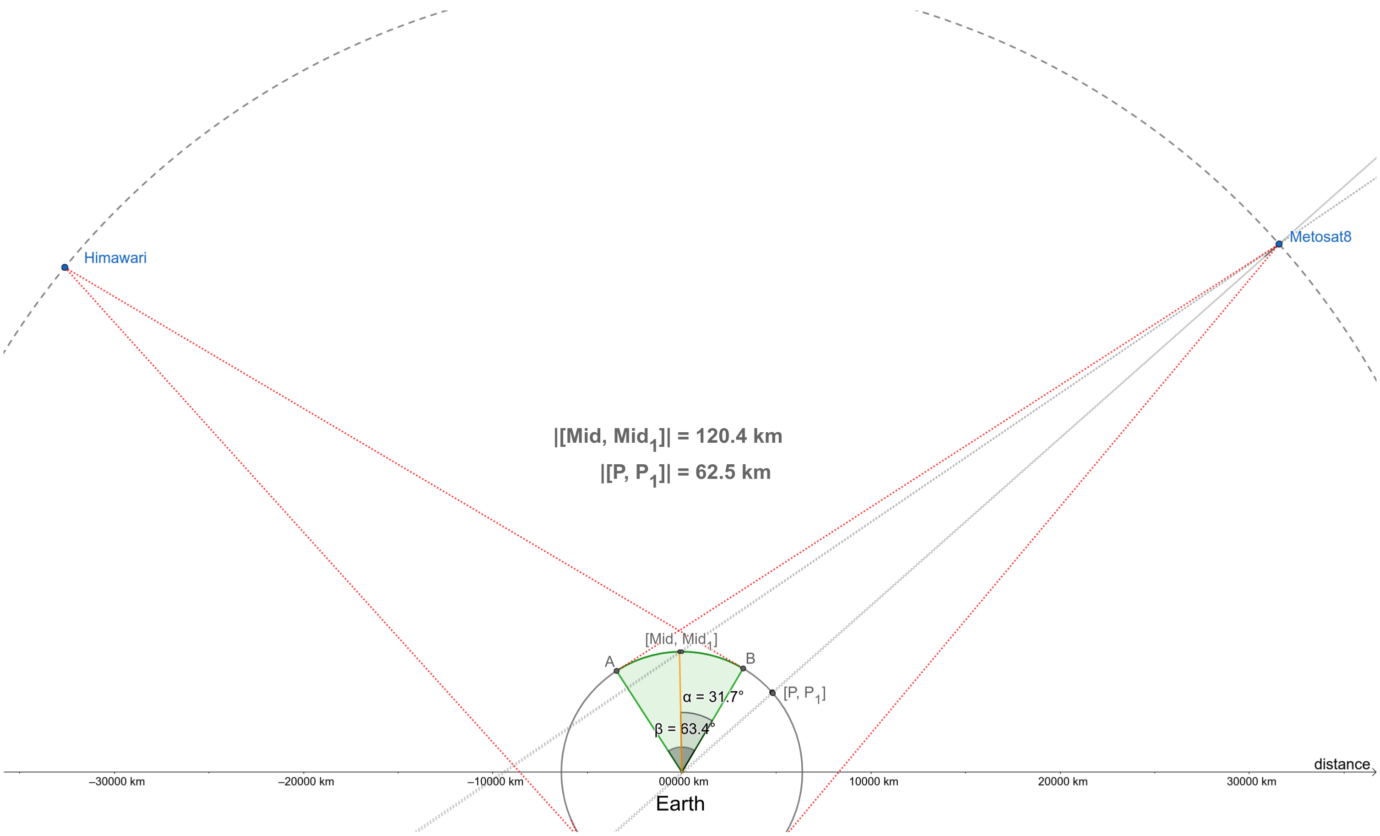

Figure 2.4: Comparison made with GeoGebra of the resolution at a perfect angle versus resolution at the worst point. Both measurement points is made of a vector with two points that is 0.1° apart from each other, from Meteosat 8's perspective. $[P, P_1]$ is the vector from a perfect angle, while $[Mid, Mid_1]$ is the vector at the worst point. The illustration also shows the length of these vectors.

In figure 2.4 above there exists two named vectors. These are named $[Mid, Mid_1]$ and $[P, P_1]$, and represents Meteosat 8's accuracy at the given location. $[Mid, Mid_1]$ is the accuracy at the worst location, which is the location in the middle of the edge of Meteosat 8 and the edge of Himawari. These locations are marked as A and B in figure 2.4, respectively. $[P, P_1]$ is the accuracy at the best location. With the use of GeoGebra, the lengths of the two named vectors can be calculated and displayed. These are displayed in the middle of figure 2.4 above. The following values is observed:

$$||[Mid, Mid_1]|| = 120.4km \tag{2.8}$$
$$||[P, P_1]|| = 62.5km \tag{2.9}$$

With the values from equation 2.8 and 2.9, the drop in accuracy (AD_{PMid}) from $[P, P_1]$ to $[Mid, Mid_1]$ can be calculated:

$$\begin{aligned} AD_{PMid} &= ||[P, P_1]|| \ / \ ||[Mid, Mid_1]|| \\ &= 62.5km/120.4km \\ &= 0.5191 \\ &= 51.91\% \end{aligned} \tag{2.10}$$

Equation 2.10 above shows that the accuracy of $[Mid, Mid_1]$ is 51.91% the accuracy of $[P, P_1]$. Dividing the resolution of Meteosat 8 with AD_{PMid} gives the resolution at the location with the worst accuracy:

$$\begin{aligned} Mid_{resolution} &= P_{resolution} \ / \ AD_{PMid} \\ &= 3000m/0.5191 \\ &= 5779m \end{aligned} \tag{2.11}$$

As calculated above in equation 2.11, the worst resolution with the satellites utilized is $5779m$. The consideration of this accuracy being good enough depends on the definition of which clouds are considered above a specific coordinate. Clouds from 30° degrees vertically, relative to the horizontal plane at any location, is in this case defined as clouds above that location. Figure 2.5 below uses the distance from equation 2.11 above, to show how low the measured clouds can be, before the measurement gets too inaccurate:

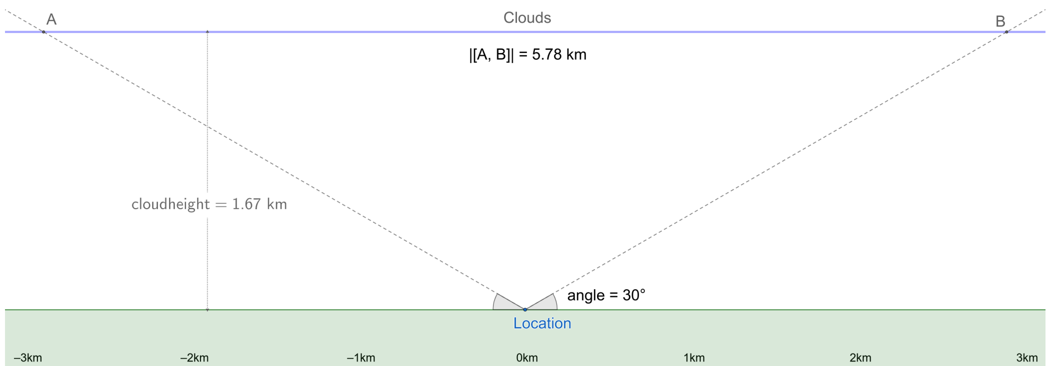

Figure 2.5: Illustration made with GeoGebra calculating the lowest cloud height that is accurate with 5.78km resolution and an angle of 30° degrees.

By inspecting figure 2.5 above, it is noticeable that the lowest cloud height at the location with the worst resolution is $1.67km$. This cloud height represents the top of the cloud. As most cloud tops has an altitude higher than $1.67km$, this resolution is considered acceptable.

Due to the fact that the location with the worst resolution is acceptable, all other locations covered must have an acceptable resolution. As mentioned earlier (see page 2.1.1) the geostationary satellites utilized in this project doesn't cover the poles. Because of this the locations above 81.3° north and below 81.3° south is not covered. Latitudes close to these values will as a side effect have lower resolution, due to the falling resolution close to the satellites' edges (illustrated in figure 2.2).

2.2 Satellite frequency bands

Satellites used for monitoring the weather on earth takes images using multiple frequency bands (ranges of frequencies). The frequencies used ranges from visible to infrared light. Each frequency band has a different area of application. Useful information can be gathered by combining these frequencies in different ways.

2.3 Obtaining satellite data

As seen in table 2.1 at page 7, the satellites is managed by three different operators. Satellites managed by the same operator can usually be implemented using the same code. This is because operators typically put the image data from satellites on the same site as the other satellites they manage. Other factors like file format and resolution is also often standardised for the satellites managed by an operator. Multiple steps must be carried out to obtain the necessary satellite data. The goal is to download the most recent data with the highest available resolution. It is beneficial if the data is separated into multiple files, so only the files needed can be downloaded. An example of this is satellite images separated based on frequency bands (mentioned in section 2.2).

2.3.1 Online resources

The two satellites operated by NASA/NOAA (GOES 16 and GOES 17) and the Himawari 8 satellite has their data available on AWS (Amazon Web Services). According to Amazon, AWS should provide the most recent data that is available for these satellites, for free [1][2]. An advantage of receiving the data from these three satellites out of the same resource, is that a lot of the code can be used for all the satellites, which reduces the time spent developing, as well as the complexity of the code. This leads to the application being less prone to bugs.

Satisfying the reasoning above, the two satellites operated by EUMETSAT/ESA (Meteosat 8 and Meteosat 11) is also downloaded from the same resource. Both of these satellites' data is obtained from directly from Eumetsat, as it is not available on AWS. The most recent data available can be obtained without any cost [5]. The only requirement is that a user is registered on Eumetsat's Earth Observation Portal.

2.3.2 Separation of data

Great separation of data at the online resource where it is received from is beneficial to reduce the amount of data to be downloaded. This is important to reduce download time. Download time is a concern, resulting from the great amount of data needed for the high resolution images, that is gathered from every satellites.

As mentioned in section 2.2 at page 12, not all frequency bands is necessary when retrieving cloud data. The data from GOES 16, GOES 17 and Himawari 8 is separated into multiple files, based on the frequency bands. By only downloading the frequency bands needed, the amount of data to be downloaded is greatly reduced.

Data from the Meteosat satellites is not separated in any way (except by time). This results in redundancy when downloading the data, as even the irrelevant frequency bands is downloaded. As there is no other resource available to download this data from directly, there is no way to get past this problem.

Chapter 3

Design and construction of software

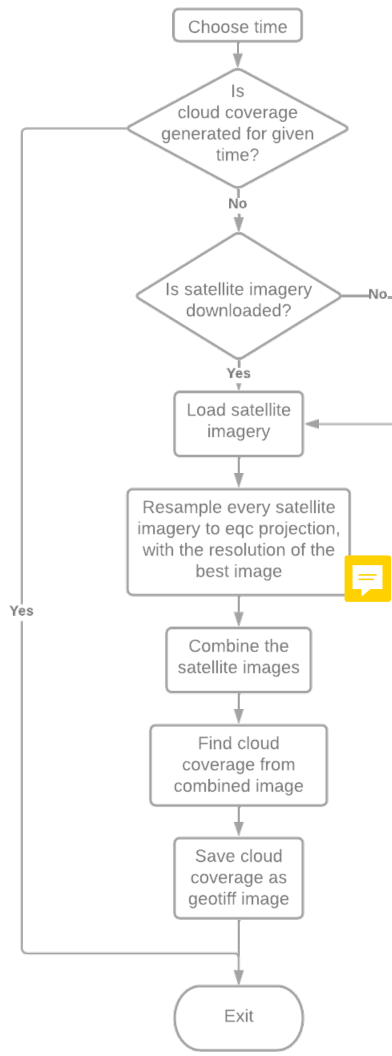

Figure 3.1: Flowchart showing a general overview of the application

Chapter 4

Results and discussion

Resultatene kan diskuteres samtidig som de presenteres: Resultat 1, diskusjon rundt resultat 1, resultat 2, diskusjon rundt resultat 2. Eller du kan først presentere alle resultatene og så diskutere dem.

Chapter 5

Conclusion

Bibliography

[1] Amazon. Jma himawari-8. https://registry.opendata.aws/noaa-himawari/, 2022.

[2] Amazon. Noaa geostationary operational environmental satellites (goes) 16 & 17. https://registry.opendata.aws/noaa-goes/, 2022.

[3] NASA Planetary Science Division. Earth - nasa solar system exploration. https://solarsystem.nasa.gov/planets/earth/in-depth/, 2018.

[4] ESA. Types of orbits. https://www.esa.int/Enabling_Support/Space_Transportation/Types_of_ 2020.

[5] Eumetsat. How to access our data. https://www.eumetsat.int/access-our-data, 2022.

[6] Eumetsat. Metosat series. https://www.eumetsat.int/our-satellites/meteosat-series, 2022.

[7] Pytroll group. Satpy documentation. https://satpy.readthedocs.io, 2021.

[8] NWS. Satellites. https://www.weather.gov/about/satellites, 2022.

[9] National Oceanic and Atmospheric Administration. Goes-r series data book. 08 2019.

[10] Meteorological Satellite Center of JMA. Himawari-8/9 spacecraft overview. https://www.data.jma.go.jp/mscweb/en/general/himawari.html, 2022.

[11] University of Twente. Itc satellite and sensor database. https://www.itc.nl/research/research-facilities/labs-resources/satellite-sensor-database/, 2022.

[12] Tomas Soler and David Eisemann. Determination of look angles to geostationary communication satellites. *Journal of Surveying Engineering*, 120:122, 08 1994.

[13] Time and Date AS. Timeanddate. https://www.timeanddate.com/company/, 2022.

Attachment A

Program list

Attachment B

Data sheet